Applied Machine Learning for Personalised Early Intervention in Autism

by

Pratibha Vellanki
(M.E.)

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Deakin University

May 2016
I am the author of the thesis entitled

Applied Machine Learning for Personalised Early Intervention in Autism

submitted for the degree of Doctor of Philosophy

This thesis may be made available for consultation, loan and limited copying in accordance with the Copyright Act 1998.

'I certify that I am the student named below and that the information provided in the form is correct'

Full Name: Prabhakar Rao Vellanki

Signed: 

Date: 31 July 2016
DEAKIN UNIVERSITY
CANDIDATE DECLARATION

I certify the following about the thesis entitled [30 word maximum]

Applied Machine Learning for Personalised Early Intervention in Autism

submitted for the degree of Doctor of Philosophy

a. I am the creator of all or part of the whole work(s) (including content and layout) and that where reference is made to the work of others, due acknowledgment is given.

b. The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

c. That if the work(s) have been commissioned, sponsored or supported by any organisation, I have fulfilled all of the obligations required by such contract or agreement.

d. That any material in the thesis which has been accepted for a degree or diploma by any university or institution is identified in the text.

e. All research integrity requirements have been complied with.

‘I certify that I am the student named below and that the information provided in the form is correct’

Pratibha Prabhakarrao Veilanki

(Please Print)

Signature Redacted by Library

Full Name: 

Signed: 

Date: 23/05/2016

Deakin University CRICOS Provider Code 00113B
This journey has been incredibly valuable and to this, first and foremost I thank my supervisors. My co-supervisor, Dr Thi Duong, I have learnt from you how to express ideas simply yet eloquently in technical writing. Your expertise in the behavioural sciences and autism fields has helped me navigate smoothly through the diversity between them and technology. Thanks for your support from the initial year of my doctoral fellowship in understanding and expressing the conclusions in the behavioural sciences terminology. My co-supervisor, Associate Professor Dinh Phung, thanks for being a constant support and helping me polish my technical skills throughout this journey. The books you left at my desk, not only increased my understanding about autism but also helped me develop empathy and critical thinking about disorders, social treatment, and acceptance. My best days of the fellowship were the ones that concluded in a meeting with you. Thanks for taking the stress off me at times when it was absolutely necessary. I extend my deepest gratitude to my principle supervisor, Professor Svetha Venkatesh. Thanks for being the inspiration that you are. You went out of your way to organise writing workshops and social gatherings for us. The reading material and wisdom you shared, made this a meaningful experience, the effect of which transcends academia. I have been fortunate to have supervisors who cared not only about my progress in research and encouraged me to build collaborative relationships with researchers inter-disciplinary fields, but also cared immensely about our social and personal lives, and mental health. Professor Svetha, I am so glad that I found a friend in a supervisor.

Carrying out work in inter-disciplinary fields cannot be complete without the valuable support of collaborators. I was fortunate to have worked with Professor Dennis Moore and Dr Angelika Anderson, from the Faculty of education in Monash University Australia. I am thankful to them for their precious time and guidance that
shaped the foundation for my dissertation (Chapter 4). I had a great opportunity to collaborate with researchers J. Godwin, K. Achary, and B. Varkey from Tamana School, a non-profit NGO for multiply challenged and autistic individuals, in Delhi, India. As a part of this work, we conducted a trial to study the benefits of having technology adapted for cultural relevance (Chapter 3). Toward the end of the fellowship, I once again had the opportunity to work with Dr Andrew Whitehouse and his team from Telethon Institute for Child Health Research, Perth, Australia. Our collaboration with them resulted in a clinical trial that was conducted in partnership with Prof. Dennis Moore’s team in Monash University. I have to especially thank Dr Joanna Granich for her time and discussions on the importance of various measures and variables as well as taking the effort to discuss integrating goals and conclusions from both perspectives. I immensely enjoyed working with her. Throughout all these collaborations, I am indebted also to Dr Stewart Greenhill, who along with my Supervisors was instrumental in the research that resulted in the development of TOBY Playpad.

I also extend my heartfelt gratitude to all my colleagues at the Centre for Pattern Recognition and Data Analytics (PRaDA). This was the first time, I moved into a new country by myself and I will always remember the moral and practical support given by Dr Sunil Gupta, Ms Shweta Gupta and Dr Santu Rana when I freshly arrived. A special thanks to my friends Shivapratap and Adham for making me feel belonged in the working place. The conversations and fun I had with you got me going at those tedious times when I was in a slump and un-slumping myself was not easily done.

Friends are the families we choose. I can’t thank Sreejoyee, Adarsha, Meylani, and Ambika enough for being my people. For allowing me to witness your journey through life and PhD and invariably consciously or not leaving examples of how to navigate the work-life balance. Thanks for being my emotional pillars through this journey.

Finally, this work wouldn’t have been possible without the support of my amazing family and extended family. Suyodhana - for being the shining beacon of my life and Tejaswi - for your practical advice on life matters. Vivek - for being my moral compass and the endless conversations about making ourselves and the world whole and a better place. My father for being patient with me. Last but not the least, to
my Mother, whose dream it was to see me graduate and study all that there was to be studied. Thanks to you Ma, I am formally in a profession that constitutes endless studies. Your passion for education has been an inspiration and your support allowed me to carve out this path. To you, I dedicate this thesis.
Relevant Publications

Part of this thesis has been published or documented elsewhere. The details of these publications are as follows:

Journal Articles:


Conference Papers:


Book Chapter:


Under Submission:

1. Greenhill, S., Duong, T., Vellanki, P., Phung, D., Venkatesh, S., Godwin, J., Achary, K., & Varkey, B. Computer Assisted Autism Interventions For India. (Submitted to a conference) (Chapter 3)

Contents

Acknowledgements iv

Abstract xxiii

Notation xxvi

1 Introduction 1

1.1 Aims ................................... 3

1.2 Significance and Contributions ...................... 5

1.3 Outline of Thesis ................................ 8

2 Background 10

2.1 Autism Spectrum Disorder: An overview ................ 11

2.1.1 ASD: Diagnosis and Prevalence ..................... 11

2.1.2 Early Intervention .................................. 14

2.1.3 Types of Intervention ............................... 14

2.1.3.1 Behavioural Interventions ....................... 15
2.1.3.2 Developmental Interventions ................. 15
2.1.3.3 Combined Interventions ...................... 16
2.1.4 Evidence-Based Practice and ASD ............... 17
2.2 Assistive Technology and ASD ........................ 18
2.2.1 Research trends in Assistive Technology for ASD Intervention 19
2.2.2 Computer-Based Intervention ...................... 20
2.3 An overview of applied machine learning in healthcare and ASD associated data ...................................... 24
2.4 Machine Learning Techniques .......................... 27
2.4.1 Overview of machine learning ....................... 27
2.4.2 Clustering techniques .............................. 28
2.4.3 Feature allocation and latent factor analysis ............. 31
2.4.3.1 Factor analysis approach ......................... 32
2.4.3.2 Non-negative matrix factorisation ................. 33
2.4.3.3 Bayesian matrix factorisation ..................... 35
2.4.3.4 Finite Bayesian matrix factorisation approach for computing the allocation matrix ...................... 37
2.4.3.5 Infinite Bayesian matrix factorisation .......... 40
2.4.3.6 The Indian buffet process ......................... 40
2.4.3.7 Stick breaking construction for IBP .............. 42
2.4.3.8 Slice sampler .................................. 43
2.4.4 Restricted Boltzmann machines and deep learning 45

2.5 Closing Remarks 49

3 Early Intervention and Data Collection 51

3.1 Therapy Outcomes By You (TOBY) Playpad 52

3.1.1 TOBY curriculum 59

3.2 Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India 66

3.2.1 Adaptation for textual content 69

3.2.2 Adaptation for cultural context of use - the classroom 72

3.2.3 Adaptation of voice content 72

3.2.4 Adaptation of video content 73

3.2.5 Adaptation of image content 74

3.2.6 Experimental Results 74

3.2.7 Discussion on the trials conducted 82

3.3 Data Dictionary 83

3.3.1 Datasets and their source 83

3.3.2 Attributes of TOBY data 84

3.3.3 Data pre-processing and features used for data analysis 90

3.4 Closing Remarks 93
5.2.1 Representation of data ...................................... 117
5.2.2 Datasets ...................................................... 117
5.3 Clustering Framework ............................................ 120
  5.3.1 The linear Poisson gamma model .......................... 121
  5.3.2 The extended LPGM model for incomplete data ........ 123
5.4 Generating Ground-Truth and Evaluation Methods .......... 126
  5.4.1 Ground-truth .............................................. 126
  5.4.2 Latent learning pattern .................................. 127
  5.4.3 Evaluation methods ...................................... 128
5.5 Results ......................................................... 129
  5.5.1 Part I - dataset A ......................................... 129
  5.5.2 Part II - dataset B ........................................ 135
    5.5.2.1 Learning Patterns .................................. 136
    5.5.2.2 Subgroup Detection ................................. 145
    5.5.2.3 Case for outliers .................................. 149
5.6 Concluding Remarks ............................................ 152

6 Understanding Factors affecting Personalisation - User Entry Profiling .................................................. 155
  6.1 Introduction .................................................. 155
  6.2 Dataset ....................................................... 158
7.5.3 Study of progress from entry profiles and performance profiles  185

7.6 Concluding Remarks ........................... 186

8 Concluding Remarks ............................. 187

8.1 Summary ............................................. 187

8.2 Future Directions ................................. 190

A Syllabus of TOBY Playpad ........................ 193

A.1 General skills ....................................... 193

A.2 Imitation Skill Category ............................ 195

A.3 Sensory Skill Category ............................... 196

A.4 Receptive Language Skill Category ................ 196

A.5 Expressive Language Skill Category ................. 197

A.6 Social Skill Category .................................. 198

B Supplementary Derivation ........................... 199

Bibliography ............................................. 202
List of Figures

2.1 Illustration of functioning of K-means algorithm. ............... 30
2.2 Graphical model of non-negative matrix factorisation. ............ 34
2.3 Graphical model for finite Bayesian matrix factorisation. ......... 36
2.4 Illustration of Indian buffet process shown as a matrix of customers and dishes. ......................................................... 41
2.5 Stick-breaking construction for Indian buffet process (IBP). ........ 42
2.6 Graphical representation of restricted Boltzmann machines (RBM). . 45
3.1 Screen shot of the Plan tab and Reports tab of TOBY Playpad. . 56
3.2 Prompt and Reinforcement delivery mechanism in TOBY Playpad. . 57
3.3 Increase and decrease in prompting level. ......................... 58
3.4 Structured syllabus of TOBY Playpad categorised by skill area. . 60
3.5 Planning and execution for a NET task. .......................... 61
3.5 Planning and execution for a NET task. .......................... 62
3.6 Example of Imitation, Sensory, Receptive Language and Expressive Language. ................................ 64
3.6 Example of Imitation, Sensory, Receptive Language and Expressive Language. ................................ 65
3.6 Example of Imitation, Sensory, Receptive Language and Expressive Language. ................................ 66
3.7 Syllabus manager, Task edit form ........................................ 70
3.8 Syllabus manager, Target view ........................................... 70
3.9 Syllabus manager, Preview of NET task .................................. 71
3.10 Imitation videos for Western and Indian versions of TOBY. ........ 74
3.11 Syllabus manager, Interface for online collection of category media. . 75
3.12 Trial 2: Total time using TOBY and proportional breakdown for Hindi (left) and English (right) users. ............. 79
3.13 Success rates for iPad syllabus skills for English and Hindi children. . 80
3.14 Success rates for NET syllabus skills for English and Hindi children. . 80
3.15 Trial 2: Task mastery for iPad (left) and NET (right ) tasks. ........ 81
3.16 Trial 2: Skill completion rates for English and Hindi children. ........ 81
3.17 Example of log data collected from one section of therapy. .......... 86
3.18 Computing a Learn Unit. ................................................... 92

4.1 Variation in response times between correct and incorrect responses for (left) solo tasks which are Sensory and Receptive Language and (right) partner tasks which are Imitation and Expressive Language. . 101
4.2 Ex-Gaussian fits to response time distribution for correct responses (left) and for incorrect responses (right). .............................................. 103

4.3 The bimodal nature of the correct response times for Receptive Language skill (left). Histogram for incorrect responses for the same skill (right). ................................................................. 104

4.4 Clusters obtained using K-means on normalised Learn Units (nLU) . 104

4.5 Skill category level cluster properties for the three groups of learners – slow, medium and fast pace. ......................................................... 106

4.6 Age distributions in different learner groups. ............................. 107

5.1 Illustration of a sparse data matrix. ................................. 118

5.2 Properties of the dataset ..................................................... 119

5.3 Graphical model for LPGM with an IBP prior .......................... 122

5.4 LPGM incorporating missing data .......................................... 124

5.5 Sample learning pattern detected using IBP The skills represented with darker shades require more LUs by the child who follows this learning pattern. ......................................................... 127

5.6 Latent Patterns obtained using IBP dataset A (left) and dataset B (right). ................................................................. 130

5.7 Membership patterns obtained using IBP on dataset A (left) and dataset B (right). ................................................................. 131

5.8 Membership of children in dataset A with the latent patterns detected using IBP. ................................................................. 132

5.9 ℓ0-norm for latent patterns using methods LPGM (left) and NMF (right) ................................................................. 132
5.10 Entropy for latent patterns using methods LPGM (left) and NMF .......................... 133
5.11 Subgroup detection performances using F-score ................................. 134
5.12 Latent factors and memberships ........................................ 135
5.13 iPad syllabus tree for TOBY. ......................................... 137
5.14 Entropy and \( l_0 \) norms for Learning Patterns .............................. 144
5.15 Memberships associated with each learning pattern. ......................... 146
5.16 Sharing of learning patterns by children ..................................... 147
5.17 Similarity matrix for the children ........................................ 148
5.18 F1 score for change in epsilon \( \epsilon \) (difference between nLUs) between children in cohort. ................................ 150
5.19 Properties of the outliers. ............................................ 151

6.1 t-SNE projection of the the posterior hidden values. Clusters are obtained by K-means. ................................. 162
6.2 Visualisation of clusters of individuals belonging to corresponding entry profiles along with collective properties of entry profile (age, sex and Learn Units required to master each entry skill) ......................... 163
6.3 Visualisation of the same individuals shown as 10 clusters correspond- ing to entry profiles. ........................................ 165

7.1 MV.RBM models used for TOBY dataset. .................................. 170
7.2 Entry Profiles and their group characteristics. .................................. 173
7.3 Properties of performance profiles (PP) discovered after progress .... 174
7.4 Progress tracking using entry profiles and performance profiles.

7.5 Entry profiles (EP) and the associated similarity matrix.

7.6 Properties of entry profiles.

7.7 MSEL measures over different entry profiles. Visual Reception, Fine Motor, Receptive Language and Expressive Language are the four categories in which each child is evaluated. very low, below average, average, and above average are the four levels that the children are given in each of the areas.

7.8 ATEC ground truths.

7.9 Performance Profiles (PP) and the associated similarity matrix.

7.10 Properties of Performance Profiles (PP).

7.11 Mapping of entry profiles (EP) to the performance profiles (PP).

A.1 Structured syllabus of TOBY Playpad categorised by skill area. Each colour represents a different skill area. The labels inside each node denote the name of the skill that is taught in it.
List of Tables

2.1 A table of some popular conjugate priors and likelihoods. Based on the application the most appropriate pair is chosen. ............ 38

3.1 Classification of skills into Solo or Partner tasks. ............ 59

3.2 Roles of parents and system in Solo, Partner and NET activities. . . 59

3.3 Examples of home routines that were modified for a school environment 72

3.4 iPad accessibility and skills in the translated version of TOBY. . . . . 77

3.5 Anonymous data description sheet ......................... 84

3.6 Attribute descriptor table for the data. .................... 87

3.7 Data representation ........................................... 92

4.1 A typical log for a child showing sequence of stimuli that TOBY presents. .................................................. 99

4.2 Median and interquartile range of response time for solo and partner tasks. (All measurements in seconds). .............. 102

5.1 Skill numbers and their categories. ......................... 119

5.2 Notations used for model ............................... 120
5.3 average ℓ0-norm and entropy values for dataset A and B. These measures are computed for LPGM and NMF methods. ................. 133

5.4 Learning Patterns Description ........................................ 138

5.4 Learning Patterns Description ........................................ 139

5.4 Learning Patterns Description ........................................ 140

5.4 Learning Patterns Description ........................................ 141

5.4 Learning Patterns Description ........................................ 142

5.5 Clusters identified and associated learning patterns. .............. 148

A.1 List of general skills ...................................................... 193

A.2 List of Imitation skills ................................................... 195

A.3 List of Sensory skills ..................................................... 196

A.4 List of Receptive Language skills ..................................... 197

A.5 List of Expressive Language skills .................................... 197

A.6 List of Social skills ....................................................... 198
Abstract

Autism Spectrum Disorder (ASD) is a developmental disorder that interferes with the early learning in children. Research has established that diagnosis and intervention at an early age can help reduce the austerity of the condition. This intervention highly relies on data that was manually collected and analysed in conventional therapy approaches. With advance technology and computing, this process has improved drastically. The data leveraged from interventions delivered through computer based applications can be analysed to increase knowledge about ASD that can be in turn used for designing personalised interventions tailored for individual needs. The children with ASD express highly individual behavioural and learning styles, with strengths and deficits that vary across the spectrum. The discovery of learning patterns and performances plays a central role in understanding how children with ASD react to intervention and are affected by it. This allows clustering of children based on their performance on the intervention and might shed light on potential alternate personalised treatments.

With this motivation this thesis aims to: a) Identify learning patterns across the cohort of children with ASD, and b) Identify and develop tools for studying task progression. One of the incidental goals is also to study the significance of translating technology based intervention to suit cultural/linguistic context. En-route to addressing these aims the main challenges encountered are: a) Acquiring a dataset that records performance of children across a structured syllabus with relevant ground-truth, that covers a broad variety of skills and to find an ideal representation framework that quantifies learning, b) Building probabilistic and machine learning framework that addresses the growing learning patterns and model complexity as the number of children and their performance data grow over time, c) Construct-
ing framework to handle complex, heterogeneous datasets that are highly correlated with missing elements.

First, in the field of ASD, it is difficult to obtain intervention data in the form of a large dataset. In the existing literature the size of the datasets used for studies is often limited to a single digit size of the data. We introduce the intervention dataset of children with ASD with more than 500 children. This is the largest dataset in the world, in this domain to the best of our knowledge. The data is collected from the application called TOBY Playpad, which is an iPad-based intervention application for children with ASD, covering skills over the categories of Imitation, Sensory, Expressive Language, and Receptive Language. The performance of these children is quantified in terms of a measure called Learn Units, where a higher number of Learn Units required to master the skill indicates a higher difficulty dealt by the child. This information is also accompanied with the child’s age and sex as indicated by the caregivers.

Alongside, we translate the framework of delivering instruction to suit the cultural context by exploring factors such as language constraints, cultural context and contrast in utility. This required adaptations in the textual, image, video, and audio content, wherein aspects of the intervention were also edited to suit the classroom environment. This was done to suit the requirements of the trial conducted in a school in India, through which it was observed that the progress of the children was slightly better in Hindi as compared with the English version. Our study also aligns with evidence based intervention principles, which suggests that interventions should be generalisable to match individual child’s needs.

Secondly, for discovering latent learning profiles we used the recent advances in the field of Bayesian nonparametrics. Specifically, we use a nonparametric, nonnegative, matrix factorisation method to obtain learning patterns and membership patterns. Learning patterns are sets of skills that a group of children find particularly difficult to master together. Deeper analysis of learning patterns shows connections between skills that show that some skills are potentially pre-requisites for others. Membership patterns reveal connections between children who follow similar learning patterns. This allows us to cluster them into groups that are similar in their performance on the given syllabus. Knowledge regarding this can assist in tailoring personalised syllabi.
Finally, as such data is generally collected anonymously from the users, sensitive information such as diagnosis and ground-truth is not available. The trade-off involved in collecting such sensitive information from a clinical trial is that it limits the size of the data. In order to study the progress, it is crucial to have information about the child’s profile before they begin the intervention. Then one can compare their profiles before and after progress in the intervention. Thus we develop a framework using mixed-variate restricted Boltzmann machines to identify entry profiles of children. Entry profiles define groups of children that are similar in their characteristics such as sex, age, and performance of beginning (easy level) skills. Once the children progress through the syllabus, we determine their performance profiles, which indicate groups of children that are similar after progress. Mapping entry profiles onto performance profiles is a way to study how children migrate away or stay together (similar) before and after intervention as contributed in this thesis. Furthermore, clinical data, by the means of ground-truth, when available increases our confidence in the results. With the data of 38 children with ASD collected from a trial conducted in Australia, we use the similar framework to observe the entry and performance profiles of this cohort and compare them against the Mullen Scales of Early Learning (MSEL) and Autism Treatment Evaluation Checklist (ATEC).

In summary, to our best of knowledge, this thesis is the first that addresses the problems of early intervention in ASD through the lens of machine learning and data analytics. Our key contribution includes the establishment of large datasets in this domain for the first time together with a systematic machine learning based approach to extract knowledge relevant to ASD. We believe the findings and contributions from this thesis form the initial foundations for a data-driven approach to personalised early intervention for children with ASD.
Abbreviations

ASD  Autism Spectrum Disorder
ABA  Applied Behaviour Analysis
DTT  Discrete Trial Training
PECS Picture Exchange Communication System
LU   Learn unit
nLU  Normalised learn unit
NMF  Non-negative matrix factorisation
LPGM Linear Poisson gamma model
IBP  Indian buffet process
RBM  Restricted Boltzmann machine
MV.RBM Mixed-variate restricted Boltzmann machine
DNN  Deep neural networks
Chapter 1

Introduction

Autism Spectrum Disorders (ASD) are pervasive neurodevelopmental disorders that cause impairment in social interactions, communication and cognitive skills. These disorders predominantly disrupt early learning and developmental process, often restricting interests and causing stereotyped behaviour patterns associated to repetitiveness, resistance to change, attachment to specific objects and inability in comprehending social conduct and emotional exchanges in affected children (American Psychiatric Association, 2000). These impairments disrupt everyday life activities, increasing the dependence of individuals with ASD on others.

Research has established that early diagnosis and beginning intervention at an early age, as young as 2 to 3 years of age, can help reduce the austerity of the condition (Dawson et al., 2008). It is suggested that a rigorous early intervention program, which includes 20 to 25 hours per week of one to one intensive training with a trained therapist, may achieve the anticipated results (Lovaa, 1987). Many such intensive interventions demand a lot of manual work for preparing material and analysing progress, requiring parents to work along with the therapists. The popular method Applied Behaviour Analysis (ABA) is an example of such an intervention. It is difficult for a therapist to engage rigorously with each child independently and the process is also time intensive. What’s more is the usual waiting time that the parents have to go through causing increased stress levels in parents (McKinney and Peterson, 1987).
Before delving into the motivation of our research, we would also like you to note that children with ASD exhibit individual learning patterns. The strengths and deficits they exhibit in each skill area vary considerably calling for a thorough analysis to identify the patterns. Once the learning pattern is identified, an appropriate intervention program can be tailored for each child. ABA (Lorimer et al., 2002) provides a general framework regarding how the process of intervention and analysis are to be administered for a broad range of skill areas. Nevertheless, the recent SCERTS (Prizant et al., 2006) model implicates the importance of identifying and understanding individual strengths and weaknesses. The SCERTS works with parents to achieve its targets.

Many technology based solutions are available that cater to the learning needs of children with ASD. Such assisted technology can be helpful to parents as well as therapists in delivering intervention. Parents may benefit from assisted technology either by appending it to the traditional intervention routine or by using it during the waiting time in the absence of a therapist. Technology-based interventions have simplified and also improved the efficacy of data collection and analysis. The approach of computer-based interventions towards the problem of individual learning patterns is two-fold: to build applications that cater to specific skill deficits, or to build applications with a broad syllabus such that it can be used by all children with ASD. In both the cases, it is essential to begin with manually determining the child’s strengths and deficits. The required intervention is usually a combination of many skills with higher training in a few identified skills. There are existing applications that analyse the data recorded to justify if the given application is an Evidence-Based Intervention (Whalen et al., 2010; Charlop-Christy et al., 2002).

An intervention which detects the requirements of individual child with ASD and adapts the syllabus can be termed as a personalised intervention. We have identified a gap in the data-based studies with regards identification of factors affecting personalisation. Traditional research conducted on children with ASD has been limited by the amount of data that can be recorded for analysis in a clinical study. The size of these datasets is relatively small and unsuitable for applying machine learning approach. On the other hand, studies conducted in the domain of computer-based ASD intervention have not leveraged the data recording capabilities of these applications to investigate the underlying patterns associated with learning. This avenue is now open for research with the increase in availability of applications that have
1.1. Aims

TOBY Playpad (Venkatesh et al., 2013) is an early intervention program that provides a structured syllabus for children with ASD containing 326 tasks covering 51 skills. It is an application that incorporates the recommendations (Wainer and Ingersoll, 2011) made for a tablet or touch-phone based interventions and thus is state-of-the-art. TOBY rigorously records the detailed report of each user through the syllabus. As a result, it allows us to analyse the data and extract the learning patterns of individual users. Using machine learning techniques, we intend to build models that automatically identify these learning patterns and is capable of grouping the children accordingly.

1.1 Aims

Unprecedented data from children offers the capacity to produce evidence-based insights into learning patterns. This has the potential for better delivery of new programs. TOBY has presented means to provide early intervention in a rigorous learning framework, with an extensive recording of stimulus, prompt and response. The rich data is complex, amounting to a record of every interaction made by a child over time in using TOBY. This presents a unique opportunity to apply novel techniques and interrogates the data, to expose useful patterns.

To realise this, the broad goals are:

- **Aim 1: Identifying learning patterns across the cohort of children (Chapters 4 and 5).** This includes understanding learning patterns using a quantified measure of learning amongst children with ASD. Learning patterns are a set of skills that a group of children will find difficult to learn together. We discover these patterns using a factor analysis approach using Bayesian non-parametric, non-negative, matrix factorisation. Identification of learning patterns has potential implications for improving early intervention through facilitating the development of evidence-based strategies tailored for different learner profiles.

- **Aim 2: Developing tools for studying task progression (Chapters 6 and 7).**
1.1. Aims

It is essential to determine the profiles of the children before the onset of intervention before one can analyse progress. *Entry profiles*, as we name them, encompass features of children such as their age, sex and their performance on beginner skills. By employing restricted Boltzmann machines on these features we identify the entry profiles that the children belong to. We then analyse the progress of the children after they have undergone intervention as compared to their respective entry profiles.

The overarching aims are broken down to specific projects associated with each chapter as follows:

**Chapter 4: Perform exploratory data analysis:** This chapter will examine the progression of children in the cohort using the rate of learning measured by Learn Units (LU). A learn unit to mastery of a skill is quantitatively defined, and slow, medium and fast learner groups are derived. These clusters are then further scrutinised with respect to the four sets of skills — imitation, sensory, receptive language and expressive language. **Open questions include:** What are the differential patterns in different learner groups across skills? What are the patterns of prompting related to learning rates? What could the extraction of different learner groups offer in term of building intervention programs, tailoring the syllabus to different learner profiles?

**Chapter 5: Use Bayesian non-negative matrix factorisation techniques for identification of learning patterns:** Chapter 4 examines the data directly. This chapter seeks to use machine learning methods, specifically non-negative matrix factorisation in an infinite Bayesian setting, to extract latent patterns and sub-groups of children in terms of usage of these patterns. Non-negative matrix factorisation (Lee and Seung, 1999) is used because it can factorise a matrix to yield part-based explanations. We generate a different view of the entire cohort by converting the data to a (child-skill) matrix, where each entry is *LUs to mastery*. Factorisation will yield two matrices; the first latent factor matrix describes groups of skills that are often found difficult to master together (as determined by a larger number of LU acquisition in the specific skills), and the second matrix describes how each child utilises the discovered factors. **Open questions include:** Can the learner profiles be built across all skills using data from users who download and use TOBY? What could this information reveal about the learner’s strengths and weakness in communication (language and social), and cognitive skill acquisition?
1.2 Significance and Contributions

Chapter 6: Understand the factors affecting personalisation: Determining profiles at the onset of intervention is integral to building ground-truth data to compare the progress of the children before and after intervention and eventually enabling suggestions regarding the course of intervention. Entry profiles, as we define, are a way of describing the state of the participant before beginning therapeutic intervention. For example, we would expect children within a certain age range, assigned a certain sex at birth and expressing difficulty in similar skill areas to belong to the same entry profile. In an online setting such as TOBY, this information becomes especially crucial for personalisation. This chapter observes features of participants such as age, sex and their performance on the entry skills to identify entry profiles and cluster children based on these profiles. Open questions include: Can a set of mixed-variate data such as binary (sex), continuous (age) and count data (LUs to mastery) be integrated into a framework to yield meaningful entry profiles? This is difficult because many variables such as their diagnosis and conditions and in some cases, even basic demographic information may be unknown.

Chapter 7: Track change in performance from entry profiles to performance profiles: Building upon our aim in the Chapter 6, where we develop an understanding that entry profiles are an alternate representation of the ground-truth, in this chapter we are motivated to derive performance profiles using a similar framework and analyse the progress relative to entry profiles. Performance profiles are the state of the children once they have made some progress along the syllabus of the TOBY. We then map the entry profiles onto their performance profiles and analyse their progress. Open questions include: How are children similar or different in entry profiles affected once the intervention is administered? Who stays together and which characteristics do they present?

1.2 Significance and Contributions

The significance of this work is to advance the understanding of early learning is children with ASD and through this offer insights into facilitating personalised intervention programs. The overarching significance and contributions are listed below:

- A thorough investigation into the therapeutic data and its representation. The
First contribution of this thesis lies in the dataset itself. The extensive data leveraged from children using TOBY as an intervention medium is to the best of our knowledge the largest dataset in this domain. As the data collection process is ongoing, there will always be children who are working their way through the syllabus. This results in missing elements in the data for parts of the syllabus that are yet to be attempted. The data is highly correlated and also challenging in limited information regarding diagnosis and presence of missing elements. This calls for a thorough investigation of the variables recorded, their representation and quantification. We conduct an exploratory study to assess the usefulness of some of the variables. We cluster the children based on Learn Units required to master each skill using a K-means. We perform a categorical analysis of the variations in the performance by each group in different skill categories. We then discuss these findings in the context of the literature in the field of ASD education and therapy. As the data is novel these steps contribute to improved understanding of how the children engage with the application in their natural environment with parent supervision. Furthermore, it assists in developing new features and cohort characteristics that will be significant for further investigation. By comparing the response time of the cohort for correct and incorrect responses, we investigate if deliberate responses and random responses can be distinguished. The findings form the foundation and motivate us toward the use of more complex models.

- **Discovery of learning patterns and clustering of autism subgroups from data.** Using Bayesian non-parametric factor analysis method (LPGM) offers several advantages: (a) In a complex model where the data is highly correlated, such a non-parametric method makes it possible to automatically determine the number of latent factors; (b) Non-negative matrix factorisation results in meaningful latent factors, which in our case are called learning patterns, can be easily translated in terms of syllabus; (c) Factor analysis approach allows us to simultaneously study learning patterns and how children’s membership to these patterns play out; (d) The model accommodates the growing data and allows for extension to incorporate the missing nature of the data. By applying a Bayesian non-parametric factor analysis method we contribute two analysis strategies: to understand the frequently co-dependent skills i.e sets of skills that are often found difficult to master possibly owing to the common pre-requisites, and to understand the groups of children who fare similarly on
these sets of skills. Additionally, we extend the framework to work in the presence of missing data, which is a common occurrence in real-world data and may not be critically imputable in the case of health data. Due to the complexity of the data adopting a Bayesian non-parametric approach allowed us to discover the number of factors (sets of skills found difficult to master) automatically by the model. The differential patterns of learner groups across skills could offer opportunities for improved therapies by presenting more learning opportunities (LUs) for difficult skills areas and utilising the learners’ strength in some skills to accelerate their progress into others. Additionally, inter-skills and crossed skills patterns are potentially useful in contributing to the developments of intervention syllabi that address the challenges of providing suitable pre-requisites and supportive sets of skills from the vast skill pool for targeting a particular deficit. That could help learners confidently progress through tasks, accelerate learning, and provide a safe net of skills to fall back on if they are struggling.

- **Developing entry profiles based on the characteristics at the onset of intervention.** Entry profiles offer a basis for evaluating the progress of individuals in any given medical or educational procedure. In a clinical/classroom setting, it is feasible to obtain the characteristics of the individual before the intervention begins. In the case of computer-based intervention however, such diagnostic information may be difficult to collect due to privacy concerns of the individual. Hence it is crucial to develop a mechanism of constructing entry profiles based on their performance on the basic skills in each skill category. Combining it with the information such as age and sex of the child we attempt to construct entry profiles by employing restricted Boltzmann machines. We leverage the ability of multi-variate restricted Boltzmann machine (MV.RBM) to model heterogeneous data and use age, sex and performance of the beginner skills in terms of LUs to determine entry profiles of the children. These profiles represent the ground-truth of data and are crucial for personalisation of intervention.

- **Analysis of progress by mapping entry profiles onto performance profiles.** In order to study the progress of the children through the intervention, we use a similar modelling framework to determine Performance profiles. A similar framework as entry profiles (MV.RBMs) are employed to determine these per-
performance profiles of the children based on the LUs required to master the skills once they have completed part of intervention. We compare the entry profiles with performance profiles to study how children migrate away or toward each other as they master parts of the syllabus. Clinical data, by the means of ground-truth, when available increases our confidence in the results. With the data of 38 children with ASD collected from a trial conducted in Australia, we compare profiles against the baseline measures recorded during the trial.

1.3 Outline of Thesis

The rest of the thesis is organised as follows:

• In Chapter 2 we present the literature and background relevant to this thesis. The two main domains we cover in this chapter are Autism Spectrum Disorder (ASD) and Machine Learning. In the first two sections, we describe the ASD condition, its diagnosis and traditional methods of intervention. In the third section, we review how machine learning has been applied in the relevant domain. Finally in the last section, we provide a background of the machine learning techniques that we use in this thesis.

• We dedicate Chapter 3 to the description of our source that is TOBY Playpad and the dataset. Knowledge about the syllabus and functioning of TOBY is crucial to interpreting the results of our experiments.

• Chapter 4 constitutes part of our 1st aim, which is to perform an exploratory analysis of the data in order to investigate the overarching performance of the children based on the Learn Units they require to master each skill. We also investigate other variables such as the response time and age and compare and contrast their effect on the children.

• In Chapter 5 we work toward completing our 1st aim. Here, we use a Bayesian non-negative matrix factorisation technique for identification of learning patterns and children sub-groups. Specifically we employ Linear Poisson Gaussian Model and extend it to accommodate missing data. We perform a factor ana-
lysis on the results obtained and discuss the learning patterns i.e. group of skills that were found difficult to master, and sub-groups of children with ASD.

- In Chapter 6, we move on to our 2nd aim. Here we are motivated to discover entry profiles of children based on their initial characteristics before beginning the intervention. Due to the heterogenous nature of the data we employ a multi-variate restricted Boltzmann machine method to draw clusters of children possessing varying characteristics.

- In our final Chapter 7, we see our 2nd aim through by understanding the progression of children from their initial entry profiles once they have mastered a part of the syllabus. Here we discover performance profiles and discuss how children belonging to different entry profiles deviate toward different performance profiles or converge into one.
Chapter 2

Background

This chapter provides a review of the literature and the background for the work investigated in the thesis. Revisiting the introduction, our main goal in this thesis is to develop a framework that helps us interrogate data and reveal patterns relevant to how children with ASD learn via therapy. This thesis is the culmination of an intersection between the disciplines of behavioural sciences and machine learning. In particular, we use machine learning techniques that help us identify patterns and look toward the behavioural sciences for implications. This chapter is thus divided into four sections. The first section is dedicated to detailing the background about Autism Spectrum Disorder and various ways in which intervention techniques are traditionally delivered. In the second section, we address how technology plays a part in interventions by accommodating to the needs of the child, by providing alternate means to administer therapy, and by assisting therapists in recording and analysing rigorous data. In the third section, we discuss the application of machine learning techniques in the field of general healthcare while identifying the lack of in-depth research on the usefulness of data harnessed in the field of ASD. Finally, the fourth section is dedicated to the background and theory on the machine learning techniques, mainly covering two approaches: the latent factor analysis models and the restricted Boltzmann machines. The aforementioned methods are of main interest to us due to their ability in successfully recognising patterns that are meaningful and learning features respectively.
2.1 Autism Spectrum Disorder: An overview

Neurodevelopmental disorders are a group of conditions that typically manifest at an early age and are characterised by developmental deficits that produce impairments of personal, social, academic and occupational functioning. Autism Spectrum disorder (ASD) is a neurodevelopmental disorder that is specifically characterised by a) Deficits in social communication and social interaction, and b) Restricted, repetitive patterns of behaviour, interests or activities.

2.1.1 ASD: Diagnosis and Prevalence

The latest edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) by the American Psychiatric Association (2013) includes disorders such as early infantile autism, childhood autism, Kanner’s autism, high-functioning autism, atypical autism, pervasive developmental disorder not otherwise specified (PDD-NOS), childhood disintegrative disorder, and Asperger’s disorder under the umbrella of autism spectrum disorder. Although the essential features of ASD remain as listed in the preceding paragraph and symptoms emerge at an early age, these features diversely manifest themselves depending on severity, developmental level, and chronological age.

Social and communication impairments result in deficits in verbal and non-verbal communication, social-emotional reciprocity, and developing, maintaining and understanding relationships (Kanner et al., 1943; Caronna et al., 2008; American Psychiatric Association, 2013). These impairments are pervasive and sustained. Verbal communication is affected in many individuals, ranging from complete lack of speech, poor comprehension, or echoed speech, and even when formal language skills such as vocabulary and grammar are intact, inability to use language for reciprocal social communication is observed.

Nonverbal communication impairments are seen in the form of absent, reduced, or atypical use of eye contact, gestures, facial expressions, body orientation, or speech intonation. Impaired joint attention is one of the prevalent features of ASD, expressed by lack of pointing, showing, bringing objects to share an interest with
others, or failure to follow someone’s gaze or pointing. Integration of aforementioned aspects and non-verbal and verbal communication is noticeably poor, even though individual aspects may be subtle. Delayed language development along with a lack of social interest or unusual social interaction (non-verbal communication with no eye contact) are often the first symptoms of ASD.

Social-emotional reciprocity is the ability to engage and share thoughts and feelings with others. Young children with ASD show deficits in these aspects, ranging from little or no initiation of interaction or sharing of emotions. Accompanied with the inability to imitate others’ behaviour, the person experiences difficulty in processing and responding to complex social cues.

Deficits in developing, maintaining and understanding relationships are manifested as rejection of others, passivity, or inappropriate approaches that seem aggressive or disruptive. Young children in whom these deficits are particularly palpable, show lack of shared social play and imagination, insist on fixed rules of game, and prefer solitary activities.

Restricted and repetitive patterns of behaviour, interest or activities are also a prominent characteristic of ASD (Watt et al., 2008; American Psychiatric Association, 2013). These features are manifested as simple motor actions (e.g., hand flapping, finger flicking), repetitive use of objects (e.g., spinning coins, lining up toys), and repetitive speech (e.g., echolalia, delayed or immediate chanting of heard words, use of “you” while referring to self). Excessive adherence to routine may result in resistance to change, rigidity in thinking, highly restricted and fixated interests. In addition, hyper- or hypo-reactivity to sensory input may be related to some interests and routines, expressed by extreme responses to specific sounds or textures or tastes, excessive smelling or touching of objects, fascination with lights, and sometimes apparent indifference to pain, heat, or cold.

The DSM states that for the diagnosis of ASD the criteria that must be met are: a) Symptoms must be present in the early developmental period, b) Symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning, c) These disturbances are not better explained by intellectual disability or global developmental delay while being aware that intellectual disability and ASD frequently co-occur.
The symptoms of ASD are typically noticeable at the age between 12 and 24 months, based on the severity, earlier or later than the specified range. Although symptoms of ASD emerge at early childhood and early school years, with therapeutic intervention developmental gains are observed in later childhood in at least some areas. ASD is not a degenerative disorder and thus learning and compensation continue throughout life. Although most individuals improve, only a minority are able to live and work independently in adulthood. Those individuals on the higher end of the spectrum (high-functioning) and with lower levels of impairment are able to find an occupation that suits their interest and skill; those who do are still susceptible to vulnerability and remain socially naive and are often prone to anxiety and depression. These differences in a person diagnosed with ASD renders them more dependent on others, making everyday life which is effortless for a non-ASD person, encumbering for a person diagnosed with ASD (Mottron, 2011).

Since the time Kanner first coined the term Autism for a set of behavioural patterns and mental conditions (Kanner et al., 1943), research in demographics shows that the count of children being diagnosed with autism has increased tremendously over the years causing a fear among the public regarding the possibility of an 'epidemic' (Weintraub, 2011). A study shows that the diagnosis of autism has increased from about 1 in 5000 in the year 1975 to about 1 in 110 in 2009 (Weintraub, 2011). There is evidence of a density as high as about 1 in 100 in localised studies in UK in the year of 2006 (Baird et al., 2006). Though a prevalence of 1 in 150 has been confidently derived within some states in the US, it is believed that even these high rates are only average estimates which could be an underestimate of the true prevalence of ASD (Caronna et al., 2008; Fombonne, 2009). ASD is diagnosed four times more often in males than in females (Williams et al., 2008; Bartley, 2006). Research also shows evidence than when diagnosed in females, ASD is manifested with higher severity and accompanied by intellectual disability (Dworzynski et al., 2012).

The risk factors and causes for ASD is an ongoing research study. Studies narrow down on environmental and genetic factors as the main causes. A variety of nonspecific risk factors such as advanced parental age, low birth weight, or fetal exposure to certain chemicals may be listed under the environmental causes. Genetic research suggests a heritability for ASD ranging from 37% to higher than 90% based on a study on twins. Scientists are still struggling to explain the causes of autism and
2.1. Autism Spectrum Disorder: An overview

the exponential increase in the diagnosis, the main speculations being attributed to shifted diagnostic accreditation and heightened awareness (Rutter, 2011; Weintraub, 2011).

2.1.2 Early Intervention

In the early stage of development, interaction between children's brain and their environment plays an important role in development of social circuitry and language processing circuitry in the brain which help the child to support more complex behaviours such as social intentional communications, imitation and other behaviours that are typically impaired in children diagnosed with ASD (Dawson et al., 2005b,a, 2008; Kuhl, 2007). Dawson et al. (2005b; 2005a) stress that early parent/caretaker-child interaction influence the development of a normal social brain, continuing onto which Kulh (2007) hypothesises that the development of language is initiated through the social brain. An intensive behavioural intervention at an early developmental period has resulted in increased positive outcomes, signifying the importance of diagnosis and starting therapy as early as 2 - 4 years of age (Lovaas, 1987; Dawson et al., 2008; Harris and Handleman, 2000; Green et al., 2002). Over 25 hours of weekly intensive training was given to the experimental group in the study conducted by Lovaas (Lovaas, 1987).

2.1.3 Types of Intervention

There are a variety of approaches towards intervention, two of the most popular are behavioural, developmental, and combined interventions. Ospina et al. (2008) review many interventions that fall under either of the category or a combination of both mentioned approaches. Each of the interventions targets specific deficits. While behavioural intervention works on teaching new behaviours and skills using a structured set of rules and techniques, developmental interventions focus on building a positive attitude, meaningful relationships with people and skills, which they can use in their daily lives and primarily constitutes teaching socio-communication skills. Combined interventions in turn combine both the approaches and use developing meaningful relationships via interactions for delivering behavioural methods and in
turn use these relationships to deliver intervention more effectively. A list of all the established practices is documented by the National Autism Centre (NAC, 2009).

2.1.3.1 Behavioural Interventions

Applied behaviour Analysis (ABA) which is also known as an Early Intensive Behavioural Intervention (EIBI), is one of the established behavioural intervention techniques (NAC, 2009). Discrete Trail Training (DTT) is one of the most studied intervention technique which is derived from ABA. DTT involves breaking down the skill teaching process into its most basic steps as giving a cue, prompting the child to give a response, giving feedback or declaring the consequence, and providing an inter-trial break (Smith, 2001). The Lovaas program (Lovaas, 1987), Pivotal Response Training (PRT) (Koegel et al., 1999) and Positive Behaviour Support (PBS) (Horner, 2000) are other well known behavioural intervention methods. These therapies can be administered by therapists as well as family members, but because ABA heavily relies on paper-based materials, it translates into significant preparation costs, suffers in explaining dynamic concepts based on actions and requires rigorous manual tracking of the outcomes, which when implemented practically cost a lot in terms of time. Behavioural interventions are the most popular methods with enormous research conducted and established results showing that they work. Applications based on these methods are also the focus of our research and in the next section, we will revisit this topic to understand why.

2.1.3.2 Developmental Interventions

The Developmental, Individual Difference, Relationship-based model (DIR) is a popular developmental intervention technique. Greenspan states that DIR is not a single rigid intervention program, it is rather a flexible program that works on understanding the uniqueness of each child and allows the clinicians to dynamical incorporate behavioural or relationship-based approaches into a broader DIR model (Greenspan and Wieder, 2006). The DIR program is motivated by the insights that language, cognitive, emotional and social skills are all learned through meaningful emotional relationships and interactions; all children are different in the way they process sens-
ory and motor signals; and progress in all areas in interconnected. Thus it attempts to teach the children skills by building emotional connections and completing loops of communication between the child and the therapist or the parent. DIR/Floortime (Greenspan and Wieder, 2006) and responsive teaching (Mahoney et al., 2006) are other practices in this category. Research suggests that a high instruction rate of about 15 hours per week at home is required to achieve promising results (Pajareya and Nopmaneejumruslers, 2011).

2.1.3.3 Combined Interventions

SCERTS (Prizant et al., 2006) is an example of a program that is influenced by the existing methods of therapy and yet stands out for its smooth integration of practices belonging to varied categories as behavioural and developmental. SCERTS Model is motivated by prioritising three core deficits that are observed in ASD: social - communication, emotional regulation and transactional support. By addressing the complex inter-dependencies between these deficits the SCERTS model forms a comprehensive evidence-based practice that supports the development of the child as well as their families/caregivers. According to the SCERTS model, the service providers assess the current disposition of the child with respect to social communication and emotional regulation and along with their caregivers develop a profile of the child documenting the child’s capabilities and later they develop a program with a list of goals which are customised according to the child’s needs and the family’s requirements. This is an essential component which builds a model that serves as a positive change that not only helps in enhancing the child’s capabilities in overcoming the deficits but also help the parents by reducing confusion and stress. Early Start Denver Model (Princiotta and Goldstein, 2013) and TEACH (Van Bourgondien and Coonrod, 2013) are other well known combined interventions. Although these methods require intensive training of on an average 15 hours per week, there is evidence suggesting that these methods show promising results (Dawson et al., 2010).
## 2.1.4 Evidence-Based Practice and ASD

The formulation of evidence-based practice (EBP) is borrowed from the field of medicine where it first emerged and was identified as evidence-based medicine (EBM). The research in the domain of ASD is conducted by researchers from diverse disciplines such as medicine, psychology and education. The community of stakeholders who consume the practices is also diverse including the therapists, caregivers, parents and the individuals with ASD. The definition of EBP was thus expanded and interpreted in multiple ways across these disciplines (Reichow and Volkmar, 2011). The concerns within the psychology community regarding a narrow definition of EBP and restricted access to care and treatment choice, led the American Psychology Association (APA) to build a team of scientists and psychologists called APA Presidential Task Force on Evidence-Based Practice in 2005. With the goal of improving quality, cost-effectiveness and accountability, the Task Force team re-defined evidence-based practice in psychology (EBPP) as: “the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences.” (Force, 2006)

We enlist the following future research emphasis suggested by the Task Force (Force, 2006) that motivate our research. We believe these factors can be positively influenced by using computer-based programs as assistive technology in delivering interventions:

1. Moderating treatment response based on patient characteristics in natural settings.

2. Providing clinicians with real-time patient feedback for monitoring progress, and support tools to adjust intervention accordingly.

3. Developing quantitative measures for measuring therapeutic progress, accessing therapeutic process and quantify diagnostic judgements.


5. Monitor the effect of feedback regarding progress in therapy to psychologist or patient.
Volkmar et al. (2011) identify Outcome Measures as the area where a major impact can be made with state-of-the-art technology increasingly being available. Especially, digitally operated hand-held devices such as smartphones and tablets offer new possibilities for data collection and can be easily transported. Volkmar et al. also emphasise the data recorded from these devices may impact our current knowledge of the ASD condition.

The capacity of making decisions based on the data and accurately implementing the intervention by integrating research developments is an essential factor that promotes evidence-based practices (NAC, 2011). Evidence-based practices are central to ASD interventions that allow researchers to assess if the given intervention is effective for a particular child, and if the intervention follows set guidelines for scientific methodology. The data collected from applications has been used to establish if the given interventions are evidence-based practices (Whalen et al., 2010; Charlop-Christy et al., 2002; Mesibov and Shea, 2010; Odom et al., 2010). But research on analysing data to obtain a deeper understanding of ASD is still limited. Such insights can potentially be used to make suitable changes to the intervention program, which will improve the efficacy of intervention for individual children with ASD.

2.2 Assistive Technology and ASD

The introduction of technology in pervasive healthcare has significantly improved the efficacy and convenience in recording and processing data. For the aid of ASD, technology has targeted four main areas: early detection, assistive tools for affect recognition, tracking and measurement tools for progress, and social skills improvement. By studying and following the trends in the field of autism and developmental disorders, the gaps and avenues left unexplored become evident and allow us to situate our research in this domain. In the following sub-sections we describe in brief some of the technological aids that have been researched and implemented in mainly assistive technology that supports delivering intervention and help tracking and measurement of progress. We would like to note that early detection is beyond the scope of our research.
2.2.1 Research trends in Assistive Technology for ASD Intervention

As mentioned in the previous section, the intervention techniques require a lot of manual assistance in data tracking and analysis. In order to address these issues Kienzt et al. (2007) explored many dimensions and came up with relevant projects including Abaris and CareLog. Abaris is a tool to support discrete trial training and has helped therapists to improve the efficiency of data collection, reduce errors and plot graphs automatically, which can potentially improve therapy delivery. Additionally it has proved beneficial for sharing information during group meetings between different therapists working together as well as with parents. Functional behaviour assessment is one of the methods used to determine the causes of certain inappropriate and harmful behaviours that children with autism are likely to exhibit. This involves observing the antecedent and consequence events of the behaviour of interest. But in natural settings tracking the incidents that happened in the past is a tedious mental process with a high possibility of reduced accuracy in implementation. This is where CareLog (Kientz et al., 2007) comes into picture; it provides a recording system with audio as well as video buffering such that once the caregiver hits the wireless record button it provides a selective archive of media after the events have occurred automatically. Hirano et al. (2010) developed and tested the efficacy of implementing a visual scheduler system vSked in a classroom for children with autism. When tested against a paper-based visual scheduler system vSked not only reduced the effort of the staff in the classroom, but also increased the quality of social interactions between the children.

Michaud and Theberge-Turmel (2002) investigated the use of mobile robotic toys to teach social interaction to children with ASD. The main goal was to draw the child’s attention and provide various stimuli for the child to interact with, with the long term goal of a more personalised robotic toy that caters to individual child’s strengths and weaknesses through a personalised approach. A machine can provide stimulus and interact with children with ASD in a controlled manner resulting in higher predictability, thus causing the child to be less intimidated by the environment. A preference for robotic toys is also reflected in the research conducted by Stanton et al. (2008), where children with autism were less likely to engage in typical ‘autistic’ behaviours when interacting with a robotic dog AIBO.
Virtual reality is another domain of interest in assistive technology for delivering therapy to children with ASD. Similar to robotic toys, features such as controllable stimuli, and the preference shown by children with ASD toward computer interaction make virtual reality desirable. Furthermore, virtual reality also provides ease of generalisation, adaptation and personalisation, absence of distracting cues, safer learning environment and possible use of head and body trackers to monitor physical activity and provide feedback to the machine to adapt according to patient’s reactions (Strickland, 1997; Parsons and Mitchell, 2002). Virtual reality goes a step further from computer programs in generating three-dimensional virtual environments where the user is allowed to interact (using devices such as a head-gear, mouse or a joystick) and cause change in the stimuli, thus simulating near real life experiences. Research using virtual reality to train children with ASD to recognise and track an object showed that once the hardware such as the head-gear was accepted the children immediately immersed themselves in the environment and showed encouraging results (Strickland, 1997; Strickland et al., 1996). The use of virtual reality to teach creative activity social skills has also been explored (Parés et al., 2005; Schmidt and Schmidt, 2008). The research in this field is still ongoing with many open queries to address (Parsons and Cobb, 2011).

The wireless sensors by Kienz et al. (2007) were an initiative in developing tools for affect recognition. The wireless sensors that can be worn by children with autism, are capable of detecting self-stimulatory behaviour and intimating the parent or the therapist. This project investigated means for children with autism to be able to communicate their state of mind to their caregivers in settings where they find themselves unable to vocalise their needs.

### 2.2.2 Computer-Based Intervention

In a research survey conducted to understand what persons with autism and their caregivers desire in terms of software and technology (Putnam and Chong, 2008), three domains were identified: social/communication, academic and organisation. The study established that there was a positive attitude and interest toward technology and it was reported as a major strength.
Computer assisted programs for therapy are helpful to reduce the distractions by providing a controlled environment and reducing inhibitions associated with social contact for students (Moore et al., 2000). They also reduce the effort required to collect material for parents and therapists when compared to paper-based methods. The repetitive nature of one-on-one instruction delivery model crucial for teaching children with ASD can be efficiently delivered on a computer platform (as a computer is more patient than a human being) using audio-visual prompts and reinforcements (Bosseler and Massaro, 2003; Williams et al., 2002). Computer-based interventions are also capable of engaging children with ASD on the device for longer as the children show less resistance towards using them (Hetzroni and Tannous, 2004; Williams et al., 2002). Furthermore, they allow data to be recorded effortlessly and with more accuracy, thus assisting the therapist in analysis and administration of the intervention.

Research suggests that children with autism prefer video presentation modes that can be easily integrated into computer assisted technologies, over live demonstration for learning dynamic concepts and daily living skills (Cardon and Azuma, 2012; Shipley-Benamou et al., 2002). Augmentative and alternative communication (AAC) involve strategies that augment the needs of persons with limited natural speech when the current speech does not meet functional communication needs and provide alternate strategies for persons with no speech. Although earliest uses of AAC were non-tech, mainly relying on manual signs (Carr et al., 1978), in recent times due to increase in the usage of hand-held devices with dynamic displays and advanced quality synthesised voice output, using technology for AAC has become efficient for implementation and convenient for usage (Shane et al., 2012; Sennott and Bowker, 2009). Along with assisting with social communication, AAC interventions have shown an increase in speech generation with modest gains (Schlosser and Wendt, 2008). Picture Exchange Communication System (PECS) is an example of an AAC system; PECS relies on behavioural principles for teaching social communication using pictures and symbols (Bondy et al., 2001; Howlin et al., 2007). Proloque2Go is another example of an AAC system; it runs on an iPhone or iPod touch (Sennott and Bowker, 2009). Video modelling has shown to work for teaching vocabulary and language to children with ASD. The children learning language via this platform were not only able to learn and recall new vocabulary, but were also able to transfer the skills and reuse them in a natural, untrained environment.
2.2. Assistive Technology and ASD

(Bosseler and Massaro, 2003). Presenting audio-visual stimuli to deliver feedback for teaching language using HCI techniques also show promising results for teaching language (Hailpern et al., 2008, 2009).

Computer-based applications show high capability in terms of the data that can be recorded and harnessed through them. Such a dataset can hold key to sensitive information about the child that can be in turn used to improve the intervention for that child. Hence, opens up the avenue for automatic personalisation of syllabus for the child using the application. In the following paragraphs, we describe the factors that lead us to the choice of the computer-assisted program for our research project.

Applied Behaviour Analysis (ABA) is one of the popular methods that are used to administer therapy to children with ASD. ABA is the study of the socially significant human behaviour and changing them by varying the environment. ABA is being increasingly used in the area of developmental disabilities and has shown promising results in improving social, intellectual and verbal functionality of children on the autism spectrum (Foxx, 2008; Remington et al., 2007). Comprehensive ABA interventions have positively (medium to high) affected the intellectual functioning, language development and adaptive behaviour of children with ASD, with superior outcomes in language-related areas (Virués-Ortega, 2010). There are multiple ways of delivering therapy within the framework of ABA, for example: Discrete Trial Training (DTT), Early Intensive Behavioural Therapy/Intervention (EIBI), Pivotal Response Training (PRT) and Verbal Behaviour Intervention (VBI). As explained in the previous section, DTT is the one of the popular techniques that use ABA principles to breakdown a skill into small teachable units that are delivered repetitively till the skill is learnt. DTT techniques can be delivered by not only trained professionals but also parents and caregivers (Smith, 2001). These features of behavioural intervention make them most suitable for adaptation into computer-based programs. As they involve delivering small teaching units and recording rigorous data, computer-assisted programs that incorporate these principles are most suitable for our study.

The approach of computer based interventions towards the problem of delivering instruction for children with ASD is two-fold: to build applications that cater to specific skill deficits (e.g. language, social or organisational skills), or to build applications with a comprehensive syllabus. Anecdotal evidence suggests more applic-
2.2. Assistive Technology and ASD

Applications targeting specific skills. Examples of such applications are: *Proloquo2Go*, *iComm*, and *TapToTalk* for communication via symbols and text-to-voice; *Grace* for sentence building via images; *iCommunicate* for storyboarding with pictures and images; *First-Then-Visual-Schedule* for daily schedules; *AutismExpress* for emotion interpretation; and *Stories2Learn* for personalised social stories about complex social situations.

Two interesting application targeting the social communication area, which is one of the most common challenge in ASD, are MOSOCO and VidCoach. MOSOCO (Escobedo et al., 2012) is application that was implemented on an android mobile platform capable of teaching children with autism to practice social-skills required in real-life social situations. This application could increase social communication between children with autism and normal children by giving them enough practice to deal with appropriate situations and by reducing social mis-steps (Escobedo et al., 2012). VidCoach is another mobile application system, which guides youth with autism in peer as well as self-modeling, which is based on evidence showing video-modeling as an effective manner of teaching social skills to individuals with autism (Ulgado et al., 2013).

The modified *Pedagogy Wheel* \(^1\) by Cherrie Pickering shows how many diverse applications deliver instruction in different skill categories and have different learning traits. *TeachTown* \(^2\), *TOBY Playpad*\(^3\), and *DT Trainer*\(^4\) are examples of programs with comprehensive syllabus (Whalen et al., 2006; Ashton, 2001; Venkatesh et al., 2012). *TeachTown* and *DT Trainer* are software programs run on a computer, while TOBY Playpad is an iPad application. TeachTown and TOBY and perhaps similar and competitive in terms of the broad syllabus they offer; *TeachTown*, however, suffers from a restricted set of stimuli and impoverished response to syllabus.

Evidence supporting use of iPad and tablets for teaching skills to children with autism symptoms, has been increasing accumulating (Hourcade et al., 2013; Clark et al., 2014; Kagohara et al., 2013; Neely et al., 2013; Williams et al., 2002; Hetzroni and Tannous, 2004). With computers it is possible to retain many benefits of the virtual reality devices by providing safe spaces for exploration, as seen in

\(^1\)https://www.autismspeaks.org/sites/default/files/pedagogy_wheel.pdf
\(^2\)http://web.teachtown.com/home
\(^3\)http://tobyplaypad.com/
\(^4\)http://www.dttrainer.com/
section 2.2.1. Computer assisted programs are generally a good mode of delivering instruction as they provide enough of incentive for the children to develop quality social interaction skills and with the help of the data recording capabilities allow us to learn more about how children with ASD relate to the world (Hourcade et al., 2012). Parents and trained professionals have shown favourable attitude toward the use of these devices, although the usage by professionals was inconsistent (Clark et al., 2014). Kagohara et al. (2013) review the usage of iPods, iPads and iPhones for effective delivery of therapy for individuals with autism and believe that applications based on ABA-based instructional procedures implemented for these devices can be favourable technological aids.

Wainer and Ingersoll (2011) review the technological applications that assist therapy and report development areas for future programs. These include extending exclusively computer-assisted programs to include both computer-led and teacher-led interactions, expanding contingency management to deliver feedback and reinforcement based on user’s need, expanding instruction in language from receptive to expressive language, and maintaining learning environments that sustain the interest of the children with ASD. TOBY Playpad includes all these aspects and thus is state-of-the-art and preferred for our study.

2.3 An overview of applied machine learning in healthcare and ASD associated data

Machine Learning techniques have played an important role in identifying patterns, grouping similar objects together and modelling data. These techniques learn from the given data and are capable of predicting of outcomes for new datasets. In applications where the data is too large and cannot be manually analysed, applying machine learning techniques has helped in automation of many processes. In general, in the field of medicine and human psychology, recent examples of applied machine learning are to analyse medical data (Jensen et al., 2012b), understand the working of human mind in conceptual learning, language acquisition and building causal relations (Tenenbaum et al., 2011), and to model transfer learning that is exhibited by human learners (Canini et al., 2010).
In the field of general healthcare, the rise of computerisation has lead to electronic health records with large patient databases. Machine learning techniques such as decision trees and random forests, Bayesian networks, neural networks and Support Vector Machines (SVMs) have been leveraged recently to extract knowledge from these vast databases (Meyfroidt et al., 2009). Mining electronic health records also has a potential to reveal unknown disease correlations and when combined with genetic data has a potential of revealing genotype-phenotype relationships (Jensen et al., 2012a). Medical diagnosis is another area where application of machine learning has been is seen as promising (Kononenko, 2001).

Data mining has been applied in medical domain for improving health in several ways. Obenshain et al. (2004) describe the methods of data mining broadly that are motivated by healthcare data in particular. They present examples of data mining applications in healthcare including hospital infection control, ranking hospitals, and identifying high-risk patients. Within the same context, they also describe factor analysis, which is one of the data mining techniques, as specifically motivated by affinity analysis. It is a branch of unsupervised learning techniques that deals with identifying relationships between events occurring in conjunctions.

Within the field of ASD, Beykikhoshk et al. (2015b) have used machine learning models, specifically hierarchical Dirichlet process, to discover the emergence and development of knowledge in ASD using the literature corpus since the time the term Autism was coined. A novel methodology examining the linguistic and semantic aspects of data collected from twitter (Beykikhoshk et al., 2015a) shows potential to bridge the gap between public-policy decision makers and medical practitioners on one side and caregivers and individuals with ASD on the other. This encourages both parties to take informed and evidence-based decisions and makes communication of latest development in treatment guidelines easier. Although twitter is a powerful platform that is accessible to all, the posts on twitter are limited to 140 characters. An in-depth study regarding the mood and behaviour of caregivers and individuals with ASD and would require longer entries in the form of documents as referred to in natural language processing. Nguyen et al. (2014; 2015) conduct a study on such an online community by analysing data from Live Journal weblog communities, using latent Dirichlet process (LDA) and LIWC package, and comparing ASD community against others over sentiment, topic and language style. They found that topic and language styles were good at predicting posts relevant to autism and the sentiment
of the ASD community were poorer than the others. We notice that driven by the data collected from the user directly, research in the field of ASD shows potential to develop novel methodologies that bridge the knowledge gap between different communities in ASD.

Discovering subgroups within autism spectrum is an ongoing research problem. Research has been conducted to identify subgroups using cluster analysis based on the behaviour patterns and developmental history (Prior et al., 1998). Subgroups have also been discovered using hierarchical clustering approach, based on comorbidities like seizures, gastro-intestinal disorders and psychiatric illness, as these comorbidities have been observed to occur unevenly over the spectrum (Doshi-Velez et al., 2014). The phenome project aims at discovering subtypes within ASD based on biomedical and behavioural aspects (Singer, 2005). Factor analysis has been used to identifying characteristics that are unique to Social Anxiety Disorder and ASD for accurate diagnosis (White et al., 2012). But research in this domain is ongoing with novel approaches and challenges.

Outside the problem of ASD, comorbidity analysis using the approach of multinomial-logit generative model based on Indian buffet process (IBP) (Ghahramani and Griffiths, 2006) has been used for psychiatric disorders (Ruiz et al., 2014). Gupta et al. (2012) propose a Bayesian nonparametric framework that is a linear Poisson gamma model (LPGM) for count based data. This model, while emulating the non-negative matrix factorisation (NMF) by providing non-negative intelligible parts based representation, also overcomes the limitation of NMF by learning the number of factors automatically from the data. The model achieves this by using IBP as a prior that allows the model to grow infinitely as the data grows. Finding clusters of similar patients can help in delivering tailored healthcare plans effectively to each group. Mixed-variate restricted Boltzmann machines have been adapted for identifying subgroups of patients with a chronic diabetes condition (Nguyen et al., 2013).

Technology intervention in Autism is most popular in the areas of early diagnosis or assisting early intervention applications. Machine learning techniques have only been recently been applied to research in autism diagnosis that helps speed up the process (Wall et al., 2012). Support Vector Machines (SVM) one of the popular supervised machine learning techniques has been used to detect anatomical differences in the brain of children with ASD (Ecker et al., 2010; Jiao et al., 2010; Calderoni
et al., 2012). They have also been used to understand learning in the mind of a person with autism (Cohen, 1994) in the past. Although there has been an increase in technology-based interventional options available, they have been used limitedly to identify peculiar learning patterns of individual children with autism in which can eventually further the knowledge required for catering personalised syllabi. One of the main challenges in this area is the availability of large datasets. Traditional research in the area of clinical study involves recruiting children and conducting experimental studies at the research institution. The data collected via this channel is highly accurate, but is limited by the number of children that can be a part of the study, which is observed to be small sometimes less than 10 (Pennington, 2010; Ploog et al., 2013). Machine learning techniques have the potential to enable detection of such patterns when applied to the data collected from a computer-based intervention. But such data-based studies require a larger dataset to rely on, which was not easily recordable till recently.

2.4 Machine Learning Techniques

Based on our motivation the focus of this thesis lies in unsupervised learning techniques, especially the latent factor analysis and the restricted Boltzmann machines. Before delving into the details of why these techniques are best suited for our problem though, we describe the overview of machine learning techniques.

2.4.1 Overview of machine learning

The knowledge of a system in terms of the input and output may be complete or incomplete depending on the application. Based on this machine learning problems can be classified broadly into supervised or unsupervised learning techniques.

When the information of the output in relation to the input provided is available and the task of the algorithm is to learn this input-output relation and predict the output for new inputs. This problem may be solved by using supervised learning. The main focus of supervised learning techniques is accurate prediction. Popular techniques
2.4. Machine Learning Techniques

in supervised learning include linear regression, logistic regression, classification and support vector machines (SVM).

In the absence of the *ground-truth*, which shows which output is generated when a certain set of inputs are given to the system, the problem becomes one of unsupervised learning. The main focus of unsupervised learning techniques is to find compact and alternate forms of description of data. These techniques are an attempt to understand the data in terms of its probability distribution. This is generally achieved by modelling the data in terms of 'latent' variables which improves our understanding of the association between the variables that represent data. Unsupervised learning techniques address problems such as clustering and data representation by dimensionality reduction. As seen in Section 2.1, ASD has a spectrum nature which makes it difficult to classify children based on ground-truth labels. A more efficient way to describe the children would be in terms of a probabilistic combination of different features. Furthermore as seen in Section 2.3, there exists a trade-off between collecting data from a large number of subjects and from a clinical trial, where additional sensitive information about the child can be availed. Thus our interest for this thesis mainly lies in unsupervised machine learning techniques.

Another broad classification of machine learning algorithms is as parametric and nonparametric models. While parametric models are faster to run, they make strong assumptions about the data and have fixed number of parameters. On the other hand, nonparametric models allow a flexible number of parameters and allows the model complexity to grow as more data arrives. The number of parameters is often decided by a prior, which is a distribution on the parameters.

### 2.4.2 Clustering techniques

Clustering is discovering meaningful partitions of the data points in multidimensional space, based on a similarity measure such that objects within the same clusters are more similar to one another than those assigned to different clusters. Clustering is opted when the labels associated with the data are absent and it is an unsupervised learning technique. In fact clustering is the most popular unsupervised learning technique. Based on the structure of the clusters they can be categorised
as flat clustering (e.g. K-means, DBSCAN, Affinity Propagation, etc) or hierarchical clustering (e.g. agglomerative clustering). In the following paragraphs, we describe two techniques: K-means and affinity propagation that are particularly useful in our analysis.

**K-means clustering** This is one of the most popular clustering techniques in a parametric setting. K-means aims to partition the data points into $K$ clusters, by assigning each data-point to the cluster whose centroid is closest to the data point. Suppose we have $N$ observations and the data set is represented as $\{x_1, x_2, ..., x_N\}$, our aim is to partition this data into $K$ clusters. We introduce a set of centroids $\mu_k$, for all clusters $k = 1 : K$ such that $\mu_k$ is associated with $k^{th}$ cluster, and a set of indication vectors $z_{nk} \in \{1, 0\}$ for each data point $x_n$ such that $z_{nk} = 1$ if the data point $x_n$ is assigned to cluster $k$. We can then define an objective function as follows

\[
J = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} ||x_n - \mu_n||^2 \quad (2.4.1)
\]

As we can see, objective function represents the sum of squares of the distance between each data point and the centroid that it is assigned to. We can determine the values of the centroids and the indication vectors by minimising the objective function (Equation 2.4.1). This is an NP hard optimisation problem (Dasgupta, 2008) but the goal can be achieved using heuristic techniques efficiently. This is done iteratively in three steps. In the first step we choose some initial values for the centroids $u_k$. In the second step we minimise $J$ with respect to $z_{nk}$ by keeping $\mu_k$ as constant. This equates to choosing $z_{nk} = 1$ for the value of $k$ that gives minimum $||x_n - \mu_k||$. In other words, we assign the data point to the closest cluster centroid. In the third step, we minimise $J$ with respect to $\mu_k$, while keeping the indication vectors constant. Setting the derivative of the objective function with respect to $\mu_k$, to zero will lead us to $\mu_k = \frac{\sum z_{nk} x_n}{\sum z_{nk}}$. This can be interpreted as evaluating the new cluster centroid values as the mean of the data points assigned to cluster $k$ in step two. And thus the algorithm is named K-means and the steps 2 and 3 are iteratively repeated until further change in assignments is not observed or till a maximum number of iterations are reached. Figure 2.1 shows an example of how
2.4. Machine Learning Techniques

Figure 2.1: Illustration of functioning K-means algorithm on rescaled Old Faithful data set in four stages. The green data points in the Figure (a) represent the euclidean separation between them and cluster centroids are initialised shown as the blue cross and the red cross. Figure (b) shows the reassignment of the values for the indicator vectors where data points are assigned to the clusters whose centroids are closest. Figure (c) shows the subsequent re-evaluation of the centroid values based on the data points assigned to each cluster. Figure (d) shows the final partitioning of the data after the algorithm converges.

K-means algorithm assigns data points to clusters.

K-means requires us to have an intuition about the number of clusters $K$ in advance. Despite its simplicity and popularity, it is thus especially difficult to use K-means in settings where the number of clusters in the data cannot be pre-determined. K-means is also sensitive to the initialisation values and the distance metric used to compute the pairwise distance between the data point and the centroid of the cluster.
2.4. Machine Learning Techniques

Affinity Propagation  Affinity Propagation (AP) is a clustering algorithm that identifies representative examples called as exemplars and uses message passing between data points to cluster data. Unlike the K-means it does not need the number of clusters to be known or estimated in advance. AP takes on the input, which is a similarity matrix between the data points. The goal is to infer the exemplar nodes for each data point in the data set \( \{x_1, x_2, ..., x_n\} \) and then choose the most similar exemplar for the data point. To this effect the algorithm introduces two matrices: responsibility matrix \( R \) with elements \( r(n, k) \) and availability matrix \( A \) with elements \( a(n, k) \), initialises them to zero and iteratively updates them till most dominant data points emerge as an exemplar in each cluster. In the first step the responsibility \( r(n, k) \) that indicates the how well \( n \) trusts \( k \) as its exemplar is sent to the candidate exemplar. The second step involves the candidate exemplar sending its availability \( a(n, k) \) to the data point as a measure of how appropriate it would be for \( n \) to choose \( k \) as its exemplar. Gradually after these messages are passed around for some iterations, groups are formed and dominant data points emerge as the exemplars for each cluster. Further details of this algorithm can be found in (Frey and Dueck, 2007).

2.4.3 Feature allocation and latent factor analysis

The clustering techniques we described in the previous section were classical approaches to clustering that are most commonly used. However, they are parametric methods which involve some knowledge about the number of parameters in advance. Even in cases that this is not required the input to these algorithms is required to take a specific representation that is rigid. In real life, the data may not match these criteria, especially so when the data has a large number of variables and they are correlated. We thus move on to describe the unsupervised learning technique of factor analysis that has advantages when it comes to such data as we will see in the following sub-sections.
2.4.3.1 Factor analysis approach

Factor analysis is technically a dimensionality reduction tool, based on the theory that measurable and observed variables that are correlated can be represented using relatively independent unobserved variables known as factors or latent variables, by removing redundancy. The number of latent variables is usually smaller than the total number of observed variables. Factor analysis has many applications. The main fields these applications fall under are for example, evaluation validity of scores, theoretical formulation about nature of constructs, summarising relationships using more economic factors and testing hypothesis. Our main interest lies in the interdependency and pattern delineation, and dimensionality reduction facilitated by clustering of objects based on their latent factor scores.

Factor analysis attempts to model $V$ observed variables of $N$ subjects denoted as $(X_1, X_2, \ldots, X_V)$ of an object in terms of $K$ factors represented as $(H_1, H_2, \ldots, H_K)$. In classical factor analysis, the variable $X_v$ is represented as

$$X_v = W_{v1}H_1 + W_{v2}H_2 + \cdots + W_{vK}H_K + E_v$$

(2.4.2)

where $v = 1, 2, \ldots, V$, and $W_{v1}, W_{v2}, \ldots, W_{vK}$ are called as the factor loadings, and $E_v$ is the modelling error. Factor loadings are the representation of the data variables in the latent subspace. Thus in formal matrix notation, if $X \in \mathbb{R}^{V \times N}$ is the data matrix, $W \in \mathbb{R}^{V \times K}$ is the matrix of factor loadings, $H \in \mathbb{R}^{K \times N}$ is the factor matrix and $E \in \mathbb{R}^{V \times N}$ is the modelling error then we have,

$$X = WH + E$$

(2.4.3)

Each matrix decomposition allows four interpretations of the data: factor interpretation, geometric interpretation, component interpretation and graph interpretation. Of special interest to us are the factor, the graph and the component interpretation. The factor interpretation is an attempt to understand the data sample of the object as a result of mixing the underlying latent factors in different proportions. The component interpretation is based on the assumption that the observed variable values are a result of contribution from different processes and these processes
maybe separated. In the case of the graph interpretation, nodes associated to each object are joined with the nodes associated to the variables with edges. The weights of the edges are matrix entries and thus nodes that are similar to each other are connected by edges with similar weights (Skillicorn, 2007).

Two of the main techniques for matrix decomposition are principal component analysis (PCA) and non-negative matrix factorisation (NMF). Unlike traditional techniques of alternate data representation such as vector quantisation (VQ), PCA allows activation of multiple semantic variables in their factor loadings. However, due to a combination of positive and negative valued factor loadings, PCA renders difficult to interpret the meaning of the factors in terms of known variables. A negative valued factor loading must be interpreted as subtracting a component to construct the data matrix. When datasets involved have attributes like count of things (e.g. words), measured intensities or quantities (e.g. images) it not evident what negative occurrences of these attributes may be interpreted as. NMF was developed for cases as these, where it is inappropriate for the attributes to take on a negative value.

2.4.3.2 Non-negative matrix factorisation

Non-negative matrix factorisation (NMF) (Lee and Seung, 1999) is based on psychological and physiological evidence that part-based representation is very useful for object or pattern recognition. It forces non-negativity on its constraints so that eventually only additive combinations of the parts-based representation is used for object representation. Although text mining and bio-informatics are the most popular areas where NMF is used, it can be used in a wide range of applications where modelling the computation of observable visible variables from hidden variables is of particular interest.

The first NMF algorithm proposed was proposed by Lee and Seung (1999). Building upon the Equation 2.4.3, this algorithm tries to minimise the objective function subject to non-negative constraint on the matrices $W$ and $H$.

$$
F = \sum_{v=1}^{V} \sum_{n=1}^{N} \left[ X_{vn} \log(WH)_{vn} - (WH)_{vn} \right]
$$

(2.4.4)
Figure 2.2: Graphical model of non-negative matrix factorisation (NMF). $X_{vn}$ are the input data points, where $v = 1 : V$ are the number of variables, and $n = 1 : N$ are the number of observations. The model factorises the input into factors loadings $W_{vk}$ and latent factors $H_{kn}$, where $k = 1 : K$. (Courtesy (Lee and Seung, 1999))

The matrices $W$ and $H$ are initialised randomly and they are updated iteratively using the following update rules.

\[ W_{vk} \leftarrow W_{vk} \left( \sum_n \frac{X_{vn}}{(WH)_{vn}} H_{kn} \right) \]  \hspace{2cm} (2.4.5)

\[ W_{vk} \leftarrow \frac{W_{vk}}{\sum_i W_{ik}} \]  \hspace{2cm} (2.4.6)

\[ H_{kn} \leftarrow H_{kn} \left( \sum_v W_{vk} \frac{V_{vn}}{(WH)_{vn}} \right) \]  \hspace{2cm} (2.4.7)

The simplistic nature of NMF decomposition can be interpreted as a set of factors and an additive mixture of those factors based on their loading values to generate the observed data. Here the factors are the parts and the factors loadings are the proportion of the parts that are mixed.

One of the main challenges in NMF is determining the number of factors $K$. Generally, the rank $K$ of the factorisation is chosen such that

\[ K < \frac{vn}{v+n} \]

The specification of the number of factors a priori is a notoriously difficult task especially when the data is complex such as those generated in health and human behaviour domain. Hence we gravitate toward Bayesian nonparametric methods to find a solution to this problem.
Bayesian nonparametric is an area that has received a lot of attention in machine learning in recent years. It is an attempt to tackle the automatic model selection problem in statistics and machine learning, which has led to computational methods and toolkits that are widely applicable to different kinds of data. Traditionally, a modeller had to specify the statistical model a priori (e.g., the number of clusters, the dimension of the parameter space); this is a notoriously difficult task especially in dealing with complex data such as those that arise in health human behaviour domain. In particular, in our case, it is close to impossible to try a parametric method since the patterns of learning in children with ASD are complex and unknown in the autism literature. In particular, since we intend the data representation to admit a feature-based format, we have used nonparametric factor analysis model. Hence it is more suitable for factor analysis than traditional clustering methods (such as K-means). Factor analysis yields meaningful latent factors, which allows us to study the subtopics. In the following sub-sections we describe the evolution of a Bayesian nonparametric factorisation model from a parametric one.

### 2.4.3.3 Bayesian matrix factorisation

We begin by explaining the data generative process behind the Bayesian matrix factorisation model.

**The data generative process** The graphical model as shown in the Figure 2.3 describes the generative process whereby an observed data element $x_{vn}$, is a noise-corrupted version of the true data $\tilde{x}_{vn}$. There are in all $N$ data samples and $V$ observable variables, thus $x_{vn}$ is a part of data matrix $X \in \mathbb{R}^{V \times N}$. This data is generated from $K$ factors (also called as features), represented by the columns in the matrix $W \in \mathbb{R}^{V \times K}$. We then have an allocation matrix $Z \in \mathbb{R}^{K \times N}$ and $z_{kn} \in \{1, 0\}$; each column of $Z$ is thus a vector of $K$ elements where each element $z_{kn}$ indicates if feature element $w_{vk}$ contributes in the generation of the observed data element $x_{vn}$. The nature and the combination of these $K$ factors with the elements of the allocation matrix $Z$ is application specific.

Suppose the features are generated from a distribution $\phi$ with hyper-parameters $\beta$, and the elements of the allocation matrix are drawn from a distribution $\pi$ with hyper-
Figure 2.3: Graphical model for Bayesian matrix factorisation. The input data matrix $X_{VN}$ is represented as generated from $K$ factors represented by the factor matrix $W_{VK}$. $Z_{KN}$ is the allocation matrix, where $Z_{kn} \in \{1, 0\}$, which indicates whether feature element $W_{vk}$ plays a part in generating observed data point $X_{vn}$. The factors $W_{VK}$ are generated from a distribution $\phi$ with hyper-parameter $\beta$, while the indicators $Z_{KN}$ are drawn from a distribution $\pi$, with hyper-parameter $\alpha$.

parameters $\alpha$, the posterior distributions $p(z_{kn} \mid z_{k,-n}, W, X, \alpha)$ and $p(w_{vk} \mid Z, X, \beta)$ of this model are given by

\begin{align}
  p(z_{kn} \mid z_{k,-n}, W, X, \alpha) &\propto p(z_{kn} \mid z_{k,-n}, \alpha)p(x_{\cdot n} \mid Z_{kn}, W, \alpha, \beta) \\
p(w_{vk} \mid Z, X, \beta) &\propto p(w_{vk} \mid \beta)p(x_{\cdot v} \mid Z, w_{vk}, \alpha, \beta)
\end{align}  

(2.4.8) (2.4.9)

where the distribution $p(X \mid Z, W, \alpha, \beta)$ is the likelihood, and $p(z_{kn} \mid Z_{k,-n}, \alpha)$ and $p(w_{vk} \mid \beta)$ are the prior distributions. Based on the application, the distributions for the priors and the likelihood are selected before inferring the posterior values. The prior for the distribution over the feature matrix is derived as a conjugate prior of the likelihood function as explained in the next paragraph, while the prior for the distribution over the collocation matrix is explained in thereafter.

**Conjugate priors** As seen from the maximum likelihood setting, the posterior distribution is directly proportional to the product of the prior and the likelihood function. It is then possible to seek a prior distribution, such that posterior distribution takes the same form as the prior distribution. Such a prior is called as a
conjugate prior to the likelihood function (Bishop, 2006).

Any exponential family distribution, say \( p(x | \eta) \), where \( x_{i=1:N} \) are the data elements and \( \eta \) is the parameter can be represented as

\[
p(x_i | \eta) = h(x_i) \exp\{\eta^T t(x_i) - a(\eta)\}
\]

where \( t(x) \) is the sufficient statistic, \( h(x) \) is the underlying measure and \( a(\eta) \) is the log normaliser given as

\[
a(\eta) = \log \int h(x) \exp\{\eta^T t(x)\}
\]

(2.4.11)

For every exponential family member there exists a conjugate prior given as

\[
p(\eta | \zeta) = h_c(x_i) \exp\{\zeta_1^T \eta + \zeta_2^T (-a(\eta)) - a_c(\zeta)\}
\]

where \( \zeta \) is the natural parameter and \( \text{dim}(\zeta) = \text{dim}(\eta) + 1 \), and the sufficient statistics are \( \eta \) and \(-a(\eta)\). And the posterior takes on the same form as the prior and can be calculated as

\[
p(\eta | x_{i=1:N}, \zeta) \propto p(\eta | \zeta) \prod_{i=1}^{N} p(x_i | \eta)
\]

\[
\propto h(\eta) \exp\{(\zeta_1 + \sum_{i=1}^{N} t(x_i))^T \eta + (\zeta_2 + N)(-a(\eta))\}
\]

(2.4.13)

where the hyper-parameters are \( \tilde{\zeta}_1 = \zeta_1 + \sum t(x_i) \) and \( \tilde{\zeta}_2 = \zeta_2 + N \)

Table 2.1 lists some of the popular pairs of likelihood distributions and conjugate prior.

2.4.3.4 Finite Bayesian matrix factorisation approach for computing the allocation matrix

Consider a stochastic process with \( N \) objects and \( K \) factors. Then each outcome follows a factor with a probability of \( \pi_k \) which is indicated by an allocation matrix \( Z \), where element \( z_{kn} \in \{1, 0\}^{K \times N} \), and the factors are represented as \( W \in \mathbb{R}^{V \times K} \).
2.4. Machine Learning Techniques

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Prior</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X \sim \text{Binomial}(n, p) )</td>
<td>0 \leq p \leq 1</td>
<td>Beta(( \alpha, \beta )) ( \alpha &gt; 0, \beta &gt; 0 )</td>
<td>Beta(( \alpha', \beta' )) ( \alpha' = \alpha + x ), ( \beta' = \beta + n - x )</td>
</tr>
<tr>
<td>( X_i \overset{IID}{\sim} \text{Poisson}(\lambda) )</td>
<td>( \lambda &gt; 0 )</td>
<td>Gamma(( \alpha, \beta )) ( \alpha &gt; 0, \beta &gt; 0 )</td>
<td>Gamma(( \alpha', \beta' )) ( \alpha' = \alpha + n\bar{x} ), ( \beta' = \beta + n )</td>
</tr>
<tr>
<td>( X_i \overset{IID}{\sim} \text{Exponential}(\lambda) )</td>
<td>( \lambda &gt; 0 )</td>
<td>Gamma(( \alpha, \beta )) ( \alpha &gt; 0, \beta &gt; 0 )</td>
<td>Gamma(( \alpha', \beta' )) ( \alpha' = \alpha + n ), ( \beta' = \beta + n\bar{x} )</td>
</tr>
<tr>
<td>( X \sim \text{Multinomial}(p_{i=1:k}) )</td>
<td>( \sum p_i = 1 )</td>
<td>Dirichlet(( \alpha_{i=1:k} )) ( \alpha_i &gt; 0 )</td>
<td>Dirichlet(( \alpha'_{i=1:k} )) ( \alpha'<em>i = \alpha_i + \sum^n</em>{j=1} x_j )</td>
</tr>
</tbody>
</table>

Table 2.1: A table of some popular conjugate priors and likelihoods. Based on the application the most appropriate pair is chosen.

Thus the parameter vector given by \( \pi = [\pi_1, \pi_2, ..., \pi_K]^T \), where the probability \( \pi_k = p(z_{kn} = 1) \). Given the parameters, the factors are independently drawn as

\[
p(Z | \pi) = \prod_{n=1}^{N} \prod_{k=1}^{K} p(z_{kn} | \pi) = \prod_{k=1}^{K} \prod_{n=1}^{N} \pi_k^{z_{kn}} (1 - \pi_k)^{1-z_{kn}}
\]

\[
= \prod_{k=1}^{K} \pi_k^{c_k} (1 - \pi_k)^{N-c_k} \tag{2.4.14}
\]

where \( c_k = \sum_{n=1}^{N} z_{kn} \) is the total number of objects having \( k^{th} \) feature.

We further assume that \( \pi_k \) are drawn from a Beta prior. It is feasible to use to hyper-parameters \( \gamma \) and \( \beta \), but we simplify things by setting the hyper-parameters as \( \gamma = \frac{\alpha}{K} \) and \( \beta = 1 \), thus the distribution is given by

\[
p(\pi_k) = \frac{\alpha}{K} \pi_k^{\frac{\alpha}{K} - 1} \tag{2.4.15}
\]

The probability model is thus defined as

\[
\pi_k \sim \text{Beta}(\frac{\alpha}{K}, 1) \tag{2.4.16}
\]

\[
z_{kn} \sim \text{Bernoulli}(\pi_k) \tag{2.4.17}
\]
The posterior is given as
\[
p(\pi_k \mid Z) = \frac{p(Z \mid \pi_k)p(\pi_k)}{p(Z)} \propto p(z_k \mid \pi_k)p(\pi_k)
\]
\[\propto \pi_k^{c_k + \alpha K - 1} (1 - \pi_k)^{N - c_k} \tag{2.4.18}\]

The marginal probability distribution is given by:
\[
p(Z \mid \alpha) = \int p(Z \mid \pi)p(\pi \mid \alpha)d\pi = \prod_{k=1}^{K} \frac{\alpha K \Gamma(\frac{\alpha}{K} + c_k)\Gamma(N - c_k + 1)}{\Gamma(\frac{\alpha}{K} + N + 1)} \tag{2.4.19}\]

A parametric solution to this equation can be obtained by using Markov chain Monte Carlo (MCMC) methods like Gibbs sampling to infer the elements of the allocation matrix. This involves the computation of conditional probability of an element \(z_{kn}\) given all others as:
\[
p(z_{kn} \mid Z_{k,-n}, \alpha) = \frac{p(z_{kn}, Z_{k,-n} \mid \alpha)}{p(Z_{k,-n} \mid \alpha)} = \frac{\Gamma(\frac{\alpha}{K} + c_k)\Gamma(N - c_k + 1)\Gamma(\frac{\alpha}{K} + m_k)\Gamma(N - m_k)}{\Gamma(\frac{\alpha}{K} + N + 1)} \tag{2.4.20}\]

where \(c_k = \sum_{i \neq n} z_{ki} + z_{kn} = m_k + z_{kn}\)

So when \(z_{kn} = 1\), and when \(z_{kn} = 0\),
\[
p(z_{kn} \mid Z_{k,-n}, \alpha) = \frac{\left(\frac{\alpha}{K} + m_k\right)}{\left(\frac{\alpha}{K} + N\right)} \quad \quad p(z_{kn} \mid Z_{k,-n}, \alpha) = \frac{(N - m_k)}{\left(\frac{\alpha}{K} + N\right)}
\]

This model follows conjugacy between the beta and binomial distributions and as can be seen from the equations. The matrix \(Z\) is exchangeable, the probability of every instance depending only on the number of objects taking each feature \(c_k\).

The marginal probability as given in Equation 2.4.19 is dependent on the value of \(K\) which is unknown. When the data are complex and highly correlated, such as those recorded in health, the cardinality of latent features \(K\) is often unknown. Furthermore, such models need to grow in complexity as the data in the system grows, which is not possible in fixed model parametric setting. A solution to this problem can be achieved by adopting a nonparametric approach.
2.4.3.5 Infinite Bayesian matrix factorisation

In theory, infinite statistical models can be built by defining priors over an infinite factor allocation matrix, that is by allowing infinite rows in $Z$. These priors can then be derived by taking a limit of priors for finite models. In order for convenience of understanding, let us say that $Z \in \mathbb{R}^{[N \times K]}$, where the rows are the objects and the columns are the number of features (as opposed to $Z \in \mathbb{R}^{K \times N}$ considered till now. This is only for the present section following which we will resume with our original configuration of the allocation matrix $Z$). We need to take an infinite limit with respect to $K$ on the $P(Z)$. This is done by first mapping the binary matrices $Z$ to left-ordered binary matrices using a function $lof(\bullet)$, followed by defining $lof$-equivalence class for binary matrices. This gives the number of equivalent matrices in each class as $\prod_{h=0}^{K} K_h!$, where $K_h$ is the number of features that are taken on by $h$ objects. Hence, $K_0$ is the number of features for which $c_k = 0$ (inactive features), and $K_+ = \sum_{h=1}^{2N-1} K_h$ is the number of features for which $c_k > 0$ (active features).

Now taking the limit $K \to \infty$ on the left ordered equivalence class $[Z]$ we have,

$$
\lim_{K \to \infty} P([Z] | \alpha) = \lim_{K \to \infty} \frac{K!}{\prod_{h=0}^{2N-1} K_h!} \frac{\prod_{k=1}^{K} \alpha \Gamma(\frac{\alpha}{K} + c_k) \Gamma(N - c_k + 1)}{\Gamma(\frac{\alpha}{K} + N + 1)}
$$

$$
= \frac{\alpha^{K_+}}{\prod_{h=0}^{2N-1} K_h!} \exp\{-\alpha H_N\} \prod_{k=1}^{K_+} \frac{(N - c_k)!(c_k - 1)!}{N!} \tag{2.4.21}
$$

where $H_N = \sum_{j=1}^{N} \frac{1}{j}$ is the $N$th harmonic number. This distribution is once again exchangeable, unaffected by ordering of objects.

2.4.3.6 The Indian buffet process

The probability distribution defined in Equation 2.4.21 can be derived from a simple stochastic process named as Indian buffet process (IBP) (Ghahramani and Griffiths, 2006) inspired from the Chinese Restaurant Process (CRP). Consider an Indian restaurant in London that offers an infinite number of dishes in the lunch-time buffet. Let us say the customers represent the objects, the dishes represent the features,
Figure 2.4: Indian buffet process (IBP) shown as a matrix of customers on the rows and dishes on the columns. The 1st customer chooses the first three dishes, the second customer chooses the 1st, 3rd and 4th dishes and so on. The first 5 dishes are already sampled by customers. When the 5th \((n + 1 = 5)\) customer arrives, he samples the first five dishes based on their popularity with a probability of \(c_k/n\). Thus he samples the first dish with the probability of \(3/4\), second with a probability of \(2/4\) and so on. Then he goes on to sample Poisson\((\alpha/n)\) new dishes.

and each customer is allowed to take more than one dish, the IBP distribution is defined over infinite binary matrices \((Z \in \mathbb{R}^{[K \times N]})\) describing the ways in which the customers choose the dishes.

Let us say that \(N\) customers come one after the other and the infinite dishes are arranged in a line. The first customer starts taking a serving from each dish till her plate becomes full after Poisson\((\alpha)\) number of dishes. As the customers keep building up the popularity of the dishes in terms of \(c_k\), the number of customers who sampled dish \(k\), is updated. Hence, the \(n\)th customer now samples dishes based on their popularity, serving himself with a probability of \(c_k/n\) and once he reaches the end of previously sampled dishes takes a serving of Poisson\((\alpha/n)\) number of new dishes. Thus within the allocation matrix \(Z\), the element \(z_{kn} = 1\) if \(n\)th customer sampled \(k\)th dish. The probability of a matrix generated by this process is as follows.

\[
P(Z \mid \alpha) = \frac{\alpha^{K_+}}{\prod_{i=1}^N K_i!} \exp\{-\alpha H_N\} \prod_{k=1}^{K_+} \frac{(N - c_k)!(c_k - 1)!}{N!} \tag{2.4.22}
\]
2.4. Machine Learning Techniques

Figure 2.5: Stick-breaking construction for Indian buffet process (IBP). Initially the stick is of unit length. At each iteration, we break-off a part of stick with length $v_k$ relative to the original length of the stick and assign the remaining length of the stick $\pi_k$ to the atom $k$. This process is shown for six iterations here.

Here $\alpha$ is the hyper-parameter controlling the number of features; greater the value of $\alpha$, more number of new features will be sampled. The matrices produced by this process are not left-ordered form and the customers (objects) in this distribution cannot be exchanged. However, the $lof$ equivalent classes of the matrices generated by this process give an exchangeable distribution such that there are $\prod_{h=0}^{N-1} \frac{K_h!}{K_h}$ number of matrices in each class. The exchangeable distribution $P([Z] \mid \alpha)$ as stated in Equation 2.4.21 can then be obtained by multiplying this value (the number of matrices in each class) with Equation 2.4.22.

2.4.3.7 Stick breaking construction for IBP

Teh et al. (2007) proposed the stick-breaking construction for IBP. From the finite Bayesian matrix factorisation approach we have seen that the prior distribution on the allocation matrix are $\pi_k \sim \text{Beta}(\frac{\alpha}{K}, 1)$, where $\pi = (\pi_1, \pi_2, ..., \pi_K)$ are the...
parameters. If the parameters are arranged in decreasing order of $\pi$ such that $\pi_1 > \pi_2 > \ldots > \pi_K$ then at the limit $K \to \infty$ the $\pi_{1:K}$ obey the stick-breaking construction of IBP which is as follows:

$$v_k \sim \text{Beta}(\alpha, 1) \quad (2.4.23)$$
$$\pi_k = v_k \pi_{k-1} = \prod_{l=1}^{k} v_l \quad (2.4.24)$$

The stick-breaking constructions is way that the model assigns features based on popularity. Independent, Beta($\alpha$, 1) distributed parameters $v_k$ are first sampled. We start with a stick of length 1 and at each iteration, we break off a part of stick of length $v_k$ relative to the current length $\pi_{k-1}$ of the stick. The remaining part of the stick $\pi_k$ represents the probability we assign to the atom $k$. We then continue to recurse on the stick we broke off. This process is shown in Figure 2.5. Thus we can see that for larger values of $K$, the number of objects taking on feature $k$ reduces exponentially; also for a large value of $\alpha$ the number features assigned increases.

### 2.4.3.8 Slice sampler

Slice sampling is an auxiliary variable based inference method that samples uniformly from the region under the probability density function (Neal, 2003). Suppose we intend to sample values for variable $\pi$ from a distribution $p(\pi)$. Then slice sampling involves introduction of an auxiliary real variable $s$ such that the underlying distribution remains the same, that is $\int p(\pi, s) ds = p(\pi)$. The auxiliary variable $s$ is conditionally drawn from $p(s \mid \pi) = \frac{1}{p(\pi)} \mathbb{I}(0 \leq s \leq p(\pi))$. Gibbs sampling is used to alternately sample for variables $\pi$ and $s$, and the samples of $\pi$ would be effectively drawn from $p(\pi)$. The marginal probability distribution for $\pi$ remains
unchanged as follows:

\[
\int_s p(\pi, s) ds = \int_s p(s \mid \pi)p(\pi) ds \\
= \int_s \frac{1}{p(\pi)} \mathbb{1}(0 \le s \le p(\pi)) p(\pi) ds \\
= p(\pi) \tag{2.4.25}
\]

The conditional probability of \( \pi \) given \( s \) is uniform over the region \( Q = \{ \pi : s < p(\pi) \} \). The slice sampler is thus named as it slices away the values of \( \pi \) that are least probable, that is all values for which \( s > p(\pi) \).

The slice sampler proposed for IBP by Teh et al. (2007) leverages the stick breaking construction as described in the previous section. Here we intend to infer the values of the parameters \( \pi \) and elements of allocation matrix \( Z \). The proposed auxiliary variable \( s \) is uniformly drawn as:

\[
s \mid Z, \pi_{1:}\infty \sim \text{Uniform}[0, \pi^*] \tag{2.4.26}
\]

where if \( K^* \) is the number of active features then \( \pi^* = \text{min}_{k \le K^*}(\pi_k) \), is a function of \( \pi_{1:}\infty \) and \( Z \), and is chosen as the length of the last active feature. The conditional distribution of \( Z \) given \( s \) is

\[
p(Z \mid s, \pi_{1:}\infty, x) \propto p(Z \mid \pi_{1:}\infty, x)p(s \mid Z, \pi_{1:}\infty) \\
= p(Z \mid \pi_{1:}\infty, x)\frac{1}{\pi^*} \mathbb{1}(0 \le s \le \pi^*) \tag{2.4.27}
\]

It can be seen from Equation 2.4.27 that for the values of \( \pi_k < s \) the slice sampler does not assign the feature \( k \) to the object \( n \), that is \( z_{kn} = 0 \). Next the slice sampler samples from the respective conditionals for \( Z \) and \( \pi \) (for details please refer to (Teh et al., 2007)).

A particular model of interest is the linear Poisson gamma model (LPGM) (Gupta et al., 2012) with an IBP prior. We will elaborate upon this model and how it can be leveraged in our problem setting in Chapter 5.
2.4. Machine Learning Techniques

Figure 2.6: Graphical representation of restricted Boltzmann machines (RBM), where $h_k$ are the hidden units and $x_n$ are the visible units.

2.4.4 Restricted Boltzmann machines and deep learning

Restricted Boltzmann machines (RBM) are building blocks for generative deep learning architectures such as the deep belief network (DBN) (Hinton et al., 2006), the deep RBM (Salakhutdinov and Hinton, 2009) and the deep sigmoid belief networks (DSB) (Gregor et al., 2013). In its original form, a RBM is a generative Markov random field with two binary layers (Hinton, 2002).

When dealing with data, the key advantage of an RBM lies in its representation power at the latent layer $h$ to represent data, which can also be repeatedly constructed to build deep features. It is this versatile representation of the RBM that we leverage to model the profiles of children before they begin interventions in Chapter 6. This is possible since the hidden layer forms a continuous-valued probability vector which uniformly transforms data into a consistent mathematical object to deal with, such as to perform projection, computing similarity, conducting retrieval and so on.

Hence it induces a bipartite structure involving a visible input layer $v = [v_1, \ldots, v_N]^T \in \{0,1\}^N$ and a binary hidden layer $h = [h_1, \ldots, h_K]^T \in \{0,1\}^K$. The input layer is connected to binary representation layer using weighted connections, but there is no connection within each layer. Specifically, the RBM defines the following joint
2.4. Machine Learning Techniques

distribution:
\[
p(v, h | \theta) = \frac{1}{Z(\theta)} \exp \{-E(v, h; \theta)\}
\]
\[
E(v, h; \theta) = - \left\{ a^T v + b^T h + v^T W h \right\}
\]
\[
Z(\theta) = \sum_{v, h} \exp \{-E(v, h; \theta)\}
\]

where \( E(\cdot) \) and \( Z(\cdot) \) are respectively the energy and normalisation functions; \( a = [a_1, \ldots, a_N]^T \), \( b = [b_1, \ldots, b_K]^T \) and \( W = [W_{nk}] \) are the parameters. The parameter \( a \) is attached to only the visible layer and hence also known as visible bias parameter. Likewise, \( b \) is the hidden bias parameter and \( W \) is the matrix weights that connects the two layers.

Due to its special bipartite structure, evaluating conditional distribution in an RBM is efficient:
\[
p(v | h, \theta) = \prod_{n=1}^{N} p(v_n | h), \quad p(h | v, \theta) = \prod_{k=1}^{K} p(h_k | v)
\]

In particular, computing the posterior \( p(h_k | v) \) involves only \( N \) links from hidden unit \( h_k \) to \( v \), hence using Equations (2.4.28) and (2.4.29):
\[
p(h_k | v) \propto \exp \left( b_k h_k + \sum_{n=1}^{N} v_n W_{nk} h_k \right)
\]

Since \( h_k \) is a binary variable, it is easy to derive:
\[
p(h_k = 1 \mid v) = \frac{1}{1 + \exp \left( -b_k - \sum_{n=1}^{N} v_n W_{nk} \right)}
\]
\[
= \tau \left( b_k + \sum_{n=1}^{N} v_n W_{nk} \right)
\]

where \( \tau(x) = [1 + e^{-x}]^{-1} \) is the logistic function. Hence, if the model parameter \( \theta = \{a, b, W\} \) is known, then the \textit{posterior representation} for data \( v \) using the latent posterior vector \([p(h_1 = 1 \mid v), \ldots, p(h_K = 1 \mid v)]\) is extremely efficient to compute with the help of the equation above. For binary \( v_k \) the data generative probability \( p(v_k \mid h) \) is computed in a similar way.
A key task with RBM is to estimate the parameter $\theta$ from the $D$ observed data $\{v_1, \ldots, v_D\}$. One can address the problem by expressing the RBM as a member of exponential family.

**Exponential form of RBM** In Section 2.4.3.3 we have seen the representation of an exponential family for a single variable. Expanding the same form for 2 variable $(x, y)$ given the parameters $\eta$ and the base measure is assumed to have a unit value (as this measure does not affect the nature of our problem), the exponential family can be represented as:

$$p(x, y \mid \eta) = \exp\{\eta^T t(x, y) - a(\eta)\} \quad (2.4.35)$$

where $a(\eta)$ is the log-partition function given as

$$a(\eta) = \log \int_{x,y} \exp\{\eta^T t(x, y)\} \quad (2.4.36)$$

and the sufficient statistic is $t(x, y)$, whose expectation, in case of a general exponential family, is equal to the derivative of the log-partition function $a(\eta)$ with respect to the parameter:

$$\frac{\partial a(\eta)}{\partial \eta} = \mathbb{E}_{x,y \mid \eta}[t(x, y)] \quad (2.4.37)$$

Now the conditional distribution $p(y \mid x, \eta)$ also takes on an exponential form and is represented as follows:

$$p(y \mid x, \eta) = \exp\{\eta^T t(x, y) - b(\eta, x)\} \quad (2.4.38)$$

where $b(\eta, x)$ is the log-partition function associated with the conditional distribution given as:

$$b(\eta, x) = \log \int_y \exp\{\eta^T t(x, y)\} \quad (2.4.39)$$
Consider a real world application where the data $x$ is the only observed value. Then our interest lies in estimating the parameters $\eta$ followed by inferring the conditional distribution $p(y | x, \eta)$. Thus estimating the parameter $\eta$ is reduced to the optimisation goal of maximising the marginal log-likelihood function $p(x | \eta)$ given as follows:

$$
p(x | \eta) = \int y p(x, y | \eta) dy = \int y \exp\{\eta^T t(x, y) - a(\eta)\} dy
$$

$$
= \frac{1}{\exp\{a(\eta)\}} \int y \exp\{\eta^T t(x, y)\} dy
$$

$$
= \frac{\exp\{b(\eta, x)\}}{\exp\{a(\eta)\}} = \exp\{b(\eta, x) - a(\eta)\} \quad (2.4.40)
$$

$$
\ln p(x | \eta) = b(\eta, x) - a(\eta) \quad (2.4.41)
$$

Now the log-likelihood can be maximised using the gradient method and by using the property (Equation 2.4.37) as follows:

$$
\frac{\partial \ln p(x | \eta)}{\partial \eta} = \frac{\partial b(\eta, x)}{\partial \eta} - \frac{\partial a(x)}{\partial \eta} = \mathbb{E}_{y|x,\eta} [t(x, y)] - \mathbb{E}_{x,y|\eta} [t(x, y)] \quad (2.4.42)
$$

A RBM can be represented in the form of an exponential family as follows:

$$
p(v, h) = \exp\{\theta^T t(v, h) - a(\theta)\} \quad (2.4.43)
$$

where $\theta$ is the vector of parameters $[a, b, w]^T$, $t(v, h)$ is the vector of the observed and hidden variables $[v, h]^T$, and $a(\theta)$ is the log-partition function, that is $a(\theta) = \ln Z(\theta)$.

Thus the partial derivative for the log-likelihood function for the individual para-
meters in $\theta$ are as follows:

\[
\frac{\partial \ln p(v \mid \theta)}{\partial a_n} = \mathbb{E}_{h \mid x, \theta} [v_n] - \mathbb{E}_{h,v \mid \theta} [v_n] \tag{2.4.44}
\]

\[
\frac{\partial \ln p(v \mid \theta)}{\partial b_k} = \mathbb{E}_{h \mid x, \theta} [h_k] - \mathbb{E}_{h,v \mid \theta} [h_k] \tag{2.4.45}
\]

\[
\frac{\partial \ln p(v \mid \theta)}{\partial w_{nk}} = \mathbb{E}_{h \mid x, \theta} [v_n h_k] - \mathbb{E}_{h,v \mid \theta} [v_n h_k] \tag{2.4.46}
\]

Since the RBM belongs to the general class of exponential family, the gradients in Equations 2.4.44, 2.4.45, 2.4.46 can be computed to be dependent on the expectations of the models (Hinton et al., 2006; Wainwright and Jordan, 2008). Although the expectations of the hidden variables given the observed variables can be efficiently computed the expectation of the complete model $p(h,v \mid \theta)$ is intractable and can be estimated using a special form of Monte Carlo integration known as Contrastive Divergence (CD) sampling scheme (Hinton, 2002).

Contrastive Divergence works by initialising the Markov chain training sample that has a distribution that is already close to the underlying data distribution. CD starts from the data and estimates the hidden variables, followed by the reconstruction step where the visible units are sampled. CD does not wait for the chain to converge and while practically using it, it has been observed that even a 1-step CD works well.

As we will see in Chapter 6, our data is not simply in the form of binary, it is mixed and heterogeneous with binary, Gaussian and count components. Traditional clustering techniques cannot naturally integrate such data into a single model. A model proposed by (Tran et al., 2011) called mixed-variate restricted Boltzmann machine (MV.RBM) can be leveraged to address data that has different components.

### 2.5 Closing Remarks

In this chapter, we have provided a comprehensive review of the relevant background required for the thesis. We begin by describing Autism Spectrum Disorder (ASD) and how it is diagnosed. Due to the rise in the prevalence of ASD among the general population, it has become crucial to find alternate methods of delivering therapy
either to append the main intervention or to assist the therapist to efficiently deliver intervention. We have found from the literature that computer-based methods are particularly useful as they are capable of seamlessly recording rigorous data and also can assist in analysing the data automatically, along with being favoured by children with ASD due to reduced inhibitions. The data leveraged through these applications has high potential for understanding the disorder in a nuanced way to understand the patterns of how children with ASD learn on a given syllabus. Machine learning techniques applied on such data can reveal patterns that may not only help improve the understanding of the community but also help in tailoring personalised intervention based on individual child characteristics. Given the nature of such data that doesn’t provide additional ground-truth information regarding the children, we are mainly focused on unsupervised machine learning techniques. Especially our focus is on latent feature detection models that help us understand the patterns of underlying subgroups. We provide the background for the factor analysis approach and the restricted Boltzmann machines based on which we intend to adapt and build models that suit the structure of the underlying data. While the feature allocation method results in a meaningful representation of the data allowing us to understand the latent features in terms of observed variables as well as the grouping of the children based on the identified factors; the RBM provides a model that can accommodate additional information about the child which helps in profiling them.
Chapter 3

Early Intervention and Data Collection

In this chapter, we lay the groundwork by explaining our data source (the application via used for collection of data) and describe the data set. The significant contribution of this chapter involve collection and description of a novel dataset as a result of iPad-based ASD intervention. As described in Chapter 1, the dataset we present is the first of its kind and the largest to the best of our knowledge. We describe how the data is collected, processed, cleaned and converted to a representation that is computationally friendly. This data then can be processed using machine learning and data mining techniques to derive knowledge about ASD and performance about children with ASD.

Another major contribution of this chapter is a framework for building early intervention techniques (delivered through iPad) that can be adapted to suit the cultural context. Our quest to leverage the generalisability of computer-based applications has potential to make therapy accessible widely over different regions, languages and cultures. We show this in the context of a case study on the basis of a trial conducted in India.

We present this chapter in three sections. We begin by describing the early intervention program that records the data from children with ASD - TOBY Playpad (Now called as TOBY Autism Therapy). We focus on the aspects of TOBY Playpad
3.1 Therapy Outcomes By You (TOBY) Playpad

TOBY Playpad is an early intervention program developed by Venkatesh et al. (2013) to help parents of children with ASD and their therapists in delivering therapy. This program uses the iPad as a learning environment. The application is developed on the principles of ABA (Lovaas, 1987) intervention framework. ABA’s large dependence on paper-based delivery methods and manual tracking methods are substituted by a machine framework. This facilitates automatic delivery of electronic stimulus in the form of images and video, which are picked from a database in a structured way. Progress is tracked by recording each detail of the child behaviour on the iPad.

Three main features of TOBY (Venkatesh et al., 2013) are:

- Flexible delivery of syllabus with a rigorous learning framework using best-practice techniques from behavioural therapy - reinforcement, prompting, and measurable criteria for skill mastery and syllabus progression;
- Utilisation of a multi-skill syllabus covering visual and auditory understanding,
receptive and expressive language, and social skills and imitation;

- Deliver stimuli in *mixed environments*, on- and off-device, integrated with this learning framework.

TOBY borrows heavily from the ABA based Discrete Trial Training (DTT) framework for delivering instruction. Smith defines a discrete trial as a “small unit of instruction implemented by a teacher who works one to one with a child in a distraction-free setting” (Smith, 2001). A discrete trial is then delivered in five basic steps: giving a *cue*, *prompting* the child to give a *response*, giving feedback or declaring the *consequence*, providing an *inter-trial break* (Smith, 2001). Within the framework, these steps are incorporated into the system such that they are accurately and automatically delivered. These basic steps with reference to TOBY on iPad tasks become:

1. *Stimulus* is given to the child in the form of an image or video accompanied by an audio instruction.

2. The child gives a *response* to the stimulus by touching one of the multiple options displayed on the iPad screen or by enacting or verbalising the desired action or word.

3. This response is fed back into the system, either automatically when the child touches the response, or when the child is required to perform an action or say a word, the parent is given options on the screen in a corner to record the response as either correct, wrong or prompted.

4. If the response given by the child is incorrect the TOBY either *prompts* the child to choose a correct response, for example by flashing light on the correct image as shown in Figure 3.2a or the machine instructs the parent to prompt the child towards the desired response.

5. If the response given by the child is correct, the child receives *reinforcement* via animations and starts as shown in Figure 3.2b, which can be traded for rewards such as game activities or a short video.

The nature of the stimulus and response varies based on the skill that the child is attempting to learn. We will come back to discuss the changes in the next section.
Prompting in TOBY occurs at three automatic levels and a fourth level. At the zeroeth level of prompting, which is also the default state, there is no prompt given but a voice instruction is played to suggest the child to make another attempt. We explain this process in case of matching tasks that are a kind of iPad tasks in TOBY curriculum. Matching tasks are used for teaching categorisation of objects on the basis of sensory and semantic properties. For example, in case of sensory matching tasks, TOBY displays the image of a target object on screen followed by three alternate images with an audio instruction that says, “Find Same”. When the child selects a correct response, TOBY gives a positive reinforcement (e.g. TOBY says “well done”). On the other hand when the child selects an incorrect response TOBY says “No, try again” while prompting the child toward the correct response; the prompt is based on the prompting level that the algorithm is in. The prompting level is increased or decreased in a definite manner as shown in Figure 3.3, such that the child is given assistance at a level which is appropriate. The levels of prompting for matching tasks are as follows:

1. Level 1: A digital image of a finger pointing at the correct response is shown on screen along with a voice instruction saying for example “Try this one!”

2. Level 2: A digital image of a finger pointing at the correct response with a pulsating image is shown on screen along with a voice instruction saying for example “This one matches!”

3. Level 3: A digital image of a finger pointing at the correct response with a pulsating image which is changed in colour is shown on screen along with a voice instruction saying for example “This one matches!” This is the highest level of automatic prompting.

4. Level 4: In situations when the user is not responding to the highest automatic prompting, a level 4 prompting may be requested, where the parent is instructed to take the child’s hand towards the correct response.

The cycle of stimulus, response and reinforcement is continued till the child achieves mastery in a skill. TOBY follows a structured syllabus tree that we will describe in depth in the next section, and a strict mastery criteria, which ensures learning before progressing to a much advanced skill in the syllabus tree. When a child
obtains 8 correct responses without prompting in 10 trials, TOBY declares that the child has mastered that skill. If the child is unsuccessful in the first turn a set of 10 stimuli is given once again till the child achieves mastery. TOBY also allows fast mastery when a child responds correctly without prompting to the first 4 stimuli, thus allowing the children who are familiar with certain parts of the syllabus to quickly proceed to the next skill-set.

The Plan tab of TOBY shows the tasks that are available at any given time. Figure 3.1a shows a screen shot of the Plan tab for a test account. The parent may wish to plan the day’s activity by selecting certain tasks which can then be seen in the My Plan section. Once a task is successfully completed it can be seen under the Completed section. The Plan tab also provides a brief summary of the child’s progress in terms of the number of tasks completed on the left side of the screen. In Figure 3.1a, for example, it can be seen that 2 tasks of Imitation iPad have been completed and 12 tasks are pending in this skill area. The skill areas and syllabus are described in detail in the next sub-section.

The Reports tab of TOBY Playpad provides a brief visualisation of the child’s performance. The Figure 3.1b and Figure 3.1c show iPad task Attempts and Attempts over Time windows within the Reports tab for the test account. iPad task Attempts shows the number of times each iPad task has been attempted successfully (in green) and unsuccessfully (in red). Attempts over Time is a plot of overall attempts i.e. accounting for individual trials that make up a task. For tasks that the test person has attempted, the number of successful trials without prompts are plotted in green, the number of successful trials with prompting are plotted in yellow, and the number of incorrect responses are plotted in red. This is a brief analysis of the individual’s performance that is made available to the parent to observe. Parents or trained professionals can observe these performance reports and make decisions about the course of the intervention.
3.1. Therapy Outcomes By You (TOBY) Playpad

(a) Plan tab

(b) Reports tab - iPad task Attempts

Figure 3.1: Screen shot of the Plan tab and Reports tab of TOBY Playpad.
Figure 3.1: Screen shot of the Plan tab and Reports tab of TOBY Playpad.

Figure 3.2: Prompt and Reinforcement delivery mechanism in TOBY Playpad.
3.1. Therapy Outcomes By You (TOBY) Playpad

Figure 3.2: Prompt and Reinforcement delivery mechanism in TOBY Playpad.

Figure 3.3: Increase and decrease in prompting level.
3.1. Therapy Outcomes By You (TOBY) Playpad

### Skills Classification

<table>
<thead>
<tr>
<th>Skills</th>
<th>Imitation</th>
<th>Sensory</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Partner</td>
<td>Solo</td>
<td>Partner</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Solo</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type of tasks</th>
<th>Partner</th>
<th>Solo</th>
<th>Partner</th>
<th>Solo</th>
</tr>
</thead>
</table>
| **Table 3.1**: Classification of skills into Solo or Partner tasks.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Role</th>
<th>Role</th>
<th>Role</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solo (iPad)</td>
<td>System</td>
<td>System</td>
<td>System</td>
<td>System</td>
</tr>
<tr>
<td>Partner (iPad)</td>
<td>System</td>
<td>Parent</td>
<td>Parent</td>
<td>System</td>
</tr>
<tr>
<td>NET</td>
<td>Parent</td>
<td>Parent</td>
<td>Parent</td>
<td>Parent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Role</th>
<th>Role</th>
<th>Role</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solo (iPad)</td>
<td>System</td>
<td>System</td>
<td>System</td>
<td>System</td>
</tr>
<tr>
<td>Partner (iPad)</td>
<td>System</td>
<td>Parent</td>
<td>Parent</td>
<td>System</td>
</tr>
<tr>
<td>NET</td>
<td>Parent</td>
<td>Parent</td>
<td>Parent</td>
<td>Parent</td>
</tr>
</tbody>
</table>

**Table 3.2**: Roles of parents and system in Solo, Partner and NET activities.

3.1.1 **TOBY curriculum**

TOBY skills can be classified into *iPad* or *Natural Environment Training* (NET). The iPad skills are delivered using the framework described in the previous section, sometimes with or without the help of a caregiver. The NET skills are carried out off-iPad, in a natural setting based on the instructions TOBY gives to the caregiver. The system describes skills that can be done through daily activities by the parents. The parent delivers the instruction through the day with some planning, giving appropriate prompts and reinforcements when required. A screenshot of the NET activity is shown in Figure 3.5. Only after completion, the parent enters the outcome of the activity into the system. For example, for a NET skill such as *BrushingTeeth* requires a mirror, toothbrush and toothpaste. The caretaker is asked to follow the instructions as shown in Figure 3.5b. The roles of parent and system during instruction delivery are listed in the Table 3.2. The NET skills are conducted in a natural environment with caregivers performing the entire instruction, which gives us less control over the variables. Hence, our main focus of study is the iPad skill syllabus.
3.1. Therapy Outcomes By You (TOBY) Playpad

Figure 3.4: Structured syllabus of TOBY Playpad categorised by skill area. Each colour represents a different skill area. The labels inside each node denote the name of the skill that is taught in it.
3.1. Therapy Outcomes By You (TOBY) Playpad

(a) Example of the adaptive task *Brushing Teeth*. The *overview* page summarises the goal and the materials required to complete this activity.

(b) The *activity* page lists instructions for the parents describing the activity, the prompting mechanism and the reinforcement.

Figure 3.5: Planning and execution for a NET task.
3.1. Therapy Outcomes By You (TOBY) Playpad

(c) The outcome page allows the parent to record the child’s performance by selecting an option (indicating the level of mastery) on the screen.

Figure 3.5: Planning and execution for a NET task.

iPad skills in TOBY are delivered in four main categories:

1. Imitation: copying an action or producing pre-speech sounds.

2. Sensory: skills related to perception and discrimination of sensory cues, such as colour, shape, same-ness and difference.

3. Receptive Language: recognising objects from their names.

4. Expressive Language: identifying names for the objects.

Each category has many skills under it. We give a list and description of all the skills in Appendix B. The Figure 3.4 shows the syllabus tree of TOBY Playpad for iPad skills. The skills in Imitation category are marked in yellow, Sensory skills in blue, Receptive Language in light green and Expressive Language in dark green. The syllabus is hierarchically arranged such that the most basic skills are at the top of the tree. Initially, only the basic skills are open for attempt and the children start the syllabus with them. Once the basic skills are mastered the next skills in the node are opened for attempt. This ensures that pre-requisite basic skills are
learnt before proceeding to advanced concepts. There are a total of 34 iPad skills on TOBY. Of these, 13 are imitation (gross motor, actions and oral) based skills, 3 are sensory skills, 9 are receptive language skills, and 9 are expressive language skills. Details about these skills are described in Appendix A.

An iPad skill could be classified as solo or partner. Caretakers share responsibility with the system for instruction delivery depending on whether the skill is a Solo or Partner skill as shown in Table 3.2. In case of a solo iPad task, the system recognises the response and provides prompts or reinforcements as well. For partner iPad tasks, the caretaker identifies the child’s response, feeds it into the system and provides prompting if required.

The stimulus for an imitation iPad task is a video model of an action. The child watches the video and tries to repeat the action. Once the child reciprocates, the parent feeds the appropriate response into the system by touching one of the buttons (“Yes”, “Prompt”, “No”) provided on the top right corner of the screen. Based on the parent’s feedback, TOBY will either suggest appropriate prompts for the task (e.g. 'No' button), or move to the next stimulus ('Yes', or child achieved with 'Prompt'). Figure 3.6a shows a screen-shot of an imitation task in progress. In case of a sensory iPad task, the stimulus is given in the form of an image accompanied with three options as shown in Figure 3.6b. The child matches the object on the top with one of the three options given at the bottom of the screen. Once the child selects one of the options, the system automatically feeds the response and decides if the response is correct or incorrect. For a receptive language iPad task, the stimulus is presented in the form of images of objects accompanied with a voice instruction. For example, Figure 3.6 shows a screen-shot of a receptive language iPad task in progress. When the three objects are displayed on the screen a voice instruction “Find dog” is played. Receptive language tasks being solo tasks, the response is automatically fed into the system when the child selects one of the three objects displayed on the screen. The expressive language tasks involve identifying and vocalising the name of the object displayed on the screen. A screen shot of expressive language task in progress is shown in Figure 3.6d. The image of the object is displayed on the screen and a voice command “What is this?” is played. The child gives a response and the parent feeds the response into the system by using the button displayed in the top-right corner of the screen as shown in Figure 3.6d.
Two terms related to the syllabus that we will use repeatedly throughout the dissertation are:

1. **Skill category/skill area**: The four categories of iPad skills in TOBY syllabus - Imitation, Sensory, Receptive Language and Expressive Language.

2. **Skill/task**: An individual concept under a category e.g. `gross-motor-imitation-visible-1 (gmi-v-1)` is the first skill also called as entry skill within Imitation category. It is the first node under Imitation in the syllabus tree as seen in Figure 3.4.

3. **Trial**: A single activity that contributes to learning the skill. For example, 'gross-motor_imitation-visible-1' is a skill that is learnt by practising multiple actions e.g. *One hang up in air, Clap hands, Cover eyes, Touch nose*, etc. Each action is called as a trial.

![Image](image_url)

(a) Imitation Skills

Figure 3.6: Example of Imitation, Sensory, Receptive Language and Expressive Language.
3.1. Therapy Outcomes By You (TOBY) Playpad

Figure 3.6: Example of Imitation, Sensory, Receptive Language and Expressive Language.

(c) Sensory Skills

(d) Receptive Language Skills
3.2 Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India

Learning in TOBY Playpad occurs at 2 levels: *Multiple Exemplar Training (MET)* and *Single Exemplar Rote Training*. The syllabus within imitation, sensory and language follow this model. When a user repeatedly fails at a task, the system falls back from MET to Rote training. MET protocol decides the stimulus sets - from simple to complex contexts, rules for defining mastery, rules for controlling prompts and reinforcement and their fading, and the fallback from MET to rote training.

**3.2 Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India**

Considering autism diagnosis averaging at 1 in 100, of the 158 million children within the age group of 0 - 6 living in India (Ministry of Statistics and Programme Implementation,, 2012), about 1.5 million children can be estimated to be living with ASD. TOBY is currently designed for use at home, in English. This makes it easily adaptable to countries and contexts where English is the primary language. How-
ever, settings in other countries differ vastly. For example, the linguistic diversity in India is great and often children speak languages other than English at home. The typical things used in therapy will also differ. For example, with respect to change in the culture and food habits, adaptation of task for identifying food-items calls for addition of “idli”. The key aspects of adaptation include alteration of contextual settings, change of primary language and adaptation of the objects used for therapy. Since early intervention needs precision, domain experts are required for translation and as a part of this study, we build systems to facilitate this process. We build frameworks to facilitate collaborative translation and creation of tasks. More specifically our goals were to construct:

- **Technologies to facilitate organisation of media, maintaining dependency on language and culture.** This includes construction of schema, separating metadata from media for each cultural context, and authoring tools for collecting and indexing media for use by a person proficient in the language.

- **Collaborative environments to facilitate a team of experts to create new tasks:** A large number of context-specific tasks need to be created. The task creation is a fragmented across therapists and skill levels. We construct systems that enforce protocols for creating content, ensuring experts validate final tasks; and task management systems, to control work flow of material creation.

- **Tools to manage translation of tasks and interface.** Language in early therapy requires precision and unambiguity, so this goal was not met by machine translation. Thus interfaces for translation and management of multiple versions of each document in different languages were built.

After the frameworks were built, 19 skills were selected from the original TOBY syllabus (as described in Section 3.1) spanning a range of imitation, sensory, social and language skills. Specifically, they include 2 tutorial skills (pointing and dragging), 5 solo iPad skills, 5 parter iPad skills, and 7 NET skills. The selected skills are at the beginning of the TOBY syllabus tree and most have both iPad and NET tasks.

TOBY uses a tree-like category schema to organise media and content such as:

- **Images:** tagged by category labels, language, and region (culture) in which the
3.2. Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India

image will be integrated;

- Audio of voice instructions, prompts, and reinforcements tagged by category label for prompt type, and language in which audio is delivered;

- Video consisting of actions to demonstrate skills: tagged by action type, and region (actors are chosen to be familiar region).

To translate the language and culture we must identify objects or tasks that are not in common use in a culture or are constrained by either target language or functional utility. The reasons for changes can be outlined as:

- **Adaptation for cultural aspects.** This involves identification of sub-categories that are not in colloquial use in a country. Between Australia and India, the categories that were identified as culturally independent were personal items and household items. The categories that were culturally specific were *food* (added 6 items, removed 7), *clothes* (added 3 items, removed 5), *toys* (added 3 items and removed 6), *animals* (added 5 items, removed 4), and *outdoor* objects (added 1 item, removed 7).

- **Adaptation to language constraints.** Some multimedia items needed to be changed due to language constraints. One example is the literal translation of the English word “ankle” is not commonly used in Hindi; instead “leg” in Hindi is used. Thus to avoid confusion, one category is removed in the Hindi version.

- **Adaptation to differences in function.** Some items have different utility in the target country - for example, a bucket can be a toy in Australia, but it is a household item in India. The subcategory was re-classified in such scenarios.

In the following sub-sections we describe the implementation of changes in specific parts of the syllabus: collaborative translation of NET tasks, adaptation of NET to the classroom, voice prompts, video for imitation tasks, and images for receptive and expressive language matching. We then proceed to show the results of the trial conducted in collaboration with Tamana, a school located in Delhi, India. The primary language spoken in this region is Hindi.
3.2.1 Adaptation for textual content

The main textual information that needs translation is instructions to the parent or teacher to do NET activities. The TOBY syllabus covers 51 skills through a mixture of play and adaptive NET activities. There are 292 NET tasks in total, roughly 6 per skill. Each task contains approximately 260 words, with a total of 77500 words for the entire syllabus.

NET tasks include a mixture of instructional text, formatting, and structure data, encoded in an XML file. Attributes of each NET task include: a title for display on the task list, an associated syllabus node, a task type (either play or adaptive), a unique identifier, and a language specifier. The text for each task includes a number of sections, which are treated specially in the TOBY app.

We built an on-line syllabus management system to manage the translation task. The writers for the original English syllabus tasks commented that XML was difficult and time-consuming to write, so we used a simplified syntax similar to Markdown. This is plain text with simple conventions to indicate section headings, list items, and emphasis. This allows task writers and translators to focus on the content rather than the formatting.

We imported the TOBY English syllabus, and created an initial version of the Hindi syllabus using a machine translation service. The translation was done automatically using the web API for the Microsoft Translation service\(^1\). While the results were generally not acceptable, they provided an initial version which could then be refined by a human translator.

Figure 3.7 shows the layout of the task view in the syllabus manager. The left panel shows the document attributes (status, title, course, ID, node, type, and language) and the translated text which can be edited via the form elements. The right panel shows the attributes and text of the original document. To simplify the task for translators, we locked all attributes except title, text, and status. The system verifies that essential items are specified. For example, each of the sections listed above must be included, and must include a list of three prompts, one for each prompt level. The task syntax includes the conventions taken from the Markdown language: section

\(^1\)Bing Translator: [http://www.bing.com/translator](http://www.bing.com/translator)
3.2. Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India

Figure 3.7: Syllabus manager, Task edit form

Figure 3.8: Syllabus manager, Target view
3.2. Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India

The syllabus manager provides *targets* for users to keep track of work. Each target includes a due date, the number of tasks to be completed for adaptive and play types, and a list of nodes. Figure 3.8 shows an example of the target view. For each node to be completed, it shows the number of accepted tasks of each type (adaptive/play), and the number of tasks for each task status (translated, imported, draft, review, and accepted).

When editors log into the system, they can see the list of targets assigned to them. For each target, they can see the list of nodes with work outstanding. When an edit is completed it enters “review” state, and must be accepted by a different user to the one who made the edits.

The syllabus manager also offers a *preview* function (see Figure 3.9), which launches TOBY in the web browser using the live syllabus data. This allows the editors to see how their tasks will appear on the iPad screen. This is useful for checking layout...
3.2. Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India

<table>
<thead>
<tr>
<th>Skill</th>
<th>Home</th>
<th>Materials</th>
<th>School</th>
<th>Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matching Objects</td>
<td>Meals</td>
<td>cups, cutlery</td>
<td>Art</td>
<td>art supplies</td>
</tr>
<tr>
<td>Expressive Nouns</td>
<td>Dressing</td>
<td>clothes</td>
<td>Music</td>
<td>musical instruments</td>
</tr>
<tr>
<td>Gaze Sharing</td>
<td>Bathing</td>
<td>bath toys</td>
<td>Reading</td>
<td>books, animal toys</td>
</tr>
<tr>
<td>Matching Colour</td>
<td>Washing</td>
<td>towels, socks</td>
<td>Dance</td>
<td>scarves</td>
</tr>
</tbody>
</table>

Table 3.3: Examples of home routines that were modified for a school environment

and readability. If a task is too wordy to fit on a single page, then the editor can see what can be simplified to improve the layout.

3.2.2 Adaptation for cultural context of use - the classroom

The above interface allows tasks in TOBY to be translated. TOBY was originally designed for home settings, and many adaptive tasks are based on familiar routines such as bath time, and meal time. This means that:

1. Adaptive routines may not be suitable for use in schools, where the daily routines are different.

2. Even in a home setting, routines may vary with culture.

The Hindi trial syllabus includes 7 NET skills. This includes 26 adaptive tasks and 19 play tasks. Of the 26 tasks, 21 were found to be related to home-based routines that may not be appropriate for school settings. Thus an additional 21 NET tasks were written by modifying the existing tasks for a different school-based routine. For example, a home-based routine for sensory matching involves matching objects such as cups and spoons at meal time. For a school setting the corresponding task involves matching art materials (paper, pencils, glue) during art activities. Table 3.3 shows examples of some of the adaptations that were made.

3.2.3 Adaptation of voice content

In TOBY voice prompts are used to:
3.2. Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India

- Introduce the names for objects during exhibit trials (e.g. “this is water”).
- Prompt the user during receptive tasks (e.g. “find same”, or “find water”).
- Provide feedback and reinforcement during discrete trial training (e.g. “well done!”, “no, try again”).

Fragments for these prompts are stored in separate sound files, and utterances are generated by concatenating the sounds. For example “this is” + “water”, or “find” + “water”. This strategy works for the languages we have attempted so far, although it may have to be adapted to cope with different types of grammars.

For the translation, we recorded approximately 220 Hindi words, which includes all of the nouns and prompts used for receptive language tasks. In addition, we recorded 19 new English words, which correspond to the categories added for the Indian region.

3.2.4 Adaptation of video content

TOBY uses video modelling for imitation tasks, which are used to teach skills in gross motor, fine motor, and oral motor imitation. Imitation tasks include an average of 13 actions, with a range of 4 to 30 actions. TOBY includes 174 imitation videos with 100 actions and 74 echoic (oral-motor) videos. The videos use actors that would look normal for the target country. We recorded 67 new imitation videos, covering only the skills included in this trial. Figure 3.10 shows examples from the standard Western and Indian version.

Echoic imitation tasks teach the building blocks of words (e.g. combinations of vowels and consonants). These are expected to vary with the phonemes present in the target language, and hence also need to be adapted. This was outside the scope of the current work, but will be handled in the future when the full syllabus is implemented.
3.2.5 Adaptation of image content

TOBY uses images for receptive and expressive language training. For the added categories we collected 73 new exemplar images. Figure 3.11 shows the media collection interface which was added to the syllabus manager for this purpose. For each category, it shows the associated media files. In this case, two images and two sound files. Users can upload images from their computer which will be tagged with the category. Alternatively, they may browse the web for creative-commons licensed material from Wikipedia or Flickr. Clicking on one of these links will execute a search on the corresponding site. The images from the search result page can be dragged into the “Explore Media” section, and all meta-data required for creative-commons attribution (e.g. authorship and title) is gathered. When satisfied, the user can click “upload” which causes the syllabus manager to import the original images for incorporation into the image database.

3.2.6 Experimental Results

The translation framework was used to develop a Hindi version. A 3-person team from India performed task construction and translation. An equivalent English version is also extracted. These two prototypes were tested in India. The first trial
3.2. Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India

Figure 3.11: Syllabus manager, Interface for online collection of category media.
3.2. Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India

Examines the possibilities of using iPad interventions in an Indian classroom setting using 32 children. After establishing the technology could be used, a subsequent second trial was conducted with 16 children over a span of 6 weeks - 10 children using a Hindi version of TOBY and 6 children using equivalent TOBY version in English.

**Trial 1 - Study of iPad acceptance**  The trial was run in April 2013 at Tamana School of Hope in New Delhi, India. We approached parents of 47 children who were (1) diagnosed with autism (the DSM-IV criteria (American Psychiatric Association, 1994)), (2) more than three years old, and (3) currently enrolled at Tamana. 32 children, aged 3 to 14 years, were recruited. The cohort had markedly different functioning levels - (13/16) are verbal, with language ability ranging from single words to some fluency, and (3/16) are nonverbal.

The children attended school every weekday and worked on TOBY either in their normal classroom or in the special TOBY research lab. Each child was given a 15-20 minute time slot depending on attention span and performance ability. On average, 15 sessions were given for each child.

The study aimed to observe the acceptance of iPad technology. Children in the early intervention group (e.g. under 6 years old), were most responsive to TOBY, and found the software’s built-in rewards motivating. But their attention spans varied widely. The preparatory junior and senior (8-15 years old), and high-functioning group (6-7 years old) quickly warmed to the new TOBY instructors when using the iPad. Some boredom was noted in the high functioning group, who wanted to play more difficult “games”. Some children recognised TOBY instructors and searched for iPads even outside TOBY sessions. Others needed external rewards to continue their work on TOBY.

**Trial 2 - Evaluation of TOBY syllabus adaptations**  The second trial was conducted for 6 weeks between July and August 2013 at the Tamana School. Both Hindi and English TOBY programs have the same syllabus. 16 children were selected for the trial - 10 to work on TOBY in Hindi, and 6 in English. This grouping was based on the child’s mother tongue. The cohort consisted of children from 3 to 16
3.2. Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India

<table>
<thead>
<tr>
<th>Child</th>
<th>L</th>
<th>Family has iPad</th>
<th>Used iPad before</th>
<th>Pointing skill Pre</th>
<th>Pointing skill Post</th>
<th>Dragging skill Pre</th>
<th>Dragging skill Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 E</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>2</td>
<td>N/A</td>
<td>2</td>
<td>N/A</td>
</tr>
<tr>
<td>2 E</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>3 E</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>5</td>
<td>N/A</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>4 E</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>5</td>
<td>N/A</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>5 E</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>2</td>
<td>N/A</td>
<td>2</td>
<td>N/A</td>
</tr>
<tr>
<td>6 E</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>5</td>
<td>N/A</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>7 H</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>4</td>
<td>N/A</td>
<td>4</td>
<td>N/A</td>
</tr>
<tr>
<td>8 H</td>
<td>Yes</td>
<td>No</td>
<td>N/A</td>
<td>5</td>
<td>N/A</td>
<td>5</td>
<td>N/A</td>
</tr>
<tr>
<td>9 H</td>
<td>Yes</td>
<td>Yes</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>10 H</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>2</td>
<td>N/A</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>11 H</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>2</td>
<td>N/A</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>12 H</td>
<td>Yes</td>
<td>Yes</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>13 H</td>
<td>No</td>
<td>Yes</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>14 H</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>5</td>
<td>N/A</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>15 H</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 H</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>2</td>
<td>N/A</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.4: iPad accessibility and skills. Point and drag skills are rated 1-5 with 1 not being able, and 5 fluent. Language is indicated in column “L” (E=English, H=Hindi)

years, with the majority (12 children) between 6 and 11 years.

**Pre- and post- trial questionnaires** Parents and/or teachers of the children were asked to fill in a questionnaire before and after the trial. Information about the child was collected including accessibility/familiarity with iPads and communication skills. Parents and teachers confidence in working with their child were also recorded. 15/16 children completed both pre- and post-trial questionnaires.

**Summary of language spoken at home and iPad access:** The distribution of languages spoken at home is shown in Table 3.4. Most children (10/16) came from homes where both English and Hindi are spoken. Only 6 children came from homes that spoke only Hindi. The children who did the TOBY version of English came from dual-language speaking homes.

**iPad Access and skill improvement:** Table 3.4 shows the access to iPads before the trial. Limited access iPads was noted - only 4 children had iPads at home and of
these children only 3 had used iPads before the trial. Among the remaining 12 children without iPads at home, only 1 had access to iPads before the trial. Hence, there were a total of 12 children who did not use iPads before the trial.

The rating of children pre- and post-trial with respect to their pointing and dragging skills is also shown (1 indicates no ability to demonstrate skill, 5 indicates fluency). After the trial, half the children (7/12) had a rating above 4, while the remainder got a 2. The children’s ability to drag items on iPads had similar results: almost half (8/12) got ratings above 3.

**Analysis of performance data** Children in the trial worked with TOBY for approximately 30 minutes each day, on tasks on and off the iPad.

Figure 3.12 shows a summary of usage time for users of the Hindi (left) and English (right) syllabus. Each chart shows total time spent using TOBY (top) and relative time spent with different functions in the app. The total times range between 5 and 12 hours. A significant amount of time, up to 50% of the total, is spent between management functions (logging in and out of the system) and on the plan screen (choosing activities, displaying report summary). Usage on NET is generally low, but this is to be expected as NET tasks are performed off the iPad, and therefore are not recorded by the system. The remaining time is distributed largely between either sensory (blue) and imitation (pink) tasks, or receptive language (cyan) tasks. Normal syllabus progression is to move from lower level sensory tasks to higher level language tasks. Therefore, we expect a predominance of sensory tasks on the right hand (lower total usage time), and language tasks on the left-hand side (higher total usage time). Exceptions may occur if the learner is particularly slow (e.g. 5) or fast (e.g. 13) relative to the group.

Figure 3.13 shows the **success rates** for different iPad syllabus tasks. This is the proportion of the children who completed the task out of the set of children who attempted the task. Note that in order to attempt a task, the child must also have completed all of the prerequisite tasks, and must have done some activity on that task. Success rates are generally higher for easier tasks at the beginning of the syllabus, like the tutorial tasks and sensory matching tasks. Values are separated into the 10 Hindi users (rightmost two columns) and 6 English users (previous two
Figure 3.12: Trial 2: Total time using TOBY and proportional breakdown for Hindi (left) and English (right) users.
3.2. Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India

<table>
<thead>
<tr>
<th>Type</th>
<th>Course Node</th>
<th>Attempts</th>
<th>Success %</th>
<th>Attempts</th>
<th>Success %</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Action with Objects</td>
<td>5</td>
<td>40</td>
<td>7</td>
<td>85</td>
</tr>
<tr>
<td>L</td>
<td>Expressive Matching, Nouns</td>
<td>5</td>
<td>20</td>
<td>7</td>
<td>71</td>
</tr>
<tr>
<td>I</td>
<td>Gross Motor Imitation, Invisible, 1 Step</td>
<td>6</td>
<td>50</td>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>I</td>
<td>Oral Imitation, No Voice</td>
<td>3</td>
<td>100</td>
<td>8</td>
<td>62</td>
</tr>
<tr>
<td>L</td>
<td>Receptive Matching, Nouns</td>
<td>6</td>
<td>50</td>
<td>9</td>
<td>66</td>
</tr>
<tr>
<td>T</td>
<td>Tutorial, Dragging</td>
<td>6</td>
<td>100</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>T</td>
<td>Tutorial, Pointing</td>
<td>6</td>
<td>100</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>S</td>
<td>Visual Sensory Matching, Colour</td>
<td>4</td>
<td>75</td>
<td>8</td>
<td>87</td>
</tr>
<tr>
<td>S</td>
<td>Visual Sensory Matching, Identical Objects</td>
<td>6</td>
<td>100</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>S</td>
<td>Visual Sensory Matching, Nonidentical Objects</td>
<td>5</td>
<td>80</td>
<td>9</td>
<td>88</td>
</tr>
<tr>
<td>L</td>
<td>Vocabulary Expansion, Expressive, Nouns</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>Vocabulary Expansion, Receptive, Nouns</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 3.13: Success rates for iPad syllabus skills (I=imitation, S=sensory, L=language, T=tutorial) for English and Hindi children.

<table>
<thead>
<tr>
<th>Type</th>
<th>Course Node</th>
<th>Attempts</th>
<th>Success %</th>
<th>Attempts</th>
<th>Success %</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Action with Objects</td>
<td>3</td>
<td>66</td>
<td>9</td>
<td>66</td>
</tr>
<tr>
<td>L</td>
<td>Expressive Matching, Nouns</td>
<td>6</td>
<td>50</td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>I</td>
<td>Gross Motor Imitation, Invisible, 1 Step</td>
<td>3</td>
<td>66</td>
<td>8</td>
<td>87</td>
</tr>
<tr>
<td>I</td>
<td>Oral Imitation, No Voice</td>
<td>2</td>
<td>100</td>
<td>6</td>
<td>83</td>
</tr>
<tr>
<td>L</td>
<td>Receptive Matching, Nouns</td>
<td>6</td>
<td>66</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>P</td>
<td>Social Receptive, Follow Adult Point</td>
<td>3</td>
<td>100</td>
<td>6</td>
<td>100</td>
</tr>
<tr>
<td>P</td>
<td>Social Receptive, Gaze Sharing</td>
<td>6</td>
<td>50</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>S</td>
<td>Visual Sensory Matching, Colour</td>
<td>2</td>
<td>100</td>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>S</td>
<td>Visual Sensory Matching, Identical Objects</td>
<td>6</td>
<td>66</td>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>S</td>
<td>Visual Sensory Matching, Nonidentical Objects</td>
<td>2</td>
<td>100</td>
<td>7</td>
<td>85</td>
</tr>
</tbody>
</table>

Figure 3.14: Success rates for NET syllabus skills (I=imitation, S=sensory, L=language, T=tutorial) for English and Hindi children.

The corresponding values for NET tasks are shown in Figure 3.14.

Figure 3.16 shows the variation of success rates between the English and Hindi groups. Each point represents one TOBY skill, and the horizontal and vertical positions are the success rates for the English and Hindi groups. Note the clustering in the upper part of the graph, indicating that relatively more success is achieved by the Hindi group. The probability that this variation is due to chance alone is 0.10 for iPad skills, and 0.082 for NET skills (single-tailed paired Student’s T scores).

Figure 3.15 shows mastery of the TOBY syllabus for iPad and NET tasks. In each chart, the vertical axis lists the trial user identifiers, and the horizontal axis lists the syllabus node names. A coloured cell at the intersection between a user number and node name indicates that user successfully completed the task. Vertical bands
3.2. Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India

Figure 3.15: Trial 2: Task mastery for iPad (left) and NET (right) tasks.

Figure 3.16: Trial 2: Skill completion rates for English and Hindi children.
3.2. Adapting Technology for Cultural and Linguistic Relevance: A Case Study in India

of colour indicate tasks that were completed by many users. Horizontal bands of colour indicate users that completed many tasks.

3.2.7 Discussion on the trials conducted

We consider the translational aims to have been achieved. From the child’s perspective, two broad outcomes were achieved: the technology and the therapy were accepted. The iPads were used without any incident and this was a significant result given the socio-economic context where there is less exposure to such technologies. No iPads were broken, and all were used appropriately. The therapy was accepted as the children warmed to people delivering therapy on the iPad.

From the therapist’s perspective, three main outcomes were demonstrated:

- The syllabus material could be translated in a decentralised way, by non-experts. The trial was completed satisfactorily in both English and Hindi. This was achieved in spite of the trying working conditions.

- Both Hindi and English users were able to progress through tasks in 6 weeks. Measurable learning outcomes demonstrated that the children progress through nodes with mastery. Children spent between 5 - 12 hours in doing iPad tasks. Children who spent more time progressed further in the syllabus.

- There is some evidence that progress was slightly better in Hindi, so it is worth completing a full Hindi version of the syllabus.

The main suggestion for future translation is that the order of Hindi words needs to be considered in terms of their complexity and cultural popularity. Some NET activities, with instructions, were sent home to parents, but no feedback was received. This suggests more support is required for parents to appreciate the value of NET. A separate module to train parents needs to be built. Furthermore, a trial with rigorous pre- and post-intervention measures and/or control group would be needed to further evaluate the effectiveness of TOBY technologies in India.
3.3 Data Dictionary

The data uploaded by the caregivers via TOBY is stored on the servers that are accessed by the developers. The data sets are confidential and require authorisation from the developers of TOBY. The data we use were collected via two sources: the data collected from the anonymous users who download the application from iTunes, and data collected via clinical trials. The application server of TOBY resides in the web and is accessible via standard web protocols. Thus, once the results are fed into the system the application provides progress reports to parents as well as therapists. De-identified data is used for analysis.\footnote{This study is covered through ethics application submitted through Deakin University - 2012-035 - Assistive Technologies for Autism Support.}

3.3.1 Datasets and their source

The data was collected via two sources. The first source is groups of anonymous users who downloaded TOBY from iTunes Store and used it in their natural location. The second source is two groups of children with ASD as a part of a clinical trial conducted by the Telethon Kids Institute in Perth, Western Australia, in collaboration with Monash University in Victoria.

The anonymous users download TOBY on their iPad and use it in their natural setting. A suitability questionnaire is made available on the TOBY website\footnote{http://tobyplaypad.com/} that enables parents of children with developmental disorders to decide if their children are eligible for the use of TOBY. Additionally, TOBY also makes available intensive description videos on YouTube and Facebook. Parents and therapists are encouraged to contact the TOBY team via email and Facebook regarding any concerns they have about the application. A LITE version of TOBY is also available allowing parents to explore before they commit to the full version. These factors allow us to safely assume that the users of TOBY are either diagnosed with ASD or developmental disorders that have symptoms similar to ASD. Caregivers are informed about the data collection and their consent is taken for using the data, before registering on TOBY. Caregivers can use the application offline and upload the data
in intervals at their will. This mode of data collection provides the child’s personal information such as age and sex that is provided by the caregivers in the profile. On the other hand, the procedure of conducting a clinical trial is cumbersome and requires subjects to be in the geographical proximity. Past and current literature on study of subjects with autism is hence limited to a small number of case studies. A data-driven approach, however requires a larger dataset. Collecting data from anonymous users thus enables us to accumulate data from a larger community (up to 4000 profiles on TOBY till date). However, as the users are anonymous we are not equipped with the means to acquire any prior knowledge about the child’s condition, such as deficit skill areas and parent’s or therapist’s observation of the usual problematic areas. In other words, this data is not tagged with any ground-truth information.

Furthermore, data was collected via clinical trials. Detailed information regarding this data is described in Chapter 7. Table 3.5 is the description sheet for Anonymous user data set. Anonymous user data set is growing over time as new users download the program. The data size and date of release as mentioned in Table 3.5, thus reflects the information about the most recent data collection.

### 3.3.2 Attributes of TOBY data

The data recorded on TOBY is saved separately in different files and are linked to each other using unique identifiers like sessionID, profileID, date and time. Four
of the main data sources that are of interest to our project are the profile, session, progress, and results files. Each account of TOBY can be used by up to 5 children, hence allowing creation of 5 profiles with each one assigned a unique profileID. At the time of registration, the caretaker records a few details like the date of birth, sex and favourite things of the child. TOBY saves this information in the profile file. The child begins a session on TOBY and it records details like type of activity performed, and time spent on the activity. Ongoing skill accumulation involves learning in discrete steps called trials as explained in the previous section. TOBY records in the progress file the stimulus set given to the child as per the MET protocol, total number of successful trials, total number of responses recorded, number of correct and incorrect responses, and the number of prompts for the successful trials. While the accumulated summary of the trials is tracked in progress, the details of each trial such as the stimulus provided, result recorded by the child (in case of Solo tasks) or caregivers (in case of Partner tasks), success indication for each trial, level of prompting for each trial, and duration of response time in milliseconds are detailed in results file. Figure 3.17 shows the screen shot of part of the data collected from activity during one session of one of the users.

The data set has about 50 attributes. Identifying 18 attributes that have proven useful for our project we detail their description in Table 3.6. It is noted that the data recorded in some attributes was redundant and can be evaluated from the attributes listed in Table 3.6.
3.3. Data Dictionary

Figure 3.17: Example of log data collected from one section of therapy.
Table 3.6: Attribute descriptor table for the data.

<table>
<thead>
<tr>
<th>Attribute name</th>
<th>Data type</th>
<th>Data subtype</th>
<th>Description</th>
<th>Examples</th>
<th>Additional notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account Id</td>
<td>Categorical</td>
<td>ID</td>
<td>Account identification number</td>
<td>1386</td>
<td></td>
</tr>
<tr>
<td>Profile Id</td>
<td>Categorical</td>
<td>ID</td>
<td>Profile identification number</td>
<td>1021</td>
<td>Each account may have up to 5 profiles</td>
</tr>
<tr>
<td>Birth Date</td>
<td>Date</td>
<td></td>
<td>Date of Birth of the user</td>
<td>13/07/2010</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Binary</td>
<td></td>
<td>Sex as entered by the parents/caregivers</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>Session Id</td>
<td>Categorical</td>
<td>ID</td>
<td>Session identification number</td>
<td>16717</td>
<td></td>
</tr>
<tr>
<td>Result Id</td>
<td>Categorical</td>
<td>ID</td>
<td>Result identification number</td>
<td>284836</td>
<td>each trial is represented by one result entry</td>
</tr>
<tr>
<td>Start Time</td>
<td>Date and Time</td>
<td>Date and Time</td>
<td>Start date and time of each session</td>
<td>2013-10-08T08:49:57.182Z</td>
<td></td>
</tr>
<tr>
<td>Finish Time</td>
<td>Date and Time</td>
<td>Date and Time</td>
<td>Finish date and time of each session</td>
<td>2013-10-08T08:50:14.089Z</td>
<td></td>
</tr>
<tr>
<td>Plan</td>
<td>Categorical</td>
<td>ID</td>
<td>Name of the skill attempted in the session</td>
<td>mr-noun</td>
<td></td>
</tr>
</tbody>
</table>
### Table 3.6: Attribute descriptor table for the data

<table>
<thead>
<tr>
<th>Attribute name</th>
<th>Data type</th>
<th>Data subtype</th>
<th>Description</th>
<th>Examples</th>
<th>Additional notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimulus</td>
<td>Categorical</td>
<td>Nominal</td>
<td>Name/s of stimulus</td>
<td>crescent red, triangle</td>
<td>3 stimuli in case of matching skills, yellow, circle green</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>single stimulus in case of other skills</td>
</tr>
<tr>
<td>Target</td>
<td>Categorical</td>
<td>Nominal</td>
<td>Name of the target (the correct</td>
<td>crescent red</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>response)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target Index</td>
<td>Categorical</td>
<td>Nominal</td>
<td>Index of the target</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Response</td>
<td>Categorical</td>
<td>Nominal</td>
<td>Index of the response</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Success</td>
<td>Binary</td>
<td></td>
<td>1 - incorrect response, 0 - correct response</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Rote</td>
<td>Categorical</td>
<td>Ordinal</td>
<td>Rote level A, B and C</td>
<td>C</td>
<td>Three levels of rote are available: A, B and C</td>
</tr>
<tr>
<td>Prompt</td>
<td>Categorical</td>
<td>Ordinal</td>
<td>the level of prompt given during the trial</td>
<td>3</td>
<td>There are five levels of prompting for system</td>
</tr>
</tbody>
</table>
Table 3.6: Attribute descriptor table for the data

<table>
<thead>
<tr>
<th>Attribute name</th>
<th>Data type</th>
<th>Data subtype</th>
<th>Description</th>
<th>Examples</th>
<th>Additional notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prompt level</td>
<td>Categorical</td>
<td>level</td>
<td>the prompting level the system is locked to during the corresponding trial</td>
<td>4</td>
<td>The system changes the prompting level based on the responses.</td>
</tr>
<tr>
<td>Duration</td>
<td>Continuous</td>
<td>milliseconds</td>
<td>The duration between stimulus presented and the response is recorded</td>
<td>2908</td>
<td>This attribute is also called the response time.</td>
</tr>
</tbody>
</table>
3.3.3 Data pre-processing and features used for data analysis

At any given point of time, each child learning via TOBY would be at a different stage of learning on the syllabus. Users download and begin using TOBY at independent time points. As the children have different pre-training and different set of strengths and deficits on the syllabus, they progress through the syllabus at a different pace. Combining this with the fact that progress through the syllabus tree is allowed only by mastering the skill-nodes on the path, each child shows a distinct learning pattern. If many failed attempts are present then the skill is put on hold.

The raw data includes data from users who have just started using TOBY and are on a very early trajectory on the syllabus. As we are interested in the syllabus that has been mastered by the users it is essential for us to draw a threshold on the users based on how much syllabus they have mastered. For our study, we include only individuals who have at least mastered one node in the syllabus.

**Learn Unit** In order to observe the behaviour of each user as they learn on TOBY, it is crucial to quantify their interactions with stimuli. Greer and McDonough (1999) defined a term Learn Unit as a “measure of teaching that explicitly describes the interaction between the teacher and their students.” They found learn unit to be the strongest predictor of effective teaching and discuss how recording learn units makes analysis and strategic instruction easier for comprehensive application of Behaviour Analysis. Moore et al. (2015) adapt this concept for TOBY framework in their research. We elaborate the translation of learn unit as described by Greer and McDonough to TOBY framework in order to measure the child’s learning. In the following paragraphs, we first present the sequence of events that help us identify learning in TOBY and then describe the quantification of learning in TOBY.

Consider the following possible events of learning on TOBY. For a single trial, one of these events is guaranteed.

1. A new stimulus is provided - the child responds - TOBY identifies the response as accurate - TOBY gives positive reinforcement to the child.
2. A new stimulus is provided - the child responds - TOBY identifies the response as incorrect.

3. A stimulus that was previously incorrectly attempted is provided with a prompt - the child responds with assistance of the prompt - TOBY identifies the response as accurate - TOBY gives positive reinforcement to the child.

4. A stimulus that was previously incorrectly attempted is provided with a prompt - the child responds with assistance of the prompt - TOBY identifies the response as incorrect.

Whenever the child gives an incorrect response, TOBY repeats the stimulus with a prompt. Based on the algorithm that decides the prompting level, upon 2 consecutive incorrect responses, TOBY increases the prompting level by one; upon 3 consecutive incorrect responses, TOBY increases the prompting level to the highest. At each event, the child experiences an interaction with the system and finds one opportunity for learning. Based on this understanding we define the Learn Unit for TOBY as:

A TOBY LU is defined as a stimulus-response pair, where response is the recorded feedback of the child’s response irrespective of whether the response is correct, incorrect or prompted.

The measure of Learn Unit is adapted as a feature and its various forms are often used in our experiments. It is computed using mainly the attributes: success and prompt, for each skill by each user. We introduce some forms below and describe them in depth in the respective chapters they are used.

1. Correct LUs: The learn units that resulted in correct responses from the children.

2. Incorrect LUs: The learn units that resulted in incorrect responses from the children.

3. Prompted LUs: The learn units that were completed with the presence of a prompted stimulus.
3.3. Data Dictionary

Figure 3.18: Computing a Learn Unit.

<table>
<thead>
<tr>
<th></th>
<th>child 1</th>
<th>child 2</th>
<th>...</th>
<th>child J</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>3.0</td>
<td>5.6</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>sex</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td></td>
</tr>
<tr>
<td>task 1</td>
<td>100</td>
<td>8</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>task 2</td>
<td>250</td>
<td>180</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>task N</td>
<td>185</td>
<td>300</td>
<td>350</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.7: Data representation

4. Completed LUs: The learn units that were completed in total by a child
   (Completed LUs = Correct LUs + Incorrect LUs + Prompted LUs).

5. LUs to mastery: The number of learn units required by a child to master a
   skill. This unit is always stated for each skill individually.

6. normalised Learn Unit (nLU): Number of LUs required to complete one per-
   cent of the syllabus.

Figure 3.18 shows the processes that encompass a single Learn Unit. Along with the
data about the LUs on 36 skills for each child, we also have information on their age
and sex assigned at birth. This data can be represented in a matrix form as shown in
Table 3.7. The age of the users can be modelled using Gaussian distribution, the sex
of the user is expressed as binary data\(^4\), and the LUs to master task \(i, i = 1 : N\) is

\(^4\)This information is available to us as entered by the caregivers or parents. We assume that
the attribute Gender conveys information about the sex assigned at birth for the user
count data. We then use the data represented in this form for various data analysis methods such as factor analysis.

3.4 Closing Remarks

TOBY Playpad is an ABA-based early intervention program delivered via the iPad. Seamless data recording helps to collect rigorous data crucial for understanding the strengths and weaknesses of the child and to modify course of intervention, especially for ASD. The syllabus of TOBY is organised in a hierarchical tree structure based on complexity and skill categories, thus the child navigates the syllabus in a controlled manner. The performance data of about 543 children on 34 skills belonging to four categories - Imitation, Sensory, Receptive Language, and Expressive Language is used in our research. Quantification of the data recorded in the form of stimulus, response and reinforcement in terms of Learn Units (LUs) and normalised LU (nLU) enables us to model, cluster, analyse and discover relevant patterns from the data in future.

In adapting technology for cultural relevance, the trials we conducted in India show that the children were able to learn using the translated prototype in Hindi effectively. We show that the therapists were able to deliver the early interventions in-spite of difficult working conditions in India, and that they formulated innovative adaptations in the intervention.

As seen from the literature, the size and structure of the datasets in the field of ASD were limited for a data-based approach to be employed. The dataset we use in our research is, to the best of our knowledge, the first of its kind and the largest dataset in the domain of ASD. The nuanced approach in our research is developed owing to the TOBY dataset.
Chapter 4

Exploratory Analysis

In Chapter 3, we detailed the essence of the syllabus and functioning of TOBY Playpad. We keep referring to these details in the future chapters while we describe our methodology and discuss the results. As seen in Chapter 2 traditional datasets usually cited in research related to autism spectrum disorder vary from 10 or lower and may go up to 30 (Peters-Scheffer et al., 2011; Virués-Ortega, 2010; Remington et al., 2007; Eldevik et al., 2009). With regards to TOBY, Moore et al. (2015) conducted a pilot study on the data collected via a clinical trial from 33 families who used TOBY. We acknowledge that clinical trials are essential to collect data in a controlled manner and more reliable conclusions can be drawn. Yet limiting ourselves to clinical trial affects the number of data that can be collected. To make a data-based decision we need a dataset larger than the traditionally available datasets on ASD. TOBY has an online community of about more than 2000 users as of January of 2016. There have been no published reports on the study of data generated by these anonymous users who navigated their way independently through the software. Thus, there is an opportunity for mining these abundant data collected from users world-wide for possible learning profiles, which may assist in developing individualised early intervention programs. The work presented here is a pilot study and can be treated as an exemplar of the kind of data that can be harnessed by such intervention applications once available to the public.

In this chapter, we study the data from 185 users from the online community. We explore the dataset to answer the following questions:
• Could the information be used to assess if a child makes deliberate or random responses? The finding may suggest if a child is engaged in the computer learning environment that TOBY offers.

• Can the children be divided into sub-groups on the basis of their overall rate of learning? If so, what are the characteristics of these sub-groups?

• Are there differential patterns with respect to learning rates across the stratified groups? If so, how do they differ across skill categories?

• What implications might this have for individualised early intervention programs?

Probing the data for these questions will improve our understanding of how the children engage with the application in their natural environment with parent supervision. We compare the response time of the cohort for correct and incorrect responses. Through this comparison, we investigate if deliberate responses and random responses can be distinguished. We use K-means to place children in sub-groups based on the Learn Units (defined in 3.3.3) required to master the skills. The characteristics of the sub-groups further our understanding of the cohort characteristics. We also investigate the implications of the discovery of such sub-groups in personalised intervention. The results presented in this chapter will be foundational and serve as motivation for more complex models to deliver personalised intervention on computer based platforms.

The chapter is organised as follows. We begin with describing the participants, the variables and the methods. In the results section, firstly we present our findings comparing the response times between the correct and incorrect responses for the cohort. Secondly, we show the sub-groups of children and their characteristics based on their learning rate that we discovered. We follow up with a discussion about what our findings mean in the context of the cohort and ASD intervention.
4.1 Data Description

4.1.1 Participants and Context

This study presents an analysis of 185 children who have been using TOBY Playpad as a learning tool. The data is derived from a larger dataset of 310 children, who downloaded TOBY Playpad from iTunes, by eliminating the ones who have not mastered at least one skill. The data captures the progress of children until February 2013. At this time the children were at different stages in the curriculum. All participants are anonymous to the research team. Our data can be viewed as a progress report on a cohort of children, each working through TOBY at an individual pace. As progress across the syllabus tree is allowed only through mastery of each node in the curriculum path, children may follow different learning sequences based on their preference within a set of skills that TOBY opens for them.

Referring back to 3.1.1, TOBY syllabus tree consists of 34 iPad skills organised into four categories: Imitation, Sensory, Expressive Language, and Receptive Language. Focusing on these skills we list the independent and dependent variables in our analysis in the next sub-sections.

4.1.2 Independent Variables

For iPad tasks, which are the focus of this thesis, learning opportunities are provided in the form of stimulus presentations (the antecedent), sometimes accompanied with a prompt, comprising audio-visual signals. The child responds (the behaviour) by vocalising the response that complies with a request, imitating an action or by simply touching one of the multiple visual options on the iPad screen. TOBY responds (the consequence) to a correct response by giving stars or praise and to an incorrect response with a red cross or a verbal 'no'. TOBY Playpad records these three-term contingency interaction sequences.

Stars from correct responses are accumulated, and 10 stars can be traded for rewards such as game activities or a short video. For incorrect responses, TOBY uses the
same stimulus set and provides prompts in the next antecedent. For example, for solo tasks, e.g. matching identical objects, TOBY could provide the audio stimulus 'find same' accompanied by a prompt 'try this one'. Venkatesh et al. (2013) provide a detailed description of TOBY.

iPad tasks consist of 2 types:

- Solo: For Solo iPad tasks (Sensory and Receptive language tasks) the system recognises the child’s response (child touches or drags correct pictures) and provides prompting if required.

- Partner: For Partner iPad tasks (Imitation and Expressive language) the parents feed their child’s response into the system and, if required, provide prompting as instructed by TOBY. For example, parents choose “yes”, “no”, or “prompt” button to provide feedbacks when the child is asked to imitate an action on the iPad screen. TOBY may suggest parents to prompts (e.g. hand over hand) if necessary.

In both solo and partner tasks, TOBY provides rewards for correct responses as described above.

TOBY Playpad records:

- Antecedent: The stimulus at three levels of abstraction - skill category, skill, and exact stimulus. Each skill category consists of hierarchies of multiple skills. For example, in the skill category *Imitation* of partner tasks, one of the skills can be *1-step gross motor imitation* for which one exact stimulus is the video of a person clapping. In a scenario where the response to the previous stimulus set is incorrect, the antecedent may include the stimulus accompanied by a prompt. For example, the exact stimulus of a person clapping is accompanied by a suggestion for the parents to prompt, e.g. *hand over hand*.

- Behaviour: Response of child (either correct or incorrect).

- Consequence: System Feedback by TOBY - (Correct, Incorrect). This is recorded with the prompt state – for example, ‘correct with prompting’ or ‘correct without prompting’.
4.1.3 Dependent variables

To analyse the recorded data, we used a quantitative measure of instruction, the learn unit (LU), introduced by Greer and McDonough (1999) as ‘a measure of teaching that explicitly describes the interaction between teachers and their students’. Moore et al (2015) used LU in a study on a TOBY Playpad dataset collected from a trial conducted by the developers of TOBY. Recalling from the description of the LU in Chapter 3, a single LU involves (i) the presentation of a stimulus or task by the instructor (in this case by TOBY) (ii) a response to the task by the learner (correct or incorrect), and (iii) provision of appropriate feedback by the instructor.

Normalisation was required to compare LUs across children who were at different stages of progress. We did this by defining the normalised Learn Unit (nLU) as the number of LUs required to complete 1% of the curriculum. The nLU thus reflects the rate of progress for each child - a smaller value indicating a higher rate of progress. Importantly, nLUs can be computed by amalgamating LUs across either each skill category or all skills to provide detailed information at different levels of abstraction on how a child is progressing.

Another dependent variable in our analysis is the Response Time. Response time is the time interval between the presentation of stimulus and the input of a response. As solo iPad tasks engage only the child, it reduces any chances of discrepancy. On the other hand, as the partner iPad tasks engage the parent as well as the child, the response time is recorded as the interval between the presentation of the stimulus and the instance when the parent records the response after the child has reciprocated. This induces a possibility of additional latency introduced by the parent in recording the observation. We wish to explore the response times to observe if this factor is reflected in the data, and if so how this effects the response time distributions.

4.2 Data Analysis

To illustrate the concept of amalgamating LUs across skills and categories, consider the log of a child who has been attempting Gross Motor Imitation and Finer Motor Imitation Skill of the skill category - Imitation. Table 4.1 shows how the LUs are
### 4.2. Data Analysis

#### Table 4.1: A typical log for a child showing sequence of stimuli that TOBY presents.

<table>
<thead>
<tr>
<th>Skill Category</th>
<th>Skill</th>
<th>Exact stimulus (Antecedent)</th>
<th>Child response (Behaviour)</th>
<th>TOBY feedback (Consequence)</th>
<th>LUs for Skill</th>
<th>LUs for Skill category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imitation</td>
<td>Gross motor imitation</td>
<td>Jump</td>
<td>Child waves</td>
<td>Wrong</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jump + prompt</td>
<td>Child jumps</td>
<td>Correct</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Touch Shoes</td>
<td>Child touches shoes</td>
<td>Correct</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fine motor imitation</td>
<td>Wriggle fingers</td>
<td>Child wriggles finger</td>
<td>Correct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Open and shut hands</td>
<td>Child pushes object</td>
<td>Wrong</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Open and shut hands + prompt</td>
<td>Child open and shuts</td>
<td>Correct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Point index finger to palm</td>
<td>Child points correctly</td>
<td>Correct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extend index finger</td>
<td>Child claps</td>
<td>Wrong</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

amalgamated across the skill and skill categories.

For each child in the cohort of 185 children, the LUs were amalgamated across all skills and then normalised to complete nLUs. These variables for all children were fed into a standard clustering algorithm K-means (MacQueen et al., 1967), which partitions data points into clusters, by reducing the distance between each point and the cluster centroids.

The Mann-Whitney-Wilcoxon (Birnbaum et al., 1956) test for statistical significance was used where necessary to analyse the behaviour of the clusters and determine if response patterns of two given clusters are statistically different from each other. We define the null hypothesis as: the data from two groups belong to the same distribution with equal medians. We set the significance value of p-values at 0.05. In each case, a p-value < 0.05 means we can safely reject the null hypothesis.
4.3 Response Time Analysis

The idea that information about mental activity and organisation could be revealed through response times has been of interest to researchers (Luce, 1991). Ex-Gaussian is a well-known distribution in psychology used for modelling response times. Palmer et al. (2011) conducted a research on the shape of the response times for visual search experiments and found that the ex-Gaussian distribution is generally believed to produce a perfect fit to the response time data.

Hohle argues that response time is made up of time required to process an exponentially distributed choice or decision latency, and motor response time or residual latency. The theory behind ex-Gaussian distribution is that the cognitive processes that enable decision making capability brain can be viewed as a result of convolution of a Gaussian and an exponential distribution, where the decision latency is represented by the exponential distribution and the residual latency by the Gaussian distribution (Hohle, 1965).

Ratcliff has also suggested truncating the response time on the right tail in order to fit the data to an appropriate distribution shape and stated that general values as low as 1500ms to 3000ms and above have been chosen as a common procedure (Ratcliff et al., 1993). Ratcliff also states that instead of choosing a generalised cut-off value for truncating the right tail, a preferable method would be to select a cut-off value as a function of the proportion of responses eliminated. The nature of the stimulus for TOBY varies significantly depending on the type of skill area, ranging from a set of images to identify to video models. Thus based on the available data we chose a generalised threshold for determining outliers and eliminate the responses which take longer than 60 seconds. These eliminated responses constitute less than 3% of the data set.
4.4 Results

4.4.1 Response Times

Figure 4.1 shows overall variation of response times for correct and incorrect attempts made while learning through TOBY. Table 4.2 shows the median differences between correct and incorrect responses, and the interquartile range - a measure of statistical dispersion for the correct and incorrect responses. Rare response times lying on the right tail are marked as outliers for this analysis. Figure 4.1(left) gives an overview of response times for solo tasks. The median difference between the correct and incorrect responses for sensory skills and receptive language skills are 0.7 seconds and 1 second, respectively. The correct responses for sensory take 3.3s while for receptive it takes 2.7s (these values are the medians). Figure 4.1(right) captures the variations in response times for partner tasks where it can be observed that the median difference between correct and incorrect responses for imitation and expressive language is 6.6 seconds and 6 seconds, respectively.
4.4. Results

<table>
<thead>
<tr>
<th>Response Time</th>
<th>Solo Tasks</th>
<th>Partner tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensory</td>
<td>Receptive language</td>
</tr>
<tr>
<td></td>
<td>correct</td>
<td>incorrect</td>
</tr>
<tr>
<td>Median (seconds)</td>
<td>3.34</td>
<td>2.56</td>
</tr>
<tr>
<td>Interquartile range (seconds)</td>
<td>3.06</td>
<td>2.45</td>
</tr>
</tbody>
</table>

Table 4.2: Median and interquartile range of response time for solo and partner tasks. (All measurements in seconds).

The observation from the median difference and the interquartile differences is that whilst partner tasks show significant variation in the response times for correct and incorrect responses, the solo tasks seem to require nearly the same time to respond, with a small interquartile range irrespective of whether the response is correct or not. (For reference the percentage of outliers with respect to data points for this analysis is on average around 8%.)

Figure 4.2 shows the histograms for response times for imitation, sensory and expressive language skill areas across each cluster of children. The response times for correct responses are separated from the incorrect ones for the purpose of examining any possible variations. The general observation is that for partner tasks the variance for the incorrect response distribution is much higher than that for the correct response distribution. As shown in Figure 4.2, all the response times closely resemble the shape of ex-Gaussian distributions.

Figure 4.3 shows the histograms of correct and incorrect response times for receptive language. This skill area is an exception in terms of response time distribution. It can be observed from Figure 4.3 that for this skill area, the correct response times follow a mixture of distributions and is bimodal in nature while the incorrect responses do not show a similar pattern.
Figure 4.2: Ex-Gaussian fits to response time distribution for correct responses (left) and for incorrect responses (right). The distributions belong to the tasks (top) Imitation, (middle) Sensory and (bottom) Expressive Language skills.
4.4. Results

Figure 4.3: The bimodal nature of the correct response times for Receptive Language skill (left). Histogram for incorrect responses for the same skill (right).

Figure 4.4: Clusters obtained using K-means. Normalised Learn Units (nLUs) (y axis) for each child (x axis). Three groups shown: slow pace (high nLUs (yellow squares)); medium pace (medium nLUs (blue circles)) and fast pace (low nLUs (green diamonds)). Outliers are red triangles.

4.4.2 Learner groups

Four sub-groups of children with similar nLUs were extracted by the algorithm. Figure 4.4 shows four clear categories – fast pace (113 children), medium pace (49
children), slow pace (20 children) groups, and outliers. It can be observed that the fast pace group required the least number of nLUs with the nLUs progressively increasing for medium and slow pace group of learners. The outliers consisted of three children who required extremely large number of nLUs each. They were excluded from further analysis.

Scrutiny of the learn unit statistics across different clusters with respect to each skill category enabled us to understand patterns of group learning and how each cluster responded to the respective skill areas. Figure 4.5 shows the median nLUs for each skill areas across the three groups of children. The following observations were made:

- Participants in the fast pace group required fewer LUs to master three skill categories – sensory matching, and expressive and receptive language - relative to those in the medium pace and especially the slow pace groups. These differences are more marked in receptive language than expressive language. For receptive language, the slower learning group required many more learning opportunities to achieve mastery while for expressive language, the differences between medium and slow pace groups are less marked.

- The differences in the rate of acquisition between participants in the slow pace group and the other two groups are greatest with sensory skills and receptive language skills.

- A different observation is made for imitation. Smaller differences are noted in acquisition of imitation skills across participants in all groups, with the medium pace group mastering this skill with fewer nLUs than the fast pace group.

A statistical significance test was performed on the three clusters for the imitation category. In all cases p-values obtained were < 0.05 [p-value (group 1, group 2) = 0.0003, p-value (group 2, group 3) = 0.0377, p-value (group 1, group 3) = 0.0001] allowing us to reject the null hypothesis. All clusters differed from one another with respect to performance in imitation.
4.4.3 Learner group and age

Analysis of the age distributions revealed some discrepancies between the different learner groups. There were some outliers in these data: four participants were over 20 years old and were excluded from this analysis. The rest had their age
distributions plotted in Figure 4.6.

The slow pace group is shown to be the youngest, while the fast pace group was the oldest. The age of children in the slow pace group had the narrowest range (2.58 to 4.14 years old – 1.56 year spread), the medium pace group had a larger age range (3.14 to 5.5 years old – 2.36 year spread), and fast pace group had the largest range (3.6 to 6.58 years old – 2.98 year spread). The age differences between the three groups were statistically significant with all p-values < 0.05 [p-value (group 1, group 2) = 0.0490, p-value (group 2, group 3) = 0.0099, p-value (group 1, group 3) = 0.0004].

Figure 4.6: Age distributions in different learner groups with median: 4.75 years old (fast pace), 4.12 years old (medium pace) and 3.33 years old (slow pace).
4.5 Discussion

Considering that TOBY as an application employed at the users’ environment, it is beyond the scope of this work to obtain clinical backgrounds of the children at the start and functional assessments after the use of TOBY. Nevertheless, critical information could be revealed through analysis of the rigorously captured data. Learner profiles and learning process could be inferred and used to assist in understanding the children and their learning styles. This could potentially benefit therapists, parents and educators in offering evidence-based intervention programs.

4.5.1 Engagement level

In solo tasks, on the first presentation of a stimulus, children spent similar amounts of time before giving TOBY a correct or incorrect response. This finding may suggest the children took time to think before responding (either correctly or incorrectly) during their solo interactions with TOBY. This result encourages the potential use of computer learning platforms that may capture and maintain these children’s interest and attention spans.

In partner tasks, the incorrect response times are noticeably longer than the correct ones. This may suggest some teaching interactions between a child and her parent before a response of failure is recorded - the parent may wait expectantly a little longer to give the child more opportunities to respond correctly or attempt to prompt their child towards the answers.

4.5.2 Learner groups, skill areas and age

The automatic extraction of learner groups (fast, medium and slow pace) provides insight into different learner profiles and may afford opportunities for improved curricula.
Fast pace group  Participants in this group acquired all skills faster (requiring fewer nLUs) than the other two groups with one exception: imitation. For imitation tasks the fast pace group required significantly more nLUs than the medium paced group. For the fast pace group, imitation was the hardest skill to master, requiring more nLUs than all the other skills. This is in contrast to the slow and medium paced groups. For the fast pace learners, the ratio of nLUs required to master imitation to that required to master expressive and receptive language is 1.69: 1 and 3.10: 1 respectively.

A possible explanation for the fast pace learning rate is that this group, who is also the oldest group, consists of children: who may have so far missed out on early intervention but have the developmental readiness to now quickly acquire basic skills; or who may have learnt these skills before coming to TOBY. They may also have the most previous experience with screen media and therefore quickly learn via this medium. Mastering imitation skills, however, requires attending to other people and some level of social interaction. These older children may have developed some aversion to social related tasks possibly due to their limited capability to interact with people in their early age.

This finding may highlight again the significance of imitation in early learning. Imitation is thought to be a cusp behaviour; A skill which, once acquired, will open doors to further learning (Rosales-Ruiz and Baer, 1997). Imitation is particularly important to enable children to learn through modelling (McCoy and Hermansen, 2007; Rayner et al., 2009). This is essential for the acquisition of spoken language as speech cannot be physically prompted. For this reason, intervention targets for this fast pace group should perhaps include social skills, such as imitation and joint attention. This group may benefit from an increase in time spent on TOBY social skills – off iPad (NET) activities, which turn everyday living situations into social learning settings with specific skills being targeted and reinforced. Highly potent reinforcers may be required. Alternatively, there may be merit in working to the children’s strengths, using the screen-based media to teach social interaction skills including imitation. Older children with good language skills could benefit from social learning app such as the Social Express app\(^1\). Further research on this is warranted.

\(^1\)website of The Social Express app: http://thesocialexpress.com/
4.5. Discussion

Medium and slow pace groups  For the medium and slow pace groups, solo tasks (Sensory and Receptive Language) proved to be more challenging than partner tasks (Expressive Language and Imitation). This was most noticeable for the slow pace group who were also the youngest of our participants. Richert, Robb and Smith (2011) observed that young children seemed to require more guidance from adults in learning from screen media, both in directing their attention to the screen information and in making sense of the information. For the younger children, guided interactions by parents may be required to make their learning experience with TOBY iPad tasks more effective.

The general pattern of skill acquisition rates for these groups was very different to that in the fast paced group: The medium and slow pace groups required more nLUs for receptive language than for any other tasks. The opposite was true for the fast paced group. The TOBY syllabus has a pre-requisite structure with different but connected skill branches. Children can either access receptive tasks directly from the language branch or after passing the pre-requisite sensory tasks. Further investigation into the order of task completion by the children may be needed to understand the role of sensory discrimination as a pre-requisite skill for receptive language. It may be possible that children need to acquire the ability to distinguish objects based on the sensory inputs (colours, shapes, other distinguishing features), in order to build a representation of a class of the object (e.g. a wooden chair or a metal chair are both chairs) and understand its label (receptive labelling).

By way of contrast, the easiest tasks for these groups to master were imitation tasks: The slow and medium paced groups required fewer nLUs for imitation than for any other task, the opposite to what was observed for the fast pace group. Given their young age, it is possible that children in the medium and slow pace learner groups have not yet become averse to social interactions. They may have enjoyed and learnt more from adult mediation. These findings may be relevant in tailoring early intervention programs for young groups by maximising their ability to learn to connect with people especially in promoting language learning. More naturalist relationship focused early interventions, e.g. the Early Start Denver Model (Princiotta and Goldstein, 2013) or TOBY NETs, may be indicated. Further research on this relationship is justified.

It is also possible that screen-based instruction requires some specific pre-requisite
4.5. Discussion

skills. Video modelling for example, has been identified as an effective intervention for children with ASD (Richert et al., 2011). Yet not all such children respond equally well to video modelling and recently research has turned to the identification of pre-requisite skills to explain the mixed findings (Shukla-Mehta et al., 2010). Perhaps similar pre-requisite skills, such as the ability to attend to screen media, are required in order for children to benefit maximally from flat screen technologies and computer-based programs such as TOBY. The question certainly warrants further investigation.

Applications like TOBY are designed to enable parents and carers to maximise their daily learning opportunities with their child either while waiting to access therapy or to complement current therapies. This data set could grow to thousands of users in the future providing a valuable rich data set for investigation of differential use patterns among the targeted users; potentially invaluable information for program refinement and on-going development. There are, however, important limitations to such data with implications for the research community. As the application is downloaded at the children’s home and its use monitored by the children’s parents / caretakers, scant participant information is available such as their diagnostic status or their level of intellectual functioning, leading to external validity issues. Direct monitoring by the researcher of the fidelity with which the program is used is also currently not possible. Additional support may vary in important ways between the children, potentially confounding intervention effects.

Similarly, while the TOBY syllabus is ordered with built-in pre-requisite tasks both within and across skill areas, it does open some choices for users. Thus while traversing the common syllabus tree users might follow slightly different sequences based on the choices. For example, they could go straight to basic receptive/expressive language tasks or travel to them via the (sensory) visual matching skill set. The possible impacts of such differential learning histories on subsequent leaning trajectories and on participants’ enjoyment of the learning experience provided are clearly important areas for future research.
4.5.3 Modelling response time

The only exception in terms of response time modelling in our data is receptive language, which appears not to resemble the shape of an ex-Gaussian distribution. The nature of the distribution for this skill area is bimodal - an unexpected result, suggesting a mixture of two distributions. The stimulus set was searched to see if the entire stimulus set presented could be divided into 2 sets of response times according to the time required to take a decision about the response. We could not find definite sets of stimuli that could indicate such a hypothesis. In fact what we found is that each stimulus, at each attempt, took a response time that could randomly belong to any one of the distributions. This may deserve further investigation.

The task receptive language can be divided into 8 sub-tasks which consist of nouns from different categories like - animals, clothes, food, house-hold items, outdoor, personal and toys. We observe that the response time distribution corresponding to each of the sub-tasks follows a similar bimodal distribution.

4.6 Closing Remarks

This study analysed Learn Units required by children to complete the syllabus on TOBY as a result of which clusters of children with similar learning patterns have been extracted. Analysing responses of participants in each cluster in terms of their performance across all skill categories helped identify noteworthy differences between the groups. This information can inform further research into the effectiveness of screen media in teaching children with ASD. Specifically, there may be some support for more in vivo interaction for younger children, and the possible role of pre-requisite skills needed for children to benefit from learning via screen media. Findings can be used to further conduct an in-depth analysis on the performance of the children and tailor the syllabus of TOBY by utilising the learners’ strengths to facilitate their progress.
Chapter 5

Discovery of Learning Patterns and Autism Subgroups using Bayesian Non-negative matrix Factorisation

5.1 Introduction

In Chapter 4, we explored the data to answer questions that are significant to our project like: Can the children be divided into sub-groups on the basis of their overall rate of learning? and what implications might this have for personalising early intervention programs? These questions were answered in affirmative and motivate us to conduct an acute enquiry. In Chapter 4, we considered the nLU (normalised LU) across the skill category as the input. Each category is built up by various skills that cover different difficulty levels and abilities. In the current chapter, we intend to use factor analysis approach on the LU required to master each skill to discover what we call Learning Pattern - a set of skills that the child finds difficult to learn.

Discovering patterns in medical data in general, and autism in particular, fastens diagnosis and aids in increasing the availability of treatment to everyone. However, when the spectrum of the disorder is continuous, as it is in ASD, it is difficult to group
patients into clusters with clear boundaries. The patients are then clusters that share one or more characteristics across the spectrum of symptoms. The spectrum of this disorder varies vastly from high functioning to low functioning individuals who are not demarked from each other in terms of their abilities in the mentioned areas. Children affected by ASD show varied strengths and deficits in one or a mix of skill areas. Hence the need of opportunities for learning varies across individual children and across skill areas. Due to the crossover between characteristics there is no strong evidence of distinction between ‘types’ of ASD.

Bayesian nonparametric is an area that has received a lot of attention in machine learning in recent years. It is an attempt to tackle the automatic model selection problem in statistics and machine learning which has led to computational methods and toolkits that are widely applicable to different kinds of data. Traditionally, a modeller had to specify the statistical model a priori (e.g., the number of clusters, the dimension of the parameter space); this is a notoriously difficult task especially in dealing with complex data such as those that arise in health human behaviour domain. In particular, in our case, it is close to impossible to try a parametric method since the learning patterns are complex and unknown in the autism literature.

In particular, since the data representation admit a feature-based format, we have used nonparametric factor analysis model. Hence it is more suitable for factor analysis than traditional clustering methods (such as K-means). Factor analysis yields meaningful latent factors, which allows us to study the subtopics: In our study, we obtain learning patterns that are a collection of skills that were together found difficult by a group of children. Additionally, studying the membership matrix along with the latent learning patterns makes it possible to answer complex and nested questions like: who found which skills difficult? What other skills do they find difficult to master? Are there other children who are similar in terms of their abilities or disabilities?

The contribution of this chapter is the non-parametric discovery of: Learning Patterns - that help identify sets of difficult skills and ASD subgroups that find a particular set of skills difficult to learn. The approach of computer-based interventions towards the problem of individual learning patterns is two-fold: to build applications that cater to specific skill deficits, or to build applications with a broad syllabus such that it can be used by all children with ASD. In both the cases, however, it is
essential to perform manual analysis in order to determine the child’s strengths and deficits. This depends on the skills and ability of therapists, which might not be uniform for everyone. Identification of learner patterns has potential implications for improving early intervention through facilitating the development of evidence-based strategies tailored for different learner profiles. Hence, there is a need for technological intervention that is capable of determining the child’s coordinates on the spectrum based on their performance across a known syllabus.

Inferring learning patterns from therapeutic performance data is important to inform evidence-based intervention. However, this data-driven task was difficult in the past due to the lack of existing datasets with sufficient number of children. In this chapter, we use unsupervised learning methods on the TOBY dataset for discovering learning patterns. But a notorious problem with existing methods is the correct specification of number of patterns in advance, which in our case is even more difficult due to complexity of the data. To this end, we employ Bayesian non-parametric factor analysis (Gershman and Blei, 2012). The Poisson-gamma pair of likelihood generation is used to explain the data since our observed data is in form of counts. The model is termed as linear Poisson gamma model (LPGM) by Gupta et al. (2012). This model uses Indian buffet process (IBP) (Ghahramani and Griffiths, 2006) as prior on a binary matrix of infinite columns to allocate groups of intervention skills to children. The optimal number of learning patterns as well as subgroup assignments are inferred automatically from data. Our experimental results follow an exploratory approach and present different newly discovered learning patterns in autism domain. To provide quantitative results, we also report the clustering evaluation against K-means and Non-negative Matrix Factorisation (NMF). In addition to the novelty of this new problem, we were able to demonstrate the suitability of Bayesian non-parametric models over parametric rivals.

We further extend the model to address the problem of incomplete and growing dataset. The term missing data refers to the case where the value for a feature in an observation is missing. The dominating approach for missing data has been imputation (Olinsky et al., 2003; Allison, 2002). In applications like recommended movies prediction based on existing ratings of a user, the values of rating that are missing are interpolated based on the available entries for ratings of the user and other similar users. Once the missing values are interpolated, data processing methods can be implemented on the now complete dataset. Research evidence
suggestion suggests that using complete data by interpolating the missing values yields better results in data analysis than using sparse data (Chueinta et al., 2000; Zhang et al., 2006). But a similar approach may not be eligible for all applications.

Applications with medical significance as well as education allow less flexibility for interpolating missing values based on existing data. We explain the term *missing data* in context with the data we present here as follows. As the children progress through the syllabus at their own pace, at any point of data collection each child could be found at a different level on the syllabus tree. Their position on the syllabus tree also depends on their capabilities at each skill category: Imitation, Sensory, Receptive Language and Expressive Language. Children have to achieve mastery of their current skills, which are generally served as pre-requisites, before new skills become available. Hence, it is noted that each child will be at a specific position based on their time of entry and their rate of learning. The dataset consists of a few children who have mastered the entire syllabus as well as children who have mastered a few skills and are on their way of progress. A more accurate way of describing such a dataset would be ‘incomplete’ data rather than ‘missing’ data. Ideally, at any future time point the child is likely to progress through the syllabus and the missing data might become available. The prediction of values that are missing in our case involves unknown variables such as the child’s capabilities. Hence, a method to bypass the interpolation step and use the data as sparse is advisable.

Bayesian matrix factorisation methods are largely dependent on the availability of data that is complete. It has been observed that when the model involves Bayesian techniques, it is possible to bypass the missing data points in the equations either by using masking variables (Cemgil, 2009) or by excluding the missing term from the likelihood equations (Schmidt and Mohamed, 2009). We thus propose to extend the LPGM model such that the model can process sparse data.

In brief, our key contribution in this chapter is a non-parametric discovery of learning patterns and clustering of autism subgroups from data by extending LPGM to suit the context of missing elements in data. To our knowledge, we are the first to do so. The chapter is organised as follows. First, we begin with describing the datasets and their representation. We then describe the LPGM model and go on extend the model to accommodate missing data.
5.2 Dataset and Data Representation

5.2.1 Representation of data

The concept of LU has been used in the context of education to measure the amount of interaction of a student in classroom with the teacher (Greer and McDonough, 1999). An LU gives a rigorous and quantitative measure for our data analysis task.

Learn Unit is defined as an interaction between the system and the child that results in one response from the child.

Let $V$ be the total number of skills in TOBY syllabus and $J$ be the number of children using TOBY in our dataset. Then, $x(v, j)$ is the number of LUs required to master task $v$ for child $j$. Matrix $X$ is formed from the collection of $x(v, j)$. Figure 5.1 shows the representation of matrix $X$ for the first 20 children in dataset C.

5.2.2 Datasets

We apply the stated framework to analyse the leading patterns of learning among children with ASD who are using TOBY Playpad. The data consists of LUs required to master 32 skills, which are distributed among the skill categories: Imitation (13 skills), Sensory (3 skills), Expressive Language (8 skills) and Receptive Language (8 skills).

We use three sets of data for the task of data analysis. Dataset A consists of the therapeutic data of 12 children with ASD using TOBY Playpad. These children have mastered all the 32 skills that we study in this work. The input thus forms a complete matrix of dimensions $[32 \times 12]$, where the rows represent the 32 skills and the columns represent $x(v, j)$ for each child.

Dataset B consists of the therapeutic data of 371 children with ASD using TOBY Playpad. These children have mastered at least one skill and are at various stages of completion within the syllabus structure. Dataset A is also a subset of children from
5.2. Dataset and Data Representation

Figure 5.1: Illustration of a sparse data matrix. The colour of the bar represents a different skill; 34 skills in all. The skills in the same category are positioned next to each other. We show the LUs for mastery for the first 20 children in the data. The height of the bar represents the number of LUs accumulated for mastering each skill. The flat bars that touch the x-y plane are the skills that have not been mastered yet.

dataset B who have completed all the 32 skills in the syllabus. The input matrix is sparse with dimensions \([32 \times 371]\), where the columns represent \(x(v, j)\) for each child conditional on whether skill \(v\) is mastered.

Dataset C consists of the therapeutic data of 542 children with ASD using TOBY Playpad. While datasets A and B are the data recorded from TOBY servers before February 2013, the number of children and the progress of the existing children has since then grown. Dataset C was collected in May 2013. In that duration of time, additional two skills were added to the syllabus (particularly one to each category of Receptive and Expressive Language) making the total number of skills 34. The input matrix of dataset C is also sparse with dimensions of \([34 \times 542]\), where the columns represent \(x(v, j)\) for each child conditional on whether skill \(v\) is mastered.

There are 34 skills included in this study out of which 15 skills form a part of imitation, 3 skills form a part of sensory and 8 skills form a part of expressive and receptive language each. Details of the skills are shown in Table 5.1.
### 5.2. Dataset and Data Representation

<table>
<thead>
<tr>
<th>Category</th>
<th>Skill numbers</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imitation</td>
<td>1-2: gross motor imitation</td>
<td>copy to jumping</td>
</tr>
<tr>
<td></td>
<td>3: fine motor imitation</td>
<td>copy pointing</td>
</tr>
<tr>
<td></td>
<td>4-12: oral imitation</td>
<td>copy blowing/making a sound</td>
</tr>
<tr>
<td></td>
<td>13: action imitation</td>
<td>copy to feed teddies</td>
</tr>
<tr>
<td>Sensory</td>
<td>14 - 15: match identical/non-identical objects</td>
<td>match a picture of a chair to another</td>
</tr>
<tr>
<td></td>
<td></td>
<td>identical (or random) chair</td>
</tr>
<tr>
<td></td>
<td>16: match colours</td>
<td>match objects of same colours</td>
</tr>
<tr>
<td>Expressive</td>
<td>17: label common objects</td>
<td>answer to question “what is it?” with</td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td>a vocalisation “chair”</td>
</tr>
<tr>
<td></td>
<td>18 - 25: expand on more objects of a common theme</td>
<td>learn to label common foods in answer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to “what is it?”</td>
</tr>
<tr>
<td>Receptive</td>
<td>26: identify common objects</td>
<td>point to a picture of shirt when being</td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td>asked “find shirt”</td>
</tr>
<tr>
<td></td>
<td>27-34: expand more on objects of a common theme</td>
<td>learn to identify animals in answer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>to “find elephant”</td>
</tr>
</tbody>
</table>

Table 5.1: Skill numbers and their categories.

Our dataset consists of 542 children, with an inter-quartile range of 3.2 to 6.4 years of age, who attempted skills on TOBY. This data was accessed in February 2014 and at the time the children have mastered at least one skill. Some of the statistics of the dataset are shown in Figure 5.2. Visual observations show that the sensory categories (skills 14 - 16) have the most number of children who achieved mastery in compared with other categories. It is followed by imitation. In addition, more children mastered the standard expressive and receptive language nodes (skills 17 and 26) than their extensions (skills 18-25 and 27-34).

![Figure 5.2: Properties of the dataset](image-url)
5.3 Clustering Framework

For the sake of convenience, the notations used in this section are defined in Table 5.2.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V$</td>
<td>number of skills in the syllabus</td>
<td>$w_{vk}$</td>
<td>weight of skill $v$ in learning pattern $k$</td>
</tr>
<tr>
<td>$K$</td>
<td>number of learning patterns discovered</td>
<td>$W_{V \times K}$</td>
<td>learning patterns constituting $w_{vk}$</td>
</tr>
<tr>
<td>$N$</td>
<td>number of children in the dataset</td>
<td>$f_{kn}$</td>
<td>weight of learning pattern $k$ in child $n$</td>
</tr>
<tr>
<td>$x_{vn}$</td>
<td>the LUs required by child $n$ to master skill $v$</td>
<td>$F_{K \times N}$</td>
<td>membership matrix constituting $f_{kn}$</td>
</tr>
<tr>
<td>$X_{V \times N}$</td>
<td>data matrix constituting $x_{vn}$ for all children and skills</td>
<td>$z_{kn}$</td>
<td>indicator ${0,1}$ suggesting presence of absence of learning pattern $k$ in child $n$</td>
</tr>
<tr>
<td>$E$</td>
<td>factorisation error</td>
<td>$Z_{K \times N}$</td>
<td>indicator matrix constituting $z_{kn}$</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>parameter for modelling $E$</td>
<td>$\alpha$</td>
<td>hyper-parameter for IBP</td>
</tr>
<tr>
<td>$\alpha_\lambda, \beta_\lambda$</td>
<td>hyper-parameters for modelling $E$</td>
<td>$\alpha_0, \beta_0$</td>
<td>hyper-parameters for modelling $W$ and $F$</td>
</tr>
</tbody>
</table>

Table 5.2: Notations used for model

Matrix $X$ with dimensions $[V \times N]$ which is formed by the elements $x_{vn}$. The matrix factorisation can be represented as

$$
X_{V \times N} = W_{V \times K} \times (F_{K \times N} \odot Z_{K \times N}) + E
$$

where $W_{V \times K}$ represents the latent factors in our case called as learning patterns, $F_{K \times N}$ shows representation of each learning pattern in every child, and $Z_{K \times N}$ is the indicator matrix takes the values $\{0,1\}$ based on whether the learning pattern $k$ is present in child $n$. The product $H_{K \times N} = F_{K \times N} \odot Z_{K \times N}$ denotes the weight with which child $n$ follows learning pattern $k$.

The current model as proposed by Gupta et al. (2012), models the count data $X_{V \times N}$ as Poisson distribution, given the parameters $W_{V \times K}$ and $F_{K \times N}$ that are determined using gamma distribution. In order to solve the dimensionality and model selection
problem the model constitutes $z_{kn}$, which uses Indian buffet process (IBP) to infer the number of factors automatically. This is done by inferring values in $Z_{K \times N}$ based on IBP prior. We present the model in brief as follows.

5.3.1 The linear Poisson gamma model

We use linear Poisson gamma model (LPGM) as proposed by Gupta et al. (2012) to factorise $X$ while ensuring the non-negativity of $W$ and $H$. The data generating process for this model assumes that the $X$ is drawn from a Poisson distribution of parameters that are modelled with gamma priors. An IBP prior is chosen for the indicator matrix $Z$ allows automatic detection of number of learning patterns while allowing the model to grow infinitely. Figure 5.3 shows the graphical representation of the model. The model is represented as follows.

\[
\begin{align*}
\text{LPGM} & \sim \begin{cases} 
Z & \sim \text{StickIBP}(\alpha) \\
w_{vk} & \sim \text{Gamma}(\alpha_0, \beta_0) \\
f_{kn} & \sim \text{Gamma}(\alpha_1, \beta_1) \\
x_{vn} \mid w_{vk}, f_{kn}, z_{kn} & \sim \text{Poisson}(w_{vk}(f_{kn} \odot z_{kn}) + \lambda) \\
E & \sim \text{Poisson}(\lambda)
\end{cases}
\end{align*}
\]

The inference equations for this model are illustrated below.

- The posterior of parameter $w_{vk}$ is given as

\[
p(w_{vk} \mid Z, F, X) \propto w_{vk}^{\alpha_0 - 1} e^{-\beta_0 w_{vk}} \prod_{n=1}^{N} (f_{kn} z_{kn} w_{vk})^{R_{vn}} e^{-(f_{kn} z_{kn} w_{vk})} \]

\[
\propto w_{vk}^{\alpha_0 + \sum_{n=1}^{N} R_{vn} - 1} e^{-(\beta_0 + \sum_{n=1}^{N} f_{kn} z_{kn}) w_{vk}}
\]

The above expression takes the form gamma distribution. Hence,

\[
w_{vk} \sim \text{Gamma}(\alpha'_0, \beta'_0)
\]
where,

\[
\begin{align*}
\alpha'_0 &= \alpha_0 + \sum_{n=1}^{N} R_{vn} \\
\beta'_0 &= \beta_0 + \sum_{n=1}^{N} f_{kn} z_{kn}
\end{align*}
\]

and \(R_{vn}\) is an auxiliary variable sampled from

\[
R_{vn} \sim \text{Binomial} \left( \frac{f_{kn} z_{kn} w_{vk}}{\sum_{i \neq k} f_{in} z_{in} w_{vi} + f_{kn} z_{kn} w_{vk}} \right)
\]

- The posterior of parameter \(f_{kn}\) is given as

\[
p(f_{kn} | Z, W, X) \propto \frac{\alpha_1^{a_1-1} e^{-\beta_1 f_{kn}}}{\prod_{v=1}^{V} \left( (f_{kn} z_{kn} w_{vk})^{T_{vn} e^{-(f_{kn} z_{kn} w_{vk})}} \right)} \times \frac{\alpha_1^{a_1+\sum_{v=1}^{V} T_{vn}-1} e^{-(\beta_1+\sum_{v=1}^{V} z_{kn} w_{vk}) f_{kn}}} \]

Once again the form allows us to sample the parameter from a gamma distribution.
\[ f_{kn} \sim \text{Gamma}(\alpha'_1, \beta'_1) \]

where,

\[
\alpha'_1 = \alpha_1 + \sum_{v=1}^{V} T_{vn} \\
\beta'_1 = \beta_1 + \sum_{v=1}^{V} z_{kn} w_{vk} 
\]

and \( T_{vn} \) is an auxiliary variable sampled from

\[ T_{vn} \sim \text{Binomial} \left( x_{vn}, \frac{f_{kn} z_{kn} w_{vk}}{\sum_{i \neq k} f_{in} z_{in} w_{vi} + f_{kn} z_{kn} w_{vk}} \right) \]

The values in \( Z \) are inferred using slice sampler as proposed by Teh et al. (2007) for the stick-breaking configuration of IBP. The slice sample reduces the approximation otherwise introduced in Gibbs sampling by deciding the truncating level adaptively.

This model requires the input to be complete and thus requires substituting the values zero at data points that are incomplete. Theoretically, however, this leads to a conflict. We recapitulate that the incomplete data-points are generated when a child has not mastered the skill at the time of data collection. Substituting zeros in the place of incomplete data-points would mean that the child has completed the skill with zero LUs. Hence, we are motivated to extend this model to accommodate the condition of incomplete data. This is done by inferring the parameters \( w_{vk} \) over only the children who have mastered the skill \( v \) and the parameters \( f_{kn} \) only over the skills that the child \( n \) has mastered.

### 5.3.2 The extended LPGM model for incomplete data

\( X \) is our data matrix where the elements \( x_{vn} \), corresponding to the number of LUs accumulated by a child \( n \) in a task \( v \), are data-points. Our objective is to derive the posteriors for the parameters by using the data-points \( x_{vn} \) that are not missing.
5.3. Clustering Framework

The inference of posterior of $w_{vk}$, for a certain value of $v$, depends on the values $x_{vn}$ for all values of $n$. Similarly, the inference of $f_{kn}$ for a certain value $n$ depends on values $x_{vn}$ for all values of $v$. Hence, we consider the data points in two sets $J_v$ and $I_n$ for each inference respectively, such that $J_v$ contains all the non-missing values from the array $x_{v,1:N}$ and $I_n$ contains all of those from $x_{1:V,n}$.

Figure 5.4 shows how the inference is done when data elements are missing. We present the inference equations for a model that incorporates missing data as follows. The derivations for this model are provided in the Appendix B.

- The posterior of $w_{vk}$ is given as

$$p(w_{vk} \mid Z, F, X) \propto w_{vk}^{a_0 - 1} e^{-\beta_0 w_{vk} \sum_{n \in J_v} (f_{kn} x_{kn} w_{vk})} e^{-\sum_{n \in I_n} f_{kn} x_{kn} w_{vk}}$$

Figure 5.4: LPGM incorporating missing data
The above expression is in gamma distribution form $w_{vk} \sim \text{Gamma}(\alpha_0', \beta_0')$ where,

$$
\alpha_0' = \alpha_0 + \sum_{n \in J_v} R_{vn} \\
\beta_0' = \beta_0 + \sum_{n \in J_v} f_{kn} z_{kn}
$$

and $R_{vn}$ is an auxiliary variable sampled for any $n \in J_v$ from

$$R_{vn} \sim \text{Binomial} \left( x_{vn}, \frac{f_{kn} z_{kn} w_{vk}}{\sum_{i \neq k} z_{in} f_{in} w_{vi} + z_{kn} f_{kn} w_{vk}} \right)
$$

- The posterior of $f_{kn}$ is given as

$$p(f_{kn} | Z, W, X) \propto f_{kn}^{\alpha_1 - 1} e^{-\beta_1 f_{kn}} \prod_{v \in I_n} \left( (f_{kn} z_{kn} w_{vk})^{T_{vn}} e^{-(f_{kn} z_{kn} w_{vk})} \right)$$

$$\propto f_{kn}^{\alpha_1 + \sum_{v \in I_n} T_{vn} - 1} e^{-(\beta_1 + \sum_{v \in I_n} z_{kn} w_{vk}) f_{kn}}$$

Once again the posterior expression takes form of gamma distribution $f_{kn} \sim \text{Gamma}(\alpha_1', \beta_1')$ where,

$$\alpha_1' = \alpha_1 + \sum_{v \in I_n} T_{vn}$$

$$\beta_1' = \beta_1 + \sum_{v \in I_n} z_{kn} w_{vk}$$

and $T_{vn}$ is an auxiliary variable sampled for any $v \in I_n$ from

$$T_{vn} \sim \text{Binomial} \left( x_{vn}, \frac{f_{kn} z_{kn} w_{vk}}{\sum_{i \neq k} z_{in} f_{in} w_{vi} + f_{kn} z_{kn} w_{vk}} \right)$$
5.4 Generating Ground-Truth and Evaluation Methods

5.4.1 Ground-truth

Using naive LUs as performance measure is problematic due to the fact that children start the program at different times and require varied number of opportunities to learn a given task. When the data is collected, children could be at different stages of progress within the syllabus structure. Hence we introduce the concept of normalised learn unit (nLU). The nLU is therefore expected to reflect performance of the child regardless of entry point. We formally define nLU as follows.

A (nLU) is the measure of interaction between the child and system in order for the child to master one percentage of the syllabus.

Let $q_j$ be the percentage of the syllabus completed by the child $j$ and $M(j)$ be the number of skills mastered by child $j$. Then $q_j$ for child $j$ is given by Equation 5.4.1 and the nLU($v, j$) for child $j$ with respect to task $v$ is given by Equation 5.4.2

$$q_j = \frac{M(j)}{V} \times 100 \quad (5.4.1)$$

$$\text{nLU} (v, j) = \frac{x(v, j)}{q_j} \quad (5.4.2)$$

where $x(v, j)$ is computed conditional on the fact whether the child has mastered the task $j$ or not, indicating that $x(v, j)$ is not always defined for all $j$. The nLU($j$) for child $j$ over all the skills is represented as

$$\text{nLU} (j) = \sum_v \text{nLU} (v, j) = \frac{\sum_v x(v, j)}{q_j} \quad (5.4.3)$$
Figure 5.5: Sample learning pattern detected using IBP The skills represented with
darker shades require more LUs by the child who follows this learning pattern.

5.4.2 Latent learning pattern

The matrix $W$ indicates the latent learning patterns identified within the cluster
of children being analysed for their performance. We explain the significance of a
learning pattern with the help of an example shown in Figure 5.5. The learning
pattern suggests that a child following this pattern may require more opportunities
to master sensory and receptive language skills as compared to the rest of the syllabus.
An analysis to this effect is capable of identifying latent relationships among
inter-category skills. From our knowledge about the syllabus, sensory skills require
the ability to distinguish between cues such as colour and shape, while receptive
language demands the ability to identify the object from a given set with the know-
ledge of the label. Hence in this example we note the abilities of a child to learn
sensory skills and receptive language skills could be correlated. A similar analysis
can be done for any learning pattern discovered.
5.4.3 Evaluation methods

Since our data is new and there is no standard method for evaluation, we adapt an evaluation method for unsupervised learning applicable to the data at hand. Evaluation in supervised learning is a well-defined task, however for unsupervised learning, this becomes challenging since there is no concept of classified label to compare against the ground-truth label. To this end, we use nLUs as a criteria to determine if two children should be grouped together in the ground-truth. We define a threshold $\epsilon$ for a difference in nLUs between two children. We then report the performance as this threshold varies.

In our application, the performance of each child is expressed as a combination of multiple learning patterns. Each learning pattern is an expression of the dominant performance characteristics with respect to skills in TOBY. Thus based on the recurrence of memberships in different learning patterns, we compute the Jaccard index (Real and Vargas, 1996) as the measure of similarity between different learning patterns. Two learning patterns are similar if the Jaccard index between them is greater than a threshold $\rho$. The difference in nLUs between child $i$ and child $j$ is represented as

$$d(i, j) = |nLU(i) - nLU(j)|$$

and if the learning pattern that child $i$ follows is given by $LP(i)$, the Jaccard index between learning patterns $LP(i)$ and $LP(j)$ is given as

$$J(LP(i), LP(j)) = \frac{|LP(i) \cap LP(j)|}{|LP(i) \cup LP(j)|}$$

where $|LP(i) \cap LP(j)|$ is the number of intersecting children and $|LP(i) \cup LP(j)|$ is the number of unique children who follow the two learning patterns. For the purpose of our experiments we consider the value of $\rho$ as zero.
5.5 Results

5.5.1 Part I - dataset A

We run LPGM model on both the datasets A and B. Latent patterns are obtained after 2000 iterations of Gibbs sampling. For dataset A, LPGM detected 8 latent factors, while for dataset B it identified 22 factors as shown in Figure 5.6. On these results we perform exploratory and qualitative analysis. The connection between children and the learning pattern for dataset A is shown in Figure 5.8.

Exploratory analysis  Based on each learning pattern discovered in dataset A using LPGM method, we make observations related to the connectivity and co-occurrence of skills in the 8 learning patterns discovered (referred to as Gi, where i = 1:8). Across all the groups, language skills (both receptive and expressive) tend to require more LUs to master; while imitation seems to be mastered with less. As seen in Figure 5.6a, G1 shows a prominence of a combination of language skills. Hence children who follow this pattern may need more learning opportunities in language skills as compared to the rest of the syllabus. Similarly in G2, it can be noticed that some receptive language skills are strongly expressed. In G3 a different set of receptive language skills are more prominent. G4 and G5 identify groups of more complex expressive and receptive language skills respectively. G6 identifies the simpler of language skills (both expressive and receptive). The last but one latent pattern broadly identifies a combination of skills from the entire syllabus. In G7 motor imitation, sensory and receptive language skills are more pronounced. The discovered learning patterns thus have definite characteristics that can be used to describe the members that follow it.
5.5. Results

Figure 5.6: Latent Patterns obtained using IBP dataset A (left) and dataset B (right).
5.5. Results

Figure 5.7: Membership patterns obtained using IBP on dataset A (left) and dataset B (right).
5.5. Results

Figure 5.8: Membership of children in dataset A with the latent patterns detected using IBP. The smaller blue circles indicate the children while the larger circles indicate the latent pattern. Each child is connected to at least one latent pattern with a thread. The thickness of this thread is directly proportional to the probability that the child follows a certain latent pattern. There are many children who follow latent pattern G7 with a low probability. Hence after thresholding to discard weak connections, we find that G7 is a redundant group.

Figure 5.9: $\ell_0$-norm for latent patterns using methods LPGM (left) and NMF (right)
5.5. Results

![Entropy plots for latent patterns using methods LPGM (left) and NMF (right) for datasets A and B.](image)

(a) dataset A

(b) dataset B

Figure 5.10: Entropy for latent patterns using methods LPGM (left) and NMF (right)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>LPGM</td>
<td>NMF</td>
</tr>
<tr>
<td>average $\ell^0$-norm</td>
<td>17.87</td>
<td>14.50</td>
</tr>
<tr>
<td>average entropy</td>
<td>1.0292</td>
<td>0.9673</td>
</tr>
</tbody>
</table>

Table 5.3: average $\ell^0$-norm and entropy values for dataset A and B. These measures are computed for LPGM and NMF methods.

Qualitative analysis We compare the performance of LPGM with two standard methods: NMF and K-means. We make this comparison by observing the $\ell^0$-norm and entropy of latent patterns (Figure 5.9, 5.10) and by computing the F1-score of the results.

We compute the $\ell^0$-norm and entropy of the latent factors to observe the sparsity and information content of the factors respectively. We make this comparison by observing the results of NMF at the same number of latent factors as estimated by the LPGM method. A sparse latent factor and lower entropy indicates a less
random, distinct selection of skills that are pronounced in each latent factor, which is desirable. We ignore K-means for this analysis as it is not a factorial approach.

Further, we compute the F1-score and compare the performance. LPGM yields the number of latent patterns ($K$) automatically. On the other hand, while applying NMF and K-means prior knowledge regarding the model dimensions is required. Hence in case of these algorithms, we compute the F1-scores for a range of values of $K$ close to the $K$ value as inferred by LPGM. We show the comparison of per-
5.5. Results

Tasks of TOBY K learning patterns

(a) Latent factors discovered by our algorithm. Each column represents a learning pattern. Within a learning pattern the skills that are represented with higher weights are shown in darker shades.

(b) Memberships of the children. Each child is allowed to follow multiple learning patterns with different weights. Each row represents a combination of learning patterns followed by the child. The learning pattern that is followed by a child with a higher weight is shown in darker shades.

Figure 5.12: Latent factors and memberships

formance as the value of threshold $\epsilon$, which is the difference between the nLUs in ground-truth, changes. For dataset A, we show the F1-score for NMF and K-means for 8 latent factors, along with the variation in F1-score for those algorithms for 4, 6 and 10 latent factors (Figure 5.11a). For dataset B the F1-score for the mean value of 22 latent factors is shown for NMF and K-means, along with the variation in F1-score for the same algorithms for 18, 20, 24 and 26 latent factors (Figure ??b).

5.5.2 Part II - dataset B

We run the model on the dataset of 542 children and after 2000 iterations of Gibbs sampling we obtain 26 latent factors. The latent factors and memberships are shown in Figure 5.12. In our study, latent factors are referred to as learning patterns. The interpretation of a learning pattern and further analysis of them is explained in this section. Each child is allowed to follow one or more learning patterns in our experiment. We study their memberships to observe clusters of children that perform similarly on the syllabus of TOBY. We further study the behaviour of the algorithm towards outliers.
5.5. Results

5.5.2.1 Learning Patterns

The matrix $W$ in our experiment denotes the latent factors. Each column in this matrix represents a factor such that each entry is a weight associated with a skill in TOBY syllabus. In our study, we use LUs to master a skill to identify the latent factors amongst the cohort. Different skills are represented with varied weights in different factors. If a child consumed higher LUs for mastering a skill, it indicates that the child might have found the skill relatively difficult to learn. Vice versa if a child consumes low LUs to master a skill, it might indicate that the skill was at a relatively easier level of difficulty for the child. The difficulty level is hence reflected in the weights of the skills in the latent factors. Therefore we refer to the latent factors as 'learning patterns' for difficult skills. These learning patterns constitute composition of skills that are found most difficult by a group of children. From our knowledge about the syllabus, an analysis of the learning pattern is capable of identifying possible latent relationships among inter-category skills.

We also observe that the skills in TOBY are arranged in hierarchy based on their complexities. Additionally, the skills are grouped into four skill categories. Each category has certain skills that are open to attempt at the start. As the child masters a skill, the next skill in the hierarchy is opened up. Thus the skills are not attempted randomly, rather they follow a predefined sequence with a limited choice given at each stage. Hence learning occurs in a sequential manner, where less complex skills are learnt before progressing towards the more complex ones. Figure 5.13 shows organisation of skills in the syllabus tree. In order for a skill to be available for learning all the pre-requisite skills must be mastered.

Each identified learning pattern is a $[1 \times 34]$ dimensional vector of weights corresponding to each skill. In each learning pattern, a difficult skill is represented by a higher weight. Similarly, the easier skills have smaller weights. A learning pattern is thus the integration of weight compositions of skills. These patterns are normalised and a minimum threshold of 0.1 is applied to eliminate negligibly represented skills. The resulting vectors are then converted to a multinomial distribution. The highest weighing skills from each learning pattern vector are then represented using tag cloud visualisation as shown in Table 5.4.

Tag cloud visualisation technique enhances the comprehensibility of information. It
Figure 5.13: iPad Syllabus tree for TOBY. The skills are hierarchically arranged based on complexity and per-requisites. This tree is a smaller version of the original syllabus tree. We only focus on the skills discussed in this study.

does so by displaying information in a font size that is directly proportional to the weights with which this information is associated with. Thus information is filtered and represented in a readily understandable form. In Table 5.4, up to four of the highest weighing skills are displayed using tag cloud visualisation. Now, further inquiry into the dominant skills within each pattern is anticipated.

Each skill falls under a skill category: imitation, sensory, receptive language and expressive language. Hence, in addition to learning patterns, a higher level interpretation of category relativity can be analysed. This opens a window to observe any latent connections between the skill categories. The algorithm we employ in this study discovers 26 learning patterns. The description of each learning pattern in tag cloud visualisation is detailed in Table 5.4.
### Table 5.4: Learning Patterns Description

<table>
<thead>
<tr>
<th>Tag Cloud Visualisation</th>
<th>Description</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Receptive matching of nouns for household items.</td>
<td>Receptive Language</td>
</tr>
<tr>
<td>mr–noun–expand–...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receptive matching of nouns for animals.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Receptive matching of nouns for toys.</td>
<td>Receptive Language</td>
</tr>
<tr>
<td>mr–noun–expand–...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receptive matching of nouns for outdoor items.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mr–noun–expand–...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receptive matching of nouns for personal items.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Receptive matching of nouns for body parts.</td>
<td>Receptive Language</td>
</tr>
<tr>
<td>mr–noun–expand–...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receptive matching of nouns for food items.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Receptive matching of nouns for clothes.</td>
<td>Receptive Language</td>
</tr>
<tr>
<td>mr–noun–expand–...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receptive matching of nouns for outdoor items.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Expressive matching of nouns for food items.</td>
<td>Expressive and Receptive Language</td>
</tr>
<tr>
<td>me–noun–expand–...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expressive matching of nouns for clothes.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mr–noun–expand–...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expressive matching of nouns for outdoor items.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>me–noun–expand–...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expressive matching of nouns for household items.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 5.4: Learning Patterns Description

<table>
<thead>
<tr>
<th>Tag Cloud Visualisation</th>
<th>Description</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td><strong>me–noun–expand–clothes</strong>&lt;br&gt;<strong>mr–noun–expand–animal</strong>&lt;br&gt;<strong>mr–noun–expand–body–part</strong>&lt;br&gt;Expression matching of nouns for clothes.&lt;br&gt;Receptive matching of nouns for animals.&lt;br&gt;Expression matching of nouns for body parts.</td>
<td><strong>Expressive and Receptive Language</strong></td>
</tr>
<tr>
<td>7</td>
<td><strong>me–noun–expand–body–part</strong>&lt;br&gt;<strong>mr–noun–expand–personal–item</strong>&lt;br&gt;<strong>mr–noun–expand–food</strong>&lt;br&gt;Expression matching of nouns for body parts.&lt;br&gt;Receptive matching of nouns for personal items.&lt;br&gt;Receptive matching of nouns for food items.</td>
<td><strong>Expressive and Receptive Language</strong></td>
</tr>
<tr>
<td>8</td>
<td><strong>vsm–colour</strong>&lt;br&gt;<strong>mr–noun–expand–personal–item</strong>&lt;br&gt;<strong>mr–noun–expand–household–item</strong>&lt;br&gt;<strong>mr–noun–expand–food</strong>&lt;br&gt;Visual sensory matching for colour.&lt;br&gt;Receptive matching of nouns for personal items.&lt;br&gt;Receptive matching of nouns for household items.&lt;br&gt;Receptive matching of nouns for food items.</td>
<td><strong>Sensory and Receptive Language</strong></td>
</tr>
<tr>
<td>9</td>
<td><strong>vsm–id</strong>&lt;br&gt;<strong>mr–noun–expand–personal–item</strong>&lt;br&gt;<strong>mr–noun–expand–household–item</strong>&lt;br&gt;<strong>mr–noun–expand–food</strong>&lt;br&gt;Visual sensory matching for identical objects.</td>
<td><strong>Sensory</strong></td>
</tr>
<tr>
<td>10</td>
<td><strong>vsm–nonid</strong>&lt;br&gt;<strong>mr–noun–expand–personal–item</strong>&lt;br&gt;<strong>mr–noun–expand–household–item</strong>&lt;br&gt;<strong>mr–noun–expand–food</strong>&lt;br&gt;Visual sensory matching for non identical objects</td>
<td><strong>Sensory</strong></td>
</tr>
<tr>
<td>11</td>
<td><strong>mr–noun</strong>&lt;br&gt;<strong>mr–noun–expand–personal–item</strong>&lt;br&gt;<strong>mr–noun–expand–household–item</strong>&lt;br&gt;<strong>mr–noun–expand–food</strong>&lt;br&gt;Receptive matching of basic nouns.</td>
<td><strong>Receptive Language</strong></td>
</tr>
</tbody>
</table>
### 5.5. Results

Table 5.4: Learning Patterns Description

<table>
<thead>
<tr>
<th>Tag Cloud Visualisation</th>
<th>Description</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 gmi-v-1</td>
<td>Gross motor imitation visible skills.</td>
<td>Imitation</td>
</tr>
<tr>
<td>14 me-noun</td>
<td>Expressive matching of basic nouns.</td>
<td>Expressive language</td>
</tr>
<tr>
<td>15 oral-b me-noun-expand-food</td>
<td>Oral imitation of blowing. Expressive matching of nouns for food items.</td>
<td>Imitation and Expressive Language</td>
</tr>
<tr>
<td>16 oral-va</td>
<td>Oral imitation involving voice with action.</td>
<td>Imitation</td>
</tr>
<tr>
<td>17 oral-nv</td>
<td>Oral imitation involving face actions with no voice.</td>
<td>Imitation</td>
</tr>
<tr>
<td>18 mr-noun-expand-body-part</td>
<td>Receptive matching of nouns for body parts.</td>
<td>Receptive Language</td>
</tr>
</tbody>
</table>
### Table 5.4: Learning Patterns Description

<table>
<thead>
<tr>
<th>Tag Cloud Visualisation</th>
<th>Description</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 me-noun-expand-toy</td>
<td>Expressive matching of nouns for toys. Basic fine motor imitation skills.</td>
<td>Imitation and Receptive and Expressive language</td>
</tr>
<tr>
<td>fmi-b mr-noun-expand-personal-item</td>
<td>Receptive matching of nouns for personal items.</td>
<td></td>
</tr>
<tr>
<td>mr-noun-expand-body-part</td>
<td>Receptive matching of nouns for body parts.</td>
<td></td>
</tr>
<tr>
<td>21 awo mr-noun-expand-clothes</td>
<td>Actions with objects. Receptive matching of nouns for clothes.</td>
<td>Imitation and Receptive and Expressive language</td>
</tr>
<tr>
<td>me-noun-expand-household-item</td>
<td>Expressive matching of nouns for household items.</td>
<td></td>
</tr>
<tr>
<td>22 oral-vowel oral-cv</td>
<td>Oral imitation of vowels. Oral imitation of consonant-vowel sound combinations.</td>
<td>Imitation</td>
</tr>
<tr>
<td>23 me-noun-expand-household-item</td>
<td>Expressive matching of nouns for household items. Oral imitation of consonant-vowel-consonant combinations.</td>
<td>Imitation and Receptive and Expressive language</td>
</tr>
<tr>
<td>oral-cvc mr-noun-expand-personal-item</td>
<td>Receptive matching of nouns for personal items.</td>
<td></td>
</tr>
<tr>
<td>oral-sym</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 oral-vowel me-noun-expand-body-part</td>
<td>Oral imitation of vowels. Expressive matching of nouns for body part.</td>
<td>Imitation and Expressive language</td>
</tr>
<tr>
<td>me-noun-expand-animal</td>
<td>Expressive matching of nouns for animals.</td>
<td></td>
</tr>
</tbody>
</table>
### Table 5.4: Learning Patterns Description

<table>
<thead>
<tr>
<th>Tag Cloud Visualisation</th>
<th>Description</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>Visual sensory matching of non identical objects.</td>
<td>Sensory</td>
</tr>
<tr>
<td>vsm-nonid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Expressive matching of nouns for personal items. Oral imitation involving face actions with no voice.</td>
<td>Imitation and Expressive language</td>
</tr>
<tr>
<td>me-noun-expand-personal-item oral-nv</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Description of Findings on Learning Patterns**  
On a category level, the following observations are made from Table 5.4. All Imitation, Receptive and Expressive language skills are quite dominant and related. They appear in most groups either by themselves or together with the other two categories: Imitation (12/26), Receptive (15/26), and Expressive (10/26). On the contrary, Sensory appears in only 4 out of 26 groups, in 3 of which it appears by itself. Skills that belong to different skill categories co-occurring as a group in learning patterns may indicate a possible connection between them. As imitation is perhaps the most common form of human learning, it is expected to see that learning language involves imitation greatly. For example, Imitation category involves skills that teach oral imitation of vowels, consonants and their combinations. Whereas Language category includes skills that involve either identifying objects with a label (receptive) or vocalising the label of an object (expressive). Hence some skills in imitation are in a way prerequisite for the skills in language category. In fact, with respect to the 6 groups involving both Imitation and Expressive Language, Table 5.4 shows that 4 of them have difficulties with oral imitation (groups 15, 23, 24 and 26).

It is also interesting to observe that Sensory mostly presents by itself, except in group 8 where it pairs with Receptive language. While Imitation is more of a social skill, Sensory perhaps belongs to the cognitive category. In Sensory, children are required to match photos of objects. For example, provided with 3 choices (apple, cat, house), they are required to match one of them to a target (e.g. another apple).
5.5. Results

Given Sensory matching is a basic starting skill in most ABA programs Venkatesh et al. (2013), and it is also open to the children at the start of TOBY, there is a possibility that some students may have struggled right at the start. That could prevent them from proceeding much further. Alternative approaches or perhaps teaching of even more basic or pre-requisite skills may be required.

Group 8 (Sensory and Receptive Language) shows that the children have difficulties with matching more abstract concepts (e.g. \textit{vsm-colour}: visually matching different objects of same colours). It is important to note that in TOBY, the matching of colour skills to have become available only after children pass the pre-requisites matching skills (\textit{vsm-id} and \textit{vsm-nonid}: visual matching of identical/nonidentical objects). These same set of children were found struggled with recognising objects based on sensory cues (Receptive Language). For example, a chair could have different textures and colours, but it can still be identified as a chair due to its structure. Trained therapists may need to find alternative strategies to guide this group of children through these abstract concepts.

**Evaluation for Learning Patterns**  So as to compare with the results obtained by using non-negative matrix factorisation (non-Bayesian) we observe the entropy and the $l_0$ norm of the learning patterns obtained from both methods, i.e., NMF and Bayesian non-negative matrix factorisation with an IBP prior. Entropy and $l_0$ norm indicate the information content and sparsity of the factors respectively. We observe these values to judge the quality of the learning patterns detected by both the methods. A desired learning pattern is sparse with entropy values greater than zero, thus selecting distinct skills in every learning pattern, reducing randomness. The results are shown in Figure 5.14. Figures 5.14a and 5.14c show the entropy and $l_0$ norm for NMF, whereas Figures 5.14b and 5.14d show the same for LPGM. It is evident from observing the results that the learning patterns detected by LPGM match the requirements of a desired learning pattern. In-depth observation of learning patterns detected by the NMF show many learning patterns with a single skill selected in each pattern. This does not concur with our aim to detect a group of skills that a cluster of children are struggling with.

We extend this analysis to observe if the learning patterns detected remain stable, by collecting the data once again in October 2015 from TOBY Playpad. As time
5.5. Results

Figure 5.14: Entropy and $l_0$ norms for Learning Patterns

passes we are likely to observe that: there are new children who begin intervention using TOBY Playpad and the children who are already undergoing the intervention experience progress compared to the last study. The data collected in October 2015 contains 620 usable samples that satisfies our initial criteria stating that each child must have mastered at least one skill from the syllabus. We use LPGM on this dataset and observe the learning patterns detected. 30 learning patterns were detected compared to 26 from the previous data. In the following paragraph, we present a category level comparative analysis of the learning patterns discovered.

A category level analysis of the discovered learning patterns from the updated dataset reveals the following observations. Similar to the learning patterns in Table 5.4, all Imitation, Receptive and Expressive language skills are dominant and related. They once again appear either by themselves or with the other two categories: Imitation (14/30), Receptive (13/30), and Expressive (10/30). Sensory follows on similar to the previous findings and appears in only 4 out of the 30 learning patterns, in which it appears by itself in 2 of them. Hence we note that the discovery of these patterns remains quite stable even with inclusion of additional data points from new children and progress is made by already participating children.
5.5. Results

5.5.2.2 Subgroup Detection

The matrix $H$ represents the membership of each child to the learning patterns. Each entry in this matrix constitutes the weight with which each learning pattern is represented in a child. In our experiment, each child owns shared membership to one or more learning patterns with varied weights. We analyse matrix $H$ to observe learning patterns that are most popular. Additionally, we find clusters of children who perform similarly in terms of learning patterns for difficult skills. Finding clusters of children with similar performance can be crucial for determining possible solutions for an individual if they happen to get stuck in the syllabus at some stage. In such a scenario, we could observe the cluster that the individual belongs to and identify other children who might have faced a similar situation. The individual can then be given more learning opportunities in skills that are found collectively difficult by the cluster to make way for progress.

Description of Findings on Membership Patterns  The connectivity between children and learning patterns can be seen in Figure 5.15. This figure is obtained by applying a threshold to matrix $H$ to a value of 0.35. The larger circles with the identification number are the learning patterns and the smaller ones represent the children. Each small circle is associated with identification for the child $c_i$, where $i = 1, 2, ..., 542$ are the number of children. The children are connected with one or more learning patterns with a string. As the 26th learning pattern is connected lightly with many children, on placing a threshold on the weight with which each child is connected to a learning pattern, the 26th learning pattern becomes disconnected and thus redundant. Figure 5.15 indicates that Language (for example, nodes 1 to 8) appears to link to the most number of children. That could be because communication deficits are one of the traits of children with ASD. That is evident in their difficulties of labelling or identifying objects.

From Figure 5.15 we also observe that there are children who are not connected to any learning pattern, grouped in the centre. Small weights distributed uniformly across many learning patterns results in the non-conformity of these children in terms of membership. This could mean two things which we elaborate in the following paragraphs.
Figure 5.15: Memberships associated with each learning pattern. The larger nodes indicated with the numbers 1 to 26 are the learning patterns discovered shown in different colours. The smaller nodes represent each child and is identified by $c_i$, where $i = 1 : 542$. Each child is connected to one or more learning patterns with a weight. The edge connect the child node to the learning pattern node takes on the colour of the learning pattern that the child is connected to with the highest weight. The nodes that are not connected to any learning pattern are gathered at the centre. The node representing learning pattern 26 is also seen disconnected at the centre.
Firstly, it is observed that the learning patterns are associated with the difficulty in approach of the skill for children. Hence the distribution of small weights over many learning patterns may indicate that the children found all the skills they attempted relatively easy. When a threshold is applied on matrix H these children thus remain disassociated with the other subgroups. Secondly, the children may find a group of skills difficult that are unique to them. If a similar behaviour is not found in many children the group of skills may not be recognised as a learning pattern. This may result in these children appear as non-connected. A third possibility maybe that the children attempted very few skills that are not significantly represented in any of the learning patterns leading to the children being represented as non-connected.

Figure 5.16 shows the amount of sharing of learning patterns in terms of memberships. It can be seen from the figure that children share learning patterns 1 to 7 more often among each other. From the syllabus it can be noted that learning patterns 1 to 7 are dominated by expanded language skills, receptive as well as expressive. The abilities required for mastering these skills are a part of language learning. While expanded receptive language skills require identifying the object based on an audio label, expanded expressive language skills involve vocalising the label of the displayed object. Hence, our observation of children sharing these patterns more often is coherent with the relativity in terms of syllabus of TOBY.
5.5. Results

Clusters of Children with ASD  We compute the cosine similarity for matrix $H$. The cosine similarity is shown in the Figure 5.17. From the figure we observe 11 major clusters of children reflecting the dominating learning patterns in terms of membership. A cluster can be interpreted as a group of children who find certain skills in TOBY difficult to master. These clusters are described in terms of highest weighing learning patterns as shown in Table 5.5. Further investigations into these clusters by trained therapists may yield important information for intervention.

To understand the clusters better let us consider cluster $a$. For convenience of notation let us represent the learning patterns as $LP_i$ where $i = 1 : 26$. The individuals that belong to cluster $a$ show the leading learning pattern as $LP3$. In other words, each individual in this group follows $LP3$ with the highest weight as can be observed from matrix $H$. It can be seen from Table 5.4 that these individuals found the skills of advanced receptive language most difficult to master. Upon further inquiry, it was observed that the individuals in cluster $a$ are also similar in terms of the second and third leading learning patterns, which are $LP1$ and $LP2$. As seen from Table 5.4, $LP1$ and $LP2$ consist of advanced receptive language skills that are more complex in nature. Children finding these skills difficult together are
as a result clustered together in cluster \( a \). Expanding on this explanation, Table 5.5 shows the leading learning patterns for each cluster in Figure 5.17.

**Evaluation of Membership of Children to Learning Patterns** When an intervention program with detailed instructions is allowed to be downloaded at the home of the child undergoing therapy, the data collected from such a program comes with a trade-off. On one hand, the families experience convenience and increase in facilities, while on the other hand researchers despite benefiting from the large size of the data are limited by the details of the medical diagnosis and condition of the child undergoing the therapy. We are thus motivated to generate ground-truth from observing other factors in overall performance of each child. With this interest, we use a measure called normalized Learn Unit (nLU) as a reference for ground-truth. A nLU is defined as the number of learn units required to master 1\% of the entire syllabus as introduced by Moore *et al.* (2015) and mathematically described by Vellanki *et al.* (2014).

We use the Jaccard index on nLUs of children to decide whether two children should be placed in the same group in ground-truth. We decide a threshold \( \epsilon \) over the Jaccard index; if the difference between the nLUs of two children is less than \( \epsilon \) then the two children are similar in ground-truth. Further, we observe how the F1 score changed when this threshold \( \epsilon \) varies. Additionally, we compare this performance to the memberships obtained by using k-means and NMF. Figure 5.18 shows this comparison. LPGM is supported by a Bayesian prior which helps us to automatically discover the number of learning patterns \( K \). This advantage cannot be harnessed when the methods used are parametric; we have to manually provide \( K \). Hence, while computing the F1 score for k-means and NMF, we do so for the value \( K = 26 \) (as automatically discovered by LPGM) and for four other values of \( K = 22, 24, 28, 30 \). The difference is shown as the error margin in the results. We observe that using LPGM gives a better performance as compared to the other methods.

**5.5.2.3 Case for outliers**

Children work on the skills of TOBY at an individual pace, demanding varied number of opportunities in order to master each skill. We noted in the preceding sections
5.5. Results

Figure 5.18: F1 score for change in epsilon $\epsilon$ (difference between nLUs) between children in cohort.

that these number of opportunities are accounted in the number of LUs required to master the skill. Instances occur when the child is unable to master a skill and the LUs start accumulating by a large number. This results in generation of outliers. For our study, we define outliers as the instances of extremely large LUs for a skill in case of an individual.

We approached the problem of identifying outliers by applying a threshold to the membership weights in matrix H. We isolated 28 children with outliers and observed their LUs across all skills. Figure 5.19a is a comparative plot of accumulated LUs across each skill for 28 children with outliers and 28 children without outliers. A significant difference in accumulated LUs can be observed between the children with outliers and without outliers. The spikes in the histogram for outlier group may indicate uneven performance, especially in the skills of sensory and receptive language. Upon further observation, we found children who individually contributed more than 500 LUs in certain skills. Instances of extremely high LUs (more than a 1000 LUs) in individual children can be noted for solo skills of sensory and receptive language. Solo skills require a response like choosing a proper answer from the options displayed on the screen in the form of images. On the contrary partner skills like imitation and expressive language require the child to either vocalise the response or enact a given action, making it necessary for the caregiver to record the response of the observation. The children may attempt solo skills without any assistance from the caregiver for reinforcement or prompting as this job is done by the program automatically. As children with ASD present symptoms of easy distraction toward specific sensory stimulus, we reckon that these children maybe feeling
5.5. Results

(a) Comparison of accumulated LUs to master a skill between outlier and non-outlier groups.

(b) Histogram of weights in matrix H for outlier group.

Figure 5.19: Properties of the outliers.
reinforced even with a display of incorrect response feedback. Without caregiver supervision, this behaviour may go uncontrolled resulting in the spikes in LUs. For these reasons the ambiguity in outlier data calls for careful further inquiry. The algorithm assigns disproportionately heavy weights to the outliers which can be noted in the histogram of H matrix before normalisation as seen in Figure 5.19b. Hence they can be easily separated by applying the threshold to the matrix.

5.6 Concluding Remarks

In this chapter, for the first time we have presented learning patterns deduced automatically from data mining and machine learning methods using intervention data recorded for children with autism. Hence, the significance of our work has been focused on the contributions it makes for the Autism Spectrum Disorder community. In particular, we discover learning patterns associated with the cohort on TOBY. A learning pattern is a group of skills that were together found difficult to master by a child. Often in autism literature, the skills are inter-related and may require similar abilities in order for them to be learnt. One example from within TOBY syllabus and our study is, we have found that the ability to distinguish between sensory cues is essential for learning skills from the Sensory as well as Receptive Language categories. Hence, a child with ASD, showing disability for sensory cues may find selective skills within both these categories difficult to master. Such findings grounded in data-based discovery not only contribute to the understanding of the inter-relativity between the skills, but also allow us to find clusters of children with similar disabilities. We hope that the autism community (researchers as well as therapists and caregivers) may benefit from this novel take.

The varied nature of syllabus that is prescribed for individual children with ASD calls for analysis of each child. Once the analysis is completed and the child is guided through the syllabus, a dynamically adapting program that recommends changes in the syllabus on the basis of individual child’s performance is most favourable. A research into methods for developing such programs is essential and unexplored. Our findings in this research contribute to the development of framework for such a program. We observe that given a syllabus that is structured and data that is tracked from children with ASD who participate in learning, Bayesian models can be
used to observe the learning patterns and the association of the participants with the learning patterns. Within these set of models is already established is that Bayesian nonparametric are preferred over parametric methods as they overcome the model selection problem. The data observed in applications under given circumstances is growing as the children are mastering the syllabus one skill at a time. Many data elements referring to the skills that are yet unavailable cause missing data to occur. It becomes necessary to build a model that can learn from such data. We use the nonparametric Bayesian matrix factorisation called the linear Poisson gamma model (LPGM) with an Indian buffet process (IBP) prior and we extend this model so that it is missing data inclusive.

The experiment indicates the presence of 26 learning patterns and gives us a membership pattern showing the association between each child and learning pattern. Further examination of the associations with the help of prior knowledge regarding syllabus of the computer assisted learning tool, reveals hidden connections between parts of syllabus. The cosine matrix indicates the similarity between the children based on how they performed on each skill. This enables us in grouping the children who behaving similarly with respect to learning skills (as far as the syllabus used is considered). In comparison with parametric counterparts, we observe from the results that not only are the learning patterns detected by LPGM qualitatively more desirable, their membership patterns are justified based on F1 score as well. Further, a category level similarity analysis of the learning patterns discovered from an incrementally updated TOBY dataset shows that the learning patterns discovered remained fairly stable. These results lead us toward a novel direction in personalised computer-assisted therapy for children with ASD.

The clustering as suggested in our work could be crucial in predicting an efficient path for syllabus completion for each individual. When a personalised course of learning is prescribed it routes additional learning opportunities in weak areas making a stronger foundation. Additionally, when an individual is stuck, with the knowledge of inter-related skills and the individuals performance in those skills, the child may be automatically routed via those skills which are pre-requisites for mastery. The discovery of clusters of children with similar difficulties also enables us to examine the course of syllabus of another child from the same cluster, which assists us in identifying the reason behind the child being stuck at a certain skill. Personalisation in syllabus for children with ASD can hopefully make the delivery of therapy with
computer-assisted tools not only easily accessible but also more efficient.
Chapter 6

Understanding Factors affecting Personalisation - User Entry Profiling

6.1 Introduction

In Chapter 5, we addressed the problem of discovering learning patterns and subgroups in children with ASD from their performance data. That was an analysis of their performance after some progress is made. Determining profiles at the onset of intervention is crucial for deciding the course of intervention. By posing as ‘ground-truth’, these profiles can also assist in comparing progress of the children after intervention with respect to before beginning the intervention. This will enable a data-driven, evidence-based analysis of the progress of children. The need for a mechanism to construct such profiles of children based on their characteristics at the onset of intervention is our motivation for this chapter. This problem broadly belongs to research in user profiling, technology in data mining and machine learning community, whose primary aim is to understand user preference (e.g., in recommender system). However, none of the existing work, to our knowledge, has considered user entry profiling under a non-clinical ASD setting as considered in this chapter.
Every patient is capable of showing a different response to standard treatment procedures even for a well-defined clinical condition. Identifying patient profiles can provide a basis for comparison to monitor the progress of the patient based on their characteristics. Howard et al. (1996) discuss the benefits of a patient profiling system for psychotherapy. In doing so, they propose a system of profiling that accommodates individuality of the patient by including an estimate of the expected course of treatment based on their initial characteristics. Mazalan et al. (2012) develop a profiling system for depressive disorders by using their initial screening information and the dosage of medication that was suggested as treatment. They aimed at replacing the manual maintenance of profiles by an automated approach that could be used to dynamically update and selectively change the course of treatment. A similar method has been approached by Payne-Murphy and Beacham (2014) for profiling chronic pain patients. They identified clusters with significant differences among them and suggest the research pathway of tailoring personalised interventions based on the cluster profiles. A recent study by Garnett et al. suggests discovery of profiles among children and adolescents with Autism Spectrum Conditions using an online questionnaire (Garnett et al., 2013). Additionally Garnett et al. mention the usefulness of such profiling in administration of personalised interventions catering to individual needs. The discovery of profiles for children with ASD using computer-based programs, is an open question that needs to be addressed.

In this chapter, we aim to cluster the individuals with ASD based on their characteristics as a means of screening their condition before intervention is administered. Entry profiles as we define, are a set of characteristics individuals posses at the onset of intervention. For example, children within a certain age range, belonging to certain sex and expressing difficulty in similar skill areas. Revisiting the literature in Chapter 2, we note that the expression of ASD in individuals has been observed to have variations based on the sex of the individual (Williams et al., 2008). It is found that the prevalence of ASD in the male population is higher than that of female population (Bartley, 2006). Nevertheless, the expression of ASD in female population is found to be more severe as compared to the male (Dworzynski et al., 2012). Additionally, research has shown that intensive therapy administered at an early age, as young as 2-3 years can improve the chances of faster learning (Dawson et al., 2008). On the contrary, at a later age, therapy has been observed to have fewer positive results (Fenske et al., 1985). We try to integrate these findings in order to
determine if individuals with ASD can be clustered to broader entry profiles.

Technology can be a breakthrough for collecting and analysing clusters of children with ASD. However, an initial discovery of profiles that we label as entry profiles is essential to evaluate the progress as the child traverses through the syllabus of the program. Entry profiles can be used as ground truth for any further analysis into the performance of individuals. Additionally, our future interest lies in automatic discovery of personalised intervention based on performance of the child on the syllabus of the program. Discovery of entry profiles is fundamental for the pursuit of progress evaluation and personalised intervention.

The data obtained from TOBY is mixed-variate (age, sex and performance of entry syllabus skills) and highly correlated in nature, which makes it challenging to deal with. Natural integration of these data is not common among traditional clustering techniques. This creates the need for complex techniques that can integrate the data. Additionally, the data consists missing elements. As the individuals progress through the syllabus, the data grows. If the individual chooses to continue, for example with one skill category and master the skills within the category before progressing to another one, the performance data on certain skills may be unavailable for the time. The framework must hence address the issue of missing data.

We propose the use of Mixed-variate Restricted Boltzmann Machine (MV.RBM) on our data to discover latent entry profiles. Mixed-variate RBMs were proposed recently by Tran et al. (2011) to jointly model various types of variables such as binary, categorical, ordinal and others. This model takes heterogenous input units, where each unit encodes type-specific information, and transforms the complex data into a homogeneous representation. As we will see in this chapter, that MV.RBM by introducing no additional connections, preserves the efficient learning and fast inference properties of RBM. We use the multi-variate information of the children to model entry profiles. The children are further clustered and visualised using K-means and t-SNE, a stochastic embedding based method of projecting higher dimensional data onto a lower dimensional subspace.

In summary, our contributions in this chapter are:

1. To discover entry profiles of children based on their age, sex and performance
on each of the four entry skills of a technology-assisted intervention program, TOBY Playpad. We use the data collected from 491 children for this purpose.

2. To apply mixed-variate restricted Boltzmann machine (MV.RBM) for modeling the heterogeneous data and clustering the children into entry profiles.

3. Qualitative analysis of the discovered entry profiles.

Machine learning of entry profile, or profiling technology, is a challenging problem, but have a wide impact. While our study focuses on autism domain, it has a much wider impact from recommendation system in information retrieval to patient profiling in biomedical research.

6.2 Dataset

The four main skill categories in the syllabus of TOBY Playpad are: Imitation, Sensory, Expressive Language and Receptive Language. When an individual begins the syllabus, they are presented with entry skills in each of the four categories, which the children begin with. As the skills are organised in a hierarchical fashion, the children are allowed to move ahead in a skill category only once the entry skill in that category is mastered. We use the performance of the children on these basic 'entry skills' as a measure of their initial capabilities before they begin actual intervention and make any progress in there.

Data consists of 491 children in this study. Our dataset captures information about the individuals with ASD, including their age, sex and their performance on the entry skills of the syllabus of TOBY Playpad. The inter-quartile range of the children is 3.5 years to 6.7 years. There are 264 boys and 84 girls; the sex of 143 children is unknown.

The number of Learn Units (LUs) required to master each of the four entry skills, one from each category, are determined and used along with the age and sex as features in our framework. We may come across scenarios where the children may opt to move ahead on selected skill category branches, showing limited or no progress in other categories. As a result of those cases, our data is sparse in LUs required to
master the entry skills. In the next section, we will describe the data types we use for each of the features in this mixed-variate dataset and the framework of MV.RBM.

6.3 Framework

In this section, we build upon the foundation of RBM as described in Chapter 2, and explain MV.RBMs.

6.3.1 Mixed-variate restricted Boltzmann machines

RBMs can be derived to model individual data types, where the visible unit layer constitutes variables of the same modality. Xing et al. (2012) introduced a model called dual-wing RBM to handle continuous and Poisson variables simultaneously. To address the multi-variate and heterogenous data at hand, we leverage on the Mixed-Variate Restricted Boltzmann machine (MV.RBM) recently proposed in (Tran et al., 2011). An MV.RBM is technically an RBM, but with heterogeneous visible vector \( \mathbf{v} \), each element \( v_n \) has its own type. The energy function from Equation 2.4.29 is now redefined to be:

\[
E(\mathbf{v}, \mathbf{h}; \theta) = - \left\{ \sum_{n=1}^{N} \beta(v_n) + \mathbf{a}^T \mathbf{v} + \mathbf{b}^T \mathbf{h} + \mathbf{v}^T \mathbf{W} \mathbf{h} \right\}
\]

where \( \beta(v_n) \) is the type-specific function; \( \mathbf{a} \) and \( \mathbf{b} \) retain the same functionality as the biases of the visible and hidden layers and \( \mathbf{W} = [W_{ij}] \) are the weights specified for connections between hidden and visible layers. We note that due to the heterogeneity of the visible unit \( \mathbf{v} \), the normalisation function \( Z(\theta) \) in Equation 2.4.30 is now more complicated to compute due to the sum over the \( \mathbf{v} \). A general strategy is to group the visible vector into \( \mathbf{v} = (\mathbf{z}, \mathbf{x}) \) where \( \mathbf{z} \) is the discrete component and \( \mathbf{x} \) is a continuous component. \( Z(\theta) \) can now be computed as:

\[
Z(\theta) = \int_{\mathbf{x}} \sum_{\mathbf{z}} \exp \left\{ \sum_{n=1}^{N} \beta_n(v_n) + \mathbf{a}^T [\mathbf{x}, \mathbf{z}] + \mathbf{b}^T \mathbf{h} + [\mathbf{x}, \mathbf{z}]^T \mathbf{W} \mathbf{h} \right\} d\mathbf{x}
\]
6.3. Framework

We note that this integration must be finite for the probability model of interest to be well-defined. Lastly the type-specific function $\beta(v_n)$ can tailored for different types of data. While the posterior for the latent variable $h_k$ in Equation 2.4.33 remains easy to compute regardless the type of $v_k$, the data likelihood $p(v_k \mid h)$ becomes complicated when $v_k$ is no longer a binary variable. Again using joint distribution and energy function for the RBM as described in Equations 2.4.28 and 2.4.29 we have:

$$p(v_n \mid h) \propto \exp \left\{ \beta_n(v_n) + a_n v_n + \sum_{k=1}^{K} v_n W_{nk} h_k \right\}$$

Again the MV.RBM belongs to the standard exponential family as a result of which its gradients can be computed as the expectations of the models and CD can be used to estimate the parameters from the data.

### 6.3.2 MV.RBM for entry profiles

The visible parameters are multivariate in our experiment. Hence we consider three types of data as specified in (Tran et al., 2011):

- **Binary** observation $v_n$ for sex. In this case we simply set $\beta_n(v_n) = 0$, hence

  $$p(v_n \mid h) = \tau \left( a_n + \sum_{k=1}^{K} W_{nk} h_k \right).$$

- **Continuous** observation $v_n$ for age. We use the Gaussian variable, setting $\beta_n(v_n) = -\frac{v_n^2}{2\sigma^2}$ where $\sigma^2$ is a fixed variance specified in advance. In this case, $p(v_n \mid h)$ can be shown to be a univariate Gaussian with mean $\sigma^2 \left( a_n + \sum_{k=1}^{K} W_{nk} h_k \right)$ and variance $\sigma^2$.

- **Count** observation $v_n$ for LUs required to master each entry skill. We employ the replicated softmax representation (Hinton and Salakhutdinov, 2009) which allows us to represent repeated counts in an observation such as words in a documents or tasks performed repeatedly within a section. In this case $v_n$ can move beyond from scalars to count vector $v_n \in \{1, \ldots, M\}^{L_n}$ where $M$ is the dictionary size and $L_n = |v_n|$ is the size of $v_n$. Hinton and Salakhutdinov (2009) employs a simple method to represent $v_n$ by ‘replicating’ the $v_n$ onto a vector of
6.4. Experimental Results

$L_n$-dim vector $v'_n = (v'_{n1}, \ldots, v'_{nL_n})$ in the visible layer where $v'_{nj} \in \{1, \ldots M\}$, each is specified by softmax unit and then ‘ties’ the parameter through them; in other words, $p(v'_{nj} = m | h) \propto \exp\left(a_n + \sum_{k=1}^{K} W_{nk} h_k\right)$. We note that due to this replication procedure and parameter typing, the posterior $p(h_k = 1 | v)$ will also be updated to account for the replicated visible units.

Our data presents itself with missing data. This occurs when the child is yet unable to master the skill. While registering with TOBY a caretaker may render the child’s information confidential and choose not to enter the age or sex of the child. This might also generate instances of missing data. Our data is complete in age, but for the other visible variables we substitute zero where the data is missing for simplicity. The latent posterior equation of RBM consists of the product term involving $v_n$, which results in the model to account for no statistics from the missing elements.

We employ CD (Hinton, 2002) for learning with learning rates varying based on the data type. The parameters are updated after every 100 children, which is the size of our mini-batch, and we terminate learning in 100 data sweeps. The number of hidden units ($K$) is determined to be 10. The reconstruction error is observed to gradually reduce.

K number of posteriors are computed for each child. Using t-SNE (Van der Maaten and Hinton, 2008) we then project the posterior values of hidden parameters to a two-dimensional space to observe the clusters of profiles. K-means is used on this projection to find clusters of children sharing similar entry profiles. We present the results in the next section.

6.4 Experimental Results

In this section, we observe the clusters and the entry profiles to explore their meaningfulness. The clusters obtained are shown in Figure 6.1. In the projected space we use K-means to discover 5 clusters. The clusters represent individuals belonging to respective entry profiles. We then observe the individual properties of each cluster. We would like to note at this point that the LUs required to master each skill are proportional to the amount of difficulty the individual faced while master-
6.4. Experimental Results

Figure 6.1: t-SNE projection of the posterior hidden values. Clusters are obtained by K-means.

In the subsequent Figure 6.2a to Figure 6.2e the properties of the individuals belonging to the entry profiles 1 to 5 are collectively shown. The number of LUs required to master each entry skill are shown as a box plot. We observe that entry profile 1 (Figure 6.2a) represents male individuals who found the entry skill of sensory skill category most difficult. The entry profile 2 (Figure 6.2b) includes females and the individuals whose sex is missing in the data who struggled most with the receptive language entry skill. The entry profile 3 (Figure 6.2c) constitutes of male individuals who struggled with sensory more than imitation. The entry profile 4 (Figure 6.2d) constitutes individuals who found sensory more difficult than imitation similar to entry profile 3. It can also be seen that these individuals collectively struggled with sensory entry skill more than their male counterparts in entry profile 3. Entry profile 5 (Figure 6.2e) includes male individuals who found sensory and receptive language entry skills most difficult to master. Additionally, the age histogram is also shown for each entry profile.
6.4. Experimental Results

Figure 6.2: Visualisation of clusters of individuals belonging to corresponding entry profiles along with collective properties of entry profile (age, sex and Learn Units required to master each entry skill)
6.5 Discussion

The advantage of using technology assisted therapeutic interventions is rigorous and effortless recording of data, which allows more freedom of analysis. Nevertheless, we are often faced with challenges. Each child entering the intervention begins at a different age and a different set of abilities in terms of skills learnt compared with one another. It is crucial to map each child onto an entry profile at the beginning if the aim is to dynamically adapt the syllabus changing the course of intervention according to the performance and capabilities of the child. Determining the entry profile of a child can enable us to allow changes in the intervention to be made by giving the child more opportunities in areas where they are struggling. An entry profile can also be used as a ground truth for monitoring the progress of the child. The work presented in this chapter is antecedent attempt to discover entry profiles that will enable comparative study of progress as presented in Chapter 7.

One of the main challenges that we continually face with the data is the incomplete nature of it. Missing elements in the current study were replaced with zero values for simplicity. A more accurate approach we are headed toward is to model each child as a separate MV.RBM. The model will then be a family of MV.RBMs of different sizes. This approach has also been suggested by Hinton and Salakhutdinov (2010; 2007) as a more accurate alternative for applications where imputation of missing data is not an option. Due to the extremely random nature of skill abilities in children with ASD, imputation of the number Learn Units required to master a skill which is not yet attempted would be unjust. Additionally due to binary assignment for sex, the individuals whose sex in currently unknown are clustered together as either male or female (in this study female). The modelling of entry profiles as a family of MV.RBMs would thus be a more accurate way of dealing with the missing data.
6.6 Concluding Remarks

Deciding the number of entry profiles is also an open question which ought to be explored in more depth. We present 5 broader entry profiles in this study. However, it is possible to define more number of entry profiles with finer differences between them. This is seen in Figure 6.3. The same children have been clustered into 10 entry profiles in this figure. We further observed finer variations in the children’s struggles with different entry skills in this clustering. This calls for further analysis.

![Figure 6.3: Visualisation of the same individuals shown as 10 clusters corresponding to entry profiles.](image)

6.6 Concluding Remarks

The profiling of individuals with ASD is a challenging and ongoing research question. Profiling is a foundational step in determining dynamic personalisation of therapeutic interventions that can be administered based on the child’s performance. Our discovery of meaningful broad entry profiles may suggest a novel approach to the problem of profiling children with ASD. We present a framework for discovery of entry profiles for children with ASD using mixed-variate Boltzmann machine (MV.RBM). The complexities involved with non-homogeneous data could be incorporated by using this framework. MV.RBM has the capacity of modelling data that is incomplete when it is designed as a family MV.RBMs, hence making it one of the best choices for the data at hand.
Chapter 7

Tracking Change in Performance from Entry Profiles to Performance Profiles

7.1 Introduction

In the preceding chapter, we used the age and the sex of the child, and their performance on a few fundamental skills the child’s profile at the onset of the intervention to determine entry profiles. We observe that determination of entry profiles, while establishing a novel way to sum up the nature of the child with autism before the begin of intervention, is also crucial in furnishing a reference point for studying progress.

In this chapter, we carry on from there, understanding that entry profiles are an alternate representation of ground-truth. We expand on this by additionally determining their performance profile using data from the children once they have progressed along the syllabus of the intervention. We ensure that a similar framework using MV.RBM is used so that the outcomes in terms of entry profiles and performance profiles are comparable. We then analyse their progress by observing their performance profiles with respect to where they were at the onset of intervention, determined by entry profiles. We do this by mapping the entry profiles onto
7.2 Dataset

In this chapter, we divide our data into two subsets. The first subset, we call entry profile data, contains the performance of the children on the entry skills of TOBY (first skill in each of the 4 skill categories: Imitation, Sensory, Receptive and Expressive Language) in terms of LU required to master the 4 entry skills along with the sex and the age of the children. The second subset, we call performance data contains the LUs required to master all the skills in the syllabus attempted so far by the children. Further, the experiments are conducted on two datasets: Anonymous dataset and Clinical dataset. Details about each are presented in the following paragraphs.
7.2. Dataset

7.2.1 Anonymous dataset

The anonymous dataset consists of 283 children using TOBY for intervention. This dataset is collected from anonymous users who download and use TOBY from iTunes, as was described in Chapter 3. This data is a subset of the larger data downloaded in Feb 2014. Children start using TOBY at random entry times and show progress based on their ability, and time spent on intervention at their convenience. Hence, in the context of the downloaded data, each child is at a different place in the intervention. Our research studies the progress of a child through the intervention. Hence, we place a condition on the original data that the child must have mastered at least one entry skill and another skill within the same category as the entry skill. Only then we consider progress to have occurred. Our data contains information regarding the child such as age, sex and their performance across the syllabus attempted. The inter-quartile range of age of the children in this study is 1.58 to 9.75 years. Our dataset studies performance of 152 boys and 44 girls; the caretakers of 87 children choose to withhold the information about the sex of the child and in unknown.

7.2.2 Clinical dataset

This dataset is collected as a part of a clinical trial that has been conducted across two Australian sites: Victoria (Monash, La Trobe and Deakin Universities) and in Western Australia (WA) at the Telethon Kids Institute. The primary coordinating site for this trial is in WA whilst Monash University is the main site for coordination in Victoria. The WA site has been funded by the Australian Children’s Trust while the site in Victoria has been funded by the Helen MacPherson Smith Trust. This study has been approved by the Human Research Ethics Committee (HREC) at the University of Western Australia (RA/4/1/5656) and in Victoria, by the HREC’s at Monash University (CF13/991-2013000482) and the La Trobe University. The study has been registered with the Australian and New Zealand Clinical Trials Registry (ACTRN12614000738628). The study is ongoing and has been conducted with the objective to test the effectiveness of TOBY as a complement to early behavioural intervention and its impact on parental empowerment and stress.
The dataset consists of 38 children between the ages of 1.9 years to 5 years who underwent TOBY intervention for a period of about 9 months. There are 29 male and 9 female participants. Similar to the anonymous dataset, the clinical dataset consists of variables indicating the age, sex, and LUs acquired to master each skill. Additionally, the clinical study also consists of outcome measures that measure the severity of ASD across a variety of skills. Particularly, we use the measures *Mullen Scales of Early Learning* (MSEL) (Mullen et al., 1995) and *Autism Treatment Evaluation Checklist* (ATEC) (Rimland and Edelson, 2000).

The MSEL is a standardised developmental test for children (from birth to 68 months) which includes interactive tasks completed by the child to measure cognitive ability and motor development. The MSEL includes 124 items that measure five specific domains: *Gross Motor*, *Fine Motor*, *Visual Reception*, *Expressive Language*, and *Receptive Language*. Four cognitive scales (Visual Reception, Fine Motor, Receptive Language, and Expressive Language) sum to represent an *Early Learning Composite Score* (ELCS) which measures overall cognitive functioning. We represent each of these four cognitive scales is the form of ordinal data such that each scale can take one of the four ranks: 1 - Very Low, 2 - Below Average, 3 - Average, and 4 - Above Average. Again the ELCS also take these rank values.

ATEC is the primary measures used in this study. It is a parent-rated measure consisting of four subtest scales: Scale I. *Speech/Language/Communication* (14 items—with scores ranging from 0 to 28); Scale II. *Sociability* (20 items—with scores ranging from 0 to 40); Scale III. *Sensory/Cognitive Awareness* (18 items—with scores ranging from 0 to 36), and Scale IV. *Health/Physical Behaviour* (25 items—with scores ranging from 0 to 75). The four subscale scores can be used to calculate a total score (with total scores ranging from 0 to 180). The higher the subscale and total score, the more impaired the participant and vice versa.

In the next section, we describe the application of the framework in context of the datasets and expand on how we use these datasets to study progress in the children undergoing intervention with TOBY.
7.3  Contextual Application of MV.RBM on TOBY Dataset

In our research, we first discover entry profiles by modelling the sex, age and LUs required to master entry skills of the syllabus using MV.RBM. This is done using an MV.RBM framework similar to the one built in Chapter 6. We then model the LUs required to master all the skills attempted in the syllabus to discover performance profiles. We then attempt to analyse the progress of the children with respect to entry profiles to the performance profiles. For this study, we identify three data types to the multivariate visible units used in our research: binary for sex, continuous for age, and Replicated Softmax for LUs required to master skills (Vellanki et al., 2015). Figure 7.1 shows the model representation for each entry profiles and performance profiles.

(a) MV.RBM model for Entry Profile data. \( H = h_{1:k} \) are the hidden units and \( V_{1:3} \) are the visible units where \( v_1 = A \), the age; \( v_2 = S \), the sex; and \( v_3 = LU \), the LUs required to master the entry skills.

(b) MV.RBM model for Performance data. \( H = h_{1:k} \) are the hidden units and \( V = LU \), the LUs required to master all the skills in the syllabus.

Figure 7.1: MV.RBM models used for TOBY dataset.

7.4  Experimental Results on the Anonymous Data

In this section, we present the results of our experiment in three parts. We use CD with a batch size of 100 children for learning the parameters. After 100 data sweeps posterior hidden units are extracted and clustered for similar profiles using K-means. In this manner, we obtain both entry profiles and performance profiles. We then use t-SNE, a dimensionality reduction technique, for visualising the results in a two-dimensional space. We begin with the entry profiles and their description,
followed by the performance profiles. In the last part, we show the mapping of entry profiles to performance that enables us to study how the members of the cohort diverge after progress on part of or entire syllabus. This analysis enables us to observe how groups that are similar at the onset of the intervention react differently to the same syllabus and end up either together or diverge from each other after intervention.

7.4.1 Entry profiles

We use MV.RBM on the dataset of 283 children for entry profiles. After 100 epochs we obtain the posteriors for 10 hidden units. We then use K-means to cluster them into 5 entry profiles. In order to visualise the results that have a dimension of [281X10], we employ t-SNE. Figure 7.2 shows the discovered entry profiles (EP). We describe the shared characteristics of the entry profiles first before listing their specific properties.

Regarding the diversity of the 4 categories (Imitation, Sensory, Receptive and Expressive Language), children in EP 1 and EP 2 mastered skills across most categories: at least 2 and at most from all 4 categories, followed by children in EP 4: at least 1 and at most 4. While children in EP 3 and EP 5 managed to mastered skills from only 1 or to 3 categories at the most.

With respect to the difficulties among skill categories, most children found Receptive Language most challenging (EP 1, EP 2, and EP 4) evident from the highest medians, followed by Sensory (EP 1 and EP 2) or Expressive (EP 4). Children from EP 3 and EP 5 did not master any Receptive Language and among the remaining 3 categories mastered, they found Sensory most difficult. It is interesting to observe that the discrepancy in the amount of work required (e.g. LUs) to master Sensory is most significant for all groups involved, followed by Receptive.

- Entry Profile 1: 85 members - all are male. Imitation was shown to be the least challenging skill - lowest LU accumulation.

- Entry Profile 2: 58 members - sex is unknown. They mastered entry skills from Imitation and Expressive Language categories with the equally least efforts
7.4. Experimental Results on the Anonymous Data

(lowest LUs). Compared to EP 1, this group achieved mastery in Receptive with lesser LUs.

- Entry Profile 3: 29 members - sex is unknown. Similar to EP 2, this group found Imitation and Expressive Language least difficult. Moreover, their ability to master skills in these two categories is more uniform (evident by small interquartile in their LUs).

- Entry Profile 4: 44 members - all are female. The entry skills in all categories were found to be mastered by around uniform number of children. While this group contains the opposite sex of EP1, they also found Imitation most manageable.

- Entry Profile 5: 67 members - all are male. Distribution of LUs for EP 5 bears similarity with that of EP 3, however, the medians and the variances are much higher for EP 5, especially for Sensory.

7.4.2 Performance profiles

In this part we use MV.RBM to discover the performance profiles of 283 children after they have made progress through the syllabus. After 100 epochs we obtain the posteriors of 40 hidden units. Similar to the entry profiles, K-means and t-SNE was used on the hidden units to discover the 9 performance profiles and visualise them. Figure 4 shows the performance profiles. We observe that the clusters are not well defined and they overlap often. We expected this behaviour because ASD is a spectrum and the children cannot be grouped into well defined clusters. The membership of children with ASD is overlapping as observed by Vellanki et al. (2014) as well. We describe the characteristics of the performance profiles (PP) as observed in Figure 7.3 in brief.

In a broader view, we observe that the children belonging to PP 1, PP 2, PP 7 and PP 8 have mastered 10 or fewer skills on the entire syllabus; children from PP 4 and PP 5 have mastered almost 20 skills; and children from PP 6 and PP 9 have mastered mostly 20 skills or more. We now analyse the performance separately.
7.4. Experimental Results on the Anonymous Data

Figure 7.2: Entry Profiles and their group characteristics. Here, I - Imitation, S - Sensory, E - Expressive Language and R - Receptive Language are the skill categories to which the each entry skill attempted belongs. In the sub-figures (b) and (c) the y-axis shows LUs.
7.4. Experimental Results on the Anonymous Data

(a) 9 profiles found after progress

(b) Accumulated LU statistics

(c) Number of children mastering each skill after progress in each PP

Figure 7.3: Properties of performance profiles (PP) discovered after progress
7.4. Experimental Results on the Anonymous Data

From PP 1, PP 2, PP 7, and PP 8 consisting of 30, 35, 47 and 40 children respectively, the children belonging to these groups mastered 10 or fewer skills in all. Their performance in Figure 7.3c shows that the Sensory skills were mastered by most. The accumulated LU statistics as shown in Figure 7.3b shows similar distribution for PP1, PP 2 and PP 7, Sensory being the most difficult to master, followed by Receptive Language, when attempted. Compared to the other profiles, PP 7 and PP 8 have found Imitation difficult to master.

Now we analyse the performance of PP 4 and PP 5 together, each consisting 18 and 35 children. Figure 7.3c shows that the children from these profiles have mainly preferred Imitation and Sensory categories of the syllabus. Children of PP 4 show little progress in the language skills. The LUs accumulated from Figure 7.3b toward mastery shows that the children in PP 4 who mastered the language skills have struggled through this category, supported by comparatively higher LUs in language areas than the other skills.

Moving on to PP 6 and PP 9, each consisting 31 and 15 children, these profiles have mastered the most number of skills from the cohort. The children from PP 6 have mastered most of the skills from the entire syllabus, while those in PP 9 have mastered most of the syllabus except a few advanced Expressive Language skills. From the accumulated LU statistics in Figure 7.3b it is observed that the children in PP 6, express most difficulty in learning the Receptive Language area, followed by the Expressive Language area. On the other hand, the children from the PP 9, express most difficulty in Receptive Language skills and Sensory. Additionally, it can be noted that the PP 9 struggled with the later skills of Imitation, which consist of learning oral imitation, as compared to PP 6. Oral imitation is linked with Expressive Language, hence a person struggling to master oral imitation (imitating sounds of vowels and consonants) subsequently may find it difficult to vocalise entire words.
7.4. Experimental Results on the Anonymous Data

7.4.3 Study of progress from entry profiles and performance profiles

We map the entry profiles to the performance profiles in Figure 7.4a and observe how groups are dispersed after progress. Figure 7.4b shows the network of how the entry profiles migrate toward the performance profiles after progress. In this figure, the probability of migration between a pair of entry and performance profiles is indicated by the thickness of the connection. For example, we observe that EP 1 is more likely to migrate to PP 2, and similarly EP 5 is most likely to migrate toward PP 7 or PP 1.

We observed the groups based on the patterns of migration: group 1 (EP 1 and EP 2), group 2 (EP 4) and group 3 (EP 3 and EP 5). We observe that EP 1 and EP 2, that bore similarity transgressed together toward PP 2, PP 3, PP 4, PP 5, PP 6, PP 8 and PP 9. Again, EP 3 and EP 5, found to be similar at onset of the intervention transitioned together to another group of PPs (1 and 7). EP 4 is the only one to have highly dispersed after the progress. These observations may be critical to determine the weakness and strengths of the children and adapt the intervention course based on their mapping. Medical practitioners might find this kind of analysis invaluable for gaining a deeper understanding of how different children with ASD behave on a standard structured syllabus.

Figure 7.4: Progress tracking using entry profiles and performance profiles.
7.5 Experimental Results on Clinical Dataset

In this section, we present the entry profiles and the performance profiles discovered from the clinical dataset (please refer to section 7.2.2). As this dataset is much smaller in size (38 children), the hyper-parameters require being finely tuned. We select a batch size of 5 children and after 100 iterations we extract the posterior hidden units. The hidden units are then clustered for similar profiles using affinity propagation (AP) algorithm. Following this, we determine the entry profiles and performance profiles for this cohort. The data being a clinical dataset, is collected in a controlled setting where various measures regarding the performance of the child before the intervention begins are recorded. These measures described in section 7.2.2 as MSEL and ATEC serve as baseline scores. This enables us to compare the characteristics of the performance of the children on the entry skills and their baseline measures, improving our confidence in the results. The results presented in this section are a result of initial analysis of clinically obtained datasets and forms the basis for encouraging future work.

7.5.1 Entry profiles

16 hidden units were extracted from the entry profile clinical dataset. AP cluster was further used to cluster these hidden units into 4 entry profiles. In order to visualise this the posterior hidden units, t-SNE was used to project the 16-dimensional data into a 3-dimensional space as shown in Figure 7.5a. Figure 7.5b shows the similarity matrix based on the Jaccard index between the entry profiles. It can be seen that the cohort consists of predominantly 3 profiles, where entry profiles 2 and 3 are by far quite similar to each other. In the following paragraph, we describe the group properties of the children belonging to each entry profiles in brief. We refer to Figures 7.5 and Figure 7.6 in order to describe these properties in each case.

- Entry Profile 1: 9 members - all are male. These children are predominantly 4 years old, ranging from 1 - 4 years of age. All the skills were about uniformly attempted by the children belonging to this entry profile. Imitation and Expressive Language were shown to be the least challenging skills - lowest LU
accumulation. On the other hand, within this group Sensory skills were shown as the most difficult to attempt followed by Receptive Language.

- Entry Profile 2: 16 members - all are female. These children are 3-6 years old. All the skills were once again uniformly attempted by the children of this profile. This profile shows the least recorded accumulated LUs in most of the entry skills. Imitation, Expressive Language, and Receptive Language were shown to be relatively easiest to master. Sensory was found as one the challenging skills within the profile, but relatively easier when compared with other profiles.

- Entry Profile 3: 7 members - all male. These children are predominantly 4 years old. None of the children from this group have mastered Receptive Language skills yet. As compared with other profiles, the children from this group have mastered fewer skills than the others. Among the attempted skills Sensory was found to be the most challenging skill.

- Entry Profile 4: 6 members - all male. The children from this profile are between 3 - 5 years of age. Imitation is shown to be the least attempted skill in this profile. The children from this profile are shown to find Sensory, Expressive Language and Receptive Language tremendously challenging as compared to the other profiles. Sensory is the most challenging skill, followed by Receptive Language and Expressive Language.

The properties as described above are deciphered from the performance on TOBY syllabus in terms of the categories of Imitation, Sensory, Expressive Language and Receptive Language. With the help of MSEL and ATEC measures in the clinical trial, we further analyse the children’s baseline scores with respect to the entry profiles. Figures 7.7 and 7.8 show the baseline measures MSEL and ATEC with respect to each entry profile. In the following paragraph we describe the entry profiles in terms of their corresponding accumulated MSEL (Visual Reception, Fine Motor, Receptive Language and Expressive Language, finally the Early Learning Composite (ELC) score) and ATEC measures (Speech/Language/Communication, Sociability, Sensory/ Cognitive and Health/ Physical Behaviour).

- Entry Profile 1: The children from this profile are shown to be mainly very low and varying between very low to average in all the areas measured in MSEL.
7.5. Experimental Results on Clinical Dataset

Figure 7.5: Entry profiles (EP) and the associated similarity matrix

(a) Entry Profiles (EP)

(b) Similarity matrix showing similarity in terms of Jaccard index between children.

Figure 7.6: Properties of entry profiles.

(a) Number of children mastering each entry skill in Entry Profiles.

(b) LUs to master distribution for the Entry Profiles.

(c) Age distribution of the children in each entry profile
On the other hand, their ATEC measures show that they have similar mean measures on all subscales.

- **Entry Profile 2:** The children following this entry profile vary between *very low* to *above average*. One of the unique observation is that all the children who fare as *above average* in any areas measured using MSEL belong to this entry profile. The ATEC scores show that the children in this profile are less impaired in Speech/Language/Communication (subscale I) area and have similar scores on other areas. The ATEC total scores also reflect low mean value with a relatively lower variance. This information matches with the observation of lowest accumulated LU statistics for this entry profile as seen in the Figure 7.6.

- **Entry Profile 3:** The MSEL measures of this profile show that most of the children are *very low* or *below average* in all the skills. The ATEC scores reflect a higher mean for the Health/Physical behaviour (subscale IV) area.

- **Entry Profile 4:** The children in this profile appear to be *very low* or *average* in the Visual Reception and Fine Motor areas while varying between *very low* to *average* in the Language skills (Receptive and Expressive). ATEC scores in this profile reflect a high mean and high variance in the Health/Physical behaviour (subscale IV) area.

For EP 3 and EP 4, the higher means on Sensory and Language skills for accumulated LU statistics are also reflected in the *very low* or *below average* MSEL scores. The ATEC total scores as shown in Figure 7.8b reflect the overall difference in the domains which are measured between the entry profiles. Based on the total ATEC scores it can be observed that the children in the entry profile 3 are most impaired followed by the children in entry profile 1. EP 2 and EP 4 show similar means with a higher variance on entry profile 4. We observe that the MSEL and ATEC scores show some coherence with the LU statistics of the entry profiles.

### 7.5.2 Performance profiles

From the performance profile dataset, using a similar framework, we extracted 25 hidden units. Upon using AP cluster algorithm these hidden units were clustered
7.5. Experimental Results on Clinical Dataset

Figure 7.7: MSEL measures over different entry profiles. Visual Reception, Fine Motor, Receptive Language and Expressive Language are the four categories in which each child is evaluated. very low, below average, average, and above average are the four levels that the children are given in each of the areas.
7.5. Experimental Results on Clinical Dataset

(a) ATEC measurement scores of children over different entry profiles in boxplot. ATEC measures the children in the areas of Speech/Language/Communication (Subscale I), Sociability (Subscale II), Sensory/Cognitive (Subscale III), and Health/Physical behaviour (Subscale IV).

(b) ATEC total scores showing the overall measure for the children.

Figure 7.8: ATEC ground truths
into 5 performance profiles. We use t-SNE once again to visualise the 25-dimensional data in 3-dimensional space as seen in Figure 7.9a. Figure 7.9b shows the similarity matrix between the children based on the Jaccard index. In the following paragraph, we explain in brief the overarching characteristics of these performance profiles as observed from their properties in Figure 7.10. The similarity matrix reveals 3 prominent groups, although the strong similarity is not observed.

The PP 1 (5 members - 4 male and 1 female) and PP 2 (14 members - 10 male and 4 female) fall in a similar range with respect to the category-wise accumulated LUs of the children. It is observed that Sensory and Expressive Language are the most challenging skills for children in both these profiles; compared to PP1, PP 2 shows higher means for both these skills. One of the notable differences between these profiles is that children of PP 1 found Imitation more challenging when compared with PP 2.

The PP 3 (9 members - 7 male and 2 female), PP 4 (3 members - 2 male and 1 female), and PP 5 (7 members - 6 male and 1 female) show a similar range in accumulated LUs. These profiles are observed to have significantly higher number of LUs and hence, the upper bound for their LU statistics is set to 5000 as compared to 600 in the case of PP 1 and PP 2. The children from PP 3 and PP 4 also seem to have mastered most of the syllabus as compared to the other performance profiles. For children in PP 3 and PP 4, we observe that Receptive Language was the most
7.5. Experimental Results on Clinical Dataset

(a) Accumulated LUs statistics for the Performance Profiles.

(b) Number of children mastering each skill across Performance Profiles.

Figure 7.10: Properties of Performance Profiles (PP)
challenging skill followed by Expressive Language, Sensory, and Imitation skills. PP 5 shows that for these children, Receptive Language and Sensory were the most challenging skills. In general, it seems that PP 5 presents most challenging learning behaviour based on performance on TOBY.

### 7.5.3 Study of progress from entry profiles and performance profiles

We track the change in membership of children between entry profiles and performance profiles. Through this, similar to the non-clinical dataset, we intend to observe how the children change after intervention, if they end up together in the same profile after progress across the syllabus, or if they move apart. Figure 7.11 shows the mapping from entry profiles to performance profiles, where the thickness of the edge connecting an EP to a PP depends on the number of children following this path. We observe that the children in EP 1 and EP 3 mainly end up in PP 2 after intervention. The children from EP 2 disperse and end up in all the performance profiles. Finally, the children in EP 4 transgress to PP 5.

As seen in the results for the anonymous (non-clinical) dataset, such information could be useful for designing the syllabus and providing learning opportunities that suit individual needs of each sub-type of children. Additional insights on such stud-
ies may be drawn with the support of medical health professionals. Integrating
ground-truth measures that evaluate children on various criteria of ASD deficits
and strengths can build confidence in the results of anonymous data. The challenge,
however, is that clinical datasets (such as the one used in this section) are very rare
and when available are limited to a smaller sample size. This limits our ability to
draw meaningful information with a strong confidence from small cluster sizes.

7.6 Concluding Remarks

Profiling and tracking the progress of children with ASD is crucial personalised inter-
vention. It is made challenging, nevertheless, by the inseparability of children with
ASD into well-defined clusters and unavailability of large datasets suitable for data-
based studies. The nuance in this study is possible due to the dataset we present,
but at the same time dealing with heterogeneous, mix-variate, and highly correlated
data with missing elements needs complex models. In this study, we present entry
and performance profiles discovered using MV.RBM on a dataset of 283 children
and map the progress after they have undergone intervention. Furthermore, we also
present the results of similar nature on a clinical dataset consisting of 38 children
who underwent TOBY intervention during a clinical trial. Their baseline measures
(including MSEL and ATEC) were measured before the intervention, which we use
as ground-truth to confirm our results. MSEL and ATEC total scores show some
coherence with the group statistics of entry profiles. Our study of progress medi-
ated by a discovery of entry profiles and performance profiles is, to the best of our
knowledge, first of its kind. Followed up with predictive analysis it can help re-
commend suitable intervention paths for children with ASD. Additionally, such an
analysis might be invaluable to medical practitioners to furthering the understanding
of learning patterns of children with ASD.
Chapter 8

Concluding Remarks

8.1 Summary

This thesis has aimed at understanding the patterns of learning in children with Autism Spectrum Disorder (ASD) on a computer-based application. In literature, studies have been able to address problems in diagnosis and genetic identifiers and such, using machine learning techniques. The problem of personalisation of intervention for children with ASD is a crucial one that qualified therapists/medical professionals address. This requires rigorous data collection and analysis which was manually conducted. With the advent of technology, building applications that support seamless and accurate data collection has become easier. These applications can be used either as a support in addition to conventional therapy, or in the absence of a therapist (during waiting times between diagnosis and finding a therapist). The data collected via these applications can also be leveraged to understand the behaviour of the children and how the disorder affects them. In the past, unavailability of a vast dataset in this domain made it difficult to conduct such a data-based study. In this thesis, we identify the research gap problem of identifying factors that can potentially benefit personalisation of intervention in ASD. To realise this aim, on a novel dataset consisting more than 500 children, collected from an application called TOBY Playpad, we use non-negative latent factorisation and restricted Boltzmann machines approaches to address problems in ASD intervention domain. To the best of our knowledge, the work presented in this thesis is a first attempt to do so.
One of the main challenges encountered in addressing our aim is to obtain a dataset that records rigorous performance of the children with ASD over a structured syllabus. We address this challenge in Chapter 3, by introducing the source of our data, TOBY Playpad, the iPad application for ASD intervention and the dataset obtained from this application. Caregivers download TOBY on their iPads and children undergo intervention at their natural environment with the caregivers’ assistance. The TOBY syllabus covers a broad range of skills out of which, iPad-based skills from the categories of *Imitation, Sensory, Receptive Language* and *Expressive Language* were the focus of our research. Understanding the mode of delivering instruction and the structure and content of the syllabus, was crucial to interpreting the results in the domain. The dataset presented was unique and novel in consisting of performance information of about 500 children. The dataset and the associated variables of interest were presented with the help of a data dictionary describing the raw data. In Chapter 3, we also presented the framework for translation of the application to a cultural context to suit personalised needs relevant to geography and language. This framework was then tested during a clinical trial conducted on about 40 children from India. We found that the children were able to effectively learn using the translated prototype in Hindi.

Owing to the novel nature of the research, we presented an exploratory approach to analysing the data by looking at various variables individually in Chapter 4. We interrogated the data using K-means to observe whether the sub-groups of children with ASD can be identified and whether they present overarching differential profiles of learning. We also studied the differences in the ways the children respond correctly and incorrectly using their response time. We presented our observations regarding the differential profiles in the context of research literature relevant to challenges faced by children with ASD across various areas. The study conducted in this chapter lay the ground work for this thesis and motivated us toward the use of more complex models to discover the inherent patterns of learning among the cohort of children with ASD.

A factor analysis approach to analysing the data is most appropriate for identifying detailed learning patterns and sub-groups of children with ASD. To this end, in Chapter 5, we used a Bayesian, nonparametric, non-negative matrix factorisation model called Linear Poisson Gaussian Model (LPGM) with an Indian Buffet Process (IBP) prior to automatically discover the number of latent learning patterns.
Using a Bayesian nonparametric approach saved us from the dimensionality curse of manually estimating the number of hidden patterns. Our dataset grows in time by new children beginning intervention and existing children progressing in the syllabus. A nonparametric approach inherently also allowed us to control the model complexity with growing data. To address the presence of missing data elements due to growing data, we extended this model to compute the patterns without having to impute the data. We discovered 26 learning patterns that are various combinations of skills with their weights indicating the level of difficulty faced by the children in terms of the number of Learn Units required to master the skills. By exploring the membership patterns among the children, we were able to find sub-groups of children with ASD, who behaved similarly over the syllabus.

Finally, in Chapter 6 and 7, we developed a framework for describing the children in terms of entry profiles and performance profiles. Entry profiles are the characteristics of the children before they begin intervention. Knowledge regarding the entry profiles is essential to determine progress during the course of therapy. This information is often available in the form of diagnostic data in traditionally delivered therapy. But in the case of computer-based intervention, which the child undertakes with help from caregivers in their natural environment, the diagnostic information may be difficult to gather due to security and privacy issues. We modelled basic information like age and sex of the child and additionally their performance on the entry skills (the basic skills they encounter at the top of the syllabus tree for each category) to define entry profiles. We use restricted Boltzmann machine to this effect which allows us to model heterogeneous data to obtain clusters of children who are similar before the actual intervention begins. We further used a similar framework to model their performance once they have made some progress on the syllabus to obtain Performance profiles. Similar to entry profiles, performance profiles suggest the characteristic grouping of the children after they have undergone some progress across the syllabus. Comparing the entry profiles and the performance profiles revealed how each group of children in entry profiles has evolved after the intervention. We did this by mapping the entry profiles onto the performance profiles and show that while some children belonging to the same entry profiles diverge into different performance profiles once they have made progress, others cluster together in the same performance profiles. One of the main challenges we faced was devising means to verify the entry profiles with ground-truth. We thus used a clinical dataset of 38
children, whose baseline measures (such as Mullen Scales of Early Learning (MSEL) and Autism Treatment Evaluation Checklist (ATEC)) were recorded that served as ground-truth about the child before intervention. The children then underwent TOBY intervention for a period of about 9 months and the data collected was used to determine their entry profiles and performance profiles. We found some coherence in the baseline measures and the entry profiles discovered.

In summary, to the best of our knowledge, this thesis is the first that addresses the problems of early intervention in ASD through the lens of machine learning and data analytics. Our key contribution includes the establishment of large datasets in this domain for the first time together with a systematic machine learning based approach to extract knowledge relevant to ASD. We believe the findings and contributions from this thesis form the initial foundations for a data-driven approach to personalised early intervention for children with ASD.

8.2 Future Directions

Overall, this thesis has presented foundational work in understanding hidden patterns, discovering ASD sub-groups, and analysing the progress of children with ASD on computer-based intervention setting. The studies undertaken in this thesis suggest several potential ideas for future work. In the following paragraphs, we speculate on a few ideas and possible extensions.

This thesis presents a heuristic approach to discovering hidden patterns where the order of the skills attempted is not taken into consideration. In Chapter 5, the learning patterns discovered attempt to explore hidden connections between skills in the syllabus. Inter-related skills potentially signify that the order of learning may affect the efficiency of progress. For example, in this thesis, one of the reappearing results is that the likelihood that the child will find Sensory and Receptive Language skills challenging together is high. As both the skills require identifying objects, based on the sensory cue (colour) or based on the audio label, the overarching skill set required to learn both the skills could be overlapping. In this scenario, it may be worth exploring how children interact with syllabus when the order of learning is changed.
The data representation that we consider in this thesis is the accumulation of the performance of the child over the entire time, since the child begins the intervention to the time the data was collected. It is possible to obtain information further in detail about their daily performance. Analysis of the data from each day may lead to an understanding about which skills the child prefers to learn together and whether the combination is a good one (child is learning both efficiently) or is it harming the child (child is not able to learn either). One of the tools that could be used for analysing such data is Hierarchical Dirichlet Process (HDP) that finds latent topics, which are combinations of skills in this scenario. One of the main challenges encountered is, however, insufficiency of data at a day level.

In Chapter 4, we briefly explored the concept of response time. Response time, in literature, is identified as a measure that can reveal information about the process of cognitive decision making and residual latency related to motor response time. In a conventional intervention, it is extremely difficult to record accurate response times. Computer-based interventions, however, are an attractive domain for recording, recovering, and analysing the response time models. Modelling response times under various skills and various scenarios could contribute to the knowledge of the cognitive functioning of the brain of children with ASD.

We describe that our dataset consists of missing elements that are a result of children still working on the syllabus at any given point. We extend the Linear Poisson Gamma Model (LPGM) in Chapter 5 to adapt the model such that it accommodates missing data. Here we assume that the data is missing at random. However, the data is missing because the skill has not been attempted yet by the child. As the syllabus is arranged in a hierarchical structure and the children cannot pass the nodes till they master the skills. The chances of encountering missing data thus increase as we go up the syllabus. As our research does not consider the order of learning skills the assumption of data missing at random may be safe. But considering the order and making the missing not at random assumption would be a worthwhile pursuit.

We use restricted Boltzmann machines (RBM), which are the building blocks of deep neural networks (DNN), in Chapter 6 and 7. Although RBMs are elegant and efficient, DNNs enable adding multiple layers of hidden units between the input and the output layers, allowing modelling of complex non-linear relationships. The composition of features from the lower layers also increases the potential of DNNs
to model complex data. Instead of adding all the variables on the same input layer (entry profiles and performance profiles), using a DNN can potentially reduce the number of hidden units within the same efficiency.

Finally, this thesis addresses the factors necessary for personalisation of the syllabus for children with ASD. The next direction evidently points toward building a predictive framework, that dynamically allow suggestions regarding skills and learning opportunities based on the profiles of the child. This framework may potentially be able to detect whether and why a child is stuck in the syllabus and provide a strategy to overcome the challenge based on the how similar children behaved in such situations. In this thesis, we have considered only two time frames - one at the onset of the intervention and the other at the time the data was collected from the server. One approach that we suggest is to split the dataset at different time frames and analyse its progress through all these time frame steps. Discovering learning patterns at various time frames may be leveraged to study unit changes in progress and better prediction of the course of the intervention.
Appendix A

Syllabus of TOBY Playpad

We present here the syllabus of TOBY under four categories - their short-names as referred to in the syllabus tree, their long names, and their content. The syllabus tree in Figure A.1 is the same syllabus tree as presented in Chapter 3. We show it here again for convenience of reference.

A.1 General skills

These skills are for the children to get acquainted with using iPads and they are pre-requisites for every other skills. With these three skills the main objective is to teach the children how to point to an answer, how to drag objects around on the iPad screen, and how to identify and use prompts.

<table>
<thead>
<tr>
<th>Serial Number</th>
<th>Skill short name</th>
<th>Skill short name</th>
<th>Skill</th>
<th>iPad</th>
<th>NET</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>tut-point</td>
<td>Tutorial, Pointing</td>
<td>yes</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>tut-drag</td>
<td>Tutorial, Dragging</td>
<td>yes</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>tut-prompt</td>
<td>Tutorial, Prompts</td>
<td>yes</td>
<td>no</td>
<td></td>
</tr>
</tbody>
</table>

Table A.1: List of general skills
Figure A.1: Structured syllabus of TOBY Playpad categorised by skill area. Each colour represents a different skill area. The labels inside each node denote the name of the skill that is taught in it.
A.2 Imitation Skill Category

For imitation tasks (iPad), stimuli are generally in the form of video models, and the child performs the task by copying the actions of the model. These actions mainly focus on teaching gross and fine motor skills, and oral imitation. Gross and fine motor skills are essential for learning how to use different objects and perform various actions in life. Oral imitation skills, on the other hand, teach oral motor movements that are a pre-requisite to expressive language. The following are the list of skills in this category that are in TOBY syllabus - their short names (as referred to in the figure A.1) and what they stand for.

<table>
<thead>
<tr>
<th>Serial No.</th>
<th>Skill short name</th>
<th>Skill description</th>
<th>iPad</th>
<th>NET</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gmi-i-1</td>
<td>Gross Motor Imitation, Invisible, 1 Step</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>2</td>
<td>gmi-v-1</td>
<td>Gross Motor Imitation, Visible, 1 Step</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>3</td>
<td>fmi-b</td>
<td>Fine Motor Imitation, Basic</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>4</td>
<td>fmi-c</td>
<td>Fine Motor Imitation, Complex</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>5</td>
<td>oral-nv</td>
<td>Oral Imitation, No Voice</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>6</td>
<td>oral-va</td>
<td>Oral Imitation, Voice with Action</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>7</td>
<td>oral-vowel</td>
<td>Oral Imitation, Vowels</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>8</td>
<td>oral-cons</td>
<td>Oral Imitation, Consonants</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>9</td>
<td>oral-sym</td>
<td>Oral Imitation, Symbolic</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>10</td>
<td>oral-cv</td>
<td>Oral Imitation, CV</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>11</td>
<td>oral-cvc</td>
<td>Oral Imitation, CVC</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>12</td>
<td>oral-cvcv</td>
<td>Oral Imitation, CVCV</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>13</td>
<td>oral-b</td>
<td>Oral Imitation, Blowing</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>14</td>
<td>awo</td>
<td>Action with Objects</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>15</td>
<td>awo-c</td>
<td>Copying</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>16</td>
<td>awo-c-blocks</td>
<td>Copying, Block Designs</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>17</td>
<td>awo-c-drawing</td>
<td>Copying, Drawings</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>

Table A.2: List of Imitation skills
A.3 Sensory Skill Category

For sensory skills (iPad), stimulus is mainly in the form of images and skills are presented as matching tasks. Matching tasks are used for teaching categorisation, where the children learn to identify objects with similar sensory properties (i.e. colour, size, shape). The following are the sensory skills in the syllabus of TOBY.

<table>
<thead>
<tr>
<th>Serial No.</th>
<th>Skill short name</th>
<th>Skill short name</th>
<th>Skill</th>
<th>iPad</th>
<th>NET</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>vsm-id</td>
<td>Visual Sensory Matching, Identical Objects</td>
<td>yes</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>vsm-nonid</td>
<td>Visual Sensory Matching, Nonidentical Objects</td>
<td>yes</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>vsm-size</td>
<td>Visual Sensory Matching, Size</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>vsm-shape</td>
<td>Visual Sensory Matching, Shape</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>vsm-colour</td>
<td>Visual Sensory Matching, Colour</td>
<td>yes</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>vsm-verb</td>
<td>Visual Sensory Matching, Verbs</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>vsm-bp</td>
<td>Visual Sensory Matching, Body Parts</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>asm-s-ns</td>
<td>Auditory Sensory Matching, Sound/NonSound</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>asm-s-s</td>
<td>Auditory Sensory Matching, Sound/Sound</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>asm-s-w</td>
<td>Auditory Sensory Matching, Sound/Word</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>asm-w-w</td>
<td>Auditory Sensory Matching, Word/Word</td>
<td>no</td>
<td>yes</td>
<td></td>
</tr>
</tbody>
</table>

Table A.3: List of Sensory skills

A.4 Receptive Language Skill Category

Receptive language skills involve learning to identify objects based on their labels. These skills are presented as matching tasks, where the audio label of the objects is played and the children must identify the appropriate object from a set of three objects presented.
### A.5 Expressive Language Skill Category

Expressive language skills teach children how to identify labels of the objects as opposed to what is learned in receptive language. The stimulus is presented in the form of image of the object and the children must identify the label of that object. The following are the skills under this category in TOBY syllabus.

<table>
<thead>
<tr>
<th>Serial No.</th>
<th>Skill short name</th>
<th>Skill</th>
<th>iPad</th>
<th>NET</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>me-noun</td>
<td>Expressive Matching, Nouns</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>2</td>
<td>me-verb</td>
<td>Expressive Matching, Verbs</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>3</td>
<td>me-adj-shape</td>
<td>Expressive Matching, Shape</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>4</td>
<td>me-adj-colour</td>
<td>Expressive Matching, Colour</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>5</td>
<td>me-adj-size</td>
<td>Expressive Matching, Size</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>6</td>
<td>me-bp</td>
<td>Expressive Matching, Body Parts</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table A.5: List of Expressive Language skills
A.6 Social Skill Category

Social skill are all activities that have to be learned in Natural Environment Training (NET). They teach children how to engage in social, non-verbal communication. As all of these skills are NET skills, they are not one of our main interests in this thesis.

<table>
<thead>
<tr>
<th>Serial No</th>
<th>Skill short name</th>
<th>Skill</th>
<th>iPad</th>
<th>NET</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>sr-gs</td>
<td>Social Receptive, Gaze Sharing</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>2</td>
<td>sr-fap</td>
<td>Social Receptive, Follow Adult Point</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>3</td>
<td>sr-fag</td>
<td>Social Receptive, Follow Adult Gaze</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>4</td>
<td>sr-fag-fe</td>
<td>Social Receptive, Follow Adult Gesture, Facial Expression</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>5</td>
<td>sr-faa-sw</td>
<td>Social Receptive, Follow Adult Attention, Single Word</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>6</td>
<td>se-dcp</td>
<td>Social Expressive, Direct Child Point</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>7</td>
<td>se-dcg</td>
<td>Social Expressive, Direct Child Gaze</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>8</td>
<td>se-dcg-yn</td>
<td>Social Expressive, Direct Child Gesture, Yes-No</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>9</td>
<td>se-dcfe</td>
<td>Social Expressive, Direct Child Facial Expression</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>10</td>
<td>se-dca-sw</td>
<td>Social Expressive, Direct Child Attention, Single Word</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table A.6: List of Social skills
Appendix B

Supplementary Derivation

In this appendix, we present the derivations for the posteriors of $w_{vk}$ and $f_{kn}$, in the scenario where the dataset has missing data. $X$ is our data matrix where the elements $x_{vn}$, corresponding to the number of LUs accumulated by a child $n$ in a task $v$, are data-points. Our objective is to derive the posteriors for the parameters by using the data-points $x_{vn}$ that are not missing. The inference of posterior of $w_{vk}$, for a certain value of $v$, depends on the values $x_{vn}$ for all values of $n$. Similarly the inference of $f_{kn}$ for a certain value $n$ depends on values $x_{vn}$ for all values of $v$. Hence we consider the data points in two sets $J_v$ and $I_n$ for each inference respectively, such that $J_v$ contains all the non missing values from the array $x_{v,1:N}$ and $I_n$ contains all of those from $x_{1:V,n}$.

Let’s represent the sum $\sum_{i=1}^{K} f_{in} z_{in} w_{vi}$ as $\eta_i$ the term including parameters for all values $i = 1 : K$, $\sum_{i=1, i \neq k}^{K} f_{in} z_{in} w_{vi}$ as $\eta_{-i}$ the term including parameters for all values of $i$ except for $k$, and $f_{kn} z_{kn} w_{vk}$ as $\eta_k$ the term for the condition when $i$ takes the value $k$. 
• The posterior of \( w_{vk} \) is

\[
p(w_{vk} \mid Z, F, X) \propto p(X \mid Z, F, W)p(w_{vk} \mid \alpha_0, \beta_0)
\]

\[
= \left( \prod_{n \in J_v} p(x_{vn} \mid \eta_k) \right) \Gamma(\alpha_0, \beta_0)
\]

\[
= \left( \prod_{n \in J_v} \frac{(\eta_k)^{x_{vn}} e^{-(\eta_k)}}{x_{vn}} \right) \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} w_{vk}^{\alpha_0-1} e^{-\beta_0 w_{vk}}
\]

\[
\propto w_{vk}^{\alpha_0-1} e^{-\beta_0 w_{vk}} \times \prod_{n \in J_v} \left( (\eta_k)^{x_{vn}} e^{-(\eta_k)} \right)
\]

\[
= w_{vk}^{\alpha_0-1} e^{-\beta_0 w_{vk}} \times \prod_{n \in J_v} \left( (\eta-k + \eta_k)^{x_{vn}} e^{-(\eta-k + \eta_k)} \right)
\]

\[
\propto w_{vk}^{\alpha_0-1} e^{-\beta_0 w_{vk}} \prod_{n \in J_v} \left( (\eta-k + \eta_k)^{x_{vn}} e^{-(\eta-k + \eta_k)} \right)
\]

In order to solve the above equation, we take the help of an auxiliary variable. Let us consider that the probability \( p(w_{vk}) \) is proportional to the unnormalised exponential function \( p^*(w_{vk}) \), where \( p^*(w_{vk}) \) is given by and can be expanded as a binomial function as follows:

\[
p^*(w_{vk}) = \left( \eta-k + \eta_k \right)^{x_{vn}}
\]

\[
= \sum_{j=0}^{x_{vn}} \binom{x_{vn}}{j} (\eta_k)^j (\eta-k)^{x_{vn}-j}
\]

Hence we have,

\[
p(w_{vk}) \propto (\eta-k + \eta_k)^{x_{vn}}
\]

Now let \( r_{vn} \) be an auxiliary variable. We aim to define a probability \( p(w_{vk}, r_{vn}) \) proportional to \( p^*(w_{vk}, r_{vn}) \) such that \( \sum_{r_{vn}} p^*(w_{vk}, r_{vn}) = p^*(w_{vk}) \). So let,

\[
p^*(w_{vk}, r_{vn}) = \left( \frac{\eta_k}{r_{vn}} \right)^{r_{vn}} (\eta-k)^{x_{vn}-r_{vn}} \]

where, \( r_{vn} = \{0, 1, 2, ..., x_{vn}\} \). Hence we have,

\[
\sum_{r_{vn}=0}^{x_{vn}} p^*(w_{vk}, r_{vn}) = \sum_{r_{vn}=0}^{x_{vn}} \left( \frac{\eta_k}{r_{vn}} \right)^{r_{vn}} (\eta-k)^{x_{vn}-r_{vn}}
\]

\[
= p^*(w_{vk})
\]
Additionally we have,

\[
p(w_{vk} \mid r_{vn}) = \frac{p(w_{vk}, r_{vn})}{p(r_{vn})} \\
\propto p(w_{vk}, r_{vn}) \\
\propto p^*(w_{vk}, r_{vn}) \\
= \frac{(x_{vn})_{r_{vn}} (\eta_k)_{r_{vn}} (\eta_{i})_{x_{vn} - r_{vn}}}{(\eta_{-i} + \eta_k)^{x_{vn}}}
\]

\[
p(r_{vn} \mid w_{vk}) = \frac{p(w_{vk}, r_{vn})}{p(w_{vk})} \\
\propto \frac{p^*(w_{vk}, r_{vn})}{p^*(w_{vk})} \\
= \frac{(x_{vn})_{r_{vn}} (\eta_k)_{r_{vn}} (\eta_{i})_{x_{vn} - r_{vn}}}{(\eta_{-i} + \eta_k)^{x_{vn}}}
\]

Hence the conditional distributions have a form of the binomial distribution. After substituting back the values of \(\eta_{-i}\) and \(\eta_k\), if we sample \(r_{vn}\) from such a distribution we can approximate the binomial expansion as follows:

\[
R_{vn} \sim \text{Binomial} \left( x_{vn}, \frac{z_{kn} f_{kn w_{vk}}}{\sum_{i \neq k} z_{in} f_{in w_{vi}} + z_{kn} f_{kn w_{vk}}} \right), \quad \forall n \in J_v
\]

\[
\left( \sum_{i \neq k} z_{in} f_{in w_{vi}} + z_{kn} f_{kn w_{vk}} \right)^{x_{vn}} \propto (z_{kn} f_{kn w_{vk}})^{R_{vn}}
\]

Hence, we have

\[
p(w_{vk} \mid Z, F, X) \propto w_{vk}^{a_0 - 1} e^{-\beta_0 w_{vk}} \prod_{n \in J_v} \left( (f_{kn z_{kn w_{vk}}})^{R_{vn}} e^{-(f_{kn z_{kn w_{vk}}})} \right) \\
\propto w_{vk}^{a_0 + \sum_{n \in J_v} R_{vn} - 1} e^{-(\beta_0 + \sum_{n \in J_v} f_{kn z_{kn w_{vk}}}) w_{vk}}
\]

The above expression is in gamma distribution form: \(w_{vk} \sim \text{Gamma}(\alpha'_0, \beta'_0)\)
where,

\[ \alpha'_0 = \alpha_0 + \sum_{n \in J_v} R_{vn} \]

\[ \beta'_0 = \beta_0 + \sum_{n \in J_v} f_{kn} z_{kn} \]

• The posterior of \( f_{kn} \) is similarly calculated as:

\[
p(f_{kn} | Z, W, X) \propto p(X | Z, F, W) p(f_{kn} | \alpha_1, \beta_1) = \left( \prod_{m \in I_n} p(x_{vn} | \sum_{i=1}^{K} f_{in} z_{in} w_{vi}) \right) \text{Gamma}(\alpha_1, \beta_1)
\]

Hence, we have

\[
p(f_{kn} | Z, W, X) \propto f_{kn}^{\alpha_1 - 1} e^{-\beta_1 f_{kn}} \prod_{v \in I_n} \left( (f_{kn} z_{kn} w_{vk})^{T_{vn}} e^{-(f_{kn} z_{kn} w_{vk})} \right)
\]

\[
\propto f_{kn}^{\alpha_1 + \sum_{v \in I_n} T_{vn} - 1} e^{-(\beta_1 + \sum_{v \in I_n} z_{kn} w_{vk}) f_{kn}}
\]

The above expression is in gamma distribution for \( f_{kn} \sim \text{Gamma}(\alpha'_1, \beta'_1) \)

where,

\[ \alpha'_1 = \alpha_1 + \sum_{v \in I_n} T_{vn} \]

\[ \beta'_1 = \beta_1 + \sum_{v \in I_n} z_{kn} w_{vk} \]

and the auxiliary variable \( T_{vn} \) is sampled similar to \( R_{vn} \) from

\[ T_{vn} \sim \text{Binomial} \left( x_{vn}, \frac{f_{kn} z_{kn} w_{vk}}{\sum_{i \neq k} f_{in} z_{in} w_{vi} + f_{kn} z_{kn} w_{vk}} \right), \quad \forall n \in I_n \]
Bibliography


Adham Beykikhoshk, Ognjen Arandjelovic, Svetla Venkatesh, and Dinh Phung. Hierarchical dirichlet process for tracking complex topical structure evolution and


Debora M Kagohara, Larah van der Meer, Sathiyaprakash Ramdoss, Mark F O’Reilly, Giulio E Lancioni, Tonya N Davis, Mandy Rispoli, Russell Lang, Peter B


R Duncan Luce. *Response times: Their role in inferring elementary mental organization*. Number 8. Oxford University Press on Demand, 1991.


Samuel L Odom, Lana Collet-Klingenberg, Sally J Rogers, and Deborah D Hatton. Evidence-based practices in interventions for children and youth with autism.


Pratibha Vellanki, Thi Duong, Svetha Venkatesh, and Dinh Phung. Nonparametric discovery of learning patterns and autism subgroups from therapeutic data. In


Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.
ICST Copyright and Consent Form

Title of Work: Data Mining of Intervention for Children with Autism Spectrum Disorder

Complete list of authors: Vellanki, Pratibha<pratibha.vellanki@gmail.com> (Deakin University), Duong, Thi<thi.duong@deakin.edu.au> (Deakin University), Phung, Dinh<dinh.phung@deakin.edu.au> (Deakin University), Venkatesh, Sverha<sverha.venkatesh@deakin.edu.au> (Deakin University)

ICST publication title: Proceedings of EAI International Conference on Personal, Pervasive and mobile Health

Transfer of Copyright Agreement

Copyright to the Work, to any supplemental material integral to the Work which is submitted with it for review and publication such as an extended proof or supplementary text and figures, and to any subsequent errata, is hereby according to the rules set in the U.S. Copyright Act, 17 U.S.C., transferred to ICST for the full term throughout the world, subject to the Author Rights (as hereinafter defined) and to the acceptance of the Work for publication by ICST. This transfer of copyright includes all material to be published as part of the Work, including but not limited to tables, figures, graphs, movies, other multimedia files, and all supplemental materials. ICST shall have the right to register copyright in the Work in its name as claimant, whether separately or as part of the journal issue, book, volume or other medium in which the Work is 101 included.

The author(s), and in the case of a Work Made For Hire, as defined in the U.S. Copyright Act, 17 U.S.C. §101, the employer named below, shall have the following rights (the "Author Rights"):

1. All proprietary rights other than copyright, such as patent.
2. The right to reuse any portion of the Work, without fee, in future works of the author(s) or employer, including books, lectures and presentations in all media, provided that a citation of the ICST -published work, notice of the Copyright, and ICST DOI are included.
3. The right to make, and hold copyright in, works derived from the Work, as long as the following conditions are met: (i) at least one author of the derivative work is an author of the Work; (ii) the derived work includes at least 30% of new material not covered by ICST’s copyright in the Work. If these conditions are met, copyright in the derivative work rests with the authors of that work, and ICST and its successors and assigns will make no claim on that copyright. If these conditions are not met, explicit ICST permission must be obtained. Nothing in this Section shall prevent ICST and its successors and assigns from exercising its rights in the Work.
4. The right to post and update author-prepared versions of the article on free-access e-print servers, including the author and/or employer’s home page and any repository legally mandated by the agency funding the research on which the Work is based. If the author wishes the ICST-prepared version to be used for an online posting, permission is required from ICST; if granted, use will be subject to ICST terms and conditions.

General Terms

• The undersigned represents that he/she has the power and authority to make and execute this assignment;
• The undersigned agrees to indemnify and hold harmless ICST from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.
• In the event that the Work is not accepted and published by ICST or is withdrawn by the author(s) before acceptance by ICST, the foregoing copyright transfer shall become null and void.
• For jointly authored works, all joint authors should sign, or one of the authors should sign as an authorized agent for the others. If the article has been prepared as a Work Made for Hire, as defined in the U.S. Copyright Act, 17 U.S.C. §101, the transfer should be signed by the employer.

Signature(s)

A. Single Author/Authorized Agent for Joint Authors

To sign the agreement, please enter your name, job title and employer in the box provided.

signature Pratibha Vellanki, Doctoral Candidate, Deakin University, Australia

Date 2016/05/16
B. Employer authorized signature (for Work Made for Hire)

signature  2016/05/16  
Date

C. U.S. Government Employee Certification (where applicable)

This will certify that all authors of the Work are U.S. government employees and prepared the Work on a subject within the scope of their official duties. As such, the Work is not subject to U.S. copyright protection. (Authors should still sign signature line [A] above to enable ICST to claim and protect its copyright in international jurisdictions.)

signature  2016/05/16  
Date

D. Crown Copyright Certification (where applicable)

This will certify that all authors of the Work are employees of the British or British Commonwealth Government and prepared the Work in connection with their official duties. As such, the Work is subject to Crown Copyright and is not assigned to ICST as set forth in the first sentence of the Transfer of Copyright Agreement above. The undersigned acknowledges, however, that ICST has the right to publish, distribute and reprint the Work in all forms and media. (Authors should still sign signature line [A] above to indicate their acceptance of all terms other than the copyright transfer.)

signature  2016/05/16  
Date

I have read and agree to the forms of the ICST Copyright and Consent Form.
This Agreement between Pratibha Vellanki ("You") and Springer ("Springer") consists of your license details and the terms and conditions provided by Springer and Copyright Clearance Center.

License Number: 3919610577957
License date: Jul 31, 2016
Licensed Content Publisher: Springer
Licensed Content Publication: Knowledge and Information Systems
Licensed Content Title: Nonparametric discovery and analysis of learning patterns and autism subgroups from therapeutic data
Licensed Content Author: Pratibha Vellanki
Licensed Content Date: Jan 1, 2016
Type of Use: Thesis/Dissertation
Portion: Full text
Number of copies: 1
Author of this Springer article: Yes and you are a contributor of the new work

Order reference number
Title of your thesis / dissertation: Applied Machine Learning for Personalized Early Intervention in Autism
Expected completion date: May 2016
Estimated size (pages): 200
Requestor Location: Pratibha Vellanki
7, Macken Ct,
Highton
Geelong, 3216
Australia
Attn: Pratibha Vellanki
Billing Type: Invoice
Billing Address: Pratibha Vellanki
7, Macken Ct,
Highton
Geelong, Australia 3216
Attn: Pratibha Vellanki
Total: 0.00 USD

Terms and Conditions

Introduction
The publisher for this copyrighted material is Springer. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com). Limited License

With reference to your request to reuse material on which Springer controls the copyright, permission is granted for the use indicated in your enquiry under the following conditions:
- Licenses are for one-time use only with a maximum distribution equal to the number stated in your request.
- Springer material represents original material which does not carry references to other sources. If the material in question appears with a credit to another source, this permission is not valid and authorization has to be obtained from the original copyright holder.
- This permission
  • is non-exclusive
  • is only valid if no personal rights, trademarks, or competitive products are infringed.
  • explicitly excludes the right for derivatives.
- Springer does not supply original artwork or content.
- According to the format which you have selected, the following conditions apply accordingly:
  - **Print and Electronic**: This License include use in electronic form provided it is password protected, on intranet, or CD-Rom/DVD or E-book/E-journal. It may not be republished in electronic open access.
  - **Print**: This License excludes use in electronic form.
  - **Electronic**: This License only pertains to use in electronic form provided it is password protected, on intranet, or CD-Rom/DVD or E-book/E-journal. It may not be republished in electronic open access.

For any electronic use not mentioned, please contact Springer at permissions.springer@spglobal.com.

- Although Springer controls the copyright to the material and is entitled to negotiate on rights, this license is only valid subject to courtesy information to the author (address is given in the article/chapter).
- If you are an STM Signatory or your work will be published by an STM Signatory and you are requesting to reuse figures/tables/illustrations or single text extracts, permission is granted according to STM Permissions Guidelines: http://www.stm-assoc.org/permissions-guidelines/

For any electronic use not mentioned in the Guidelines, please contact Springer at permissions.springer@spglobal.com. If you request to reuse more content than stipulated in the STM Permissions Guidelines, you will be charged a permission fee for the excess content.

Permission is valid upon payment of the fee as indicated in the licensing process. If permission is granted free of charge on this occasion, that does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

- If your request is for reuse in a Thesis, permission is granted free of charge under the following conditions:
  - This license is valid for one-time use only for the purpose of defending your thesis and with a maximum of 100 extra copies in paper. If the thesis is going to be published, permission needs to be reobtained.
  - includes use in an electronic form, provided it is an author-created version of the thesis on his/her own website and his/her university’s repository, including UMI (according to the definition on the Sherpa website: http://www.sherpa.ac.uk/romeo/);
  - is subject to courtesy information to the co-author or corresponding author.

Geographic Rights: Scope
Licenses may be exercised anywhere in the world.

Altering/Modifying Material: Not Permitted
Figures, tables, and illustrations may be altered minimally to serve your work. You may not alter or modify text in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s).

Reservation of Rights
Springer reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction and (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

License Contingent on Payment
While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full
payment is received from you (either by Springer or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received by the date due, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Springer reserves the right to take any and all action to protect its copyright in the materials.

Copyright Notice: Disclaimer
You must include the following copyright and permission notice in connection with any reproduction of the licensed material:
"Springer book/journal title, chapter/article title, volume, year of publication, page, name(s) of author(s), (original copyright notice as given in the publication in which the material was originally published) "With permission of Springer"
In case of use of a graph or illustration, the caption of the graph or illustration must be included, as it is indicated in the original publication.

Warranties: None
Springer makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity
You hereby indemnify and agree to hold harmless Springer and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you without Springer's written permission.

No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of Springer, by CCC on Springer's behalf).

Objection to Contrary Terms
Springer hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction
All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of Germany, in accordance with German law.

Other conditions:
V 12AUG2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
Learning Entry Profiles of Children with Autism from Multivariate Treatment Information Using Restricted Boltzmann Machines

Pratibha Vellanki

Springer eBook

Springer

Jan 1, 2015

Order Completed

Thank you very much for your order.

This is a License Agreement between Pratibha Vellanki ("You") and Springer ("Springer"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

Get the printable license.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3850060180910</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Apr 15, 2016</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Springer</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Springer eBook</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Learning Entry Profiles of Children with Autism from Multivariate Treatment Information Using Restricted Boltzmann Machines</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Pratibha Vellanki</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Jan 1, 2015</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Portion</td>
<td>Full text</td>
</tr>
<tr>
<td>Number of copies</td>
<td>1</td>
</tr>
<tr>
<td>Author of this Springer article</td>
<td>Yes and you are a contributor of the new work</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>Applied Machine Learning for Personalized Early Intervention in Autism</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>May 2016</td>
</tr>
<tr>
<td>Estimated size/pages</td>
<td>200</td>
</tr>
<tr>
<td>Total</td>
<td>0,00 USD</td>
</tr>
</tbody>
</table>

CLOSE WINDOW
Title: Nonparametric Discovery of Learning Patterns and Autism Subgroups from Therapeutic Data

Conference: Pattern Recognition (ICPR), 2014 22nd International Conference on

Author: P. Vellanki; T. Duong; S. Venkatesh; D. Phung

Publisher: IEEE

Date: 24-28 Aug. 2014

Copyright © 2014, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author’s approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author’s university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]’s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.