

From sensitivity and specificity to confirmation and occurrence

Tatiana Kiseliova

University of Dortmund,
Department of Computer Science I,
D-44221 Dortmund, Germany
Ivane Javakhishvili Tbilisi State University
University St. 2, 0145, Tbilisi, Georgia
Tatiana.Kiseliova@udo.edu

Claudio Moraga

University of Dortmund,
Department of Computer Science I,
D-44221 Dortmund, Germany
Claudio.Moraga@udo.edu

Abstract

In this paper we show that some applications of a fuzzy methodology in medical diagnosis were "inspired" by probabilistic approaches. Although probability theory and fuzzy set theory describe different facets of uncertainty. In this paper we trace the transformation of statistical measures of diagnostic accuracy to fuzzy occurrence, exclusion and confirmation relations, applied in several fuzzy computer-assisted systems. Applications of likelihood ratios, parallel and serial tests in fuzzy expert systems are also mentioned.

Keywords: medical diagnosis, fuzzy relations.

1 Introduction

It is known that probability theory and fuzzy set theory describe different facets of uncertainty. "Unlike fuzziness, probability dissipates with increasing information" [10]. This difference can rather clearly be observed in medical applications, in particular, in medical expert systems. In this field of research fuzzy set theory appears to be a more flexible mechanism in solving the problems, where the probability theory was already applied [16, 17]. This especially concerns the knowledge acquisition in medical knowledge based systems. In this paper we trace the transformation of statistical measures of diagnostic accuracy to fuzzy occurrence, exclusion and confirmation relations, applied in several fuzzy computer-assisted systems: in particular, CADIAG-like systems [2, 9, 4, 3] are representative examples. Applications of likelihood ratios, parallel and serial tests in fuzzy expert systems are also mentioned.

Since this paper intends to be an extended abstract, some deep comparisons between the two approaches are left out and completeness of description is not at its final stage. The main purpose of the paper is to describe the contour of commonalities between statistical and fuzzy approaches in medicine, and to show in some examples, that several successful ideas, practically realized in fuzzy medical expert systems, were "inspired" by probabilistic approaches.

2 Basic notations and definitions

Let us introduce the basic notations and definitions used in this paper. Let

- symptoms, signs, test results and findings - be denoted as S and called symptoms;
- diseases - D .

A fuzzy set F is defined as $F : U \rightarrow [0, 1]$, where U is a universe of discourse. Note, that we identify F and its membership function μ_F to simplify the notations. Let $\Pi = \{p_1, \dots, p_r\}$, $\Sigma = \{s_1, \dots, s_n\}$ and $\Delta = \{d_1, \dots, d_m\}$ be the crisp sets of patients, symptoms and diseases under consideration. For example, Δ can denote rheumatic diseases, Σ are symptoms of these diseases, and Π are investigated patients for rheumatic diseases [14, 3]. Fuzzy set $S_p : \Sigma \rightarrow [0, 1]$ describes the information about a patient, in particular, each element of the fuzzy set S_p shows to which degree it is true, that a patient p has symptom s_i . $D_p : \Delta \rightarrow [0, 1]$ is a fuzzy set of possible diagnoses for a patient. $R_{PS} : \Pi \times \Sigma \rightarrow [0, 1]$, $R_{PD} : \Pi \times \Delta \rightarrow [0, 1]$, $R_{SD} : \Sigma \times \Delta \rightarrow [0, 1]$ describe patient-symptom,

patient-disease, symptom-disease fuzzy relations. A composition of fuzzy relations, introduced by L. Zadeh and adopted by Sanchez for medical diagnoses [13]:

$$R_{PD} =_{\text{def}} R_{PS} \circ R_{SD} \quad (1)$$

is used in CADIAG-II-like systems [4] as a max – min composition: $\forall d_j \in \Delta, \forall p_q \in \Pi$

$$R_{PD}(p_q, d_j) =_{\text{def}} \max_{s_i \in \Sigma} \min\{R_{PS}(p_q, s_i); R_{SD}(s_i, d_j)\} \quad (2)$$

where $R_{PD} : \Pi \times \Delta \rightarrow [0, 1]$ are inferred diagnoses for the patient(s).

For one patient, schema (1) takes the following form:

$$D_p =_{\text{def}} S_p \circ R_{SD} \quad (3)$$

and

$$D_p(d_j) = \max_{s_i \in \Sigma} \min\{S_p(s_i); R_{SD}(s_i, d_j)\} \quad (4)$$

corresponds to the max – min composition (2).

In several fuzzy computer-assisted diagnosis systems such as, e.g., CADIAG-II, MedFrame/CADIAG-IV [2, 15], different kinds of symptom-disease relations - occurrence, confirmation, exclusion - constitute the inference mechanism. These types of relations have their roots in statistical methods.

3 Statistical diagnostic accuracy

3.1 Sensitivity and specificity

In the following we assume, that there are only two mutually exclusive states of diseases and symptoms: present or absent. A symptom, indicating the disease’s presence is called *positive*; indicating its absence, *negative*. A positive symptom s is defined as $S_p(s) = 1$, negative $S_p(s) = 0$. Analogically, for the present disease $D_p(d) = 1$, for the absent $D_p(d) = 0$.

Two basic measures of diagnostic accuracy are *sensitivity* (SE) and *specificity* (SP) [18]. Under accuracy we understand the ability of a present

symptom to indicate a disease and rule out a disease when the symptom is absent. The concept of accuracy comprises aspects of sensitivity and specificity, which will be explain below.

The sensitivity of a symptom is its ability to detect the disease when it is present. Sensitivity can be written as the probability P that the symptom s is present, $S_p(s) = 1$, given that the disease d is present, $D_p(d) = 1$, i.e.,

$$SE =_{\text{def}} P(S_p(s) = 1 | D_p(d) = 1) \quad (5)$$

From Table 1 among $a + b$ patients with disease d , a patients have symptom s . Thus, $SE = \frac{a}{a+b}$ (Table 2).

The specificity of a symptom is its ability to exclude the disease when this symptom is absent. Specificity SP is the probability P that the symptom s is absent, $S_p(s) = 0$, given that the disease d is absent, $D_p(d) = 0$, i.e.,

$$SP =_{\text{def}} P(S_p(s) = 0 | D_p(d) = 0) \quad (6)$$

From Table 1 among $c + e$ patients without disease d , e have no symptom s . Thus, $SP = \frac{e}{c+e}$ (Table 2).

	$S_p(s) = 1$	$S_p(s) = 0$	<i>Total</i>
$D_p(d) = 1$	a	b	$a + b$
$D_p(d) = 0$	c	e	$c + e$
<i>Total</i>	$a + c$	$b + e$	$a + b + c + e$

Table 1: The number of patients in various categories.

$S_p(s) = 1$	$S_p(s) = 0$
$SE(\text{or } TPR) = \frac{a}{a+b}$	$FNR = \frac{b}{a+b}$
$FPR = \frac{c}{c+e}$	$SP(\text{or } TNR) = \frac{e}{c+e}$

Table 2: From the middle part of Table 1: sensitivity and specificity.

Sometimes SE is also called true-positive rate (TPR), SP is called true-negative rate (TNR). Similarly, $FNR = \frac{b}{a+b}$ is called a false negative rate, and $FPR = \frac{c}{c+e}$ is a false positive rate (see Table 2). Absolute values a, b, c, e can be called by analogy TP, FN, FP, TN , correspondingly.

If the information about patients is collected in the form of numerical measurements, SE and SP are calculated based on a particular decision threshold, say t . All measurements bigger than t are considered as positive cases, the others - negative, and corresponding sensitivity and specificity are calculated.

The information about patients can be classified not only numerically, but due to the some rating scale, for example, as in mammography used, "normal, benign, probably benign, suspicious, malignant". In this case a decision threshold is linguistic. For example, only "suspicious and malignant" cases are called positive, others are negative.

Sensitivity and specificity are measures of intrinsic accuracy [18] because for them it is does not matter, for example, 30 patients without cancer among 60 or 30 among 3000, (i.e., with different prevalence rates, 50% and 1% accordingly) were considered. Therefore, estimated from a study sample, they are applicable to other populations with different prevalence rates.

3.2 Receiver Operating Characteristic

Only one threshold, one separation into positive and negative cases can lead to lost of information. A special method allows to overcome the limitations connected with one threshold. It is called the Receiver Operating Characteristic or ROC curve on a plane. Each point on the curve is generated by a different decision threshold, and on the x and y axis FPR and SE are plotted correspondingly.

3.3 Probability of the correct test result

Sometimes sensitivity and specificity are summarized in one number, a probability of the correct test result. From the Table 1 this number is equal to $\frac{a+e}{a+b+c+e}$. But it depends on the prevalence.

3.4 Likelihood ratio

Another single index of diagnostic accuracy is the likelihood ratio LR :

$$LR(i) = \frac{P(S_p(s) = i | D_p(d) = 1)}{P(S_p(s) = i | D_p(d) = 0)} \quad (7)$$

If a symptom s is present, i.e., $i = 1$, the corresponding likelihood ratio, $LR(+)$, is called positive LR . In the case of $i = 0$, the negative likelihood ratio is defined as $LR(-)$. One can see that

$$LR(+) = \frac{SE}{FPR} \quad (8)$$

and

$$LR(-) = \frac{FNR}{SP} \quad (9)$$

The likelihood ratio reflects the magnitude of evidence that a particular symptom provides in favor of the presence of the disease relative to the absence of the disease. A likelihood ratio of 1.0 indicates that the symptom is equally likely in patients with and without disease; a likelihood ratio > 1.0 indicates that the symptom is more likely among patients with the disease than without the disease; and a likelihood ratio < 1.0 indicates that the symptom is more likely among patients without the disease.

For example, $LR(+) = 1.53$ means that a positive test result (symptom s) is 1.53 time more likely in patient with cancer (disease d) as compared patients without cancer. But not necessarily that given the test result, a patient is 1.53 times more likely to have a cancer that not to have it.

3.5 Positive and negative predictive values

All above described measures do not answer the question, what is the probability that a patient with or without a symptom has or not a disease, i.e.,

$$P(D_p(d) = 1 | S_p(s) = 1) \quad (10)$$

or

$$P(D_p(d) = 0 | S_p(s) = 0) \quad (11)$$

respectively. The values in (10), (11) are called positive and negative predictive values PPV , NPV , respectively. These values are calculated as shown in Table 3. Determination of the probabilities depends not only on the intrinsic accuracy of the test, but also on the probability of the diagnosis before the test is performed (a priori diagnosis): the Bayes' theorem here comes into play [12].

	$S_p(s) = 1$	$S_p(s) = 0$
$D_p(d) = 1$	$PPV = \frac{a}{a+c}$	$1 - PPV$
$D_p(d) = 0$	$1 - NPV$	$NPV = \frac{e}{b+e}$
<i>Total</i>	$a + c$	$b + e$

Table 3: From Table 1: Positive and negative predictive values.

3.6 Parallel and serial tests

More than one diagnostic tests can be performed in parallel (at the same time and interpreted in combination) or serially (the results of the first test determine whether the second test is performed)

For example, for two parallel tests s_1 and s_2 with independent measures of diagnostic accuracy SE_{s_1} , SE_{s_2} , SP_{s_1} , SP_{s_2} , connected by AND or OR, the sensitivity and specificity of the combined results are defined as follows [5]:

$$SE_{s_1 \text{ OR } s_2} = SE_{s_1} + SE_{s_2} - SE_{s_1} \times SE_{s_2}$$

$$SP_{s_1 \text{ OR } s_2} = SP_{s_1} \times SP_{s_2}$$

$$SE_{s_1 \text{ AND } s_2} = SE_{s_1} \times SE_{s_2}$$

$$SP_{s_1 \text{ AND } s_2} = SP_{s_1} + SP_{s_2} - SP_{s_1} \times SP_{s_2}$$

4 Fuzzy diagnostic accuracy

Let us now describe the fuzzy counterparts of the statistical characteristics presented in the previous section.

4.1 Occurrence and confirmation

CADIAG-II like systems [4] use two types of symptom-disease relations: occurrence R_{SD}^o and confirmation R_{SD}^c . These relations are interpreted statistically and linguistically in the systems.

If a statistical way is taken, $R_{SD}^o(s_i, d_j)$ and $R_{SD}^c(s_i, d_j)$ are derived from relative frequencies:

$$R_{SD}^o(s_i, d_j) = F(s_i|d_j)$$

and

$$R_{SD}^c(s_i, d_j) = F(d_j|s_i)$$

where

$$F(s_i|d_j) = \frac{F(d_j \cap s_i)}{F(d_j)}$$

$$F(d_j|s_i) = \frac{F(d_j \cap s_i)}{F(s_i)}$$

$F(s_i|d_j)$ is a conditional frequency of s_i given d_j , $F(d_j|s_i)$ is a conditional frequency of d_j given s_i , $F(d_j \cap s_i)$ is the absolute frequency of $d_j \cap s_i$. The absolute frequencies are included to the corresponding tables where number of patients in different categories are collected (see, for example Table 1).

The similarities between the sensitivity and occurrence, PPV and confirmation can be established. Instead of a probability, a frequency is applied to avoid the critical problem of Bayes approach - finding of an a priori probability.

A linguistic representation often needs additional clarifications. Not everything, described linguistically, can be considered as fuzzy sets. Sometimes words like *always*, *often*, *medium*, *seldom*, *never* are "coded" by numbers, e.g., 1, .75, .5, .25, 0, that take part in the calculations (4) together with S_p . The values of S_p are numbers from [0, 1].

Although information about a patient symptoms in CADIAG-II-like systems is given by a fuzzy set, i.e., assuming all values between 0 and 1, a physician (not a patient himself!) uses at most three possibilities to estimate patient's symptoms practically. Thus, investigated patient symptoms are considered as *present*, *not present*, *non-applicable*, that may be "coded" by, e.g., 1, 0, $\frac{1}{2}$, correspondingly. *Non-applicable* often has several meanings: for example, "a symptom has not been examined", "my experience tells me nothing about this symptom", etc.

Assigning only unique numbers to words seems to be a rather naive way of medical knowledge qualification. More realistic is to consider intervals that represent the verbal physician opinion. If a physician says that symptom s_i often meets d_j , he may mean, that approximately in 95% s_i meets disease d_j . This medical knowledge can be represented as interval [0.92, 0.98] or a fuzzy number. The interval representation is similar to the predefined threshold, described in the sections 3.1 and 3.2.

A linguistic way opens a possibility to estimate relations R_{SD}^o and R_{SD}^c using fuzzy sets such as, e.g., *almost always*, *very seldom*, *often*, *medium*,

seldom, very seldom, almost never. These fuzzy sets are defined as mappings from $[0, 1]$ to $[0, 1]$.

Another interpretation of these linguistic terms is that they define the possibility distribution of the intervals of time within which a symptom may be observed. For example, in *influenza (always) causes fever in day 1 to day 3* the "always" indicates that indicates that if fever is not observed in 1–3 days, the disease can not be influenza [6, 7].

The numerical and linguistic definitions of symptom-disease relations seem to belong to different types of fuzzy sets. If numerical values represent the membership degrees of fuzzy sets R_{SD}^o or R_{SD}^c in points (s_i, d_j) , linguistic are close to the fuzzy sets type-2, where degrees of membership of a fuzzy relation are fuzzy sets themselves [11]. To visualize this, a symptom-disease relation can be represented in a tabular form (see Table 4), where each element of the table f_{ij} , describing connections between symptoms and diseases, is a fuzzy set.

To unify the representation of symptom-disease relations, the statistical way can also be represented by fuzzy relations type-2, where numerical values $F(s_i|d_j)$ and $F(d_j|s_i)$ are fuzzy singletons. This way the numerical and linguistic representations can be uniquely described.

4.2 Exclusion relations

In Conorm-CADIAG-II [4] the exclusion relation $R_{SD}^e : \Sigma \times \Delta \rightarrow [0, 1]$ was introduced. The value $R_{SD}^e(s_i, d_j)$ indicates the degree in which the present symptom excludes (or disconfirms) the disease d_j ,

$$R_{SD}^e = 1 - R_{SD}^c = 1 - PPV$$

In MedFrame/CADIAG-IY four fuzzy relations are used: frequencies of occurrence of the antecedents with the consequents, strengths of confirmation of the antecedents for the consequent, frequencies of occurrence of the antecedents with *not* the consequents, strengths of exclusion of the antecedents for the consequent [3]. They are "inspired" by sensitivity, specificity, PPV and NPV.

4.3 The importance of ratios

Likelihood ratios are used for example, in the system Disco [1, 8] to build confirmation and exclusion relations from occurrence relations. From the initial table, which rows and columns represent symptoms and diseases correspondingly, and sensitivities as elements f_{ij} (see Table 4), new two tables, with elements describing confirmation and exclusion relations are deduced. The higher the likelihood ratio, the likelier the symptom among patients with the disease relative to patients without disease.

	d_1	d_2	\dots	d_n
s_1	f_{11}	f_{12}	\dots	f_{1n}
s_2	f_{21}	f_{22}	\dots	f_{2n}
\vdots	\vdots	\vdots	f_{ij}	\vdots
s_m	f_{m1}	f_{m2}	\dots	f_{mn}

Table 4: An initial table for symptom-disease connections.

4.4 Combination of rules

Combinations of statistical measures of diagnostic accuracy (Section 3.6) also have found their counterparts in fuzzy medical expert systems. For example, $SE_{s_1 \text{ OR } s_2}$ aggregates two relations between symptoms and the diseases similar to the max operator in the max – min composition (2). In particular, $SE_{s_1 \text{ AND } s_2}$ and $SE_{s_1 \text{ OR } s_2}$ are aggregation operators, based on the t -norm product and its dual t -conorm.

5 Conclusions

Many computer-assisted medical systems rely on the tables of conditional probabilities and decisions are taken using the Bayes formula. But inconsistency, incompleteness of data are the main crucial points associated with probabilistic approaches. To avoid many of data collection problems, fuzzy technology is applied in medical diagnosis. In this paper we have considered some counterparts of probability and fuzzy in medical diagnosis context, such as, for example, occurrence and sensitivity, conformation and positive

predictive value. The spectrum of possible similarities between probability and fuzzy approaches in medical diagnosis is much wider. An experience, especially negative, collected by probabilities during centuries, can be helpful to find the practical application of fuzzy logic in medicine.

References

- [1] J. F. Baldwin, B. W. Pilsworth, and D. Norris. Medical diagnoses from patient records. *Fuzzy Sets and Systems*, 23:73–87, 1987.
- [2] K. Boegl. *Design and Implementation of a Web-Based Knowledge Acquisition Toolkit for Medical Expert Consultation Systems*. PhD thesis, Technische Universität Wien, 1997.
- [3] K. Boegl, K.-P. Adlassnig, Y. Hayashi, T. Rothenfluh, and H. Leitich. Knowledge acquisition in the fuzzy knowledge representation framework of a medical consultation system. *Artificial Intelligence in Medicine*, 30:1–26, 2004.
- [4] M. Daniel, P. Hájek, and H. Nguyen. CADIAG-2 and MYCIN-like systems. *Artificial Intelligence in Medicine*, 9:241–259, 1997.
- [5] P. Griner, R. Mayewski, A. Mushlin, and P. Greenland. Selection and interpretation of diagnostic tests and procedures. *Ann. Intern. Med.*, 94:553–592, 1981.
- [6] P. R. Innocent and R. I. John. Computer aided medical diagnoses. *Information Sciences*, 162(2004):81–104.
- [7] P. R. Innocent, R. I. John, and J. Garibaldi. Fuzzy methods and medical diagnoses. *Applied Artificial Intelligence*, 19(1):69–98, 2005.
- [8] T. Kiseliova. *DISCO-A computer assisted method in medical diagnosis using new information technologies*. PhD thesis, Tbilisi State University, Georgia Technical University, 1995.
- [9] G. Kolousek. *The System Architecture of an Integrated Medical Consultation System and Its Implementation Based on Fuzzy Technology*. PhD thesis, Technische Universität Wien, 1997.
- [10] B. Kosko. *Neural Networks and Fuzzy Systems: A Dynamical Approach to Machine Intelligence*. Englewood Cliffs, New Jersey: Prentice Hall, 1992.
- [11] J. M. Mendel. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall PTR, Upper Saddle River, 2000.
- [12] C. Naylor. How to build an inferencing engine. In R. Forsyth, editor, *Expert Systems, Principles and Case Studies*, chapter 5, pages 85–105. Chapman and Hall, 1989.
- [13] E. Sanchez. Medical diagnosis and composite fuzzy relations. In M. M. Gupta, R. K. Ragade, and R. R. Yager, editors, *Advances in Fuzzy Set Theory and Applications*, pages 437–444. North-Hollands, New York, 1979.
- [14] R. Seising, C. Schuh, and K.-P. Adlassnig. Medical knowledge, fuzzy sets and expert systems. In *Workshop on Intelligent and Adaptive Systems in Medicine*, Prague, March 31 - April 1, 2003.
- [15] F. Steimann. Fuzzy set theory in medicine. *Artificial Intelligence in Medicine*, 11(1997):1–7.
- [16] L. Zadeh. The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy sets and Systems*, 11:199–227, 1983.
- [17] L. Zadeh. Probability theory and fuzzy logic are complementary rather than competitive. *Technometrics*, 37:271–276, 1995.
- [18] X.-H. Zhou, N. A. Obuchowski, and D. K. McClish. *Statistical Methods in Diagnostic Medicine*. Wiley Series in Probability and Statistics. A John Wiley and Sons, Inc., Publication, New York, 2002.