Indoor Location Prediction Using Multiple Wireless Received Signal Strengths

Kha Tran, Dinh Phung, Brett Adams, Svetha Venkatesh

Department of Computing
Curtin University of Technology,
GPO Box U 1987, Perth, WA, Australia,
{k.tran@postgrad,d.phung,b.adams,s.venkatesh}curtin.edu.au

Abstract

This paper presents a framework for indoor location prediction system using multiple wireless signals available freely in public or office spaces. We first propose an abstract architectural design for the system, outlining its key components and their functionalities. Different from existing works, such as robot indoor localization which requires as precise localization as possible, our work focuses on a higher grain: location prediction. Such a problem has a great implication in context-aware systems such as indoor navigation or smart self-managed mobile devices (e.g., battery management). Central to these systems is an effective method to perform location prediction under different constraints such as dealing with multiple wireless sources, effects of human body heats or mobility of the users. To this end, the second part of this paper presents a comparative and comprehensive study on different choices for modeling signals strengths and prediction methods under different condition settings. The results show that with simple, but effective modeling method, almost perfect prediction accuracy can be achieved in the static environment, and up to 85% in the presence of human movements. Finally, adopting the proposed framework we outline a fully developed system, named Marauder, that support user interface interaction and real-time voice-enabled location prediction.

Keywords: Indoor positioning, WiFi signal, Naive Bayes, Hidden Naive Bayes, indoor navigation.

1 Introduction

The increasing number of mobile devices has called for a new framework to exploit mobile computing power and to support more intelligent information services. To this end, context-aware applications that model information from users and their surrounding environments have been developed to provide value-added services. Information about context is multi-dimensional: positioning data, proximate people, communication and utility usage. Outdoor positioning is more or less a solved problem for devices equipped with GPS receivers. Indoor positioning, however, offering a myriad potential applications in indoor navigation and social pattern extraction, remains an open research problem, and is our focus in this study.

There are two main approaches to solving the indoor positioning problem: (i) installation of specialized indoor positioning systems, and (ii) use of existing radio-frequency infrastructures such as GSM, 802.11 and Bluetooth. Methods in the first category have high accuracy, but are expensive and unsuitable for large scale deployment. Methods using the latter approach are more economical, but suffer from signal instability and noise due to hardware characteristics, exacerbated by environmental factors, such as people in motion. We will focus on methods using 802.11 infrastructures. At the early, RADAR (Bahl & Padmanabhan 2000) applies the Nearest Neighbor algorithm to estimate location but a poor performance is obtained because it could not cover the nature of the variance of WiFi signals. Current approaches (Roos et al. 2002, Ladd et al. 2002, Krumm & Horvitz 2004, Xiang et al. 2004) get a better performance by viewing the problem in terms of probabilistic model which is well dealing with the the uncertainty. In these probabilistic approaches Bayes' rule is used for prediction and WiFi signals is in different form such as histogram (Youssef et al. 2003) and smoothed histogram (Roos et al. 2002), exponential functions (Xiang et al. 2004) and Gaussian (Kaemarungsi 2005). However, applying probabilistic model in recent works is empirical and there is no systematical investigation in terms of parameter estimation, prediction model selection as well as experiment environment. We will cast them as cases of Naive Bayes and discuss more in an unified framework. Furthermore, all recent approaches are fully-supervised and therefore the degree of calibration required is also a limiting factor to usability.

Motivated by the potential usefulness of indoor positioning systems to an array of applications, such as navigation of office workspaces, we desire a system generic enough to leverage existing WiFi access points found in an urban environment. Importantly, we examine the practical case where both the training and testing signals are acquired in a mobile fashion. We implement and compare two probabilistic models under a set of different conditions: Naive Bayes, where the signal at each WiFi access point is considered to be independent, and the Hidden Naive Bayes (Zhang et al. 2005), which models the joint relationship among the WiFi access point signals to estimate location by embedding the physical proximity of access points in an environment. We also make use of a Boolean adjacency matrix to impose constraints among moving paths. We perform experiments in different scenarios, including where the wireless device is fixed and in motion, both with and without the presence of humans. Our results demonstrate these models can be potentially deployed in complex environments by design and implementation of the real indoor positioning framework and an applications upon this framework.

The significance of this work is in using available

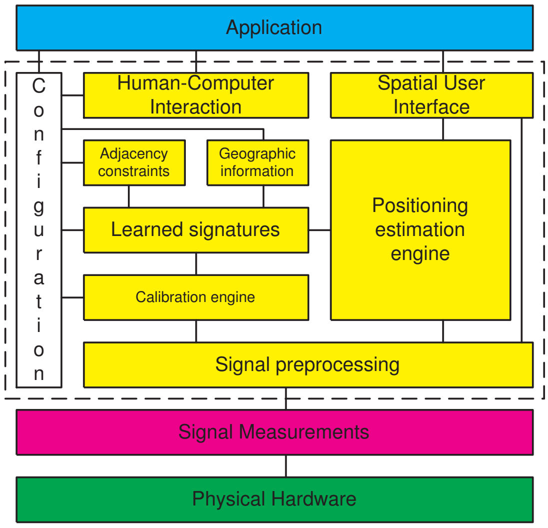


Figure 1: The architecture of Marauder.NET

low-cost infrastructures for location detection in a robust fashion. Importantly, as good performance is obtained for the case where both training and testing data is acquired in a mobile fashion, the model is suitable for general use in urban spaces, and in particular, for fine-grain indoor positioning for the visually impaired.

The layout of the remainder of the paper is as follows. The framework and its principal components are introduced in Section 2. Section 3 discusses about experiments and results. A indoor navigator prototype is demonstrated in Section 4. Session 5 provides a concluding summary.

2 Architecture

We first briefly outline each module in the proposed framework and then discuss in detail two principle components, namely database of learned signatures (signature representation) and positioning estimation engine (prediction model).

2.1 Proposed framework

Figure 1 outlines the architecture of the proposed framework in which the higher the layer, the more abstract the module. Layer Application sits on the top of schema with built-in indoor positioning functions is designed for user-oriental application such as indoor navigator, blind assistant, etc.. The two lowest modules of WiFi hardware and measurement are widely available in the market where most of WiFi adapter is integrated in recent wearable devices such as notebooks and smartphones and its its software drivers including signal measurement freely provided to popular operating systems such as PlaceLab [1] and OpenNetCF [2]. The heart of this system is the core engine inside the dashed rectangle which separates into several sub-modules and their relationships are represented as lines between components.

2.1.1 Signal pre processing

There are time-based techniques used to collected and bundled WiFi signals as a collection such as non-overlap window and overlap window (Figure 2). Normally, the window size is in order of seconds for daily office activities.

[1] www.placelab.org
[2] www.opennetcf.com

2.1.2 Calibration Engine

Our system is supervised so that the system requires training data collected in the calibration stage. The steps to collect the data at one location is very simple: user with mobile device is standing at that location and recording the WiFi signals for a given time interval. We introduce three approaches to labeling locations of text, voice and map-click in which two first approaches are positoningless (voice is suitable for people with blind while text is absolutely simple and can be automatically transferred to voice using available text-to-speech frameworks) and the last one is supporting the offset coordinates with related to provided partial vector/raster maps. Moreover, the process of calibration can be done incrementally and help the system more flexible and updated.

2.1.3 Trained Signatures

This database is the product of discussed calibration engine. One signature, which is represented for each location, consists of a set of W distributions of signal strengths of W access points and a distribution representing the number of appearance of W access points received at this location. Moreover, weak access points with infrequent number of appearance are also detected and eliminated out of final signatures. It requires mechanism to optimal organize and structure those signature in this database when the number of location is large. We propose a simple method of partitioning the whole database into cluster using access point MAC and geographic relationship. While access points MAC is available in signature and can be computed efficiently, the information of geographic relationship needs to be imported from user and service providers.

2.1.4 Geographic Information

This optional module takes the constraints among physical construction components in urban workspaces such as buildings, levels, sections and areas covered by access points. From that, the large number of locations is partitioned into sub-groups which reduce query processing time from estimation engine. This geographic information is usually stable and could be easily collected by user or service provider.

2.1.5 Adjacency Constraints

We introduce an optional module to keep a set of neighbor locations for particular location for faster retrieval. They are logical constraints that user can only move from a location to its neighboring locations. Once current location is known with a high probability, movement is constrained by topology around a given location, and hence only neighboring locations need be considered. This Boolean adjacency matrix is taken into account in our experiments.

2.1.6 Positioning Estimation Engine

Given a set of access points and their signal strengths, the estimation engine will query the a set of locations, calculate the posterior probabilities and the location with the highest probability is returned us predicted location.

2.1.7 Spatial User Interface

Its roles are for receiving the requests from utilities in layer Application and returning the corresponding location from estimation engine.

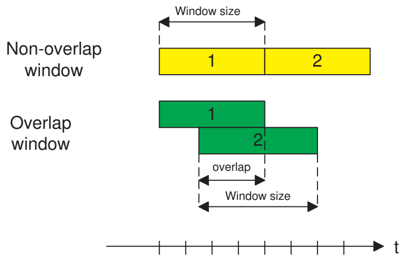


Figure 2: Non-overlap window and overlap window.

2.1.8 Human-Computer Interaction

Besides indoor location returning, Marauder framework also provides rich-informative meta-data warehouse such as vector/raster maps as background and voice guidance library which provides more relaxing for higher layers of application. While background map helps to provide more fancy and friendly to normal enduser, voice function relaxes users out of the device's monitor. Moreover, every piece of voice information about around context is trivial for normal people but is significantly meaningful to disable one so that this module aims to provide superior support to the blind.

2.2 Signature representation

Given a particular location, observed WiFi signals consist of the WiFi access point identifiers (MAC address) and corresponding received signal strength (RSS). We define the*location signature* as distributions of signal strengths over a finite set of access points received at that location. Precisely, the location signature consists of a set of W distributions of signal strengths over W access points and a multinomial distribution representing the number of appearances of these W access points at this location.

While the frequency of appearance of W access points is often modeled as discrete distribution of size W, there are different methods to model the signal strengths over each access points and the chosen method can affect the prediction accuracy significantly. Figure 3 shows the plot of measured WiFi signals of one access points in 5-minute interval at a particular location when mobile device is hold stay still at a position. The blue bar and red curve show the empirical histogram and the estimated Gaussian distribution respectively.

Three methods of modeling, namely histogram, smoothed histogram with kernel Gaussian function and Gaussian are investigated in this works. With a small bin of 1dBm, the signal strength is discrete into $V = 100$ values from -100dBm to 0dBm and counts over all received signals. The histogram signature is the distribution of V normalized values. Kernel Gaussian function $\mathcal{K}(y) = \frac{1}{2\pi\sigma_k^2}\exp(\frac{(y-\mu_k)^2}{2})$ where (μ_k, σ_k^2) is kernel parameters and y is signal strength, is introduced to smooth the histogram. The number of parameters needs for storing a signature in histogram as well as smoothed histogram distributions are the same and equal to $W(1+V)$ parameters if the pin is $1dBm$. Gaussian distribution captures the signals with just two parameters, mean and variance.

2.3 Prediction model and signature parameter estimation

Location prediction in our work is cast as a classification problem. Most previous works has used the Naive Bayes (NB) which the critical assumption is the independence of received signals among access points conditionally on the current location. One model to deal with correlation among attributes is Hidden Naive Bayes (HNB) (Zhang et al. 2005). It creates a hidden parent node for each attribute node, capturing the influence from other nodes. Below we briefly outline both the NB and HNB.

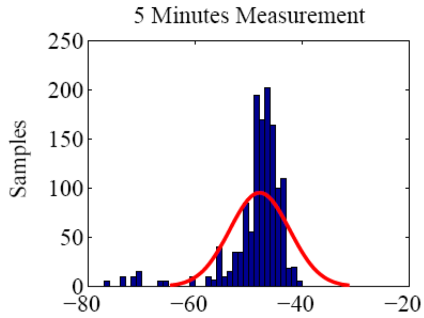


Figure 3: The variance of RSS at a investigated location.

Let $C \in \{1, ..., K\}$ be the location random variable where K is number of locations, $X_m \in \{1, ..., W\}$ represents the m-th access point random variable, $Y_m \in \{1, ..., V\}$ represent the signal strength corresponding to m-th access point where W is number of access points, M is number of access points of an observation and V is number of discrete values of signal strength. Signature parameters are estimated from a set of training data of N observations $D = \{\mathbf{o}_1, ..., \mathbf{o}_N\}$ where $\mathbf{o}_n = (c^{(n)}, x_1^{(n)}, y_1^{(n)}, ..., x_M^{(n)}, y_M^{(n)}), n = 1, ..., N$. In the prediction phase, the predicted location c^* is inferred based on current observation $\mathbf{o} = (x_1, y_1, ..., x_M, y_M)$.

2.3.1 Naive Bayes

Figure 4 shows the NB. The joint distribution $P(C, X_1, Y_1, \ldots, X_M, Y_M)$ is given by:

$$P(C)\prod_{m=1}^{M} P(X_m|C)P(Y_m|C, X_m)$$

where the distribution of access point x given a location c, $P(X_m = x|C = c)$ is multinomial (W-size parameter π_c), the probability of signal strength y given location c and access point x, $P(Y_m = y|C = c, X_m = x)$, is normalized histogram (V-size vector parameter $\gamma_{c,x}$), smoothed histogram (V-size vector parameter $\eta_{c,x}$), Gaussian (two parameters $\mu_{c,x}$ and $\sigma_{c,x}^2$) respectively. Without any prior knowledge about the current location c, the distribution $P(C = c)$ could be assigned as uniform.

Let the identify function $\mathbb{I}(a, b) = 1$ if $a = b$ else $= 0$, in maximum likelihood estimation framework, the sufficient statistics are:

$$n_c = \sum_{n=1}^{N}\sum_{m=1}^{M} \mathbb{I}(c^{(n)}, c)$$

$$n_{c,x}^{(y)} = \sum_{n=1}^{N}\sum_{m=1}^{M} \mathbb{I}(c^{(n)}, c)\mathbb{I}(x_m^{(n)}, x)\mathbb{I}(y_m^{(n)}, y)$$

$$n_c^{(x)} = \sum_{n=1}^{N}\sum_{m=1}^{M} \mathbb{I}(c^{(n)}, c)\mathbb{I}(x_m^{(n)}, x)$$

The parameters of $P(X_m|C)$ are estimated as

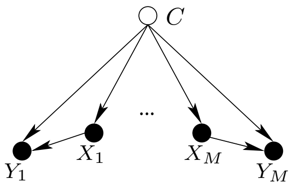


Figure 4: Naive Bayes model.

$$\hat{\pi}_c^{(x)} = \frac{n_c^{(x)} + 1}{n_c + W}$$

The parameters of $P(Y_m|C, X_m)$ are differently estimated according to three methods of representation. In histogram case, the parameters are as follows

$$\hat{\gamma}_{c,x}^{(y)} = \frac{n_{c,x}^{(y)} + 1}{n_c^{(x)} + V}$$

In smoothed case, the parameters are estimated as:

$$\hat{\eta}_{c,x}^{(y)} = \frac{m_{c,x}^{(y)}}{\sum_{y=1}^{V} m_{c,x}^{(y)}}$$

where

$$m_{c,x}^{(y)} = \sum_{n=1}^{N}\sum_{m=1}^{M} \mathbb{I}(c^{(n)}, c)\mathbb{I}(x_m^{(n)}, x)\mathcal{K}(y - y_m^{(n)})$$

In the Gaussian case, the mean and variance are

$$\hat{\mu}_{c,x} = \frac{m_{c,x}}{n_{c,x}}$$

$$\hat{\sigma}_{c,x}^2 = \frac{m_{c,x}^2}{n_{c,x}}$$

where

$$m_{c,x} = \sum_{n=1}^{N}\sum_{m=1}^{M} \mathbb{I}(c^{(n)}, c)\mathbb{I}(x_m^{(n)}, x)y_m^{(n)}$$

and

$$m_{c,x}^2 = \sum_{n=1}^{N}\sum_{m=1}^{M} \mathbb{I}(c^{(n)}, c)\mathbb{I}(x_m^{(n)}, x)(y_m^{(n)} - \mu_{c,x})^2$$

At the prediction step, the location is found by finding the location having the highest likelihood:

$$\begin{aligned} c^* &\propto \arg\max_{c\in\{1,\ldots,K\}} P(\mathbf{o}|c)P(c) \\ &= \arg\max_{c\in\{1,\ldots,K\}} P(c)\prod_{m=1}^{M} P(x_m|c)P(y_m|c, x_m) \end{aligned}$$

where Bayes' rule is used.

2.3.2 Hidden Naive Bayes

HNB relaxes the independent assumption in the NB by letting attributes depends on each other. In our case the HNB approximates the full correlation of access points by creating a hidden parent variable H_m for each variable Y_m and then linearly simplifies the conditional probabilities. Figure 5.a and 5.b show fully connected node Y_m and its HNB approximation. The joint distribution $P(C, X_1, Y_1, \ldots, X_M, Y_M)$ is defined as:

$$P(C)\prod_{m=1}^{M} P(X_m|C)P(Y_m|Y_{-m}, X_m, C)$$

where distribution of access point x_m given location c $P(X_m = x|C = c)$ is multinomial and the distribution of signal strength y_m of access point x_m given location c and a set of signal strengths $Y_{-m} = \{Y_1 = y_1, ..., Y_{m-1} = y_{m-1}, Y_{m+1} = y_{m+1}, ..., Y_M\}$ of other access points $P(Y_m = y_m|Y_{-m}, X_m = x_m, C = c)$ is also a multinomial.

In (Zhang et al. 2005), $P(Y_m|Y_{-m}, X_m, C)$ is represented by $P(Y_m|H_m, X_m, C)$ and is formulated as:

$$\sum_{j=1, j\neq m}^{M} w_{x_m, x_j|c} P(Y_m|Y_j, X_j, X_m, C)$$

where $\sum_{m=1,j\neq i}^{M} w_{X_m,X_j|C} = 1$.

The weight $w_{x_i x_j|c}$ of two access points x_i and x_j conditional on location c is defined in (Zhang et al. 2005):

$$w_{x_i,x_j|c} = \frac{I_P(x_i, x_j|c)}{\sum_{j=1,j\neq i}^{M} I_P(x_i, x_j|c)}$$

where $I_P(x_i, x_j|c)$ is the conditional mutual information

$$I_P(x_i, x_j|c) = H(x_i|c) + H(x_j|c) - H(x_i, x_j|c)$$

and $H(x_i|c)$ is the entropy of access point x_i and $H(x_i, x_j|c)$ is the joint entropy of x_i and x_j:

$$H(x_i|c) = -\sum_{y_i=1}^{V} P(y_i|x_i, c)\log P(y_i|x_i, c)$$

and

$$\begin{aligned} H(x_i, x_j|c) &= -\sum_{y_i=1}^{V}\sum_{y_j=1}^{V} P(y_i, y_j|x_i, x_j, c) \\ &\quad \log P(y_i, y_j|x_i, x_j, c) \end{aligned}$$

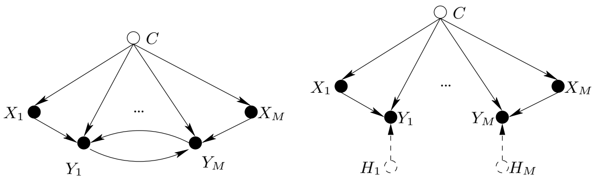
Defining the all distributions in HNB as multinomial where the parameters of $P(y_i|y_j, x_j, x_i, c)$ is V-size vector $\tau_{x_i|x_j,y_j,c}$, the parameter of $P(x_i|c)$ is W-size vector π_c, the parameter of $P(y_i|x_i, c)$ is V-size vector $\gamma_{x_i|c}$ and the parameter of $P(y_i, y_j|x_i, x_j, c)$ is $V \times V$-dimension matrix $\Phi_{x_i,x_j|c}$.

The sufficient statistics in this case are:

$$n_c = \sum_{n=1}^{N} \mathbb{I}(c^{(n)}, c)$$

$$n_c^{(x_i)} = \sum_{n=1}^{N}\sum_{m=1}^{M} \mathbb{I}(c^{(n)}, c)\mathbb{I}(x_m^{(n)}, x_i)$$

$$n_{c,x_i}^{(y_i)} = \sum_{n=1}^{N}\sum_{m=1}^{M} \mathbb{I}(c^{(n)}, c)\mathbb{I}(x_m^{(n)}, x_i)\mathbb{I}(y_m^{(n)}, y_i)$$



(a) (b)

Figure 5: (a) Model when received signals are fully dependent and (b) its approximation, the HNB.

Figure 6: Layout of office space used in the experiments.

$$n^{(y_i,y_j)}_{c,x_i,x_j} = \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{l=1,l\neq m}^{M} \mathbb{I}(c^{(n)},c)\mathbb{I}(x_m^{(n)},x_i) \mathbb{I}(y_m^{(n)},y_i)\mathbb{I}(x_l^{(n)},x_j)\mathbb{I}(y_l^{(n)},y_j)$$

The parameters of $P(y_i|x_i,c)$, $P(y_i,y_j|x_i,x_j,c)$ and $P(y_i|y_j,x_j,x_i,c)$ are estimated as follows (Zhang et al. 2005):

$$\hat{\pi}_c^{(x_i)} = \frac{n_c^{(x_i)}+1}{n_c+W}$$

$$\hat{\gamma}^{(y_i)}_{x_i|c} = \frac{n^{(y_i)}_{c,x_i}+1}{n_c+V}$$

$$\hat{\Phi}^{(y_i,y_j)}_{x_i,x_j|c} = \frac{n^{(y_i,y_j)}_{c,x_i,x_j}+1}{n_c+V^2}$$

$$\hat{\tau}^{(y_i)}_{x_i|x_j,y_j,c} = \frac{n^{(y_i,y_j)}_{c,x_i,x_j}+1}{n^{(y_j)}_{c,x_j}+V}$$

Similar to NB model, at the classification step, the location is found by finding the location having the highest likelihood:

$$\begin{aligned} c^* &\propto \arg\max_{c\in\{1,\ldots,K\}} P(c|o)P(c) \\ &= \arg\max_{c\in\{1,\ldots,K\}} P(c)\prod_{i=1}^{M}P(x_i|c) \\ &\quad \sum_{j=1,j\neq i}^{M} w_{x_i,x_j|c}P(y_i|y_j,x_j,x_i,c) \end{aligned}$$

Again, without any prior knowledge about current location, the probability $P(c)$ is assigned to uniform distribution and have no effect during classification step.

3 Experiments

We conducted experiments comparing the NB with the HNB. In the case of NB, we consider three cases wherein the RSS is represented as a histogram (Model I, NB+H), smoothed histogram using a kernel Gaussian function (Model II, NB+K), and the Gaussian (Model III, NB+G). In the case of HNB, the RSS is represented by a histogram (Model IV, HNB+H). In order to investigate realistic settings, three environments were defined: A–no humans present, B–humans present but not moving, and C–humans moving during testing and training. Investigated results illustrate that the system performance will be significant affected with human presence and especially human in moving (Xiang et al. 2004). RSS is processed using both with and without overlap window. The system is predicted in a time slot of every 2 seconds. The non-overlap window is 2s while overlap window size is 4 seconds with 2 second overlap.

The system was set up in a corridor area whose layout is depicted in Figure 6 (corridor is indicated in yellow). The building is equipped with an IEEE 802.11b wireless network with 2.4 GHz frequency bandwidth consisting of three Cisco Aironet 1200 Series access points. The calibration and testing program was run on a Sony Vaio VGN-UX17GP under Windows XP with a built-in wireless card (Intel(R) PRO/Wireless 3945ABG). We modeled the environment as 12 locations (1-12) with the distance between two neighboring locations being 4 meters. For each scenario described above, training data was collected for each location in 5 minutes intervals with approximately 300 observations. The calibration data $12\text{locations} \times 15\text{minutes} \times 3\text{environments} = 9\text{hours}$ is randomly divided to 3 parts, 1 for training and 2 testing.

The system performance is evaluated by using two measures of accuracy (meters) and precision (percentage) adapted in (Liu et al. 2007). While the accuracy is predefined according to the calibration data, the precision is the distribution of distance error between the estimated location and the true location. The data collected for each location belong to a region with the radius of $2m$, therefore the precision with accuracy $2m$ is the recall rate of which is measured as the ratio of estimated location and ground truths.

Table 1 shows the precision for four models in twelve scenarios. Overall, model NB+H is marginally better compared to the other models, especially in noisy environments. The rate decreases gradually as noise is introduced as a result of allowing moving humans and objects in the environment, and increases when tuning techniques are integrated.

In terms of tuning techniques, overlap window yields improved results of approximately 10% in scenarios where humans are moving. While the use of an adjacency matrix improves only 2% in performance, it does reduce computation time considerably in case of large-scale environment because of its role of clustering.

Surprisingly, the HNB with more complex and computational model, does not demonstrate superior performance compared to simpler models. On the other hand, although obtained performance is slightly lower, model NB+G shows potential opportunity to be deployed as large-scale system in real environment because of its useful characteristics such as compressed signature and simple prediction engine.

Table 1: Precision rate (%) when accuracy is 2 meters.

Environment	A (no humans)				B (humans static)				C (humans moving)			
Window(overlap)	2s		4s(2s)		2s		4s(2s)		2s		4s(2s)	
Adjacency constraint	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
I (NB+H)	99.09	99.19	99.34	99.39	93.25	93.92	96.88	97.10	74.22	77.14	85.36	85.64
II (NB+K)	99.19	99.29	99.14	99.14	92.06	93.84	95.77	96.07	70.16	72.89	79.98	80.54
III (NB+G)	95.22	95.62	96.83	96.89	90.88	92.14	94.51	94.81	67.51	68.90	77.43	78.73
IV (HNB+H)	99.39	99.19	99.39	99.39	87.69	89.54	92.95	93.03	70.25	72.61	78.94	80.92

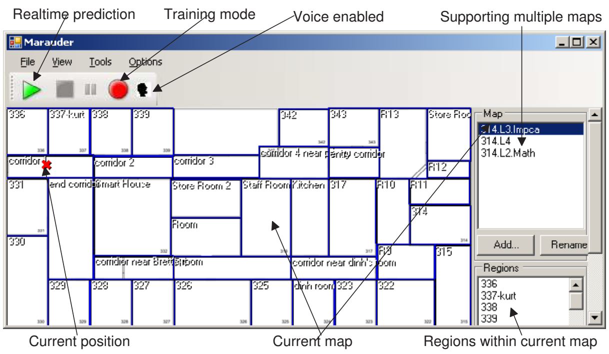


Figure 7: Indoor navigator prototype is developed using Marauder.NET framework.

4 Marauder

Several potential applications can be developed based on our proposed framework such as indoor navigator and blind positioning assistant. Figure 7 shows an indoor navigator application named Marauder, working under Windows XP platform and running in portable device Sony Vaio VGN-UX17GP. This application support end-users to import background maps, setup calibration regions in preparation step and locate where we are in the building. Existing background map is listed in top-right panel where their names represent hierarchical relationship such as building, floor, section so so on. In the bottom-right panel, set of calibrated regions of the current map are easy to adjust/add/remove. The wide center area shows the current map where the red crossing sign tells us where we are in this map. While the green triangle button is enable for realtime location prediction, the red circle button supports for recording the signatures in training phase. To release user out of application monitor or support visual impaired, voice guidance assistant could be triggered with black human button on the toolbar. Besides, there are several other functions such as navigating and zooming (menu View) and prediction engine mode (menu Tools) and managing parameter and reporting (menu Options) built in the menubar on the top of GUI.

5 Conclusion

A framework is proposed to provide indoor positioning capabilities for an array of potential applications such as indoor navigator, visual-impaired assistant or indoor surveillance system. We present a systematic study of two probabilistic models, the Naive Bayes and Hidden Naive Bayes, for positioning classification using WiFi signals. We also have experimented with various methods of modeling signal strengths, histogram, smoothed histogram and Gaussian in several different conditions of real environments. The results show that simple Bayesian models can be used to provide a reliable location detection accuracy. The precision is nearly perfect in non-human environment, around 95% while people are still and 85% in moving circumstance. Surprisingly, HNB model only shows same performance in non-human case and slightly less accuracy in most of remaining scenarios.

References

Bahl, P. & Padmanabhan, V. (2000), RADAR: an in-building RF-based user location and tracking system, *in* 'Proceedings of The 19th Annual Joint Conference of The IEEE Computer and Communications Societies (INFOCOM)', Vol. 2.

Kaemarungsi, K. (2005), Design of Indoor Positioning Systems Based on Location Fingerprinting Technique, PhD thesis, University of Pittsburgh.

Krumm, J. & Horvitz, E. (2004), Locadio: Inferring Motion and Location from Wi-Fi Signal Strengths, *in* 'Proceedings of International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous)'.

Ladd, A., Bekris, K., Rudys, A., Kavraki, L. & Wallach, D. (2002), Robotics-Based Location Sensing Using Wireless Ethernet, *in* 'Proceesings of The 8th ACM International Conference on Mobile Computing and Networking (MOBICOM)'.

Liu, H., Darabi, H., Banerjee, P. & Liu, J. (2007), 'Survey of Wireless Indoor Positioning Techniques and Systems', *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews* **37**(6), 1067–1080.

Roos, T., Myllymäki, P., Tirri, H., Misikangas, P. & Sievänen, J. (2002), 'A Probabilistic Approach to WLAN User Location Estimation', *International Journal of Wireless Information Networks* **9**(3), 155–164.

Xiang, Z., Song, S., Chen, J., Wang, H., Huang, J. & Gao, X. (2004), 'A Wireless LAN-based Indoor Positioning Technology', *IBM Journal of Research and Development* **48**(5/6), 617–626.

Youssef, M., Agrawala, A. & Udaya Shankar, A. (2003), WLAN location determination via clustering and probability distributions, *in* 'Proceedings of the 1st IEEE International Conference on Pervasive Computing and Communications (PerCom)', pp. 143–150.

Zhang, H., Jiang, L. & Su, J. (2005), Hidden Naive Bayes, *in* 'Procedings of the 20th National Conference on Artificial Intelligence (AAAI)', pp. 919–924.