Beliakov, Gleb and Warren, Jim 2001, Appropriate choice of aggregation operators in fuzzy decision support systems, *IEEE transactions on fuzzy systems*, vol. 9, no. 6, pp. 773-784.

Appropriate Choice of Aggregation Operators in Fuzzy Decision Support Systems

Gleb Beliakov and Jim Warren

***Abstract*—Fuzzy logic provides a mathematical formalism for a unified treatment of vagueness and imprecision that are ever present in decision support and expert systems in many areas. The choice of aggregation operators is crucial to the behavior of the system that is intended to mimic human decision making. This paper discusses how aggregation operators can be selected and adjusted to fit empirical data—a series of test cases. Both parametric and nonparametric regression are considered and compared. A practical application of the proposed methods to electronic implementation of clinical guidelines is presented.**

***Index Terms*—Aggregation operators, clinical guidelines, monotone splines, restricted least squares.**

I. INTRODUCTION

EXPERT and decision support systems are common in the areas where the alternatives are selected based on combined support of a number of factors, none of which could determine the alternative by itself. An example of such an area is medicine, where diagnosis or management are almost never decided based on individual criterion. A weighted combination of many criteria is used instead, each criterion may support various alternatives, and the alternative with the strongest support is selected as the decision.

This is a typical problem of multicriteria decision making (MCDM), various approaches to which have been discussed in [16], [30], [37], [38], [44], [70]. One important class of methods in MCDM is based on constructing a utility or value function $u(x)$, which represents the overall strength of support in favor of the alternative x. This approach is known as multiattribute utility theory (MAUT).

In MAUT, one can represent the preference relation $\succeq$ on a set of alternatives X with a single-valued function $u(x)$ on X, called utility, such that for any $x, y \in X x \succeq y \iff u(x) \geq u(y)$. Maximization of $u(x)$ over X provides the solution to the problem of selecting x.

Essentially, there are two conditions that guarantee the existence of the utility function. One is that the relation $\succeq$ is a complete preorder, that is, all alternatives in X are comparable and can be either preferred one to another $(x \succ y)$ or be indifferent $(x \sim y)$, both $\succ$ and $\sim$ are transitive, $\succ$ is asymmetric, and $\sim$ is symmetric and reflexive. The second condition guarantees that there are sufficiently many real numbers to represent the preference relation. Formally, it states that there must exist a countable subset $A \subset X$, such that for any two alternatives $x, y \in X, x \succ y$, there exists $z \in A : x \succ z \succ y$. An example when the utility function does not exist is when two or more criteria are ordered lexicographically [27], [37]

$$(x_1, x_2) \succ (y_1, y_2) \iff [x_1 > y_1 \text{ or } (x_1 = y_1, x_2 > y_2)].$$

The utility function is defined up to an increasing monotone transform, which preserves the ordering of the alternatives. A positive linear transform $v(x) = au(x) + b, a > 0$, preserves not only the ordering but also the order of preference differences. Consequently, the utility function can always be scaled to a suitable interval (the unit interval $[0, 1]$ for convenience) with a simple change of variables. In this paper, without loss of generality, the range of the utility functions will be presumed to be $[0, 1]$.

In fuzzy set theory (FST), membership functions of fuzzy sets play the role similar to the utility functions—the role of degrees of preference [10]. Many authors, including Zadeh himself, refer to membership functions as "a kind of utility functions" [13], [23]–[25], [42], [58], [68], [70]. The equivalence of utility and membership functions extends from semantical to syntactical level [8], [9]. Although, this is not the only possible interpretation of membership functions [7], [24], it allows one to formulate and solve problems of MCDM using the formalism of FST.

The vector maximization problem of MAUT

$$\text{Maximize } (u_1(x_1), u_2(x_2), \ldots u_n(x_n)) \text{ over } X$$

where $u_i(x_i)$ are the utility functions of the corresponding attributes x_i, takes the form

$$\text{Maximize } u(\mathbf{x}) = U(u_1(x_1), u_2(x_2), \ldots, u_n(x_n))$$

where U is some function of n real variables which aggregates the individual utility values into the overall utility of the alternative $\mathbf{x}$. An essential assumption is that the individual utility functions $u_i(x_i)$ exist, in other words, the attributes are utility independent of the other attributes.

In FST this problem takes the form

$$\text{Maximize } \mu_\Omega(\mathbf{x}) = A_{\text{gg}}(\mu_{A1}(x_1), \mu_{A2}(x_2), \ldots, \mu_{An}(x_n))$$

where A_{gg} stands for an appropriate aggregation operator. It combines the membership values in the sets $A_1, A_2, \ldots A_n$, into the membership value of the set Ω, formed by some operation

Manuscript received October 10, 2000; revised May 21, 2001. This work was supported in part by an Australian Research Council Small Grant awarded through the University of South Australia.

G. Beliakov is with the School of Computing and Mathematics, Deakin University, Rusden campus, Clayton 3168, Australia (e-mail: gleb@deakin.edu.au).

J. Warren is with the Health Informatics Research Group, School of Computer and Information Science, University of South Australia, Levels Campus, Mawson Lakes S 5095, Australia (e-mail: warren@cs.unisa.edu.au).

Publisher Item Identifier S 1063-6706(01)09618-7.

on the sets $A_1, A_2, \dots A_n$, such as intersection, union, or their combination. For example Ω could be $A_1 \cap A_2 A_3$, or $A_1 \cup A_2 \cup A_3$, or $(A_1 \cap A_2) \cup A_3$, etc.

In MAUT, the combining function U is usually additive or multiplicative

$$U(\mathbf{x}) = a_1 u_1(x_1) + a_2 u_2(x_2) + \cdots + a_n u_n(x_n), \text{ or}$$
$$U(\mathbf{x}) = \frac{1}{\mu}[(1 + \mu a_1 u_1(x_1))(1 + \mu a_2 u_2(x_2)) \cdots (1 + \mu a_n u_n(x_n)) - 1]$$

although, other functions are not explicitly prohibited. Additive aggregation, as well as max, min and product operators, are traditionally used in neuro-fuzzy systems [15], [40], [41], but again, more general forms of aggregation are allowed.

In FST the aggregation operators take a large variety of forms, e.g., $\min, \max$, Yager, Dubois & Prade, Schweizer & Sklar, Hamacher, Frank, Dombi families, averaging operators, ordered weight aggregation (OWA), compensatory operators, operators based on Choquet and Sugeno integrals, etc. [22], [23], [31], [32], [42], [52]–[55], [70]. Aggregation operators have been extensively studied and their application to MAUT problems provides a good theoretical setting.

The choice of aggregation operators in FST is not simple. Initially, only $\min$ and $\max$ operators were used to model fuzzy set intersection and union, primarily because of their strong algebraic properties: these are the only operators that preserve mutual distributivity. With the development of the theory other operators have emerged. They provided compensatory properties and better fit to empirical data and, therefore, seemed to model human decision making better. Nowadays, there are some 90 different families of aggregation operators used in various applications. Theoretically, they are all equivalent and can be mapped to a class of metrics in which the similarity to the Ideal alternative is measured [8], [9]. Therefore, other criteria have to be used to select the appropriate aggregation operator [70]. Among these criteria we will emphasize the empirical fit, adaptability and semantical clarity.

This paper will discuss various methods of selecting and adjusting aggregation operators based on empirical data and expert opinion. We consider the situations where no *a priori* knowledge about the properties of operators is available, and therefore they have to be built using exclusively empirical data, as well as the situations where the expert opinion dictates the form of the operator and the available free parameters are adjusted to fit the data. We illustrate this process on model problems, and then on a real application of a medical decision support system that incorporates clinical guidelines.

II. Clinical Guidelines in Medical Decision Support System

Almost any medical computer application can be classified as a medical decision support system—a computer program designed to help health professionals make clinical decisions [6]. In this paper we are considering only those applications that provide clinicians with some form of active customized advice, based on symptoms and signs from the electronic patient record. This, of course, does not exclude other forms of advice, such as consultation by the clinician of general information about diseases, treatments, protocols, guidelines, etc. But our primary goal is to develop mechanisms of customising the advice to the specific problems of a given patient.

This problem has two faces: 1) generation of the advice and (2) its delivery. We have discussed the problem of the delivery of the advice, including fuzzy advice, elsewhere [10], [64], [65]. This paper addresses generation of the advice based on the electronic patient record (EPR). Also, we do not discuss the problem of diagnosis and the associated problem of representing general medical knowledge. We limit the scope of this paper to a specific niche of generating advice on treatment and patient management options, based on clinical practice guidelines.

Clinical practice guidelines (hereafter, *clinical guidelines* or simply *guidelines*) are standardized specifications for care developed by a formal process that incorporates the best scientific evidence of effectiveness with opinions of experts in the fields [47]. In general, they have been developed in an effort to reduce escalating health care costs without sacrificing quality and have been shown to improve health care outcomes when followed [34]. To be effective, guidelines need to be integrated into the physician's decision-making process in daily practice [45]. It has been recognized that the guideline statements should be linked to the actual patient data, and therefore be integrated with the EPR. The most predictable impact is achieved when "the guideline is made accessible through computer-based, patient-specific reminders that are integrated into the clinician's workflow" [45], [69]. That is, ideally one provides guidance "just in time" in a clinical workstation environment—for example, as with a drug interaction alert at the time a doctor writes a prescription.

Most current guidelines are not represented in the form of algorithms. Instead, they are implemented in the form of text narratives, describing possible medical conditions and signs with the appropriate recommendations. This fact creates a significant obstacle for computerising clinical guidelines, their electronic exchange and assessment. Despite recent progress in developing formal syntax for guideline representation [45], [47]–[50], [57], [62], [69], in the computerised form the guidelines are mostly translations of text-based narratives [69].

It is the task of knowledge engineers to extract knowledge from health professionals and to represent guidelines in more suitable (for computers) form, such as the collection of *If…then…* rules. It turns out, however, that even if formulated as *If…then…* rules, clinical guidelines are still not suitable for computer implementation. There are different sources of uncertainty present, among which are: lack of information, nonspecificity, probabilistic nature of data and outcomes, vagueness of recommendations, strife and fuzziness in determination and interpretation of clinical signs [1], [7], [36], [43], [49], [65].

In this paper we only deal with the problem of fuzziness and vagueness, regardless of their source. Suppose that the guideline has been formulated as

$$\textbf{If}\, x \text{ is } A \textbf{ AND } x \text{ is } B, \textbf{THEN } C.$$

Here x is a linguistic variable, A and B are fuzzy sets and C is the rule consequent. For example, one such guideline could be

If x is obese **AND** x is smoker,
THEN check blood pressure every visit.

Both sets *obese* and *smoker* are fuzzy, and the numerical variables body mass index (BMI), defined as the ratio of weight and the square of height of the patient, and the number of cigarettes per day determine the membership values of x in these sets. To calculate the strength of the recommendation, the membership values of both antecedents must be combined using an aggregation operator, and then the result is used as the argument of the implication operator.

This process seems straightforward, and generally $\min$, product, or other simple operators are used for aggregation in such simplified examples. When it comes to practical implementation of the system, however, it turns out that its behavior is quite different from that of the experts who provided the rules. Firstly, despite having identical form of the antecedent, like A AND B AND C, different rules seem to require different aggregation operators. Secondly, the rule antecedents are usually more complicated. Combination of conjunction and disjunction, like (A OR B OR C) AND (D OR E), are common. Thirdly, the meaning of the connectives "and" and "or" is not fixed. In some places they correspond to disjunction and conjunction, in other places the meaning is reversed (consider the statement "the drawer contains forks AND knives" [36]), and in some cases they do not correspond to either. This is the case when various criteria support each other and trigger the rule collectively (what Kasabov calls synergism ([41], p. 357), or when the criteria display conjunctive and disjunctive behavior simultaneously (compensatory operators [70]), or when aggregation depends on their values (i.e., conjunctive behavior for small values and disjunctive behavior for big values [54]). Consequently, it does not look feasible to predefine the form of the aggregation operators and use them throughout the rule base, one for "and", the other for "or." Instead, each rule should be examined separately and the appropriate operator for it should be found based on the feeling of the experts and on their actual decision pattern over the set of typical examples. The following sections describe this process in detail.

III. Aggregation Operators

When the rules in the decision support system contain more than one antecedent, the degrees of strength of the antecedents need to be combined to determine the overall strength of the rule consequent. In the language of fuzzy sets, the membership values of the linguistic variables in the rule antecedents have to be combined using an aggregation operator. Formally, a general aggregation operator is a real function $f:[0,1]^n \to [0,1]$, nondecreasing in all arguments, with the properties $f(\mathbf{0})=0$ and $f(\mathbf{1})=1$ (a number in bold denotes an n-vector) [42], [55], [70]. This is the most general mathematical representation of aggregation, because the monotonicity requirement is essential for order preservation (if one criterion increases the support of the alternative, the overall support cannot decrease), and the normalization conditions guarantee consistency with classical logic in the limiting case.

General aggregation operators display the whole range of behavior: disjunctive, conjunctive, averaging, mixed, commutative, mutually reinforcing or otherwise, and correspond to vague and loosely defined "and" and "or" connectives, or synergism [41].

All aggregation operators are equivalent to the distance to the ideal $\mathbf{1}$ or anti-ideal $\mathbf{0}$ in the relevant metric [8], [9], and therefore are equivalent among themselves. Particular families of aggregation operators have been identified by enforcing some other properties, such as commutativity and associativity. Triangular norms and conorms and averaging operators are well known examples [23], [42], [55], [70]. However, many other operators have emerged: compensatory operators, uninorm operators, ordered weighted aggregation, operators based on Choquet and Sugeno fuzzy integrals, piecewise continuous operators, etc. [31], [32], [52]–[54]. Families of aggregation operators overlap, some include other families, some form equivalence classes [31].

Because different classes of aggregation operators display substantially different behavior, it is not logical to use any particular class to provide generic representation of aggregation. Therefore, we will use general aggregation operators to model aggregation of rule antecedents in decision support systems. They will provide the highest degree of adaptability and excellent empirical fit. However, if there are strong reasons to restrict the selection to a particular family of operators, we will impose the relevant constraints.

Consider general aggregation operator $f(\mathbf{x})$. The function f can have a simple algebraic form, such as

$$
\begin{aligned}
f(\mathbf{x}) &= \min(x_1, x_2, x_3, \dots, x_n) \quad \text{or}\\
f(\mathbf{x}) &= x_1 \times x_2 \times x_3 \times \cdots \times x_n \quad \text{or}\\
f(\mathbf{x}) &= \min(1, x_1 + x_2 + x_3 + \cdots + x_n) \quad \text{or}\\
f(\mathbf{x}) &= (x_1 + x_2 + x_3 + \cdots + x_n)/n.
\end{aligned}
$$

f can also be a combination of simple operators, like

$$
f(\mathbf{x}) = \alpha \min(x_1, x_2, x_3, \dots, x_n) + (1-\alpha)\max(x_1, x_2, x_3, \dots, x_n).
$$

The degrees of importance of rule antecedents (vector a) can be easily incorporated into aggregation operators in a variety of ways [8], [67], [68]. For example

$$
\begin{aligned}
f(\mathbf{x}) &= \min(x_1, a_1) \times \min(x_2, a_2)\\
&\quad \times \min(x_3, a_3) \times \cdots \times \min(x_n, a_n) \text{ or}\\
f(\mathbf{x}) &= \min(1, a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n).
\end{aligned}
$$

f can incorporate permutations of the antecedents (OWA) or can be expressed as Choquet or Sugeno integrals [32], [42], [43]. Finally, f might not have a meaningful algebraic representation. In this case we think of it as a function $f(\mathbf{x};\mathbf{p})$, which depends on m parameters $\mathbf{p}$, that could be adjusted to fit empirical data. In all cases, f needs to be nondecreasing in all arguments, and to satisfy $f(\mathbf{0})=0$ and $f(\mathbf{1})=1$.

Our task is to select one or another operator for a given decision rule based on observed decision pattern of the experts. We assume that empirical data of the form $\{(\mathbf{x}_i, y_i)\}_{i=1}^{I}$, where $\mathbf{x}_i$ are the vectors of membership values of the antecedents and y_i are the strengths of response of the expert, is available. In addition, the expert's opinion about the algebraic form of the operator or its properties, and the initial guess about the degrees of importance of the criteria might also be available.

Essentially, we have two established techniques we can use (1) parametric and (2) nonparametric regression. In parametric regression, we fix the algebraic form of the operator, and fit the available free parameters to the data. The main advantage of this approach is its clarity: the parameters have semantical meaning, properties of the operator are known, its behavior is predictable, and, if necessary, calculations can be performed with pen and paper. Moreover, if the algebraic form of the operator has meaning in the specialist's domain (as discussed in the next section), then not only the end result but also intermediary steps of aggregation are clear.

In contrast, in nonparametric regression parameters are meaningless,[1] and aggregation operator behaves like a black box. However, the nonparametric representation is much more adaptable and versatile (in modeling functions of different shapes), it provides much better fit to empirical data and does not rely on the correctly selected class of aggregation operators. Moreover, the technique of splines we present in a later section allows one to balance the generality of representation with the need to specify particular classes (e.g., commutative operators or triangular norms).

Let us formulate the problem formally.

A. Problem

Given empirical data $\{(\mathbf{x}_i, y_i)\}_{i=1}^{I}$, and possibly experts' opinion about the aggregation operator, find such representation of the aggregation operator that

1) provides good approximation to empirical data;
2) is flexible to model various classes of aggregation operators;
3) is able to confine to a particular class of operators, or a particular property;
4) is semantically clear.

IV. Fitting to Data and Expert Opinion: Semantical Clarity

In this section, we consider the situation where the experts have a clear idea about the form of the aggregation operator, but are not sure about certain parameters, such as the relative importance of rule antecedents. These parameters can be adjusted to fit the empirical data $(\mathbf{x}, y)$ by using nonlinear regression techniques.

A typical example in medical decision making is aggregation of risk factors or indicators. For instance, consider the situation where bone density studies (BDS) has indicators as per Table I. Some indicators are considered strong indicators (e.g., postmenopause), whereas other indicators are mild. None of the indicators is sufficient by itself. If the total score is greater than five, BDS is recommended.

TABLE I
Indicators for Bone Density Studies (BDS)

Factor	*Weight*
Postmenopausal female	2
Multiple hospital admissions	2
Low body weight	1
FEV1 <60% of predicted	2
Poor physical activity	2
Smoker	1

Such a system of scores is modeled by the (weighted) bounded sum operator (member of Yager family of triangular conorms [42])

$$f(\mathbf{x}) = \min(1, a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n).$$

The importance factors a_i are simply the indicators' scores divided by 5. Therefore, the algebraic form of the aggregation operator is given. However, the importance factors might not be correct: given values reflect the guess of the experts about their relative importance. "Guess" may seem a slightly unfair term; this is to say that it is the expert opinion based on consideration of a variety facts possibly external to their experience (e.g., through reading research literature) as well as from their personal observation, however, lacking an explicit derivation of the weighting factors from the evidence. For example, in the case of the BDS guideline, a literature review [63] reveals that some risk factors have been clearly established as having a significant, independent contribution to loss of bone mass by randomized controlled trials (RCTs) in clinical contexts very similar to the area of intended application for this guideline—these are given a weight of 2. Other factors are less well established by relevant RCTs but have some empirical evidence or analytical feasibility—these receive a weight of 1. Empirical findings are not available to firmly establish the strength of all correlations among the factors or to provide direct evidence for a particular aggregation method. The concept of a simple threshold for BDS based on the weighted sum is as much a concession to having a computationally simple and easily-expressed criterion as a firmly-held belief of the experts.

One way of determining the importance factors is to use the analytical hierarchy process (AHP) by Saaty [61]. However, the ratios of relative importance may not be available from the experts, besides, the weights of importance need to be somehow normalized. The normalization criteria $\sum a_i = 1, \sum a_i = n$, or $\max a_i = 1$ [67] do not always make sense (e.g., in the bounded sum operator). Of course, the problem of normalization can be somehow resolved, but the fact that AHP requires knowledge elicitation procedure makes it unusable when experts' knowledge is not readily accessible. In contrast, the em-

[1]The adjective "nonparametric" is misleading. Nonparametric methods (also called distribution-free in statistical literature) actually do involve parameters, either fixed *a priori* or estimated from the data. Unlike in parametric approach, these parameters are not meaningful in relation to the original problem. A comprehensive discussion of the method of splines in nonparametric regression is given in [28].

pirical data—expert responses to given situations, can be routinely (and automatically) collected.

The alternative way of adjusting importance factors is to observe experts' decision pattern on a set of model situations and to find the values via nonlinear regression. Nonlinear least squares algorithms, such as Levenberg–Marquardt method [46], are extensively used for this purpose and could be found in commercial software (e.g., [19]). As the initial approximation, the experts' guess about the importance factors can be taken.

In our model problems we used the Levenberg–Marquardt method implemented in DataFit software by Oakdale Engineering, Inc., Oakdale, PA. [19]. As the operator we wanted to reconstruct on the basis of empirical data, we have chosen the bounded sum operator with 6 antecedents and the importance factors $\mathbf{a} = (0.5, 0.3, 0.25, 0.3, 0.4, 0.3)^t$ (this was the experts' consensus in the BDS example—see last section). As the initial guess the vector $\mathbf{a}_0 = (0.4, 0.4, 0.2, 0.4, 0.4, 0.2)^t$ was taken. It represents experts' initial opinion about the aggregation operator (Table I). To test the method, we randomly generated 20 cases (vectors $\mathbf{x}$) and modeled the experts' solutions to them using the selected aggregation operator, with added random noise, uniformly distributed in $[-0.1, 0.1]$ (values y). This model of the empirical data $(\mathbf{x}, y)$ was used to reconstruct the aggregation operator. We replicated this numerical experiment 20 times and obtained a very good fit (mean error in coefficients <0.05) in all cases, despite the noise in the data.

Confident in the robustness of this method, we applied it to the real data: 20 actual cases obtained during the development of care plan on-line (CPOL) system described in the last section.

V. Fitting to Data: Adaptability and Empirical Fit

In many cases it is hard to identify the appropriate form of the aggregation operator before adjusting it to the data. Besides, an expert's "feeling" about it sometimes could be misleading. It is usually based on the simplicity argument, which may result in oversimplifying. For instance, the scores system for indicators or risk factors is designed for calculations by medical practitioners with pen and paper, and therefore semantical and numerical simplicity are the most important. For computer implementation of a decision support system these criteria, specifically the numerical simplicity, are of secondary importance. The quality of fit becomes the dominating factor. This section describes the use of nonparametric methods to approximate aggregation operators.

Let us represent the aggregation operator as the linear combination

$$f(\mathbf{x}) = \sum_{j=1}^{J} p_j b_j(\mathbf{x})$$

where $b_j(\mathbf{x})$ is a collection of basis functions and p_j are coefficients. As our basis functions we choose products of univariate B-spline functions. The aggregation operator becomes a tensor-product polynomial spline whose coefficients can be found using the least squares method.

The advantages of B-splines for approximating functions are widely documented [14], [15], [20], [28]. These functions are easy to calculate, they provide enormous flexibility, excellent fit and good smoothness. Linear and quadratic splines are frequently used to represent membership functions of fuzzy sets, including neuro-fuzzy systems [2], [15], [41]. In fact, tensor-product spline approximation can be seen as a neural network system (lattice-based B-spline network [15, Ch. 3], [8], which is equivalent to the adaptive neuro-fuzzy system ANFIS [40, Ch. 12]. However, we prefer to use traditional spline approximation terminology in this paper.

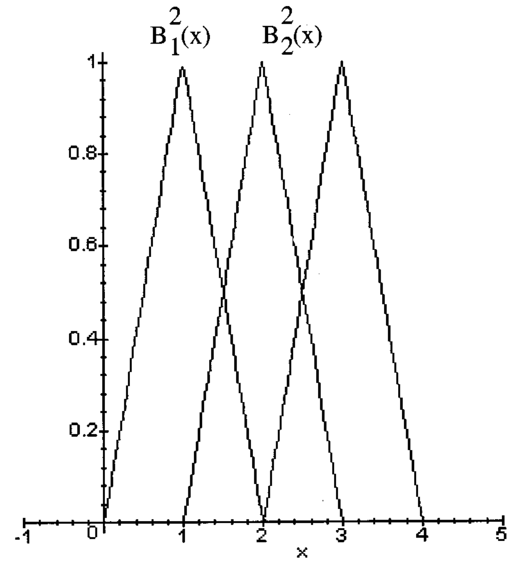


Fig. 1. Second order B-splines.

We will use second-order B-splines $B_j^2(x)$ on a uniform mesh because of their simplicity and mild smoothness constraints on the aggregation operator (we require only continuity). Graphically $B_j^2(x)$ splines are shown in Fig. 1. We split the $[0, 1]$ interval into three regions, and therefore we have 4 (appropriately scaled) B-splines which are not 0 on $[0, 1]$. These splines form the basis for each variable. For an aggregation operator taking K arguments, the 4^K basis functions are given as tensor products of univariate B-splines $B_{j_1 \ldots j_k}(\mathbf{x}) = B_{j_1}(x_1) B_{j_2}(x_2) \ldots B_{j_k}(x_K)$. The coefficients p_j can be numbered in the similar way, that is $p_{j_1 \ldots j_k}$ and the formula for f becomes

$$f(\mathbf{x}) = \sum_{j_1, j_2, \ldots, j_k = 1}^{4} p_{j_1, j_2, \ldots, j_k} B_{j_1}(x_1) B_{j_2}(x_2) \ldots B_{j_k}(x_K).$$

The vector of coefficients $\mathbf{p}$ can be determined from scattered data using least squares method [20], [46], and the spline is referred to as the least squares spline.

When used to approximate aggregation operators, however, this technique does not give satisfactory results: it fails to take into account the monotonicity of $f(\mathbf{x})$ and boundary conditions $f(\mathbf{0}) = 0$ and $f(\mathbf{1}) = 1$. The boundary conditions can be dealt with easily: the coefficients $p_{11\ldots1}$ and $p_{44\ldots4}$ must be set to 0 and 1, respectively. The monotonicity, which is essential from semantical point of view, requires a more careful consideration [59], [60].

Even though the data to which the spline is adjusted is monotone, the splines themselves are not necessarily monotone. This is illustrated in Fig. 2 in the univariate case. For splines

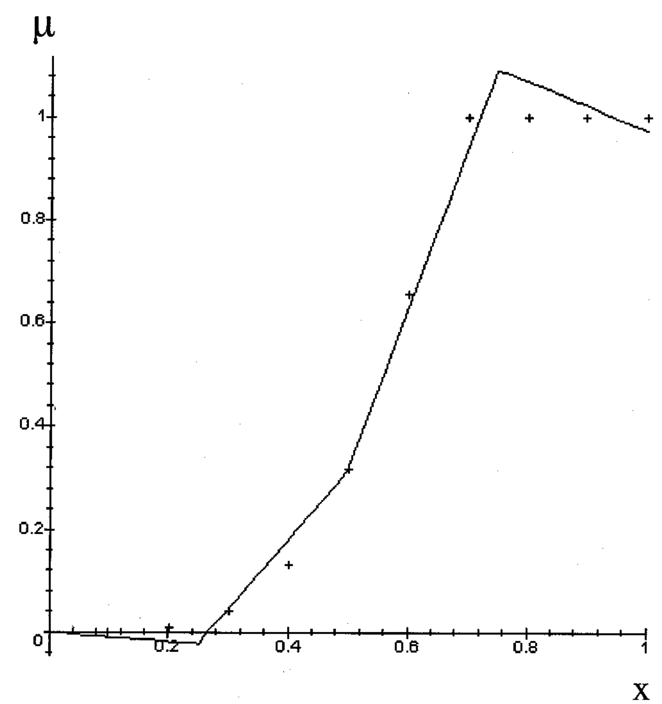


Fig. 2. Least squares spline does not preserve monotonicity of the data.

whose nodes coincide with data points, the solution is to use constrained splines [2], [3], [5], [20], [26], [28], [33], [39]. For multivariate case this would require data to be given on a rectangular mesh, which is not the case in most problems—the data is scattered. One approach to monotone approximation of scattered data is Powell–Sabin splines [66]. This approach is numerically expensive, and is used mostly for surface approximation. Thin-plate splines [20] with the appropriate restrictions can also be used, but again, this approach is rather complicated [59].

In this section, we will present an approach based on restricted least squares splines. It is semantically and numerically simple: tensor products of univariate splines are used as the basis functions and general restricted least squares problem has been thoroughly studied [4], [18], [35], [46], [51], [59], [60].

Consider first the univariate case. The monotonic function $f(x)$ is represented as

$$f(x) = \sum_{j=1}^{J} p_j B_j(x). \tag{1}$$

It is well known that B-splines form a partition of unity, that is

$$\sum_{j=-\infty}^{\infty} B_j(x) = 1.$$

In the case of second-order splines $B_j^2(x)$, only 2 neighboring splines are not 0 at any point x. The partition of unity property implies that the derivative $(B_j^2(x) + B_{j+1}^2(x))' = 0$, and therefore $(B_j^2(x))' = -(B_{j+1}^2(x))'$. The derivative of the spline

$$\begin{aligned} f'(x) &= (p_j B_j^2(x) + p_{j+1} B_{j+1}^2(x))' \\ &= (B_{j+1}^2(x))'(p_{j+1} - p_j). \end{aligned}$$

Since $B_{j+1}^2(x)$ is increasing and $B_j^2(x)$ is decreasing, the condition that $f'(x) \geq 0$ translates into $p_{j+1} \geq p_j$. In other words, the coefficients of the monotonically nondecreasing spline p_i form a nondecreasing sequence. To comply with $f(0) = 0$ and $f(1) = 1$, the sequence must start with 0 and terminate with 1.

Let us represent the coefficients slightly differently. Let $p_1 = v_1$, and $p_j = p_{j-1} + v_j = \sum_{k=1}^{j} v_k, v_j \geq 0$. (1) becomes

$$f(x) = \sum_{j=1}^{J} \left(\sum_{k=1}^{j} v_k \right) B_j(x) = \sum_{j=1}^{J} v_j \sum_{k=j}^{J} B_k(x). \tag{2}$$

The matrix of the system of normal equations $\mathbf{Av} = \mathbf{r}$ is given by

$$a_{jk} = \sum_{i=1}^{I} \left(\sum_{l=j}^{J} B_l(x_i) \right) \left(\sum_{m=k}^{J} B_m(x_i) \right)$$

whereas the components of the right-hand side are

$$r_j = \sum_{i=1}^{I} \left(\sum_{l=j}^{J} B_l(x_i) \right) y_i$$

where I is the total number of data points (x_i, y_i) and J is the number of basis functions. The components of the vector $\mathbf{v}$ must be nonnegative.

The naïve approach outlined above can be generalized into the following.

Proposition: The necessary and sufficient condition for monotonicity of linear and quadratic splines ($k = 2$ and $k = 3$)

$$S(x) = \sum_{j=1}^{J} v_j T_j^k(x)$$

with

$$T_j^k(x) = \sum_{l=j}^{J} B_l^k(x)$$

is $v_j \geq 0, j = -1, \ldots, J$. For higher-order splines, this is only a sufficient condition [11], [12].

Here the functions $T_j^k(x)$ (trapezoidal, or T-splines) are linear combinations of the usual B-splines and they are chosen as an alternative basis to B-splines, because they express monotonicity restriction in a very simple form, as nonnegativity of the coefficients. Simple relation between B- and T-splines allows one to calculate the new basis immediately. Calculation of spline coefficients is performed by solving the system of normal equations.

Thus, we arrived to the classical problem of restricted least squares [46]. The solution is guaranteed to exist either inside the admissible domain (positive components of $\mathbf{v}$) or on the boundary (some components are 0). Various methods of solution are known. The method of Lagrange multipliers is one of them [46], [51]. Alternatively, branch-and-bound algorithms can be employed [4]. One such method takes advantage of the special form of the constraints and permits the reduction of the search space (the size is $2^J, J$ is the length of $\mathbf{v}$) significantly [17], [18]. The nonnegative least squares algorithms NNLS and LSEI

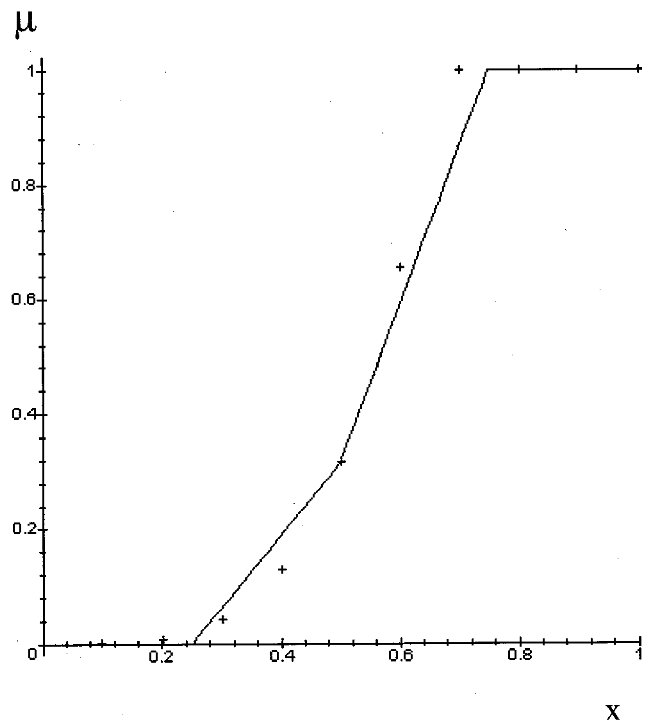


Fig. 3. Constrained least squares spline.

are described in [35], [46] and are available from NETLIB [21], [56]. Fig. 3 shows the monotonic least squares spline calculated using LSEI.

Let us now turn to multivariate case, important for approximating aggregation operators. The function $f(\mathbf{x})$ has to be nondecreasing in all arguments. Since

$$f(\mathbf{x}) = \sum_{j_1,j_2,\ldots,j_k=1}^{4} p_{j_1,j_2\ldots j_k} B_{j_1}(x_1)B_{j_2}(x_2)\ldots B_{j_k}(x_K)$$

it can be shown that the monotonicity condition translates into a set of linear restrictions on $p_{j_1\ldots j_k}$. To avoid algebraic difficulties, we will use T-spline basis again; this will give the system of normal equations almost immediately.

As earlier, let

$$f(x) = \sum_{j=1}^{J} v_j T_j(x), \quad \text{where } T_j(x) = \sum_{k=j}^{J} B_k(x).$$

All new basis functions are linearly independent and nondecreasing on the interval $[0,1]$, therefore, their positive linear combination is also nondecreasing. They are related to integrated I-splines from [59], but are defined slightly differently. The tensor product of nondecreasing basis functions is also nondecreasing in all arguments in every point of $[0,1]^K$. Therefore we express $f(\mathbf{x})$ as

$$f(\mathbf{x}) = \sum_{j_1,j_2,\ldots,j_k=1}^{4} v_{j_1,j_2,\ldots,j_k} T_{j_1}(x_1)T_{j_2}(x_2)\ldots T_{j_k}(x_K).$$

If all the coefficients $v_{j_1\ldots j_k}$ satisfy

$$\sum_{j_1=1}^{V_1}\cdots\sum_{j_K=1}^{V_K} v_{j_1\ldots j_u\ldots j_K} \geq 0 \quad j_u = 1,2,\ldots 4$$
$$V_l = 1,2,\ldots,4 \quad l \neq u \quad u = 1,2,\ldots,K$$

the function $f(\mathbf{x})$ is nondecreasing in all arguments. The sums are taken over all combinations of upper index limits $V_l, l \neq u$. Altogether, there are at most $K \times 4^K$ inequalities (some are redundant). They can be easily represented in matrix form, and the matrix consists of 0s and 1s arranged in a fashion consistent with the indexing system.

For example, in two-dimensional (2-D) case we have

$$\sum_{j_1=1}^{t} v_{j_1 j_2} \geq 0$$
$$\text{for every } j_2 = 1,\ldots,4 \quad \text{and} \quad t = 1,\ldots,4 \quad \text{and}$$
$$\sum_{j_2=1}^{t} v_{j_1 j_2} \geq 0$$
$$\text{for every } j_1 = 1,\ldots,4 \quad \text{and} \quad t = 1,\ldots,4.$$

The next step is to build the system of normal equations and to solve the problem using any restricted least squares algorithm, such as LSEI [21], [35], [56] or BVLS [17], [21], [46], [56]. Consider LSEI method. It consists in solving the following system

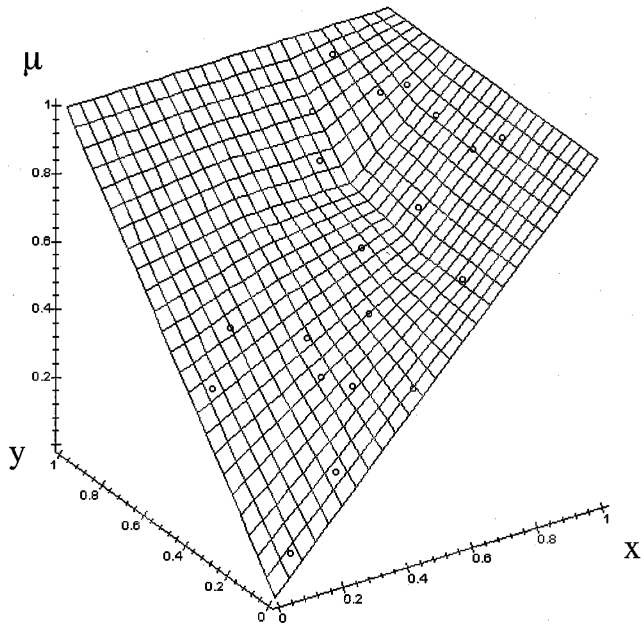
$$\text{Solve } \mathbf{Tv} \approx \mathbf{y}, \quad \text{given that } \mathbf{Gv} \geq \mathbf{0} \quad \text{and} \quad \mathbf{Ev} = \bar{\mathbf{y}}$$

where the first system of equations describes approximation conditions, the system of inequalities ensures monotonicity and the last system describes additional interpolation conditions (it could be empty).

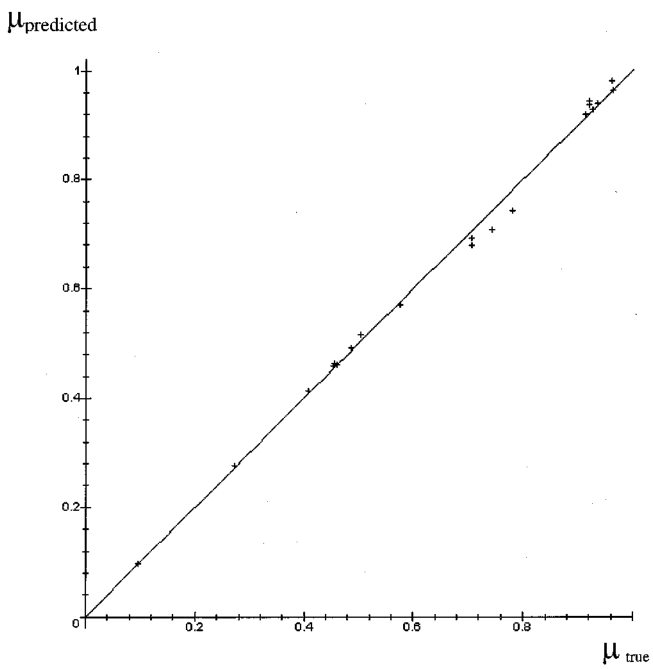
The ability to incorporate interpolation conditions (besides fitting the data) turns out to be very useful when approximating aggregation operators. Recall that aggregation operators require $f(\mathbf{0}) = 0$ and $f(\mathbf{1}) = 1$. Rather than imposing complicated restrictions on the coefficients $\mathbf{v}$, one can simply add two interpolation conditions into the system $\mathbf{Ev} = \bar{\mathbf{y}}$ and the required boundary conditions will be satisfied.

This technique can be extended further to impose other restrictions on the aggregation operators. For instance, triangular conorms satisfy $f(0,0,\ldots,x,\ldots 0) = x$ and $f(1,1,\ldots x,\ldots 1) = 1$ for any x at any position, (triangular norms have a similar property). These restrictions can be enforced by adding the relevant interpolation conditions to the system $\mathbf{Ev} = \bar{\mathbf{y}}$ (namely at $x = 0, 1/3, 2/3, 1$). The same method is employed when the idempotency property $f(x,x) = x$ is needed. Commutativity of the aggregation operator is enforced by making the matrix $\mathbf{b}$ symmetric, as discussed in [11], [12]. Finally, the associativity property is enforced by approximating the additive generators [11], [12].

Fig. 4 shows the result of approximating a triangular conorm with monotone tensor product least squares spline. The (noisy) data are given in the form of 20 randomly scattered points,



(a)



(b)

Fig. 4. (a) Hamacher sum operator $f(x, y) = (x + y - 2xy)/(1 - xy)$ approximated with monotone least squares spline. The 20 data points are randomly generated and are marked with circles. The data contains random noise uniformly distributed in $[-0.1, 0.1]$. (b) True versus predicted values (straight line corresponds to perfect fit).

marked with circles. Restrictions $f(x,0) = f(0,x) = x$ and $f(x,1) = f(1,x) = 1$ are imposed. Figs. 5 and 6 illustrate the use of monotone splines in empirical studies of aggregation operators. The data is taken from [70] and represent the empirically determined membership values of 20 objects in the fuzzy sets "metallic object", "container" and "metallic container". The aggregation operator that models intersection of the first two sets was constructed based on the data and the boundary conditions $f(0,0) = 0$ and $f(1,1) = 1$. It is shown on Fig. 5. Fig. 6 shows the relationships between empirical and predicted grades of

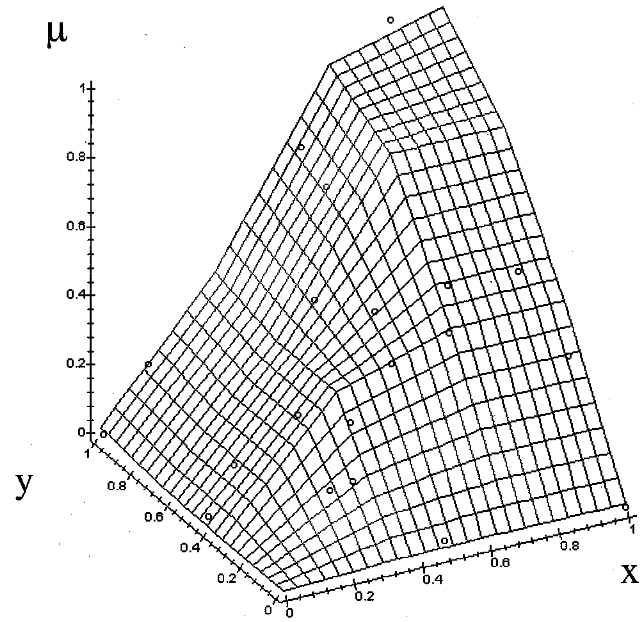


Fig. 5. Empirical data used to approximate fuzzy set intersection in "metallic container" example [59]. Monotone spline aggregation operator.

membership (the straight line indicates perfect prediction for error-free data). Following [70], we tested the acceptability of the constructed operator statistically using Student t-test for the differences between empirical and predicted grades of membership. The mean of the sample is 0.000 03 (versus 0.052 for min operator [70]), the observed standard deviation is 0.036 (versus 0.067) and the result is $t = 0.00$ (versus 3.471), which means that at a given confidence level $(\alpha = 0.025)$ the proposed aggregation operator should be accepted (min is rejected). Increasing confidence level to $\alpha = 0.000\,01$ does not change this result.

The proposed approach of constrained least squares splines compares favorably with other methods of approximation of aggregation operators, such as using neural networks or nonlinear regression [29]. First of all, the method of splines is linear and consequently the solution is reached in the first iteration. Secondly, it allows one to represent monotonicity explicitly, as nonnegativity of the coefficients, whereas in other methods this is not feasible. Splines are very flexible to model functions of any shape, and their quality of fit is controlled by the number of approximation knots. Finally, many semantically important properties of certain families of aggregation operators, such as commutativity and idempotency, can be enforced by using interpolation conditions at key points, whereas other methods are less adaptable. Thus, constrained splines satisfy the first three requirements we set in the problem of approximation of aggregation operators.

Disadvantages of this approach are also clear. There is no semantic interpretation of spline coefficients, and the constructed operator behaves like a black box. Tensor-product spline requires many coefficients to be determined, and hence substantial amount of empirical data (that may not be available). Even though LSEI method can solve underdetermined systems (less data than coefficients), for quality approximation the data is essential. Consequently, in case of few data, the previously described parametric approach should be used.

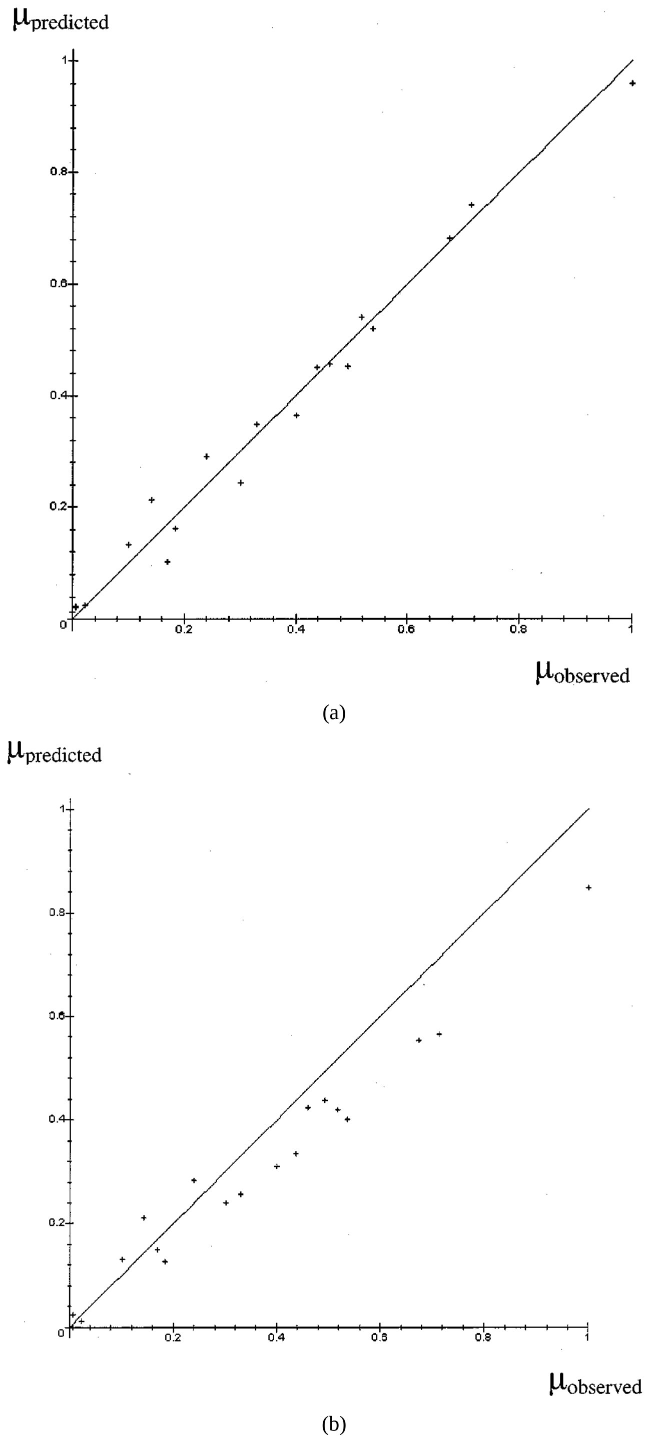


Fig. 6. Observed versus predicted membership values. (a) Monotone spline operator. (b) Min operator.

VI. CPOL SYSTEM

CPOL is an intranet-based medical decision support system that offers both solicited and unsolicited advice to a general practitioner (GP) during care planning. Care planning is a decision making step in the process of *Coordinated Care*, a trial of which is run by the SA HealthPlus division of the South Australian Health Commission. CPOL provides a single coherent source whereby the GP can review a HealthPlus patient's EPR in the context of devising a plan of prospective services and medications. In the same application environment CPOL provides access to clinical practice guidelines tailored for SA HealthPlus. It employs multiple methods of active decision support by linking clinical guidelines with the routine observations from the electronic medical record. Unsolicited advice is provided "just in time" without disrupting the GPs workflow. The mechanisms for doing this include using nonverbal clues, facilitating implementation of the decision consistent with the advice (and providing critical advice otherwise), structuring data entry forms in a way that conveys the advice, and also providing narrative advice at critical steps [10], [64].

Fuzzification of clinical guidelines is an important step in their electronic implementation, and the problem of selecting, or approximating, aggregation operators plays a major role in this context. It is not always clear what kind of logical operations is behind "and" and "or" connectives, whether these operators are compensatory and what are the importance factors. Experts are ambiguous in answering these questions, and rather than torturing them with knowledge elicitation procedure, we offered them a selection of real and dummy cases, where they had to make a decision. This data was used to fit the aggregation operators (either in parametric or nonparametric form) to experts' decision pattern. We illustrate this process on guidelines from SA HealthPlus as implemented in CPOL.

The first guideline is for bone density studies with indicators as discussed earlier (as per Table I). If the weighted sum of the indicators applicable to a given patient (based on their EPR) is more than 5, then the procedure is recommended (CPOL places a red "!" attention flag by the BDS service as well as describing the service as "recommended" in its guideline). If the sum is between 3 and 5, BDS is suggested as an option (and receives a yellow "?" attention flag). As we mentioned earlier, this rule corresponds to bounded sum operator

$$f(\mathbf{x}) = \min(1, a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n)$$

with a_i given by importance factors divided by 5. The membership functions (for individual antecedents) are piecewise linear and they are given in Table II. These membership functions have been chosen for simplicity reasons, however they still adequately describe the corresponding fuzzy sets. More advanced methods of membership function estimation from the data are described in [2], [7], [58], [70].

As the data we took 20 cases, for which three experts (respiratory specialists involved in the framing of the guideline for use in SA HealthPlus) gave their opinion on applicability of BDS. We fitted the coefficients of the operator using DataFit nonlinear regression software. The vector of resulting coefficients was $\mathbf{a} = (0.3, 0.36, 0.31, 0.25, 0, 0.02)^t$. It may appear that at least within the bounded sum model, the "smoker" and "poor physical activity" indicators do not play any major role in doctors' decision making. We realize, however, that we used a very small number of cases, and the data were not evenly spread across its domain. All the patients participating in the study were either smokers or exsmokers and, although, we formally accounted for the time since they quit smoking, it introduced a bias into experts' opinions. As to the "poor physical activity" indicator, it appears that doctors use more sophisticated criteria than just

TABLE II
MEMBERSHIP FUNCTIONS FOR BONE DENSITY STUDY INDICATORS

Indicator	Numerical parameter x	Membership function	Comment
Postmenopausal female	Age (years)	$\mu_1(x)=\begin{cases}(x-45)/(50-45)\\ 0,\text{if } x<45\\ 1,\text{ if } x>50\end{cases}$	0 if male; is overwritten if the PM status is known
Multiple hospital admissions	Number of admissions	$\mu_2(x)=\begin{cases}0.5,\text{if } x=1\\ 0,\text{if } x=0\\ 1,\text{ if } x>1\end{cases}$	
Low body weight	Body mass index (kg/m^2)	$\mu_3(x)=\begin{cases}(20-x)/2\\ 0,\text{if } x>20\\ 1,\text{ if } x<18\end{cases}$	
FEV1 <60% of predicted	FEV1, % of predicted	$\mu_4(x)=\begin{cases}(100-x)/50\\ 0,\text{if } x>100\\ 1,\text{ if } x<50\end{cases}$	
Poor physical activity	Walking distance (m)	$\mu_5(x)=\begin{cases}(500-x)/400\\ 0,\text{if } x>500\\ 1,\text{ if } x<100\end{cases}$	
Smoker	Daily packs × years (x), or cigarettes per day (y) and years since quitting (z) if x is unavailable	$\mu_6(x)=\begin{cases}(x-30)/10\\ 0,\text{if } x<30\\ 1,\text{ if } x>40\end{cases}$, or $\mu_6(y,z)=\mu_{6a}(y)\times\mu_{6b}(z)$, $\mu_{6a}(y)=\begin{cases}(y-10)/20\\ 0,\text{if } y<10\\ 1,\text{ if } y>30\end{cases}$, $\mu_{6b}(z)=\begin{cases}1-(z-5)/10\\ 1,\text{if } z<5\\ 0,\text{ if } z>15\end{cases}$	

the walking distance. The fact that the data does not support the experts' opinion on the importance of the last two indicators deserves further investigation both from decision making and user interface perspective.

Other guideline examples from CPOL that can be readily fuzzified are

IF Diastolic Blood Pressure >90 mmHg **OR** Systolic Blood Pressure ≥ 140 mmHg,
  THEN Check BP every visit
IF BMI <20 **OR** BMI >30 **OR** $>10\%$ weight loss in 3 months,
  THEN Consider Dietician service
IF Total Cholesterol >4.5 **OR** LDL >3.5 **OR** HDL <1 **OR** Triglyzerides >2,
  THEN Consider dietary therapy
IF Received dietary therapy **AND** after 2 readings (TotChol >5.5 **OR** (HDL <1 **AND** TotChol >5) **OR** (HDL <1 **AND** Triglyzerides >2.5)),
  THEN Consider drug therapy.

There are no indicators about the form of aggregation operators, and nonparametric B-spline approach seems to be most appropriate. However, the total number of basis functions is 4^K, where K is the number of indicators, and hence large amounts of empirical data are needed to find the spline coefficients. These data can be routinely collected when employing an on-line health information system such as CPOL, and our research team intends to develop further empirical models in this fashion as the system is used further.

VII. CONCLUSION

In absence of set-theoretical criteria, semantical clarity, flexibility and goodness of fit to empirical data become the decisive factors in selecting aggregation operators. Both parametric and nonparametric regression can be employed to find the functional form of the aggregation operator. Parametric regression provides better semantical clarity, but is not as flexible and precise as nonparametric methods. In the case of nonparametric regression, multivariate linear least squares splines provide excellent flexibility, but may not preserve monotonicity, which is semantically essential for general aggregation operators. To enforce this property, linear constraints on spline coefficients have to be introduced.

An appropriate choice of basis functions in the space of linear splines results in an especially simple form of the constraints that can be used in standard algorithms. The availability of such algorithms makes the problem of monotone approximation straightforward. In addition, important properties of certain classes of aggregation operators can be translated into interpolation conditions, and therefore easily incorporated into the algorithm. Thus, splines are not only flexible to model general aggregation operators, but are also adaptable to model particular families.

The major application of the proposed methods is seen in expert and decision support systems where little is known about how the criteria (risk factors, indicators, etc.) should be aggregated. Experts in the field can provide their guess about relative importance of such criteria, and their "feeling" about mathematical properties of aggregation operators (compensatory or accumulative behavior, or lack of thereof), and can also provide empirical data by considering various cases themselves. These data are subsequently used for regression analysis. An example of practical application of our method is the CPOL system, which incorporates fuzzy versions of several clinical guidelines.

ACKNOWLEDGMENT

Special thanks are extended to Dr. B. Smith, Dr. P. Frith, and Prof. D. Ruffin for making their expertise available to us, and to J. Noone, H. Frankel and B. van der Zwaag of the CPOL team.

REFERENCES

[1] K.-P. Adlassnig, "Fuzzy set theory in medical diagnosis," *IEEE Tran. Syst., Man, Cybern.*, vol. 16, pp. 260–265, 1986.
[2] G. P. A. Cruz and G. Beliakov, "Approximate reasoning and interpretation of laboratory tests in medical diagnostics," *Cybern. Syst.*, vol. 26, pp. 713–729, 1995.
[3] L.-E. Andersson and T. Elfving, "Interpolation and approximation by monotone cubic splines," *J. Approx. Theory*, vol. 66, pp. 302–333, 1991.
[4] R. D. Armstrong and E. L. Frome, "A branch-and-bound solution of a restricted least squares problem," *Technometrics*, vol. 18, pp. 447–450, 1976.
[5] R. K. Beatson and Z. Ziegler, "Monotonicity preserving surface interpolation," *SIAM J. Numer. Anal.*, vol. 22, pp. 401–411, 1985.
[6] J. H. van Bemmel and M. A. Musen, Eds., *Handbook of Medical Informatics*, Berlin, Germany: Springer-Verlag, 1997.
[7] G. Beliakov, "Fuzzy sets and membership functions based on probabilities," *Inf. Sci.*, vol. 91, pp. 95–111, 1995.
[8] ——, "Definition of general aggregation operators through similarity relations," *Fuzzy Sets Syst.*, vol. 114, pp. 437–454, 2000.
[9] ——, "Aggregation operators as similarity relations," in *Information, Uncertainty, Fusion.*, R. Yager, B. Bouchon-Meunier, and L. Zadeh, Eds. Norwell, MA: Kluwer, 1999, pp. 331–342.
[10] G. Beliakov, J. Warren, and J. Noone, "Decision support for coordinated care planning via unsolicited advice based on clinical guidelines," presented at the Proc. 5th Int. Conf. of Int. Soc. for Decision Support Systems, Melbourne, Victoria, July 1999.
[11] G. Beliakov, "Approximation of aggregation operators using splines," in *Proc. of the 8th Intl. Conference Information Processing and Management of Uncertainty*, vol. 2, Madrid, 2000, pp. 680–685.
[12] ——, "Shape preserving approximation using least squares splines," *Approx. Theory Appl.*, vol. 16, pp. 80–98.
[13] R. E. Bellman and L. A. Zadeh, "Decision-making in a fuzzy environment," *Manage. Sci.*, vol. 17, pp. 141–164, 1970.
[14] C. de Boor, *A Practical Guide to Splines*. New York: Springer-Verlag, 1978.
[15] M. Brown, *Neurofuzzy Adaptive Modeling and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[16] S.-J. Chen and C.-L. Hwang, *Fuzzy Multiple Attribute Decision Making*, Berlin, NJ: Springer-Verlag, 1992.
[17] D. Clark, "An algorithm for solving the restricted least squares problem," *J. Austral. Math. Soc.*, ser. B, vol. 21, pp. 345–356, 1980.
[18] D. Clark and M. Osborne, "On linear and interval least-squares problems," *IMA J. Numer. Anal.*, vol. 8, pp. 23–36, 1988.
[19] [Online]. Available: http://www.oakdaleengr.com
[20] P. Dierckx, *Curve and Surface Fitting with Splines*, Oxford: Clarendon press, 1995.
[21] J. J. Dongarra and E. Grosse, "Distribution of mathematical software via electronic mail," *Comm. ACM*, vol. 30, pp. 403–440, 1987.
[22] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. New York, NY: Academic , 1980.
[23] ——, "A review of fuzzy set aggregation connectives," *Inf. Sci.*, vol. 36, pp. 85–121, 1985.
[24] ——, "The three semantics of fuzzy sets," *Fuzzy Sets Syst.*, vol. 90, pp. 141–150, 1997.
[25] D. Dubois, H. Fargier, and H. Prade, "Refinements of the maximum approach to decision-making in a fuzzy environment," *Fuzzy Sets Syst.*, vol. 81, pp. 103–122, 1996.
[26] T. Elfving and L.-E. Andersson, "An algorithm for computing constrained smoothing spline functions," *Numer. Math.*, vol. 52, pp. 583–595, 1989.
[27] *Encyclopaedia of Mathematics*. Norwell, MA: Kluwer, 1990, vol. 5.
[28] R. L. Eubank, *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker, 1988.
[29] D. Filev and R. Yager, "On the issue of obtaining OWA operator weights," *Fuzzy Sets Syst.*, vol. 94, pp. 157–169, 1998.
[30] P. Fishburn, *Decision and Value Theory*. New York: Wiley, 1964.
[31] M. Grabisch, H. Nguyen, and E. Walker, *Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference*. Norwell, MA: Kluwer, 1995.
[32] M. Grabisch and M. Roubens, *Fuzzy Measures and Integrals*, M. Grabisch, T. Murofushi, and M. Sugeno, Eds, Germany: Physica-Verlag, 2000. Application of the Choquet integral in multicriteria decision making.
[33] H. Greiner, "A survey on univariate data interpolation and approximation by splines of given shape," *Math. Comp. Modeling*, vol. 15, pp. 97–106, 1991.
[34] J. Grimshaw and I. Russel, "Effect of clinical guidelines on medical practice: A systematic review of rigorous evaluations," *Lancet*, vol. 324, pp. 1317–1322, 1993.
[35] R. J. Hanson and K. H. Haskell, "Two algorithms for the linearly constrained least squares problem," *Trans. AMC Math. Software*, vol. 8, pp. 323–333, 1982. Algorithm 587.
[36] E. Hisdal, *Logical Structures for Representation of Knowledge and Uncertainty*. Hiedelbery, Germany: Physica-Verlag, 1998.
[37] C.-L. Hwang and K. Yoon, *Multiple Attribute Decision Making*, Heidelberg, Germany: Springer-Verlag, 1981.
[38] C.-L. Hwang and A. S. Masud, *Multiple Objective Decision Making—Methods and Applications*, Heidelberg, Germany: Springer-Verlag, 1979.
[39] R. Iqbal, "A one-pass algorithm for shape-preserving quadratic spline interpolation," *J. Sci. Comput.*, vol. 7, pp. 359–376, 1992.
[40] J.-S. Jang, C.-T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
[41] N. Kasabov, *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. Cambridge, MA: MIT Press, 1998.
[42] G. Klir and T. Folger, *Fuzzy Sets, Uncertainty, and Information*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
[43] G. Klir and M. Wierman, *Uncertainty-Based Information*, Germany: Physica-Verlag, 1998.
[44] Y.-J. Lai and C.-L. Hwang, *Fuzzy Multiple Objective Decision Making*, Heidelberg, Germany: Springer-Verlag, 1994.
[45] S. Lam, "Implementation and evaluation of practice guidelines," in *J. Amer.. Med. Inf. Assoc., Symp. Suppl.*, 1994, pp. 253–263.
[46] C. Lawson and R. Hanson, *Solving Least Squares Problems*. Philadelphia, PA: SIAM, 1995.
[47] L. Leape, "Practice guidelines and standards: An overview," *Quart. Rev. Bulletin'90*, vol. 16, pp. 42–49, 1990.
[48] E. Liem, J. Obeid, E. Shareck, L. Sato, and R. Greenes, "Representation of clinical practice guidelines through an interactive world-wide-web interface," in *J. Amer. Med. Inf. Assoc., Symp. Suppl.*, 1995, pp. 223–227.

[49] J. Liu and R. Shiffman, "Operationalization of clinical practice guidelines using fuzzy logic," in *Proc. Amer. Med. Inf. Assoc. Annu. Fall Symp.*, 1997, pp. 283–287.

[50] D. Lobach, "A model for adapting clinical guidelines for electronic implementation in primary care," in *J. Amer. Med. Inf. Assoc., Symp. Suppl.*, 1995, pp. 581–585.

[51] D. G. Luenberger, *Introduction to Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1973.

[52] J.-L. Marichal, *Fuzzy Measures and Integrals*, M. Grabisch, T. Murofushi, and M. Sugeno, Eds, Heidelberg, Germany: Physica-Verlag, 2000. On Choquet and Sugeno Integrals as aggregation functions.

[53] ——, "On Sugeno integral as an aggregation function," *Fuzzy Sets Syst.*, vol. 114, pp. 347–366, 2000.

[54] ——, "On the associativity functional equation," *Fuzzy Sets Syst.*, vol. 114, pp. 381–390, 2000.

[55] M. Mizumoto, "Pictorial representations of fuzzy connectives, part I: Cases of t-norms, t-conorms and averaging operators," *Fuzzy Sets Syst.*, vol. 31, pp. 217–242, 1989.

[56] NETLIB [Online]. Available: http://www.netlib.org

[57] L. Ohno-Machado *et al.*, "The guideline interchange format: A model for representing guidelines," *J. Amer. Med. Inf. Assoc.*, vol. 5, pp. 357–372, 1998.

[58] A. M. Norwich and I. B. Turksen, "A model for the measurement of membership and the consequences of its empirical implementation," *Fuzzy Sets Syst.*, vol. 12, pp. 1–25, 1984.

[59] J. O. Ramsay, "Monotone regression splines in action (with comments)," *Stat. Sci.*, vol. 3, pp. 425–461, 1988.

[60] P. A. Ruud, Restricted least squares subject to monotonicity and concavity constraints, in Advances in Economics and Econometrics: Theory and Applications, D. M. Kreps and K. F. Wallis, Eds., Cambridge, vol. 3, pp. 166–187, 1997.

[61] T. L. Saaty, "Exploring the interface between hierarchies, multiple objectives and fuzzy sets," *Fuzzy Sets Syst.*, vol. 1, pp. 57–68, 1978.

[62] R. Shiffman, "Representation of clinical practice guidelines in conventional and augmented decision tables," *J. Amer. Med. Inf. Assoc.*, vol. 4, pp. 382–393, 1997.

[63] B. Smith, P. J. Phillips, and R. F. Heller, "Asthma and chronic airways disease are associated with osteoporosis and fractures: A literature review," *Respirology*, vol. 4, pp. 101–110, 1999.

[64] J. Warren, G. Beliakov, J. Noone, and H. Frankel, "Chronic disease coordinated care planning: Flexible, task-centered decision support," *Topics Health Inf. Manage.*, vol. 20, pp. 52–68, 1999.

[65] J. Warren, G. Beliakov, and B. van der Zwaag, "Fuzzy logic in clinical practice decision support systems," presented at the Proc. 33rd Hawaii Int. Conf. Syst. Sci., Piscataway, NJ, 2000.

[66] K. Willemans and P. Dierckx, "Smoothing scattered data with a monotone Powell-Sabin spline surface," *Numer. Algorithms*, vol. 12, pp. 215–232, 1996.

[67] R. Yager, "Fuzzy decision making including unequal objectives," *Fuzzy Sets Syst.*, vol. 1, pp. 87–95, 1978.

[68] ——, "A new methodology for ordinal multiple aspect decisions based on fuzzy sets," *Decision Sci.*, vol. 2, pp. 589–600, 1981.

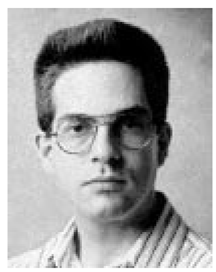
[69] R. D. Zielstorff, "Online practice guidelines," *J. Amer. Med. Inf. Assoc.*, vol. 5, pp. 227–236, 1998.

[70] H.-J. Zimmermann, *Fuzzy Set Theory—And Its Applications*. Norwell, MA: Kluwer, 1996.

Gleb Beliakov received the Ph.D. degree in physics and mathematics from Russian Peoples' Friendship University, Moscow, Russia, in 1992.

He is currently a Lecturer with the School of Computing and Mathematics, Deakin University, Australia. His research focus is in mathematics (numerical analysis, spline functions, optimization, fuzzy set theory) and computing (medical infomatics, decision support systems). He has developed a number of software products, among which is the Care Plan On-Line system. He is the author of 40 research papers in mathematics and computing.

Dr. Beliakov received the European Academic Software Award in 1994.



Jim Warren received the B.S. degree in computer science and the Ph.D. degree in operations analysis from the University of Maryland, Baltimore County (UMBC), MD.

He is currently Associate Professor in the School of Computer and Information Science (CIS) with the University of South Australia, one of the largest computing departments in Australia in terms of staff and students. He is Director of the research arm of CIS, the Advanced Computing Research Centre. He also heads the University's Health Informatics Research Group. His particular research interest is in human–computer interaction and user interface design for health information system, especially where there is an element of automated decision support or applied artificial intelligence. He has developed a number of production systems and research prototypes for general practitioners.

Dr. Warren Co-Chaired the First Australasian User Interface Conference in 2000.