Fuzzy sets in modeling of patient’s disease states

A. AMELJANČZYK
andrzej.ameljanczyk@wat.edu.pl

Military University of Technology, Faculty of Cybernetics
Institute of Computer and Information Systems
Kaliskiego Str. 2, 00-908 Warsaw, Poland

The paper concerns the mathematical modeling of patient’s disease states and disease unit patterns for the needs of algorithms supporting medical decisions. Due to the specificity of medical data and assessments in the modeling of patient’s disease states as well as diseases, the fuzzy set methodology was used. The paper presents a number of new characteristics of fuzzy sets allowing to assess the quality of medical diagnosis. In addition, a definition of a multi-aspect fuzzy set is presented, which may be useful in supporting medical diagnostics based on multi-criteria similarity models. The presented results can be used in the construction of algorithms for assessing the patient's state of health and mainly in the construction of algorithms for supporting diagnostic processes.

Keywords: fuzzy set, multi-aspect fuzzy set, fuzzy sets similarity, fuzzy pattern of disease unit, medical diagnosis.

GICID: 01.3001.0013.6596

1. Introduction

An attempt to describe the patient’s state of health or his condition involves the need to define certain reference patterns in the form of human health patterns or the patterns of so-called disease units (diseases) [4, 6, 20]. In each of these cases it is necessary to identify certain medical features (including measurement of their value) that are identifiable as a result of observation (physical examination) or as a result of specialist medical tests. These are called symptoms of the patient’s health or illness. Examples of disease symptoms of the first group are: lymphadenopathy, skin lesions, fever, lack of appetite, diarrhea, night sweats, weight loss, dizziness, headaches, abdominal pain, bleeding, coughing, chills, palpitations, etc. [13, 14, 24].

The second group of symptoms are the so-called hidden symptoms, which can be found only in the result of specialist medical tests, for example: body temperature, blood pressure, blood glucose, cholesterol, PSA level, creatinine level, level HGB, ECG, X-ray, CT scan, ultrasound, etc. Additional diagnostic information also carry so-called risk factors, which are mainly due to external circumstances rather than the patient’s organic characteristics, such as old age, smoking, inactivity physical stress, permanent stress, overweight, alcohol abuse, family burden of a given disease, fat diet, sedentary life, abdominal obesity, diabetes, a tendency to depression, etc. The diagnostic procedure is usually iterative (multi-stage) procedure. The first stage is medical history, as a result of which the doctor determines the patient’s symptoms (disease symptoms) of the first group and risk factors. Then taken is a preliminary diagnosis, which is to detail and authenticate the specialist tests selected by the doctor. This step, depending on the result, can be repeated several times. The process of medical diagnosis is therefore an iterative, extremely complex undertaking, on which the method of treatment and the patient’s final state of health depends. Generally, this is a task in the area of pattern recognition. Such a task is to determine from a set of patterns of disease units, a disease unit most similar to a properly defined disease state of the patient.

The specificity of recognizing diagnostic medical patterns is mainly due to the fact that the patient’s state of health (for various reasons) is difficult to reliably and precisely determine and the number of patterns of disease units reaches several or even several thousand. The patient’s state of health at a given time may be determined by a dozen or so symptoms and many risk factors as well as hundreds of other medical parameters values which can be determined often as a result of time-consuming, complicated and expensive specialist tests performed in medical laboratories. The severity of the occurrence of specific symptoms or risk factors
is difficult to assess both by the doctor and the patient himself mainly due to the patient’s subjective perception and individual characteristics. In addition, the patient may be sick at the same time of several concomitant diseases, the symptoms of which may interfere or even endure. In this case, the specificity of recognizing medical patterns is mainly due to the fact that you need to look for similarity of the patient’s health to several diseases. Therefore, the diagnostic process is a sequential process, which begins with a medical history during the patient’s first visit. Initial medical diagnosis is generally the basis for further diagnostic steps involving additional specialist tests. The sequence of these tests, their number and scope are extremely difficult to determine correctly. This is undoubtedly a very responsible, complex and difficult optimization task that the doctor must perform. It affects the effectiveness of treatment, duration of treatment and cost. The following considerations will be devoted to the possibility of defining medical patterns in conditions of uncertainty, subjectivity of both doctor and patient assessments for the first phase of diagnosis leading to initial diagnosis.

2. Fuzzy sets – basic definitions and characteristics

The methodology (philosophy of approach) of fuzzy set theory [35, 36] – like no other, fits the specifics of medical data modeling [1, 26, 28, 32]. The basis of this approach are the following definitions. Let $P$ finite set.

**Definition 1 [35]**

The fuzzy set $A$ in space $P$ is the set of ordered pairs:

$$A = \{ (a, \mu_\alpha(a)) | a \in P \}$$

where $\mu_\alpha(a)$, $a \in P$ is the degree of belonging element $a$ to the set $A$. The function $\mu_\alpha$ is called the function of belonging and takes values from the range $[0, 1]$.

**Definition 2 [35]**

The support of the fuzzy set $A$ is the classic (sharp) set:

$$\text{supp}(A) = \{ a \in P | \mu_\alpha(a) > 0 \}$$

**Definition 3 [35]**

Core of the fuzzy set we call the classic (sharp) set in the form of: $\text{core}(A) = \{ a \in P | \mu_\alpha(a) = 1 \}$

**Definition 4 [35]**

The height of fuzzy set $A$ is the number in the form: $\text{hgt}(A) = \max \{ \mu_\alpha(a), a \in P \}$

**Definition 5**

The threshold of the fuzzy set $A$ is the number in the form: $\text{thres}(A) = \min \{ \mu_\alpha(x), x \in \text{supp}(A) \}$

**Definition 6**

The extension of the fuzzy set $A$ is the number in the form: $\text{exten}(A) = \text{hgt}(A) - \text{thres}(A)$

**Definition 7**

The roof of the fuzzy set $A$ is the set in the form:

$$\text{roof}(A) = \{ x \in \text{supp}(A) | \text{no exists } y \in \text{supp}(A), \text{ such that } \mu_\alpha(y) > \mu_\alpha(x) \}$$

**Definition 8**

The floor of fuzzy set $A$ is the set in the form:

$$\text{floor}(A) = \{ x \in \text{supp}(A) | \text{no exists } y \in \text{supp}(A), \text{ such that } \mu_\alpha(y) < \mu_\alpha(x) \}$$

**Definition 9**

The sharpness of the fuzzy set $A$ will be called the factor:

$$\text{sharp}(A) = \frac{\sum_{a \in \text{supp}(A)} \mu_\alpha(a)}{\text{supp}(A)}$$

**Definition 10**

The fuzzyness of the set $A$ will be called the factor:

$$\text{fuzze}(A) = 1 - \frac{\sum_{a \in \text{supp}(A)} \mu_\alpha(a)}{\text{supp}(A)}$$

**Example 1**

Let $P$ – space of elements, in the next form:

$$P = \{ a, b, c, d, e, f, g, h \}$$

while $A$ is a fuzzy set as follows:
A = \{(a, 0), (b, 0), (c, 0.6), (d, 0.5), (e, 0.5), (f, 1), (g, 0), (h, 1)\} – human body weight (mass) \ a \in P

We can say about the fuzzy set \( A \) that the elements \( f \) and \( h \) belong to it with certainty, while the elements: \( a, b, g \) certainly do not belong to it. The other element belongs to it with the degree of certainty 0.5 or 0.6. We also say about the set \( A \) that it is normal fuzzy set [35].

We will obtain the following characteristics of fuzzy set \( A \) in the space \( P \):

- support of the set – \( \text{supp}(A) = \{c, d, e, f, h\} \)
- core of the set – \( \text{core}(A) = \{f, h\} \)
- height of the set – \( \text{hgt}(A) = 1 \)
- threshold of the set – \( \text{thres}(A) = 0.5 \)
- extension of the set – \( \text{exten}(A) = 1 - 0.5 = 0.5 \)
- roof of the set – \( \text{roof}(A) = \{f, h\} \)
- floor of the set – \( \text{floor}(A) = \{d, e\} \)
- sharpness of the set – \( \text{sharp}(A) = 0.72 \)
- fuzzyness of the set – \( \text{fuzze}(A) = 0.28 \)

The classic definition of fuzzy set (Definition 1) can easily be extended to fuzzy sets of a multi–aspect nature. We will understand the belonging of an element to a multi–aspects fuzzy set as belonging to this set in the sense of many of its (features) aspects. Assuming that \( P \) is a set of elements (objects) each with \( N \) characteristics, numbered index.

**Definition 11**

A multi–aspect fuzzy set \( A \) in space \( P \) will be called a set of ordered pairs:

\[
A = \{(a, \mu_s(a)) | a \in P\}
\]

where \( \mu_s(a), a \in P \) is the multi–aspects degree of belonging to the element \( a \) to set \( A \). The function \( \mu_s \) is called the membership function which assumes normalized values from the area \([0,1] \times [0,1] \times \cdots \times [0,1]\). It is a vector function of the form:

\[
\mu_s(a) = (\mu_s^1(a), \mu_s^2(a), \ldots, \mu_s^N(a)), a \in P
\]

\( \mu_s^1(a) \) – the degree of belonging element \( a \) to set \( A \) in terms of the feature \( n \in N \).

**Example 2**

Let \( P \) set of people. Set \( A \) of “big fat people” is defined as a fuzzy set

\[
A = \{(a, \mu_s(a)) | a \in P\}
\]

where \( \mu_s(a) = (\mu_s^1(a), \mu_s^2(a)), a \in P \)

\( \mu_s^1(a) \) – human growth \( a \in P \)

All qualitative characteristics for multi–aspects fuzzy sets (see Definitions 2–10) are constructed analogously.

### 3. Fuzzy sets in modeling of patient’s disease states

The patient’s state of health during the visit to the doctor is assessed as a result of visual inspection and interview. The doctor determines first of all the set of occurring disease symptoms and their severity, for example on a scale of 1 to 10. The effect of such interview is presented in the Table 1. The set of patients will be marked as follows \( X = \{x_1, \ldots, x_i, \ldots, x_K\} \).

<table>
<thead>
<tr>
<th>No.</th>
<th>Name of Symptom Indicating on Disease</th>
<th>( e^m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>arthralgia</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>headaches</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>palpitations</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>balance disorders</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>neck stiffness</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>dizziness</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>tinnitus</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>diarrhea</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>dyspnea</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>blurred vision</td>
<td>3</td>
</tr>
</tbody>
</table>

This table lists the symptoms and their intensities, subjectively assessed by a doctor. The patient’s \( x \in X \) condition \( W(x) \) can be recorded as a typical fuzzy set;

\[
W(x) = \left\{ \left( s, \mu_{w(s)}(s) \right) | s \in S \right\}
\]

(1)

Where \( S = \{s_1, \ldots, s_i, \ldots, s_n\} \) the set of determinated disease symptoms and \( \mu_{w(s)} \) the function the degree of belonging the symptoms to the fuzzy set \( W(x) \), defined as follows:
\[ \mu_{W_x}(s) = \frac{e_s}{10} \]

Therefore

\[ W(x) = \{ (1, \mu_0), (2, \mu_0), (3, \mu_0), (4, \mu_0), (5, \mu_0), \\
(6, \mu_0), (7, \mu_0), (8, \mu_0), (9, \mu_0), (10, \mu_0) \} \]

We will receive the following characteristics of the fuzzy set, which additionally describe the patient’s state of health:

- support (basic information) of the patient’s state health \( W(x) : \text{supp}(W(x)) = \{1, \ldots, 10\} \)
- core (golden symptoms [2, 17]) of the fuzzy set: \( \text{core}(W(x)) = \{4, 8\} \)
- thres \( (W(x)) = 3/10 \) – threshold of \( W(x) \)
- exten \( (W(x)) = 7/10 \) – extension of \( W(x) \)
- roof \( (W(x)) = \{4, 8\} \) – the most important symptoms in the state of health \( W(x) \)
- floor \( (W(x)) = \{7, 9, 10\} \) – the least important symptoms in the state of health \( W(x) \)
- sharp \( (W(x)) = 65/100 \) – sharpness of \( W(x) \)
- fuzz \( (W(x)) = 35/100 \) – fuzzyness of \( W(x) \)

All these fuzzy set characteristics have a very specific medical interpretation [14, 15, 28, 32, 36].

4. Modeling of disease unit patterns – fuzzy patterns

The problem of modeling disease unit patterns in application to medical diagnostic support procedures is extremely important. The pattern of the disease unit must be a model (picture) of the disease. Patterns of disease units defined by medical specialists (medical diagnostics) [6, 8, 11, 14, 19, 20, 24, 27], are based mainly on medical symptoms of a given disease, understood in terms of external symptoms found during physical examination and on “symptoms” found as a result of additional specialist examinations [2, 16, 17, 18, 24].

Table 2 presents a typical pattern of disease unit – borreliosis disease. The presence of certain symptoms and risk factors is additionally characterized by information on their significance (frequency) in diagnosing this disease, in the form of scores on a scale of 1 to 10 or on a scale [0,1]. For the sake of simplicity, we will assume in the following considerations that the initial diagnosis process will be limited only to taking account of disease symptoms. The risk factors as well as the results of specialist tests will be taken into account at later stages of the work.

Let \( M = \{1, \ldots, m, \ldots, M\} \) – a set of disease unit numbers in the diagnostic repository. The set of possible symptoms (for simplification of their numbers) will be presented in the form: \( S = \{1, \ldots, s, \ldots, S\} \). Data from the third column in Table 2, determined by an expert, denoting the significance (weight) of a specific symptom in the definition of this disease unit. They mainly result from the frequency of this symptom in the disease in question and testify to the degree of belonging to this pattern (they can also easily be converted to probability distribution estimated by an expert or normalized).

Let \( e_s^w \) be it the “point – value”, determined by the expert for the symptom \( s \in S \) in the disease \( m \in M \) \((1 \leq e_s^w \leq 10)\).

| Tab. 2. Description of the disease unit pattern |
|---|---|
| **Name of disease unit** | **SYMPTOMS** |
| **BORRELIOSIS** | **SYMPTOMS** |
| No. (s) | Name of Symptom Indicating on Disease | \( e_s^w \) |
| 1 | erythema migrans | 10 |
| 2 | facial nerve paralysis | 7 |
| 3 | arthralgia | 7 |
| 4 | headaches | 7 |
| 5 | palpitations | 7 |
| 6 | balance disorders | 5 |
| 7 | stiff neck | 5 |
| 8 | dizziness | 5 |
| 9 | convulsions | 5 |
| 10 | tinnitus | 3 |
| 11 | diarrhea | 3 |
| 12 | dyspnoea | 3 |
| 13 | blurred vision | 3 |

The symbol \( e_s^w \) will mean their normalized values \( e_s^w = \frac{e_s}{10} \) (from 0 to 1). The numerical values \( e_s^w \) are the subjective assessments of experts in the field of medical diagnostics. They can be treated as a certain approximation of real values obtained on the basis of a suitably large set of training data such as: (symptoms found – disease unit), in the machine learning process in a supervised learning model.

The interpretation of the data from Table 2 (borreliosis disease) is as follows: the most important (having the highest diagnostic value) symptom is a symptom called erythema migrans,
the next important (a bit smaller importance) are: facial nerve paralysis, arthralgia, headaches, palpitation heart. The following symptoms are slightly less important: balance disorders, neck stiffness, dizziness and convulsions. Symptoms are the least important: tinnitus, diarrhea, shortness of breath, blurry vision diarrhea, dyspnoea. The same meaning (although these are only circumstances) have so-called risk factors, whose presence also has a specific diagnostic value. It is worth noting that in addition to subjectivity in determining the significance of symptoms in patterns, assessment (also subjective) of severity is of great importance observed symptom (or risk factor) during patient examination. For many reasons, reading, registering and evaluating symptoms is a very complex and subjective process. The exposure of other interfering factors ("noise") and characteristics of the patient as well as many other interfering factors ("noise") and the subjective assessment of the doctor can be significantly distorted. An additional difficulty in the classification of a given symptom for a particular disease may be the fact that the patient suffers from more than one disease. All this causes that the set of symptoms defining a specific disease entity should be treated in terms of uncertain or approximate data [1, 26, 25, 28, 35, 36]. The theory of fuzzy sets [35] provides a lot of tools helpful in modeling this type of "objects" as the disease unit pattern. The disease unit number \( m \in M \) can be written as a fuzzy set as follows:

\[
W(m) = \left\{ (s, \mu_{W(m)}(s)) \right\}_{s \in S}
\]

where \( S \) the set of symptoms and \( \mu_{W(m)}(s) \) the function of belonging the symptoms \( s \in S \) to the pattern \( W(m) \). Having expert data on patterns or frequency data (see Table 2), the function of belonging the symptoms to the pattern will be written: \( \mu_{W(m)}(s) = \varepsilon^m \). For the example presented in Table 2, the borreliosis disease pattern will be written as follows:

\[
W(m) = \left( \{1,1\}, \{2,7/10\}, \{3,7/10\}, \ldots, \{13,3/10\} \right).
\]

Typical fuzzy set characteristics such as the support, core, height, extension, roof, floor, sharpness, fuzziness have a direct interpretation when applied to diseases as a medical standards. For example, support of the borreliosis pattern is defined as follows:

\[
\text{support}_\varepsilon(W(m)) = \left\{ s \in S \mid \mu_{W(m)}(s) > \varepsilon \right\}
\]

is a set of symptoms belonging to the disease unit pattern for which the value of the membership function is bigger than \( \varepsilon \), (the given threshold value - in practice often assumed that \( \varepsilon = 0 \)). Analogical (see Table 2)

\[
\text{core}(W(m)) = \left\{ s \in S \mid \mu_{W(m)}(s) = 1 \right\}
\]

is the so-called core of the pattern of borreliosis disease. In this case, \( \text{core}(W(m)) = \{1\} \) (it is a set of specific symptoms, so-called golden symptom, which are decisive in the process of diagnosing the disease). For example, for borreliosis disease, the core is a one-element set. Another characteristic of a fuzzy set is its sharpness (sharpness factor):

\[
\text{sharp}(W(m)) = \frac{\sum_{s \in S} \mu_{W(m)}(s)}{\text{support}(W(m))}
\]

or used opposite – fuzzy (coefficient of fuzzy set, fuzziness factor):

\[
\text{fuzzy}(W(m)) = 1 - \frac{\sum_{s \in S} \mu_{W(m)}(s)}{\text{support}(W(m))}
\]

The idea of expert description of the disease pattern in the form presented in the example table leads to fuzzy models. Similar to machine learning systems, you can also teach a pattern system in a supervised mode with a sufficiently large training data set. The direct effect of learning will be the functions of belonging the symptoms to individual disease patterns. The diagnostic task in this case is the diagnosis consisting in matching the most similar disease pattern \( W^*(m) \) to the established disease state \( W(x) \) of a patient from the Repository, containing all the patterns of disease units \( REP = \{ W(m) \mid m \in M \} \). This boils down to the task of determining such \( m \in M \) that

\[
p\left(W(x), W^*(m)\right) = \max_{m \in M} p\left(W(x), W(m)\right)
\]

where \( p\left(W(x), W(m)\right) \) is the similarity (matching) function of a fuzzy set \( W(x) \) to a fuzzy set \( W(m) \). More details can be found in [1, 4, 5, 6, 12, 14, 15, 26, 28, 35, 36].
5. Final conclusions

The paper presents a comprehensive opportunity to use the fuzzy set philosophy (approach) and tools in the construction of algorithms to support medical diagnostics at the introductory stage. A number of new concepts (definitions) have been defined, based on the fuzzy set theory, which can be used to determine and evaluate medical diagnosis such as support, core, threshold, extension, roof, floor or such as: “fuzzy patient’s condition”, “fuzzy pattern of the disease unit” or “fuzzy diagnosis”. The proposed approach does not, of course, exclude the possibility of using “fuzzy results” to define typical, base classifiers or build complex meta classifiers [6, 9, 12]. The defuzzification process serves this purpose [10, 26, 28, 36].

The presented example has mainly demonstrative aspects, because it is based on a very small Repository of disease unit patterns. The results obtained fully confirm the possibility of practical implementation of the support system, working on real data. Further work will be focused on the use of more precise similarity models to define the indicators of fuzzy set matching. Multi-aspect models seem promising and multi – criteria that allow determining the similarity of fuzzy sets [4, 7, 8, 10, 11, 12, 33]. They allow building more accurate, multi-aspect algorithms for recognizing fuzzy patterns of diseases.

6. Bibliography

Zbiory rozmyte w modelowaniu stanów chorobowych pacjenta

A. AMELJAŃCZYK

Praca dotyczy modelowania matematycznego stanów chorobowych pacjenta oraz wzorców jednostek chorobowych na potrzeby algorytmów wspomagania decyzji medycznych. Z uwagi na specyfikę danych i ocen medycznych w modelowaniu stanów chorobowych pacjenta, a także chorób zastosowano metodologię zbiorów rozmytych. W pracy przedstawiono wiele nowych charakterystyk zbiorów rozmytych pozwalacych ocenić jakość uzyskanej diagnozy. Dodatkowo zaprezentowano definicję wieloaspektowego zbioru rozmytego, która może być przydatna we wspomaganiu diagnozy medycznej, opartej na wielokryterialnych modelach podobieństwa. Uzyskane wyniki mogą być wykorzystane w budowie algorytmów oceniania stanu zdrowia pacjenta, a głównie w budowie algorytmów wspomagania procesów diagnostycznych.

Słowa kluczowe: wieloaspektowy zbiór rozmyty, podobieństwo zbiorów rozmytych, rozmyty wzorzec jednostki chorobowej, diagnoza medyczna.