LDDS – A Fuzzy Rule Based Lung Diseases Diagnostic System
Combining Positive and Negative Knowledge

Nguyen Hoang Phuong\textsuperscript{a}, Le Huu Nghia\textsuperscript{a}, Hune Cho\textsuperscript{b}, Yun Sik Kwak\textsuperscript{b}

\textsuperscript{a} Institute of Information Technology, National Center for Natural Science and Technology, Vietnam
\textsuperscript{b} Dept of Medical Informatics, Kyungpook National University School of Medicine, Daegu, Korea

Abstract

In this paper, we present the structure of a Fuzzy Rule Based Lung Diseases Diagnostic System Combining Positive and Negative Knowledge called LDDS. This rule-based expert system has been developed using five levels of knowledge representation with fuzzy linguistic labels to express the uncertainty and the approximate reasoning combining positive and negative knowledge which act effectively with unclear information. We focus mainly on the five levels (contexts, facts, rules, modules and strategies) of knowledge representation used to represent medical entities and relationships between them for diagnosis of lung diseases and on inference mechanism combining positive knowledge and negative knowledge for medical diagnosis. Example of reference process and illustration of the implemented system LDDS are given.

Keywords:
Lung diseases, diagnostic systems, combining positive and negative knowledge.

Introduction

Lung diseases are now still popular and very dangerous diseases in many countries in the world (for example: lung abscess, collapses of respiration, pleurisy, bronchitis...). However there are many effective medicines which can treat those lung diseases, the danger is that the patients always ignore some initial symptoms of lung diseases and they only afraid when those symptoms are likely clear. But at that time, it seems to be more difficult to treat those diseases. In another case, when the patients live in countryside, there is not enough of facilities to diagnosis and treat them diseases soon and they always afraid of going to doctors in high level hospitals. So the principal cause is that patients always go to hospital later than they can and the treatments are more difficult that time.

In diagnosing lung diseases, doctors have to cope with many difficulties: the patients' symptoms are usually unclear, the similarities in some lung diseases' symptoms are difficult distinguished. Doctors always have to test many times before making a decision. The more long testing time, the more hard treatment. So the diagnosis' result depends on not only patients' symptoms but also the doctors' experiences. Wrong decision means wrong treatment and the patient would be died.

One way to deal with this problem is to build a intelligent decision support system which can mimic the reasoning of medical experienced doctors in diagnosis of lung diseases. Up to now, there are many famous expert systems in medicine such as MYCIN – a computer based medical consultation [16], INTERNIST 1 – a computer-based diagnostic consultant for general internal medicine [17], CADIAG-2 – Computer-Assisted Medical Diagnosis Using Fuzzy Subsets [15] etc have already been built and applied in clinical applications.

In this paper, we extend the features of fuzzy IF-THEN rules in TUBERDIAG – an expert system for pulmonary Tuberculosis diagnosis [4] by using the five levels of knowledge representation of MILORD (see[1]) and combining positive knowledge and negative knowledge [2] for lung diseases. (see Appendix). The organization of the paper is as the following: Second section presents a structure of the system LDDS. Third section shows a performance of the experimental implemented system. Finally, some conclusions are given.

A Structure of LDDS system

An Expert System (ES) has two main components: knowledge base (KB) and inference engine (IE). While the KB's structure depends on the problems to be solved by the ES, the same IE structure can support for many problems. But for more effective, the KB should be designed according with the IE. So it's necessary to change the structure of KB for each type of IE and for each specific type of problem to have the more accurate conclusion [12].

The usual structures of KB are frames, semantic nets and production rules. The knowledge representation of LDDS based on the KB's structure with Production Rules in the form below [1,7]:

\[
\text{IF} \ (\text{Condition}) \ \text{THEN} \ (\text{Linguistic Degree}) \ (\text{Action})
\]

where the condition can be one or several facts such as [Frequent contacts with animals] and [\textgreater \text{age 70}].

That rule can be presented in another structure:

\[
\text{IF} \ (\text{Condition}) \ \text{THEN} \ (\text{Action}) \ \text{WITH \ DEGREE} \ (\text{Linguistic Degree})
\]
The most usual action of a rule is to conclude other facts and giving to them an associated certainty value. The deduced facts can be used in the condition of other rules, so that make an inferential knowledge network which is the essence of the KB and the IE's task is using this network to make conclusion from some initial facts. In LDDS, the initial facts are the patient's symptoms and the conclusions are some diseases with their certainty values.

**Knowledge representation in LDDS**

LDDS consist of five types of objects: facts, rules, meta-rules, modules, strategies and contexts.

**Contexts**
The KB is partitioned into a set of contexts. Each context is specialized in a particular task and composed of several modules. In LDDS, the following task is considered:
- Intermediate: find out more symptoms
- Diagnostics: diagnosing diseases
- Meta: controlling the search strategy.

**Facts:**
Facts are the most elementary objects of KB. For example, they are the patients' symptoms: chest pain, age > 60, throbbing nose; or diseases: Pleurisy, Pneumothorax, ...; or some another values.

**Rules**
Rules can show the relation from facts to facts in the KB network. Which rule in LDDS is in one of two states: Positive (confirm the conclusion) and Negative (exclude the conclusion). There are two types of rules in LDDS: Production rules and Meta rules.

- **Production Rules:**
The rules used to produce new fact are called production rules. There are two types of production rules in LDDS depending on its task: Intermediate Rules, Diagnostic Rules. The destination of Intermediate rules is a symptom while the destination of Diagnostic rules is a disease.
- **Meta rules:**
Meta rules is the special type of rules used to control the reasoning process by determining which production rules are applicable and which are not [1,7]. Meta rules can be build based on the experiences of medical experts to solve the problem and can be generated from the sets of positive productive rules using some appropriate learning algorithm. So, from the input symptoms, the system can determine which diseases should be examined and which should not, then all of unwanted production rules can be reject from the system process. In addition, this representation of meta rules is also useful for the explanation.

- **Structure of rules:**
The structure of Production Rules in LDDS is:

```
IF <Premise>
THEN <Conclusion>
WITH <Degree>
```

It’s also similar with the structure below:

```
IF <Premise>
THEN <Degree> <Conclusion>
```

where "premise" can be one or several symptoms (include only positive facts, not negative facts) and "conclusion" is one symptom (in intermediate rules) or one disease (in diagnostic rules). Each fact in premise associates with another by "AND" operator. Each rule is called "fired" when all symptoms in premise are matched. So the rule means if a patient surely infects all symptoms of premise, the disease in conclusion will be found with a certainty degree equal to the rule's degree.

Positive Rules' structure is:

```
IF <Premise>
THEN CONFIRM <Conclusion>
WITH <Degree>
```

Negative Rules' structure is:

```
IF <Premise>
THEN EXCLUDE <Conclusion>
WITH <Degree>
```

The structure of Meta Rules is slightly different:

```
IF <Premise>
THEN VISIT MODULE <Name>
WITH <Degree>
```

where premise is only one symptom and the module's name is also the disease's name which its all diagnostic rules direct to. The Meta Rule means if the patient has the symptom in the premise, maybe he has been infected the disease directed by the module with the degree of rule.

**For example:** some rules in LDDS:

**Intermediate Rule 3:**
```
IF <Chest pain, High fever>
THEN CONFIRM <Possible>
INFECTED WITH <Coughing>
```

**Diagnostic Rule 19:**
```
IF <Gaunt face, High fever, Mat complexion, Oliguressis>
THEN CONFIRM WITH DEGREE <Lung abscess>
```

**Diagnostic Rule 102:**
```
IF <Coughing up purulent sputum, Yellow sputum>
THEN CONFIRM WITH DEGREE <Lung abscess>
```

**Diagnostic Rule 59:**
```
IF <Dyspnoea at supine position, Nocturnal dyspnoea>
THEN CONFIRM WITH DEGREE <Cardio pulmonary>
```

**Meta Rule 2:**
```
IF <Yellow Sputum>
```
THEN VISIT MODULE <Lung abscess>
WITH DEGREE <Possible>

**Modules**

A module is a set of rules grouped by several criteria. In DLLS we group the rules by the same conclusion. Each module has a set of meta rules that supervise the applicability of the rules in that module, so the diagnosis becomes the oriented action.

**Strategies**

A strategy is a list of modules to be evaluated and a set of modules to be rejected. Strategies are built dynamically by Meta rules at run-time depending on which meta rules can be fired. The modules of a strategy usually be sort descending by their certainty degree which evaluated by meta rules and the module with highest certainty degree will be visited first, then the other will be visited.

**Processing of uncertain information and logical inference**

**Linguistic label**

In LDDS as well as many medical expert systems, to present vague medical information, we use "fuzzy linguistic labels" which can be know as a certainty degree and takes values in [0;1]. Sometimes it is represented by a linguistic number or a linguistic label to make friendly interface. We use seven linguistic certainty values which correspond to seven ranges from 0 to 1:

- [0 ; ε ] : Impossible
- (ε ; 0.1] : Slightly Possible
- (0.1 ; 0.4] : Moderately Possible
- (0.4 ; 0.6) : Possible
- [0.6 ; 0.8) : Quite Possible
- [0.8 ; 1 - ε ) : Very Possible
- [1 - ε ; 1] : Sure.

Where “ε” is a inconsiderable value which approximate to zero.

**Defining some medical entities in LDDS**

- P : Patient
- S={S_1,S_2,...,S_m} : Set of symptoms
- D={D_1,D_2,...,D_n} : Set of diseases
- μ : degree value

- Value μ_{S_i}, taking value in [0;1], indicates the degree to which the symptom S_i is exhibited for the patient P_q:
  - μ_{S_i}=1 means symptom S_i surely present for patient P_q;
  - μ_{S_i}=0 means symptom S_i surely absent for patient P_q;
  - 0<μ_{S_i}<1 means symptom S_i possibly present for patient P_q.

- Value μ_{D_j}, taking value in [0;1], indicates the degree to which the disease D_j is confirmed for the patient P_q:
  - Value μ_{D_j}^{C_{D_j}} taking value in [0;1], indicates the degree to which the disease D_j is confirmed for the patient P_q. The combination will be described as the part below:
    - μ_{D_j}=1 means disease D_j surely present for patient P_q
    - μ_{D_j}=0 means disease D_j is unknowable with patient P_q
    - μ_{D_j}∈(0;1) means disease D_j possible present for patient P_q
    - μ_{D_j}=-1 means disease D_j surely absent for patient P_q

The target of LDDS is finding the diseases' certain degree (μ_{D_j}) from the initial patient's symptoms based on approximate reasoning which accepts fuzzy descriptions of patient's symptoms. One patient may be infected with one or some diseases and we have to find each disease's certain degree (or CF - certain Factor) to make conclusion. (see [16]).

With disease D_j, we have a positive module and a negative module which relate to D_j. Suppose that rule R_k fire and type of rule R_k is:

\[ \text{IF } <S_1, ..., S_t> \text{THEN } <D_j> \text{WITH } \mu_{R_k} \]

We have:

\[ \mu_{C_{D_j}} = \text{Min}(\mu_{S_1}, \mu_{S_2}, ..., \mu_{S_t}) \land \mu_{R_k} \]

where \land is a suitable t-norm, in LDDS, \land is a min operation.

When the positive module is visited, we have the confirmed degree:

\[ \mu_{C_{D_j}} = \mu_{1}^{C_{D_j}} \lor \mu_{2}^{C_{D_j}}, \lor, \lor, \mu_{m}^{C_{D_j}} \]

When the negative module is visited, we have the excluded degree:

\[ \mu_{E_{D_j}} = \mu_{1}^{E_{D_j}} \lor \mu_{2}^{E_{D_j}}, \lor, \lor, \mu_{n}^{E_{D_j}} \]

where \lor stands for a suitable t-conorm, one example is the Probabilistic Operator ⊕ (as we use in LDDS):

\[ x \oplus y = x + y - xy \quad \text{where } x, y \in [0;1] \]

So, we combine the confirmation and exclusion and found the certainty degree of disease D_j as following:

\[ \mu_{D_j} = \mu_{C_{D_j}} \oplus -\mu_{E_{D_j}} \]

where \oplus is an extended ordered Abelian group operation on [-1;1] (see [4]).

- Value μ_{D_j}, taking value in [-1;1], μ_{D_j} combined from μ_{C_{D_j}} and μ_{E_{D_j}}, indicates the degree to which the disease D_j is exhibited for the patient P_q. The combination will be described at the part below:
Finally we have $\mu_{D_j}$ which proclaims the degree of disease $D_j$ infected by the patient $P_q$.

$\mu_{D_j}$ take value in number from $-1$ to $1$.

a) If $\mu_{D_j} = 1$ means the disease $D_j$ is surely infected by the patient $P_q$.

b) If $0.6 < \mu_{D_j} < 1$ means the disease $D_j$ is almost possible infected by the patient $P_q$.

c) If $0.2 < \mu_{D_j} \leq 0.6$ means the disease $D_j$ is possible infected by the patient $P_q$.

d) If $-0.2 \leq \mu_{D_j} \leq 0.2$ means unknown about the disease $D_j$.

e) If $-0.6 \leq \mu_{D_j} < -0.2$ means the disease $D_j$ is possible not infected by the patient $P_q$.

f) If $-1 < \mu_{D_j} < -0.6$ means the disease $D_j$ is almost not infected by the patient $P_q$.

g) If $\mu_{D_j} = 1$ means the disease $D_j$ is surely not infected by the patient $P_q$.

Rule Checking

The conception of rule checking is very large. There are many criteria to check the rule set. In LDSS, we apply two simple types of checking:

- **Redundant rules**: two rules are redundant if they are identical, ignoring the certainty values.
- **Contradictory rules**: two rules are contradictory if they have the same If-conditions and opposite conclusions.

The checking is applied when a new rule is prepared to add to the KB. If the new rule and another rule are matched with at least one of these two types, the system will send back an error message and that new rule will not be able to add to the KB.

Diagnosing Process

**Step 1:**
Use the Intermediate Rules to find out more symptoms of the patient.

**Step 2:**
Use all appropriate rules in the Meta Rules to create strategy-oriented-diseases based on the initial symptoms. It means that a doctor consider some diseases which the patient may be infected and the doctor will make diagnosis belong to those diseases.

After this step we have a list named VISITATION which direct to all suspected diseases whose degrees are higher than a predefined threshold. This list is sorted descending on the diseases’ degree which indicates the order to visit to each disease module.

If there are no modules to evaluate in the VISITATION list, go to step 4 to make conclusion that no lung diseases are found. In other case, resume next step.

**Step 3:**
Examine each disease module of diagnostic rules in VISITATION list, use backward chaining and consider each suspected disease in VISITATION list according to the diseases module’s order list and then calculate its degrees.

In the case, if the current symptoms in working memory are not match the premise of the rule, then the system poses some questions to patient to gather more symptoms in the premise of the rule considered if necessary.

**Step 4:**
Show the result of diagnosing.

The result of diagnosing are a list of diseases that the patient may be infected. That list is sorted descending on the diseases’ degree and the first disease in list is the most suspect disease which the physician must consider about.

*Note:* In LDSS, we use a threshold (default=0) for the conclusion of diagnosis. If all obtained disease's degrees are less than the threshold, the system will list only one disease with highest disease’s degree.

Explanation

An expert system must be able to explain for what it is doing. LDSS supports two types of question: Why and How. During the processing, patient can prompt to ask "Why a question is asked?" or "How a conclusion is given?"

The "why" question is asked when the system gives to patient a question to know whether he is infected with some symptom, then LDSS will show the present suspected disease to explain.

The "how" question is asked when the system gives a conclusion, so LDSS will list all symptoms infected by the patient and all diagnostic rules which infer to the conclusion. This explanation is still simple, but we still conduct some research on this topic for further version of the system.

Implementation and the System Performance

LDSS is implemented in VB 6.0 and designed in two languages: English and Vietnamese. Here we will present for English interface in our system.

In this testing version, our knowledge base comprises 140 symptoms, more 300 rules and 19 lung diseases. The knowledge base will be expanded in future for more accurate in diagnosing. See examples in Appendix.

Knowledge acquisition

**Update symptoms and diseases:**

As in Figure 1 user can uses Add, Update, Delete and Save button to update easily the system's symptoms and diseases. He only need to choose the English name and the Vietnamese name (the disease allows user to use English or Vietnamese options). For example, in Figure 1, 16 Diseases were updated as Pleurisy, Acute Bronchitis, Bronchial Asthma ..
The information which user must input for a new production rules is premise, conclusion and rule’s degree. Premise includes one or several symptoms and user can choose from list of all symptoms and add to the premise of new rule. Conclusion of new rule can be a symptom (intermediate rules) or a disease (diagnostic rule) and user can choose from "Then" combo box. It's also with the degree. (see Figure 2). In figure 2, for example if the knowledge engineer choose the symptoms: Damp skin and Rapid pulse and Ruddy face and Rapid respiration for confirmation of the disease “Acute Bronchitis”, the rule 67: <IF Damp skin and Rapid pulse and Ruddy face and Rapid respiration > 25 times/min.> THEN CONFIRM <Possible = 0.4> INFECTED WITH <Acute Bronchitis> established and stored in the Rule Base of the system.

The initial information for diagnosing is a list of symptoms which infected by the patient. They can be chose from list of all symptoms to add to the list of patient's symptoms, for example, high fever, coughing weight losing etc. Then the user (in this case user is a medical doctor) can provide his symptoms' degree (as possible, possible and moderately .., respectively) and LDDS start to diagnose. Finally, the diagnosing result is show as Figure 3 with all diseases that the patient maybe infected and their degree sorted descending. This list of diseases is useful for the physician to choose therapeutic. The disease with maximum degree (the first disease in list) will be determined first and so on. For example, Pulmonary scierosis with “Very Possible = 0.83”, Alveolus stretch with “Very Possible = 0.82”, Chronic pneumonia with “Very Possible = 0.81” etc. (we may use the threshold to limit the number of the diagnosed diseases).

The explanations are also given and show each step of the reasoning. User can choose “Full explanation” or “Disease Explanation”. If user want to detail only one disease, he can select “Disease explanation”. Default value of explanation is Full.

Conclusions

We have presented the structure and result of the Lung Diseases Diagnosis System. The special features in LDDS are the hierarchical representation with Fuzzy linguistic labels, the positive/negative knowledge and the control structure with meta rules. In addition, LDDS is designed to diagnose for multiple diseases of Lung diseases in a single patient. Some examples to illustrate the system are given and its result indicates that the proposed reasoning handling by Meta Rules and negative rules with the hierarchical presentation can give more effective accurate diagnosis. The system is evaluated by medical doctors. The obtained results seem suitable to medical experts. But knowledge base is also checked and updated.
To improve the LDDS system, we now focus on explanation mechanism and also control of the system (meta rules i.e. the rules handle how to use the knowledge of the system).

Acknowledgements

This work is partially supported by the project CS’03, Institute of Information Technology, NCNST, Vietnam. This project is in collaboration with Dept of Medical Informatics, Kyungpook National University School of Medicine, Daegu, Korea.

References


Appendix

List of 140 symptoms

1. Increased pulmonary artery pressure
2. Headache
3. Matinal headache
4. Chest pain
5. Spumous sputum
6. Purulent sputum
7. Rusty sputum
8. White sputum
126. Coughing up sputum
127. Curved nail
128. White tongue
129. White tongue
130. Sunken Chest
131. Ruddy face
132. Damp skin
133. Increased vugular vein pressure
134. Water losing
135. Fever at evening
140. Flat half of rib-cage

List of 19 diseases

1. Lobe pneumonia
2. Bronchitis
3. Lung abscess
4. Collapses of respiration
5. Pleurisy
6. Acute Bronchitis
7. Bronchial asthma
8. Tracheal emphysema
9. Primitive lung cancer
10. Nediastinal excrescence
11. Secondary pleural cancer
12. Pleural tumour
13. Alveolus stretch
14. Pulmonary sclerosis
15. Chronic cardio pulmonary
16. Cardio pulmonary
17. Pneumothorax
18. Chronic pneumonia
19. Chronic Bronchitis

A part of rule base

DR 1:
IF  < Sudden >
THEN CONFIRM < Slightly Poss.=0.1>
INFECTED WITH < Lobe pneumonia>

DR 2:
IF  <General weakness, Shiver, High fever >
THEN CONFIRM < Moderately Poss.=0.3>
INFECTED WITH < Lobe pneumonia>

DR 3:
IF  < Coated tongue, Dry lips, Lip herpes >
THEN CONFIRM < Possible.=0.4>
INFECTED WITH < Lobe pneumonia>

DR 17:
IF  < Chest pain, High fever >
THEN CONFIRM < Moderately Poss.=0.3>
INFECTED WITH < Lung abscess>

DR 18:
IF  < Coughing up purulent sputum >
THEN CONFIRM < Slightly Poss.=0.1>
INFECTED WITH < Lung abscess>

DR 22:
IF  < Dyspnoea, Fever >
THEN CONFIRM < Possible.=0.4>
INFECTED WITH < Pleurisy>

DR 23:
IF  < Chest pain, Fever >
THEN CONFIRM < Moderately Poss.=0.3>
INFECTED WITH < Pleurisy>

DR 235:
IF  < Ruddy face, Rapid respiration>25 times/min >
THEN CONFIRM < Moderately Poss.=0.2>
INFECTED WITH < Acute Bronchitis>

DR 236:
IF  < Local Chest pain, Weak motion of chest >
THEN CONFIRM < Moderately Poss.=0.3>
INFECTED WITH < Acute Bronchitis>

DR 237:
IF  < Repatomegaly, Limbs vedema >
THEN CONFIRM < Moderately Poss.=0.2>
INFECTED WITH < Chronic Bronchitis>

Address for correspondence
Professor Nguyen Hoang Phuong, Ph.D.
Dept of Expert Systems and Soft Computing
Institute of Information Technology, National Center for Natural Science and Technology,
18 Hoang Quoc Viet Rd, CauGiay Dist., Hanoi, Vietnam.
Tel. (84)(4) 5371284, Email:nhphuong@fmail.vnn.vn