Developing a Supporting Expert System for Lung-Disease

Diagnosis Using Fuzzy Logic

Nguyen Hoang PHUONG*, Hung H. DANG**, Nadipuram R. PRASAD**

*Institute of Information Technology, National Center for Natural Science and Technology, Vietnam
**Department of Computer and Electrical Engineering, New Mexico State University, USA

(Received 20 December 1999, revised and accepted 20 April 2000)

Abstract: This paper summarizes our research on developing a supporting expert system for pulmonary tuberculosis and lung-disease diagnosis, which has a fuzzy reasoning engine and a fuzzy-rule knowledge-base. The system has undergone through the theoretical and practical developments, which can be found in our earlier papers, and is now completely built. It has two new features as compared to those mentioned in our previous papers: capability of diagnosing more than just one disease, and a strengthened knowledge-base containing doctors' knowledge on diagnosing lung diseases: Pulmonary Tuberculosis, Lung Abscess, Lung Cancer, Asthma, Pneumonia, Bronchiectasis, etc.

Keywords: Supporting Expert Systems, Lung-Disease Diagnosis, Fuzzy Logic

1. Introduction.

1.1 About Pulmonary Tuberculosis and lung diseases.

Pulmonary Tuberculosis in particular and lung diseases in general, are very dangerous diseases. During the past few decades, in many industrialized countries, money, resources, high standards of living, and chemotherapy treatment has helped reduce these diseases to a relatively minor problem. But in poorer countries, these diseases remain almost as big as ever and Tuberculosis is considered the biggest killer in the world as a single pathogen.

1.1.1 Problems in diagnosis

There are medicines that can effectively treat most forms of lung diseases including Pulmonary Tuberculosis. However, problems still occur in diagnosis.

During the process of diagnosing, doctors have to cope with many difficulties: patient's symptoms usually are unclear, specific bacteria are not found, and there are similarities between diseases in terms of clinical symptoms.

In these cases, tests for tubercular bacilli in patients' provide negative results while the doctors can only treat a patient as a probable PT case if the bacillus is found. Unfortunately, "Negative-BK" cases are still very prevalent. In these cases, doctors have to use their knowledge and experience to arrive at reasoning based on clinical symptoms and determine whether or not to treat the patient as a PT patient. The accuracy of diagnosis therefore, depends heavily on the experience and knowledge of the doctor.

1.1.2 A solution: a diagnostic supporting expert system

To deal with the problems, we proposed a solution: a medical expert system, which can support doctors in diagnosis. This expert system comprises a knowledge-base of knowledge acquired from many doctors experienced in lung disease diagnosis and treatment. Therefore, with a suitable reasoning engine, it can carry out the diagnosis and generate conclusions about the patient's pathologic status, which is as accurate as or even more accurate than conclusions made by an experienced skilled doctor. Normally, the conclusion from the diagnostic system should be used as a secondary reference to support in the decision-making process. Another advantage is that the diagnostic system can serve anywhere and at anytime. That is, it can serve as an emergency doctor in case a human-doctor is absent or there is a lack of medical experts, which is very likely in...
areas that are not easily accessible, as is the case in many Third World countries.

1.2 Paper Organization

In this paper, we discuss an expert system called DoctorMoon - a diagnostic supporting expert system for pulmonary tuberculosis and lung diseases. In Sections 2 and 3 we discuss the first two stages in building DoctorMoon, namely, designing and developing stages. Results of tests and evaluations using DoctorMoon are provided in Section 4. Some conclusions and discussions about future development of DoctorMoon are given in Section 5.

2. Designing

2.1 General structure

Like any other expert system, the two most important components of DoctorMoon are its knowledge base and reasoning engine. Through an interface, users can edit the knowledge base or enter the necessary information the system needs to diagnose and receive the output: diagnostic result.

![DoctorMoon Diagram]

The knowledge base of DoctorMoon is a fuzzy-rule-based system with the reasoning engine powered by Fuzzy Logic.

2.2 Need for Fuzzy Set Theory

While computers normally require and can process only crisp and exact data, it is recognized that data needed for diagnosis is generally fuzzy. It is very hard to model symptoms such as “hurt”, “slightly hurt”, “long coughing”, “quite long coughing”, “likely infected”, “unlikely infected”, etc., using classical set theoretical concepts. To overcome the difficulty of representing imprecise data we had to apply Fuzzy Set Theory in building DoctorMoon. Fuzzy Set Theory [2] is an effective tool to deal with fuzzy data in that it provides methods for modeling fuzzy concepts and a number of special operators for working on those modeled data.

2.3 Designing the KNOWLEDGEBASE

Knowledge acquisition is always a very important stage in building an expert system. How the doctors' knowledge is acquired and handled will determine the quality of the system. In DoctorMoon, the experience and knowledge of doctors were represented by a set of fuzzy rules. These rules are very much like how the human would think and are consequently easy to maintain.

The general form of the rules are as follows:

IF <Condition> THEN <Conclusion> WITH <Grade> (0<Grade<1)

in which:
<Condition> represents the clinical status of the patient. This is a combination of the presence of certain symptoms, which are useful for the diagnostic process.
<Conclusion> represents the conclusion about the patient’s infection status: infected or not.
<Grade> represents the level at which the patient is infected or not infected.

Doctors actually have two types of diagnosis, namely, affirmative diagnosis for affirmation and exclusive diagnosis for exclusion. To parallel the diagnosis of a human doctor, DoctorMoon also uses two types of rules, namely, positive rules for affirmative reasoning and negative rules for exclusive reasoning. For positive rules, the “Conclusion” is “INFECTED BY DISEASE” and for negative rules, the “Conclusion” is “NOT INFECTED BY DISEASE”. In DoctorMoon, Grade is a fuzzy value.

In fact, the conclusions made by a doctor are never simply “YES” or “NO”. They can think of many possibilities: “certainly infected”, “likely infected”, “unlikely infected”, etc. Similarly, in DoctorMoon, the infection level is placed on a scale from 0 to 1.

Consider for example a Positive rule given by:

IF <“Long Cough” and “Having slight fever in the afternoon”> THEN
In this example, the rule expresses that a patient having two symptoms, namely, “Long Cough” and “Having slight fever in the afternoon” is infected by pulmonary tuberculosis at grade 0.3

A Negative rule could be of the form:

IF <<High fever” and “Response to antibiotic’> THEN < NOT INFECTED BY PULMONARY TUBERCULOSIS> WITH <0.6>

In this example, the rule expresses that a patient having two symptoms, namely, “High fever” and “Response to antibiotic” is not infected by pulmonary tuberculosis at grade 0.6.

Doctors also have another type of reasoning in which, from the presence of a set of symptoms, they can determine if another symptom, which is not included in the set, is also present. To emulate this type of diagnosis, DoctorMoon uses a third type of rule – the intermediary rule. In the intermediary rules, the general form is unchanged, but in the <Conclusion>, instead of “INFECTED BY DISEASE” or “NOT INFECTED BY DISEASE” is the affirmation of a certain symptom at the grade of <Grade>.

Consider for example the following intermediary rule:

IF <<“Long cough” and “Fatigue”> THEN <<INFECTED BY Weight loss”> WITH <0.2>

The rule expresses that a patient has “Long cough” and “Fatigue” should also be suffering from “Weight loss” at grade 0.2.

Evidently, the symptom in the <Conclusion> must not appear in the <Condition>.

2.4 Designing the REASONING ENGINE

After the knowledge base has been created, the next step is to build a reasoning engine which can reason upon a set of symptoms as the clinical status of a patient using all the suitable rules in the knowledge base. This process allows the diagnostic system to make conclusions about the patient’s infection.

To design a reasoning engine, it is necessary to consider three classes of medical entities:

- Symptoms, signs, lab test results, and finding (S), where S includes both General symptoms (Sg) and Distinguishing symptoms (Sd). Therefore, Sg, Sd ⊆ S.
- Elementary conjunctions of symptoms ( Ei)
- Diseases: D

where

- Value υS indicates the degree to which a patient exhibits symptom S, or in other words, the severity of the symptom S. υS = 1 means symptom S is surely present for patient Pq, υS= 0 means symptom S is surely absent for patient Pq, symptom S takes value υS in (0,1) means possible hypothesis of presence of symptom S for patient Pq, and value υS = ε means that the presence of symptom S is undefined (ε takes a value closer to 0). Let S = {S1, S2, S3} denote the set of symptoms.

- Symptoms Ej takes values υEj in [0,1]. Computing the weight of E using truth functions of fuzzy logic over {0,1} : NEG(x) = 1-x for negation, CONJ(x,y) = min (x,y) for conjunctions. Let E = {E1, E2, ... En} denotes the set of all elementary conjunctions of symptoms, i.e., conjunction of some symptoms or some negated symptoms.

Disease D takes values μD (Pq, D), μD (Pq, D), where the value μD (Pq, D) affirms D the value μD (Pq, D) excludes D and the value μD (Pq, D) is the total grade for affirmation and exclusion of D by patient Pq from the observed symptoms.

The values μD (Pq, D), μD (Pq, D) takes values in [0,1] and the value μD (Pq, D) takes values in [-1,1].

The following example illustrates the symptoms for Pulmonary Tuberculosis:

Group 1. General symptoms (Sg): symptoms or signs exhibited by most Pulmonary tuberculosis patients. They are:

Respiratory:

General

+++ Cough
+++ Sputum
++ Loss of weight
++ Fever and sweating
++ Blood-spitting
++ Tiredness
+ Chest wall pain
+ Loss of appetite
+ Breathlessness
+ Localized wheeze
+ Frequent colds.

(The number of “+” signs indicate the importance of the symptom)
We use these symptoms to suspect or exclude tuberculosis in the first step of the diagnosing process.

Group 2. Distinguishing symptoms ($S_i^D$): symptoms that distinguish Pulmonary tuberculosis from another disease.

These symptoms would help us to make sure that the patient is suffering from tuberculosis or another disease which causes the same set of symptoms e.g. pneumonia, lung cancer, lung abscess, bronchiectasis, asthma, etc. Although there are many symptoms that can be attributed to two or more diseases at the same time, we can still find out very particular signs to distinguish pulmonary tuberculosis from others: They are: X-ray appearance of lung, number of leucocyte, tuberculousness in medical history, living near tuberculous patient, etc.

Thus, this group contains special symptoms:
+ Symptoms that belong to only pulmonary tuberculosis, not any other similar disease. These symptoms would help us affirm pulmonary tuberculosis.
+ Symptoms that belong to a similar disease and never exhibited by a tuberculous patient. These symptoms would help us to exclude pulmonary tuberculosis.

For example, Pulmonary tuberculosis and pneumonia have similar symptoms: coughing up phlegm, high fever, lesions appear on X-ray of the lung, etc., but the number of leucocyte is normal in tuberculous patient while it increases highly in patient having pneumonia.

The relationships between the above entities may be the following:

$$E_i \Rightarrow D \ (\mu_{SD}^i (E_i, D)), \ E_i \Rightarrow \neg D \ (\mu_{SD}^i (E_i, D))$$

in which $E_i$ is a symptom or elementary conjunction of symptoms, $D$ is a lung disease $\neg D$ is auxiliary lung disease. The values $\mu_{SD}^i (E_i, D)$ indicate degrees at which present symptom or elementary conjunction of symptoms $E_i$ affirm the lung disease. Similarly, the values $\mu_{SD}^i (E_i, D)$ indicate degrees at which present symptom or elementary conjunction of symptoms $E_i$ exclude the lung disease. We make the following consistency assumptions:

$$\mu_{SD}^i (E_i, D) \ast \mu_{SD}^i (E_i, D) = 0.$$

Because it is impossible that symptom or elementary conjunction of symptoms $E_i$ both affirms and excludes $D$.

DoctorMoon accepts fuzzy descriptions of the patient’s symptoms and infers approximate descriptions of the patient’s lung disease by means of the fuzzy relationships (including positive and negative knowledge) described in the previous section.

The inference rules are used to deduce lung disease $D$ suffered by patient $P_q$ from the observed symptoms $S_i$:

$$R_{PD}^D = R_{P_S} \circ R_{SD}^D$$

defined by

$$\mu_{PD}^D (P_q, D) = \vee_{i \in S} \min \{\mu_{PS} (P_q, E_i), \mu_{SD}^i (E_i, D)\}$$

and the inference rules are used to deduce the negation of suffered by patient $P_q$ from the observed symptoms $S_i$:

$$R_{PD}^D = R_{P_S} \circ R_{SD}^D$$

defined by

$$\mu_{PD}^D (P_q, D) = \vee_{i \in S} \min \{\mu_{PS} (P_q, E_i), \mu_{SD}^i (E_i, D)\}$$

where $\vee$ stands for a t-conorm (see the definition of t-conorm in [4]).

One example of t-conorm (or s-norm) is the Probabilistic Operator: $A_t(a, b) = a + b - a \cdot b$

Thus, we define a relation of total degree for affirmation and exclusion of lung disease $D$ as the following:

$$\mu_{PD}^D (P_q, D) \ast \mu_{PD}^D (P_q, D) \ast \mu_{PD}^D (P_q, D)$$

where $\ast$ is an extended ordered Abelian group operation on $[-1, 1]$.

One can see the notion of an extended ordered Abelian group in [4],[5]. An example of the extended ordered Abelian group operation on $[-1, 1]$ as the combining function of EMYCIN is the following [4]:

$$x \oplus_{AC} y = x + y - x \cdot y \quad \text{for} \ x, y \geq 0,$$

$$x \oplus_{AC} y = x + y + x \cdot y \quad \text{for} \ x, y \leq 0,$$

$$x \oplus_{AC} y = (x + y)(1 - \min \{x, y\}) \quad \text{for} \ x, y \leq 0.$$

Where $\mu_{PD}^D (P_q, D)$ is a degree for diagnosis lung disease $D$.

To obtain a conclusion, the following criteria for diagnosis of lung disease $D$ are proposed:
Criteria for diagnosis of lung disease D

1. If the value \( \mu_{PD}^{1} (P,D) = 1 \) then absolutely affirmed lung disease.
2. If the value \( \mu_{PD}^{1} (P,D) = -1 \) then absolutely disaffirmed lung disease D.
3. If the value \( \mu_{PD}^{1} (P,D) \) such that \( -\varepsilon < \mu_{PD}^{1} (P,D) \) \( \leq \varepsilon \) then an "unknown" conclusion is suggested.
4. If the value \( \mu_{PD}^{1} (P,D) \) such that \( -1 < \mu_{PD}^{1} (P,D) \) \( \leq 1 \) then a "unlikely infection" conclusion.
5. If the value \( \mu_{PD}^{1} (P,D) \) such that \( 0 \leq \mu_{PD}^{1} (P,D) < 1 \) then a "likely infection" conclusion.

3. Developing

3.1 Software

DoctorMoon has been programmed in Borland Delphi 4.0 and run on Microsoft Windows 9x. It's easy to install and has a friendly interface.

The knowledge base of DoctorMoon is managed by a Borland Paradox Database consisting of 700 records, each represents a rule.

a. Knowledge acquisition.

The goal of this stage is to provide DoctorMoon with a brain of an experienced doctor. We used two methods of acquisition:

- Most of the rules in DoctorMoon were provided by doctors in the Vietnam National Institute of Tuberculosis and Lung Diseases (VNIITLD). We listed all the popular lung diseases symptoms (about 30 symptoms) and sorted them by their importance. This importance was determined by the doctors, e.g., how often the symptom is observed from a patient suffering from a certain lung disease. After sorting, the most important combinations of the most important symptoms were formed. This means most of popular clinical status would be considered. These combinations would be used as <Condition> in the rules. For each combination, the doctors then based on their knowledge and experiences to make conclusion about a patient's infection if he is suffering from all symptoms in the combination. This conclusion includes <Conclusion> and <Grade> of a rule.

- Rules are automatically formed. A program will browse the patient database to summarize the common syndromes that affirm or exclude a certain lung disease and then create new rules. This is done by applying suitable statistical theories as shown in [3]. A big number of rules can be created very fast this way, but rules' accuracy is not high.

![System's interface](image)

List of symptoms
changes can be made automatically or manually. Normally, this module is activated whenever a new rule is created.

3.3 A patient database

A patient database was created and has been updated. This database plays a crucial role in the automatic knowledge acquisition. Besides, it can ease the administrative work and management in VNITLD. As it stores all information about patients as well as their medical records (clinical status, doctor's diagnosis, treatment, etc.) it can effectively help to refine and improve DoctorMoon by giving comparison between diagnoses of the system and doctor.

3.4 Developing reasoning engine

DoctorMoon has a friendly interface which allows user to enter input data: the symptoms observed from patient and their severities. Once the diagnostic process takes place, all rules that match the patient i.e., rules of which the <Condition> is included in the input data will contribute to the reasoning process and then to final conclusion.

The upper left panel lists patients in the database. We can browse through this list and simultaneously see the list of all symptoms corresponding to each patient in the upper right panel. The bottom panel shows the diagnosis given to the patient by Doctor Moon

3.5 Diagnosing more than one disease

Theoretically, DoctorMoon is able to diagnose an unlimited number of diseases. The number of diseases that the system can diagnose
absolutely depends on the knowledge base. To enable DoctorMoon to diagnose a new disease, all we have to do is to make a new entry in the disease list and acquire the necessary rules to upgrade the knowledge base. So far, the system is familiar with Pulmonary Tuberculosis, Lung Abscess, Lung Cancer, Asthma, Pneumonia, Bronchiectasis.

3.6 Explaining the diagnostic results

In this medical expert system, a indispensable feature is the ability of explaining the diagnostic results: why an how the results are generated. During the diagnostic process, DoctorMoon records all the reasoning steps: getting patient’s symptoms, matching rules, diagnosing, etc., for generating a report when the diagnosis has been done.

Each patient is diagnosed for all available diseases. The diagnostic process is recorded in details in the patient’s record and stored in the patient-database. The recorded process includes as many sections as number of available diseases. In each section, we can keep track of which rules in the rule base were fired, and how the conclusion was made.

4. Testing and evaluation

DoctorMoon had undergone several tests in VNITLD. In these tests, the system was given a set of symptoms as clinical status of a patient. TUBEDIAG diagnosed that patient and returned the conclusion. The conclusion was judged by a group of experienced doctors in VNITLD to evaluate the diagnostic capability of the system.

First, DoctorMoon was given clinical status of real patients, which were provided by the patients records storage. These patients are the popular D patients. In most cases, DoctorMoon had the same conclusion as the last conclusion of doctor in the records.

Next, experts gave DoctorMoon some special combinations of symptoms as some rare, special patients. After diagnosing, DoctorMoon sometimes returned too strong affirmative or exclusive conclusion as compared to the expected conclusion given by doctors. The reason was the knowledge base was not large enough to cover most possible cases and some rules had to be corrected.

The evaluation’s result is: DoctorMoon’s diagnoses were acceptable and in order to improve system’s performance in special cases, the knowledge base needs to be strengthened. The reasoning engine is good.

5. Conclusion

DoctorMoon is a good example showing that diagnostic supporting expert system would be a good solution to solve problems occurring in medical diagnoses, especially in places which lack specialized doctors. Besides, DoctorMoon can be used as an experienced teacher in education since it has knowledge and experiences gained from a lot of high-ranking medical experts. In the next stages, we will carry out research to improve the diagnostic ability of the system: enlarge the knowledge base, improve the reasoning engine, building new module which can diagnose patient by his X-ray images, etc.

Acknowledgements

The authors are thankful to medical doctors Bui Duc Duong, Pham Tien Thinh, Nguyen Phuong Hoa and Nguyen Nhat Linh of the National Institute of Tuberculosis and Lung Diseases for their cooperation during developing and providing medical knowledge of the system.

References


NGUYEN Hoang Phuong

Nguyen Hoang Phuong was born in Hanoi, Vietnam on April 5, 1955. He received the B.S (equi. M.S.) degree in Applied Mathematics from Taskent State University. Ex. USSR in 1979 (1974-1979) and Dr. Tech. degree in Computer Science from Technical University of Vienna, Austria in 1996.

From 1979 he was with the Institute of Information Technology, National Center for Natural Science and Technology, Hanoi, Vietnam. He is now a principal researcher of the Institute of Information Technology and is teaching as a professor in Computer Science of the Hanoi Thang Long private University. Since 1999, he is a president of the Vietnamese Fuzzy Systems Society (VFSS). His research interests include fuzzy systems, soft computing, image understanding, expert systems, and multimedia intelligent communication, especially, applying soft computing techniques for integrated Western and Eastern medicine. Dr. Nguyen is a member of IFSA (International Fuzzy Systems Association (Vietnam Chapter President of IFSA)).

Hung H. DANG

He received the B.S in Computer Science from the National University of Vietnam in 1998. He has been pursuing his Master’s degree and working as research assistant in the Department of Electrical and Computer Engineering, New Mexico State University, U.S.A. His interests include: Artificial Intelligence, Control System, Digital Signal Processing and Telecommunications.

Nadipuram R. PRASAD

Nadipuram R. Prasad received the Bachelor of Engineering degree in 1966 in Electrical Engineering from Mysore University, India. In 1971, he received a Master’s degree in Electrical & Computer Science from the Massachusetts Institute of Technology. From 1971 to 1985 he worked for American Electric Power in Columbus, Ohio, where he was Manager of a Research and Development section. In 1986 he joined New Mexico State University and received a Master’s and Ph.D., both in Electrical and Computer Engineering, in 1988 and 1989, respectively. Dr. Prasad is currently an Associate Professor in Electrical Engineering where his research and teaching interests are in the area of Soft Computing. He is also the Director of the newly formed Rio Grande Institute for Soft Computing (RioSoft).