Extending the Medical Concept of Reference Intervals using Fuzzy Predicates

Mario Drobics
Biomedical Engineering - eHealth systems
Austrian Research Centers GmbH - ARC
A-1220 Vienna, Austria
mario.drobics@arcsmed.at

Klaus-Peter Adlassnig
Section on Medical Expert
and Knowledge-Based Systems
Medical University Vienna
A-1090 Vienna, Austria
klaus-peter.adlassnig@meduniwien.ac.at

Abstract

Expert systems for medical applications have to deal with medical concepts such as "normal range", "elevated", or "reduced". These concepts, although backed by a profound medical background based on reference intervals, are defined manually by physicians using interval-based representation. This approach is usually not feasible in large-scale applications. In the present study we describe a method to generate fuzzy-logic-based predicates founded on historic medical data, using a combination of established statistical methods and cluster analyses to generate concepts that correspond to established laboratory standards and the physician's interpretation. We also describe visualization techniques which help the physician to analyze and adapt the results according to clinical needs. Finally, a case study using actual laboratory data from 562 hepatitis patients is presented.

1 Introduction

Physicians' definitions of concepts such as "normal range", "elevated", or "reduced" are based on a large body of background knowledge. An alanine aminotransferase (GPT) level of 43 U/l may, for instance, be "normal" for an adult man, but may well be termed "elevated" for a young woman. However, apart from sex and age, several factors significantly influence the interpretation of laboratory data, such as the patient's race, the time of measurement or additional medication, to name a few. Due to these individual influences it is not always possible to clearly state whether a measurement in an individual patient is still in the normal range or is already pathological. A comprehensive appraisal of the patient is needed to resolve this problem.

In expert systems, medical knowledge is typically represented in terms of rules which describe the relationships between medical entities for a general population. When physicians are asked to formulate rules for such an expert system, they frequently use imprecise medical concepts such as "mild fever" to indicate the individual context of these expressions. In order to incorporate these vague concepts into medical expert systems, methods to describe imprecise expressions and draw further (logical) conclusions based on these expressions are needed.

As a possible solution to this problem, we present a novel approach based on historical medical data to automatically define the underlying (imprecise) expressions using fuzzy predicates. This is done in accordance with the recommendations on reference intervals to define normal ranges issued by the International Federation of Clinical Chemistry's (IFCC) [10, 11] and by the use of a clustering technique to define application-dependent predicates for values outside the normal range. This is done to ensure compatibility with established standards and enable physicians to interpret the data. A case study in which this method was applied to a set of 562 case records of hepatitis patients is presented.

The concept of reference intervals was introduced in a series of publications by the IFCC in 1987 [10, 11]. It is currently well established and widely used in laboratory medicine. In order to provide a feasible alternative to the concept of normal values (which is an abstract concept of the ideal state of health), a theoretical concept and a set of recommendations for collecting reference values were presented. Since this time, several modifications and extensions have been proposed, such as techniques to identify outliers [14] or bootstrapping methods to reduce the size of confidence intervals [3]. The methods recommended by the IFCC have been implemented in H.E. Solberg's *RefVal* program [12, 13].

Reference intervals are, however, not always a good choice as decision boundaries. Moreover, due to intra-individual differences, a smooth transition from "normal" to "pathological" states would be desirable. This urges one to define these predicates using fuzzy logic, where a fuzzy predicate u *in* F describes the *degree of compatibility* of u

with the (medical) concept F. Mathematically, u *in* F is computed using the membership function $\mu_S(x) \mapsto [0,1]$ of an associated fuzzy set S.

2 Previous Work

Previous work on this subject has been focused on two areas. First, extensive efforts have been made in the medical community to define (crisp) decision boundaries. Secondly, for several years now, the data-mining community has been developing methods to determine optimal splits. In the medical community, fuzzy logic approaches have been mainly used in expert systems in which the sets are defined manually. The data-mining community is also using approaches in which the underlying fuzzy sets are defined automatically. This is not surprising in view of the fact that the underlying concepts are typically well defined in medicine. In most other applications, the semantics of the predicates is defined in a problem-specific manner.

2.1 Defining Decision Boundaries

The use of reference intervals is well established in laboratory medicine. They are, as mentioned earlier, not always a good choice as decision boundaries. Therefore, receiver operating characteristics (ROC) curves are used to determine the optimal decision point. As the actual decision boundary is usually defined manually by a domain expert, it is possible to achieve an optimal balance between the sensitivity and specificity of the test. Various supervised methods, such as the minimum description length (MDL) of Fayyad and Irani [2], have been proposed to determine the optimal splitting point automatically. However, it is common practice to define crisp decision boundaries, which gives rise to a number of practical problems [7].

2.2 Defining Linguistic Variables

The simplest approach for defining fuzzy sets automatically is to form a partition for each dimension and distribute the data evenly over the range (equi-distance binning) or to define the sets such that all have the same cardinality (equi-frequency binning). Although these approaches are sufficient for basic calculations, they are associated with significant limitations in terms of semantic soundness and accuracy. Several approaches to fit the fuzzy sets to the given training data have been used to overcome these limitations [1, 4, 8]. Although the results of these two approaches are promising, the resulting fuzzy sets do not always directly correspond to linguistic expression.

The work presented in this paper is based on a study published by Schürz and Adlassnig [9]. They used a purely statistical approach together with decision rules to construct fuzzy sets for medical concepts. We enhanced this approach by using a more sophisticated definition of fuzzy sets for pathological results and by introducing enhanced visualization techniques.

3 Generation of Fuzzy Predicates

When defining linguistic variables for medical concepts, we start with the foremost concept, which is to define the fuzzy predicate for "$normal$" in close concurrence with the specifications of the Industrial Mathematics Competence Center (IMCC). We then define the concepts of $reduced$ and $elevated$, below and above $normal$, respectively. Optionally, concepts for *highly elevated* and *very highly elevated* are defined using an agglomerative clustering technique.

3.1 Mathematical Formulation

We define fuzzy sets as piecewise linear functions according to the following formula:

$$\mu_F(x) = \begin{cases} \frac{x - c_L + w_L}{w_L} & \text{if } c_L - w_L \leq x \leq c_L \\ 1 & \text{if } c_L < x < c_U \\ \frac{c_U + w_U - x}{w_U} & \text{if } c_U \leq x \leq c_U + w_U \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where c_L and c_U define the ranges in which the membership function is one, and w_L and w_U are the width of the fuzzy set on the lower and upper bound, respectively. Other types such as bell-shaped or tri-cubic fuzzy sets might be used accordingly.

3.2 Data Preparation

To obtain a small overlap of normal and pathological results, the data should ideally be split into subgroups. We use Lahti et al.'s approach [5, 6] to identify reasonable subgroups with different characteristics (e.g., male/female, child/adult, etc.). This minimizes the overlap of the resulting fuzzy sets without introducing unnecessary complexity. Reducing the vagueness in the underlying concepts eventually enhances the expressiveness of the final decision model. The criteria to determine whether a set should be split into two subgroups are summarized as follows: *"If at least one of the four proportions of the subgroups outside the common reference limits exceeds or is equal to 4.1%, or lies below or is equal to 0.9%, partitioning is recommended. In all other cases, the common reference limits could be used for both subgroups. This rule is applicable to both Gaussian and non-Gaussian distributions.*" [5].

To assist physicians in reviewing the data, the latter should be checked for possible outliers. Although no completely satisfactory method has yet been presented [14],

we use the Dixon test—as recommended by the National Committee for Clinical Laboratory Standards (NCCLS)—to identify the most likely outliers. Potential outliers should be carefully reviewed by physicians before finally removing them from the data set. The Dixon test is as follows: let $R = max(X) - min(X)$ be the range of the values and let D be the absolute difference between the most extreme (largest or smallest) value and the next most extreme value. If the ratio D/R exceeds 1/3, the extreme value in question is deleted.

3.3 Finding the Normal Range

When defining the fuzzy membership function μ_{normal} identifying for every measured point the degree of compatibility with the medical concept "normal", one should obtain a value of one when all patients whose laboratory data demonstrate this value have no pathology (with respect to this measurement), and a membership degree of zero when all patients with this value definitely show pathological results. Intermediate values reflect the vagueness and individuality of normality. It should be emphasized that these are graded truth values rather than probabilities or possibilities because the exact measurement is known but its belonging to the medical concept under consideration is gradual. This is especially important when drawing logical conclusions with such predicates in decision support or expert systems.

We define the kernel of μ_{normal}, i.e., the range of values definitely being "normal", as the 95% confidence interval of the reference group, in accordance with the recommendations of the IFCC [11] in order to ensure highest concordance with the established meaning of the concept. The support of μ_{normal} is defined by the minimum and the maximum values of the pre-processed data set (i.e., after removing outliers). The final fuzzy set is then defined by interpolating between these limits. According to Equ. (1), μ_{normal} is then defined using the following parameters:

$$\begin{aligned} c_L^{\text{normal}} &= q_{2.5}(X_R) \\ c_U^{\text{normal}} &= q_{97.5}(X_R) \\ w_L^{\text{normal}} &= q_{2.5}(X_R) - min(X_R) \\ w_U^{\text{normal}} &= max(X_R) - q_{97.5}(X_R) \end{aligned} \tag{2}$$

X_R being the data of the reference group.

3.4 Defining Standard Predicates for Values Outside the Normal Range

Having defined the fuzzy predicate "normal" already suffices to define predicates for the concepts "reduced" and "elevated". They can easily be defined by mirroring the lower and upper limits of the "normal" predicate, respectively; i.e., the new fuzzy set is constructed using Equ. (1) with the parameters:

$$\begin{aligned} c_L^{\text{elevated}} &= c_U^{\text{normal}} + w_U^{\text{normal}} \\ w_L^{\text{elevated}} &= w_U^{\text{normal}} \end{aligned} \tag{3}$$

The right margin of the fuzzy set is either defined as open or is limited by the upper bounds of the distribution of the joined data set X:

$$\begin{aligned} c_U^{\text{elevated}} &= q_{97.5}(X) \\ w_U^{\text{elevated}} &= max(X) - q_{97.5}(X) \end{aligned} \tag{4}$$

For the fuzzy predicate "reduced" the parameters are defined accordingly.

3.5 Defining Additional Predicates for Values Outside the Elevated and Reduced Ranges

The existence of further information within the laboratory values which might be of interest to the physician (e.g., to distinguish between different diseases) gives rise to the need for a finer distinction within pathological values. In practice, this refinement is based on a distinction between different groups of samples. Usually data from patients with different diagnoses are used.

The basic concept underlying the following approach is to identify intervals which allow further distinction, but also keep the fuzzy sets well distributed. In the following we will restrict ourselves to further differentiation of elevated values because this is sufficient in many cases. The same approach can be used for reduced values. The potential limits are determined by collecting the maxima and the 97.5% quantiles of all diseases in accordance with the definition of the "normal" concept. Although it would be possible to use an optimization method involving a special cost function to determine the optimal shape of the fuzzy sets, we decided in favor of this traditional approach in order to ensure compatibility with other clinical definitions.

Specifically, the approach is as follows: first the 97.5% quantile and the corresponding maxima are computed for all diseases and added to a list $B = \{b_1, \ldots, b_n\}$. Then the number of cases $c_i = |\{x \in X \mid max(X_R) < x <= b_i\}|$ above the maxima of the normal range and below the according b_i values are counted and added to a list C. Next, the elements $c_i \leq 2$ are removed from C together with the corresponding b_i's from B. The remaining elements in B and in C are then used to perform a Ward clustering [15] to identify groups within these values.

Ward clustering is an agglomerative clustering method. Each cluster i is characterized by its center b_i and a counter c^i, indicating the number of samples covered. The distance

of two clusters i and j is then defined as follows:

$$d_{\text{WARD}}(i,j) = \begin{cases} 0 & \text{if } c_i = c_j = 0 \\ \frac{c_i c_j}{c_i + c_j}\, d(b_i, b_j) & \text{otherwise} \end{cases} \quad (5)$$

In each step, the two closest clusters are merged. The new cluster center is then defined as the weighted average of the two cluster centers:

$$b_k = \frac{c_i b_i + c_j b_j}{c_i + c_j} \quad (6)$$
$$c_k = c_i + c_j$$

By tracking the so called *Ward index*, it is possible to obtain a degree of likelihood for each number of clusters. Assuming i_t and j_t to be the two clusters merged in the t-th step, the Ward index is computed as follows:

$$I(t) = \frac{1}{|B|} \frac{d_{\text{WARD}}(i_{t-1}, j_{t-1}) - d_{\text{WARD}}(i_t, j_t)}{d_{\text{WARD}}(i_{t-2}, j_{t-2}) - d_{\text{WARD}}(i_{t-1}, j_{t-1})} \quad (7)$$

The Ward index (7) is then used to decide how many predicates outside the normal range will be defined. Usually physicians have a clear understanding as to whether predicates besides normal, elevated, and reduced should be defined. In our approach we used the Ward index to decide whether two or three predicates should be defined for these values.

As Ward clustering is an agglomerative clustering technique, it joins clusters until only one single cluster remains. By knowing the number of clusters, this agglomeration process can be determined at the desired level. Each cluster is then uniquely characterized by its largest member. This value is then used together with its corresponding maximum/97.5% quantile to construct the upper bound of the corresponding fuzzy set. The lower bound of the set is constructed by mirroring the next smaller set. This approach ensures that close decision boundaries are merged, while a finer differentiation of the final decision is still supported.

3.6 Visualization

When computing fuzzy sets for well established terms, it is important to incorporate the possibility of user feedback. As validating all results on a numerical basis can be a cumbersome task, an efficient interface is required. We combine different visualization techniques to enable the user to comprehend all aspects of his/her decision.

As mentioned previously, physicians use sensitivity and specificity to measure the performance of a given test. Typically, ROC curves are employed to visualize this information. Although ROC curves provide a good overview of the situation, they are not optimal when additional information has to be visualized. Therefore we use plots in which sensitivity and specificity is plotted against the domain range and histograms and the computed fuzzy sets are also shown.

Typically, two plots are used to illustrate the results and guide the physician when modifying the set boundaries. If the data have been split into subgroups, individual plots for subgroups are presented. An example is shown in Fig. 1. As this parameter has significantly different reference intervals for men and women, we distinguished between these two groups of patients when computing the respective predicates. The fuzzy set defined for the normal range is shown in the left plots, while the predicates for elevated and highly elevated are shown in the right plots.

In the first plot the "normal" range together with a histogram plot is shown. The histogram is plotted in two colors: red for the reference group and gray for groups with diseased patients. This enables the physicians to easily grasp the distribution of parameters for the reference population and the general population. Furthermore, sensitivity and specificity are plotted as blue lines over the domain to visualize the trade-off between coverage and accuracy. Finally, the membership functions of the fuzzy sets are shown.

For the second plot which covers the entire domain, we decided to use a logarithmic scale because many distributions have a very long tail. The class distributions are visualized using box-and-whisker plots for each group of data, making it easier for the physician to understand the meanings of the different predicates. The fuzzy sets are again visualized by the corresponding membership functions.

These graphics can now be used to illustrate the definition of the predicates and discuss the results with a physician. In an interactive setting, it is possible to modify the underlying fuzzy sets interactively.

The graphics shown in Fig. 1 are discussed in detail in the following section.

4 Experiments

For the present study, we selected 562 case records of adult hepatitis patients and 231 case records of non-infected patients. The patients received in-patient treatment at the Vienna General Hospital (AKH Wien) between 1976 and 1986. In each case the patient's clinical diagnosis was verified by serology, which was regarded as the gold standard. A list of the diseases included in this survey is given in Table 1.

The patients of the last group had no liver disease and therefore constituted the *reference group*. As the data were collected between April 1, 1976 and March 31, 1986, no distinction was made between hepatitis C, D, E, F, and G. Instead, the term "hepatitis non-A non-B" was used.

A detailed laboratory analysis was performed for each patient. The following parameters were included: albumin, alkaline phosphatase, alpha 1 globulin, alpha 2 globulin, beta globulin, gamma globulin, gamma-glutamyl transpep-

Table 1. Data set of hepatitis patients

disease	abbrev.	ICD-9 code
hepatitis A	A	070.1
hepatitis B	B	070.3
hepatitis non-A non-B	NAB	070.5
chronic hepatitis	CH	571.4x
alcoholic hepatitis	ALK	571.1
hepatitis carrier	B_CAR	V02.6
psycho-physiological disorder	-	306.9

tidase, aspartate aminotransferase (GOT), GPT, lactate dehydrogenase, bilirubin, age, and sex.

In the following, we show how fuzzy predicates for GPT were created using these data. As the first step, the fuzzy predicates *normal*, *reduced*, and *elevated* were defined. The statistics for this parameter is shown in the first section of Table 2, where "R" indicates the samples of the reference group, and "D" those suffering from a disease. A statistical comparison with the employed default parameters (45 U/l for male, and 34 U/l for women) is also shown. The parameters obtained using the available data correspond to the default value for male patients, but are significantly lower for women. In both cases, the distribution has long tails. The computed fuzzy sets as well as the underlying distributions and a sensitivity/specificity plot (blue lines) are shown in Fig. 1.

In the plot for male patients, the majority of samples from the reference group are covered by the "normal" predicate with a degree of membership equal to one. The overlap with the "elevated" predicate might appear slightly large because there is a significant gap between 70 and 90 U/l. According to the Dixon test, however, this does not justify elimination of values above 90 U/l.

In the actual plots, data sets from the five different diseases are shown together with the data from the reference population (marked as "-"). Evidently, the predicate "normal" covers the reference population. The predicate "elevated" adds to the right of the "normal" predicate and its upper bound is defined according to the 97.5% quantile of the data classified as "NAB". The overlap looks rather small in this plot because of the logarithmic scale. Finally, the upper bound of the "highly elevated" predicate is defined according to the maximum of the entire data set.

5 Conclusion

We present a method for creating mathematical representations of medical concepts using fuzzy predicates, in accordance with the specifications published by the IFCC. These predicates can be used to express domain knowledge in medical expert systems as well as increase the interpretability of results in medical data-mining applications. The proposed algorithm is supplemented by specific visualization methods that aid the physician in grasping the structure of the data and modifying the results according to his needs.

This approach should help to reduce the time needed to set up an expert system involving fuzzy predicates and also help to determine the meaning of certain predicates when presenting results from a data-mining application.

Future work will focus on the integration of the presented method in a general analysis and inference system for medical data. This system will include data-mining algorithms to generate diagnostic models along with a rule induction mechanism to draw conclusions for new data.

References

[1] M. Drobics. Choosing the best predicates for data-driven fuzzy modeling. In *Proceedings of the 13th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2004)*, IEEE Press, Piscataway, pages 245–249, 2004.

[2] U. M. Fayyad and K. B. Irani. Multi-interval discretization of continuousvalued attributes for classication learning. In R. Bajcsy, editor, *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI 93)*, Morgan Kaufmann Publishers, San Mateo, pages 1022–1027, 1993.

[3] E. K. Harris and J. C. Boyd. *Statistical bases of reference values in laboratory medicine*. Marcel Dekker, Inc., New York, 1995.

[4] F. Höppner and F. Klawonn. Improved fuzzy partitions for fuzzy regression models. *Internat J Approx Reason*, 32:85–102, 2003.

[5] A. Lahti. Partitioning biochemical reference data into subgroups: comparison of existing methods. *Clin Chem Lab Med*, 2(7):725–733, 2004.

[6] A. Lahti, P. H. Petersen, J. C. Boyd, P. Rustad, P. Laake, and H. E. Solberg. Partitioning of nongaussian-distributed biochemical reference data into subgroups. *Clin Chem*, 50(5):891–900, 2004.

[7] R. S. Lord. Definitions of clinical laboratory reference limits. *Townsend Letter for Doctors and Patients*, 246:81–85, 2004.

[8] R. Mikut, J. Jäkel, and L. Gröll. Automatic design of interpretable membership functions. In D. Reichel, editor, *Proceedings of the 8th Zittau Fuzzy Colloquium*, pages 103–111, Hochschule Zittau/Görlitz, September 2000.

[9] M. Schürz, K.-P. Adlassnig, C. Lagor, B. Schneider, and G. Grabner. Definition of fuzzy sets representing medical concepts and acquisition of fuzzy relationships between them by semi-automatic procedures. In *Proceedings (CD-ROM-Version) of the European Symposium on Intelligent Techniques (ESIT 99)*, Crete, Greece, 12568_p.pdf, 1999.

[10] H. E. Solberg. Approved recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. *J Clin Chem Clin Biochem*, 25:337–342, 1987.

Table 2. Statistics of alanine aminotransferase

	male			female		
	All	R	D	All	R	D
#	410	92	318	367	139	228
Min	2.	2.	2.	1.	1.	4.
5%	5.	3.	7.95	4.	3.	5.
Median	50.	12.	116.	19.	9.	75.5
Mean	293.949	14.4565	374.808	169.515	10.7554	266.303
StdDev	527.893	13.1336	574.689	329.205	12.6261	387.039
95%	1888.75	46.4	2035.45	1200.62	20.525	1405.2
Max	3094.	94.	3094.	1970.	145.	1970.

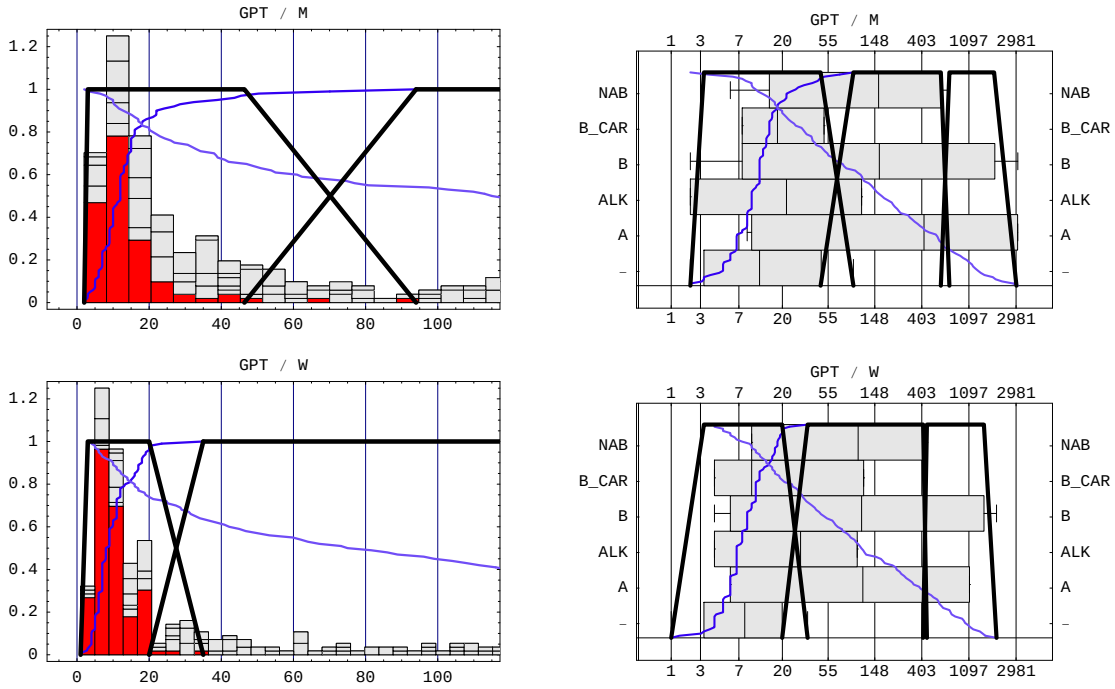

Figure 1. Fuzzy sets generated for alanine aminotransferase (male/female)

[11] H. E. Solberg. Approved recommendation (1987) on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. *J Clin Chem Clin Biochem*, 25:645–656, 1987.

[12] H. E. Solberg. RefVal: a program implementing the recommendations of the International Federation of Clinical Chemistry on the statistical treatment of reference values. *Comput Methods Programs Biomed*, 48(3):247–256, 1995.

[13] H. E. Solberg. The IFCC recommendation on estimation of reference intervals. The RefVal program. *Clin Chem Lab Med*, 42(7):710–714, 2004.

[14] H. E. Solberg and A. Lahti. Detection of outliers in reference distributions: performance of Horn's algorithm. *Clin Chem*, 51(12):2326–2332, 2005.

[15] J. H. Ward Jr. Hierarchical grouping to optimize an objective function. *J Amer Statist Assoc*, 58:236–244, 1963.