Elicitation of Time Scale-Oriented Expertise from Multiple Experts

Conference Paper · October 2007

DOI: 10.1109/ICICIC.2007.281 · Source: IEEE Xplore

2 authors:

Hui-Chun Chu
Soochow University, Taiwan
82 PUBLICATIONS **3,478** CITATIONS

SEE PROFILE

Gwo-Jen Hwang
National Taiwan University of Science and Technology
455 PUBLICATIONS **20,085** CITATIONS

SEE PROFILE

ELSEVIER

Expert Systems with Applications 31 (2006) 299–308

Expert Systems with Applications

www.elsevier.com/locate/eswa

A time scale-oriented approach for building medical expert systems☆

Gwo-Haur Hwang [a,1], Jun-Ming Chen [c,2], Gwo-Jen Hwang [b,*], Hui-Chun Chu [c,2]

[a] *Information Management Department, Ling Tung University, Taichung 40852, Taiwan, ROC*
[b] *Department of Information and Learning Technology, National University of Tainan, 33, Sec. 2, Shulin Street, Tainan 700, Taiwan, ROC*
[c] *Information Management Department, National Chi Nan University, Puli, Nantou 545, Taiwan, ROC*

Abstract

Knowledge acquisition is known to be a critical bottleneck in building expert systems. In past decades, various methods and systems have been proposed to efficiently elicit expertise from domain experts. However, in building a medical expert system, disease symptoms are usually treated as time-irrelevant attributes, such that much important information is abandoned and hence, the performance of the constructed expert systems is significantly affected. To cope with this problem, in this paper, we propose a time scale-oriented approach to eliciting medical knowledge from domain experts. The novel approach takes the time scale into consideration, such that the variant of disease symptoms in different time scales can be precisely expressed. An application to the development of a medical expert system has depicted the superiority of our approach.

Keywords: Medical expert systems; Knowledge acquisition; Artificial intelligence; Expert systems

1. Introduction

Expert systems have been applied to many problem-solving activities such as decision making, designing, planning, monitoring, diagnosing, and training activities (Feigenbaum, 1982; Liebowitz, 1997). Subject domains that are supported by experts systems include bioengineering, defense, education, engineering, finance, and medical diagnosis. Among those domains, medical diagnosis is one of the most popular and successful applications of expert systems. Dombal et al. (1972) presented the acute abdominal pain differential diagnosis system, which has been shown to be successful via an experiment on 304 cases. One year later, the MYCIN project was conducted in Stanford University, which has became a well-known medical expert system for diagnosing infectious diseases (Buchanan & Shortliffe, 1985). The success of MYCIN project has encouraged the advent of medical expert systems. In the past decades, a number of expert systems have been developed to cope with medical diagnosis problems; for example, CASNET was designed to diagnose Glaucoma (Weiss, Kulikowski, & Safir, 1978); ISODEPOR was developed to evaluate the muscle strength of Spanish top-competition athletes (Barreiro, Caraca, Fernández, López-Illescas, Montes and Olmo, 1997); FRBS-GP is a fuzzy rule-based system for diagnosing aphasia's subtypes and the classification of pap-smear examinations (Jantzen, Axer, & von Keyserlingk, 2002); neural ensemble-based detection (NED) is used to identify lung cancer cells in the images of the specimens of needle biopsies (Zhou et al., 2002). Other medical applications of expert systems include FuzzyARDS/STUDY (Steltzer, Trummer, HoÈltermann, Kolousek, Fridrich and Lewandowski, 1999), HPES (Mahaman, Harizanis, Filis, Antonopoulou, Yialouris and Sideridis, 2002) and DIARES-IPM (Mahaman, Passam, Sideridis, & Yialouris, 2003).

Those successful cases not only demonstrated the benefits of applying expert system approach to medical diagnosis problems, but also depicted the difficulty of applying it. In building an expert system, the critical bottleneck is to obtain the knowledge of the special domain from the domain experts, which is called knowledge acquisition. Although various knowledge acquisition methods have been proposed in the past decades, in building medical expert systems, disease symptoms are usually treated as time-irrelevant attributes, such that much important information is abandoned and hence, the performance of the constructed expert systems is significantly affected. To cope with this problem, we shall propose a time scale-oriented approach to eliciting medical knowledge from domain experts. The novel approach takes the time scale into

☆ This study is supported in part by the National Science Council of the Republic of China under contract numbers.
* Corresponding author. Tel.: +886 915396558; fax: +886 6 2606132.
E-mail addresses: ghhwang@mail.ltc.edu.tw (G.-H. Hwang), gjhwang@mail.nutn.edu.tw (G.-J. Hwang).
[1] Tel.: +886 915392150.
[2] Tel.: +886 910219890.

0957-4174/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2005.09.050

consideration, such that the variant of disease symptoms in different time scales can be precisely expressed. Experimental results show that our approach can achieve much better performance than conventional knowledge acquisition approach.

2. Relevant researches

In the past decades, many knowledge acquisition systems were developed to build rapid prototypes and to improve the quality of the elicited knowledge, e.g. ETS (Boose, 1984; 1985), TEIRESIAS (Davis, 1979), MORE (Kahn, Nowlan, & McDermott, 1985), SALT (Marcus, McDermott, & Wang, 1985; Marcus, 1987), NeoETS (Boose & Bradshaw, 1986; Kitto & Boose, 1986), KNACK (Klinker et al., 1987), AQUINAS (Boose & Bradshaw, 1987; Shema & Boose, 1988), KRITON (Diederich, Ruhmann, & May, 1987), Student (Gale, 1987), RuleCons (O'Bannon, 1987), MOLE (Eshelman, Ehret, McDermott, & Tan, 1987), KITTEN (Shaw & Gaines, 1987), KSSO (Gaines, 1987), ASK (Gruber, 1988), WordNet (Millar, 1990; Navigli, Velardi, & Gangemi, 2003), KADS (Wielinga, Schreiber, & Breuker, 1992; Schreiber, Wielinga, & Breuker, 1993), MCRDR (Kang, 1996), KAMET (Cairó, 1997), MedFrame/CADIAG-IV (Boegl, 1997; Kolousek, 1997; Leitich, Kiener, Kolarz, Schuh, Graninger and Adlassnig, 2001). Most of these systems were developed based on the repertory grids method originated from Kelly's personal construct theory (Kelly, 1955), which assists in identifying different objects in a domain and distinguishing among these objects.

A single repertory grid is represented as a matrix whose columns have elements labels and whose rows have construct labels. In a sense, a grid is represented for a class of objects, or individuals. In addition, the value assigned to an element–construct pair needs not be Boolean. Grid values have numeric ratings, probabilities, and other characteristics, where each value reflects the degree. After the set of constructs is ready, the expert is asked to fill the grid with ratings. A 5-scale rating mechanism is usually used in filling the grid; i.e. each rating is an integer ranging from 1 to 5, where '1' represents that the element is very likely to have the trait; '2' represents the element may have the trait; '3' represents 'unknown' or 'no relevance'; '4' represents that the element may have the opposite characteristic of the trait; '5' represents that the element is very likely to have the opposite characteristic of the trait. The whole concept of repertory grid approach can be described by the following example:

Step 1 Elicit all of the elements from the expert. The elements (e.g. E_1, E_2, E_3, E_4 and E_5) provided by the expert are placed across the top of a grid (see Table 1).

Table 1
Illustrative example of the elements in a repertory grid

	E_1	E_2	E_3	E_4	E_5

Table 2
Illustrative example of the constructs in a repertory grid

	E_1	E_2	E_3	E_4	E_5	
C_1						C_1'
C_2						C_2'
C_3						C_3'
C_4						C_4'

Table 3
Illustrative example of a repertory grid with ratings

	E_1	E_2	E_3	E_4	E_5	
C_1	5	1	5	1	1	C_1'
C_2	4	4	4	1	4	C_2'
C_3	1	4	5	1	4	C_3'
C_4	1	4	4	5	5	C_4'

Step 2 Elicit constructs (traits and their opposites) from the expert. Each time three elements are chosen to ask for a construct to distinguish one element from the other two. The constructs obtained are listed down the side of the grid. For example, if the traits C_1, C_2, C_3, C_4 and their opposites C_1', C_2', C_3', C_4' are given by the experts, the repertory grid in Table 2 will be constructed.

Step 3 Rate all of the [element, construct] entries of the grid. The values range from 5 to 1, where 1 means that the element is very likely to have trait C_i'; 2 means the element may have the trait C_i'; 3 means unknown; 4 means that the element may have the trait C_i and 5 means that the element is very likely to have the trait C_i. An illustrative example is given in Table 3.

As repertory grid approach has been widely used by researchers, some extensions have been made to enhance its representative ability. For example, Hwang extended the repertory grid technique to the fuzzy table (Hwang, 1995), in which constructs were fuzzy attributes that could be rated by means of fuzzy linguistic terms from a finite set. Jose et al. developed a technique using a fuzzy repertory grid for acquiring the finite set of attributes or variables that the expert uses in a classification problem, characterizing and discriminating a set of elements (Jose et al., 2003). Furthermore, Gonzalez and Dankel proposed a self-optimization approach based on different comparison tables for knowledge acquisition (Gonzalez & Dankel, 1993). In addition, several models have been proposed to generate more meaningful rules from the repertory grid-oriented approaches, such as the EMCUD method, which can generate embedded meanings from repertory grids by defining the impacts of the constructs to each element (Hwang & Tseng, 1990).

3. Challenges in eliciting medical diagnosis knowledge

One critical problem in applying the existing knowledge acquisition methods to elicit medical expertise is the treatment of those diseases as time-irrelevant elements. In the real world,

Table 4
Illustrative example of a repertory grid for gastrointestinal diseases

	Appendicitis	Enteritis	
Diarrhea	2	2	Not diarrhea
Lower abdominal pain	3	5	Not lower abdominal pain
Abdominal distension	4	4	Not abdominal distension
Fever	5	4	Not fever

the features of diseases are likely to change from time to time. For example, in the earlier stage of having appendicitis, the symptoms of the patients might be diarrhea, lower abdominal pain, nausea, vomit, abdominal distension, stomache and fever. Several days later, the symptoms become Teulerness R.L.Q. of abdominal, Teulerness R.L.Q to return abdominal, Leukocytosis, etc. One or two days later, these uncomfortable symptoms might gradually ease off. Sometimes, the second stage of a disease might looks similar to the first stage of other diseases; therefore, it would be improper to ask a medical expert to describe the disease symptoms without considering the time scale.

In building medical expert systems, previously proposed knowledge acquisition methods (including the repertory grid approach) only pay attentions to the relationships between diseases and symptoms; nevertheless, the variant of the symptoms in different time scales of the diseases are not taken into account. Consider the repertory grid given in Table 4, which is an example of eliciting knowledge for diagnosing various kinds of gastrointestinal diseases. Note that the rating of the (Appendicitis, Abdominal distension) entry is four, which implies highly tendency for appendicitis to have abdominal distension. However, in practical situation, influenza has significant appearance of abdominal distension in the early time scale. What has been addressed in the repertory grid is not happened in the last time scale of appendicitis. For later time scale, the abdominal distension symptom will become not so significant. Such variant of disease symptoms with respect to different time scales cannot be precisely presented by those conventional knowledge acquisition approaches.

Therefore, it would be more reasonable to express each disease by defining several time scales, such that the variant of symptoms of that disease can be clearly presented. To cope with these problems, in the following sections, a novel approach that can capture embedded meanings under time scale consideration is proposed.

4. Multi-dimensional repertory grid approach

To precisely elicit time-variant medical diagnosis knowledge from domain experts, a multi-dimension repertory grid (MDRG) is proposed in this section, which takes time scale as a new dimension in the extended repertory grid. In addition to time scale, MDRG takes importance degree for each construct to each element in different time scales into consideration, such that more embedded knowledge can be explicitly presented.

Let e_i^tdenote tth stage period of element (or disease) e_i and c_j denote a construct (or symptom), where i ranges from 1 to n, and j ranges from 1 to m. Each MDRG entry is a triplet that consists of three values: a rating to indicate the relevance of e_i^t and c_j, a certainty degree for giving the rating and an impact factor to represent the importance of c_j to e_i^t, which are represented by the following three functions:

- Rating (e_i^t, c_j): the degree of relevance for disease e_i in tth time scale to symptom c_j, ranging from 1 to 5: '1' represents that the element is very likely to have the opposite characteristic of the trait; '2' represents the element may have the opposite characteristic of the trait; '3' represents 'unknown' or 'no relevance'; '4' represents that the element may have the trait; '5' represents that the element is very likely to have the trait.
- Certainty (e_i^t, c_j): the degree of certainty for giving the rating Rating (e_i^t, c_j), which is either 'S' or 'N' representing 'sure' or 'not sure'.
- Impact_factor (e_i^t, c_j): the degree of importance for symptom c_j to disease e_i in tth time scale. Impact_factor (e_i^t, c_j) can be one of the following values:

(1) 'X': no relationship between the disease and the symptom;
(2) 'D': the symptom dominates the disease, i.e. if the value of the symptom is not matched, it is impossible for the disease to be implied;
(3) An integer, ranging from 1 to 5, to indicate that the symptom is of some degree of importance to the disease, but does not dominate the disease. That is, if the value of the symptom is not matched, it is still possible for the disease to be implied. A larger impact_factor value indicates that the symptom is more important to the disease.

In addition, a Time_Stage (e_i^t) function is used to represent the number of days for disease e_i in tth time scale. In the following subsections, we shall present the procedure of constructing a MDRG in detail.

4.1. Defining candidate diseases

Consider the diagnosis of gastrointestinal diseases. Initially, two diseases are given by the medical expert, say appendicitis and enteritis. Assume that appendicitis has two time scales, i.e. time_stage (Appendicitis, T_1) = 'within 24 hours' and time_stage (Appendicitis, T_2) = '2–3 days', while enteritis has one time scale, i.e. Time_Stage (Enteritis, T_1) = '1–2 days'. Consequently, the structure of MDRG is constructed as shown in Table 5.

4.2. Eliciting relevant symptoms

After the diseases and the corresponding time scales have been defined, the expert is asked to gives some symptoms to distinguish diseases in different time scales. An illustrative example is given in Table 6.

Table 5
Illustrative example of a MDRG structure

Disease	Appendicitis		Enteritis
Stage	T_1	T_2	T_1

Table 6
Illustrative example of symptoms given by the expert

Disease	Appendicitis		Enteritis
Stage	T_1	T_2	T_1
Diarrhea			
Lower abdominal pain			
Abdominal distension			
Fever			

4.3. Rating each entry of MDRG

Each entry of the MDRG is a triplet, consisting of a Rating (e_i^t, c_j) to indicate the relevance of the disease and the symptom, a Certainty (e_i^t, c_j) for giving the rating, and an impact_factor (e_i^t, c_j) to represent the importance of the symptom to the disease. An example of a MDRG is given in Table 7, where appendicitis in the kth time scale is recorded as T_k of appendicitis or appendicitis^k. For example, the second time scale of influenza is recorded as T_2 of influenza or influenza^2.

It can be seen that the disease in different time scales may have difference symptoms. For example, rating (Appendicitis^1, Abdominal distension) = 4 while rating (Appendicitis^2, Abdominal distension) = 1 implies that the symptom 'abdominal distension' is apparent in the first time scale but is not apparent in the second time scale of influenza.

Moreover, certainty (Enteritis^1, Fever) = 'N' indicates that the domain expert is not very certain while giving rating (Enteritis^1, Fever) = 4; certainty (Enteritis^1, Diarrhea) = 'S' indicates that the domain expert is very certain while giving rating (Enteritis^1, Diarrhea) = 4.

In addition, impact_factor (Appendicitis^1, Lower abdominal pain) = 'D' indicates that the symptom 'lower abdominal pain' dominates the disease 'appendicitis in the first time scale'; that is, if the symptom 'lower abdominal pain' is not apparent (since, rating (Appendicitis^1, Lower abdominal pain) = 5), it is impossible for the disease 'appendicitis in the first time scale' to be implied. Impact_factor (Appendicitis^1, Fever) = 4 > impact_factor (Appendicitis^1, Abdominal distension) = 2 implies that the symptom 'fever' is more important than the symptom 'abdominal distension' to the disease 'appendicitis in the first time scale'; therefore, to conclude 'the disease is appendicitis in the first time scale', the negation of the fact 'the patient has fever' will be affected the degree of certainty more than that the negation of the fact 'patient has abdominal distension'.

Table 7
Illustrative example of a MDRG

Disease	Appendicitis		Enteritis
Stage	T_1	T_2	T_1
Diarrhea	2,S,3	1,S,X	4,S,5
Lower abdominal pain	5,S,D	2,N,3	5,S,1
Abdominal distension	4,S,2	1,S,X	4,S,4
Fever	5,S,4	4,S,3	4,N,2

4.4. Generating the original and the partially matched rules

The original rules of the application domain are generated from the acquisition table, that is, the entries that are not labeled 'X' are jointed together to imply the disease in the same column. For example, in Table 7, the first time scale of appendicitis includes MDRG (Appendicitis^1, Diarrhea) = (2, S, 3), MDRG (Appendicitis^1, Lower abdominal pain) = (5, S, D), MDRG (Appendicitis^1, Abdominal distension) = (4, S, 2), MDRG (Appendicitis^1, Fever) = (5, S, 4); and hence to the following original rule is generated:

$RULE_{1,1}$: **If** (Diarrhea = true) and (Lower abdominal pain = true) and (Abdominal distension = true) and (Fever = true)
Then GOAL = Appendicitis in first time scale

The certainty factor of $RULE_{i,t}$ is determined by the certainty factor of the associated entries according to the following formula:

$CF_{i,t}$ = MIN(certainty factor of ⟨$Disease_i$, $Sympton_j$, $Time\_stage_t$⟩) for each symptom
For example, the certainty factor of $RULE_{1,1}$ is
$CF_{1,1}$ = MIN(certainty factors of ⟨Appendicitis, $Sympton_j$, 1⟩)
= MIN(mapping values of the ratings of ⟨Appendicitis, $Sympton_j$, 1⟩)
= MIN(1.0, 1.0, 1.0, 1.0) = 1.0
Consequently, $RULE_{1,1}$ can be represented as:
$RULE_{1,1}$: **If** (Diarrhea = true) and (Lower abdominal pain = true) and (Abdominal distension = true) and (Fever = true)
Then GOAL = Appendicitis in first time scale with $CF_{1,1} = 1.0$

As part of the symptoms (Diarrhea and fever) do not dominate the implication of appendicitis; therefore, it is possible that only part of the symptom values are matched, while the disease can still be implied. Such partially matched rules can be generated from the corresponding original rule by negating some symptom expressions. Consider the previous example concerning appendicitis in first time scale, the following partially matched rules are generated:

$P\_RULE_{1,1,1}$: **If** (Lower abdominal pain = true) and (Diarrhea = true) and (Abdominal distension = true) and NOT

(Fever = true)
Then GOAL = Appendicitis in first time scale
P_RULE$_{1,1,2}$: **If** (Lower abdominal pain = true) and (Diarrhea = true) and NOT (Abdominal distension = true) and (Fever = true)
Then GOAL = Appendicitis in first time scale
P_RULE$_{1,1,3}$: **If** (Lower abdominal pain = true) and (Diarrhea = true) and NOT (Abdominal distension = true) and NOT (Fever = true)
Then GOAL = Appendicitis in first time scale
P_RULE$_{1,1,4}$: **If** (Lower abdominal pain = true) and NOT (Diarrhea = true) and (Abdominal distension = true) and (Fever = true)
Then GOAL = Appendicitis in first time scale
P_RULE$_{1,1,5}$: **If** (Lower abdominal pain = true) and NOT (Diarrhea = true) and (Abdominal distension = true) and NOT (Fever = true)
Then GOAL = Appendicitis in first time scale
P_RULE$_{1,1,6}$: **If** (Lower abdominal pain = true) and NOT (Diarrhea = true) and NOT (Abdominal distension = true) and (Fever = true)
Then GOAL = Appendicitis in first time scale
P_RULE$_{1,1,7}$: **If** (Lower abdominal pain = true) and NOT (Diarrhea = true) and NOT (Abdominal distension = true) and NOT (Fever = true)
Then GOAL = Appendicitis in first time scale

4.5. Generate certainty factors for the partially matched rules

In this step, MDRG will generate certainty factors for the partially matched rules. To decide certainty factors of the partially matched rules generated from the original rule, the lower bound and the upper bound of the certainty factors must be determined first. A trivial upper bound is the certainty factor of the original rule, since it is impossible for any partially matched rule to have a greater certainty factor than its original rule. If such a situation happens, there may be something wrong in the knowledge provided by experts and backtracking is needed. We shall assume that the knowledge derived in the previous steps is correct and hence, the certainty factor of each original rule can be adopted as an upper bound. The lower bound is determined by the certainty factors for the partially matched rule with maximum certainty sequence (CS) value. The CS value for P_RULE$_{i,t,k}$ is defined as

$$\mathrm{CS}_{i,t,k} = \sum \mathrm{impact\_factor}(e_i^t, c_j)\eta$$

where $\eta = 1$, if symptom c_j is negated in the partially matched rule; 0, otherwise.

In some situation, the domain expert might reject some partially matched rules with large CS values under the circumstances that too many important symptoms are not matched. For example, it is obvious that P_RULE$_{1,1,7}$ has the maximum CS value among the partially matched rules generated from RULE$_{1,1}$. Assume that P_RULE$_{1,1,7}$ is rejected and other rules are accepted. In that case, P_RULE$_{1,1,5}$ has become the partially matched rules with maximum CS value among the accepted partially matched rules generated from RULE$_{1,1}$.

Four candidate levels suggested by the previous investigations are given, including 'confirm', 'strongly support', 'support' and 'may support', which are mapped to 1.0, 0.8, 0.6 and 0.4, respectively (Hwang & Tseng, 1990). The certainty factor of partially matched rule P_RULE$_{i,j,k}$ is then computed by using the following formula:

$$\mathrm{CF}_{i,k} = \mathrm{Upper-Bound}_i - (\mathrm{CS}_{i,t,k}/\mathrm{MAX}(\mathrm{CS}_{i,t,k})) \times (\mathrm{Upper-Bound}_i - \mathrm{Lower-Bound}_i)$$

MAX($\mathrm{CS}_{i,t,k}$): the maximum CS value of the partially matched rules generated from RULE$_{i,t}$.

The formula spreads the certainty factors of the partially matched rules in the possible range according to their CS values. Consider the previous example of RULE$_{1,1}$, the processes of deciding certainty factors for those accepted partially matched rules is given as follows:

(1) The upper bound is equal to 0.8, i.e. the certainty factor of RULE$_{1,1}$.
(2) The partially matched rule with maximum CS value is RULE$_{1,1,4}$ (since, RULE$_{1,1,1}$ is not accepted).
(3) The lower bound is determined by asking the following question:

MDRG: if the premise (Lower abdominal pain = true) and (Diarrhea = true) and (Abdominal distension = true) and (Fever = true) strongly supports that GOAL = Appendicitis in first time scale, what about the premise
(Lower abdominal pain = true) and NOT (Diarrhea = true) and NOT (Abdominal distension = true) and NOT (Fever = true) to GOAL = Appendicitis in first time scale?

1. Supports; 2. May support
EXPERT: 1.
The lower bound is then equal to 0.6.

(4) Finally, MDRG generates the certainty factors for the partially matched rules RULE$_{1,1,2}$ and RULE$_{1,1,4}$:
$\mathrm{CF}_{1,1,2} = 0.8 - (2/5) \times (0.8 - 0.6) = 0.72$
$\mathrm{CF}_{1,1,4} = 0.8 - (3/5) \times (0.8 - 0.6) = 0.68$

The final generated rules for Influenza in second time scale are shown as follows:

RULE$_{1,1}$: **If** (Lower abdominal pain = true) and (Diarrhea = true) and (Abdominal distension = true) and (Fever = true)
Then GOAL = Appendicitis in first time scale with CF = 1.0
RULE$_{1,1,2}$: **If** (Lower abdominal pain = true) and (Diarrhea = true) and NOT (Abdominal distension = true) and (Fever = true)
Then GOAL = Appendicitis in first time scale with CF = 0.72
RULE$_{1,1,4}$: **If** (Lower abdominal pain = true) and NOT (Diarrhea = true) and (Abdominal distension = true) and (Fever = true)
Then GOAL = Appendicitis in first time scale with CF = 0.68.

5. Implementation and evaluation

Based on our novel approach, a medical expert system development environment, time stage-oriented environment for medical expert systems (TOMES) has been implemented on Windows 2000 Server. It provides both knowledge acquisition and expert system execution interfaces. Fig. 1 shows the homepage of TOMES, which consists of user account authentication function, introduction of online medical diagnosis system, and the patients Guide.

Fig. 2 shows TOMES's knowledge acquisition interface. Initially, the expert is asked to enter the domain (topic) of the medical diagnosis problem. A medical problem classification menu is provided to assist the expert in defining the problem domain.

The expert is then asked to enter the name of each disease. Assume that the three diseases, 'Gastric ulcer', 'Duodenal ulcer' and 'Appendicitis', are given. Consequently, the expert is asked to define the time stages for each disease. For example, there are three time stages for 'Gastric ulcer'; that is, 'earlier stage', 'middle stage' and 'terminal stage' (as shown in Fig. 3).

After the diseases and their relevant constructs (pairs of symptoms and antonyms) are elicited, the expert is asked to entering the degree of relevance for each disease and each construct. Consequently, the degrees of importance for the symptoms to each disease are determined (see Fig. 4).

After several steps, the rules with certainty factors are generated and represented with XML format that can be accepted by DRAMA (an expert system shell developed by CoreTec co.). Fig. 5 shows the interface that presents the diagnosis result of the medical expert system based on the symptoms given by the patient. The diagnosis result includes the following messages:

(1) The diseases that might infect the patient.
(2) The stage of each possible disease.
(3) The probability of having disease and the relevant stage.
(4) The treatment to each kind of disease that the patient might be infected.

To evaluate the performance of our novel approach, two knowledge bases constructed by employing traditional repertory grid approach and the new approach, respectively. The application domain is 'the diagnosis of gastrointestinal diseases', which includes thirteen diseases: (1) gastroenteritis, (2) gastritis, (3) gastric ulcer, (4) gastroesophaged reflux disease (GERD), (5) maldiggestion, (6) malabsorption, (7) duodenal ulcer, (8) appendicitis, (9) lower drago and montoneri, (10) irritable colon, (11) ulcerative colitis, (12) hiatus hernia, and (13) amebic dysentery. Nineteen symptoms are used to identify those diseases, including 'diarrhea', 'flatulence', 'lower abdominal pain', 'nausea', 'vomit', 'abdominal distension', 'stomache', 'fever', 'Teulerness R.L.Q. of abdominal', 'Teulerness R.L.Q to return abdominal', 'abdominal pain', 'aqueous stool', 'mucinous stool', 'stool is blood-stained', 'leukocytosis', 'air belching', 'melena', 'bloody stool' and 'acid regurgitation'.

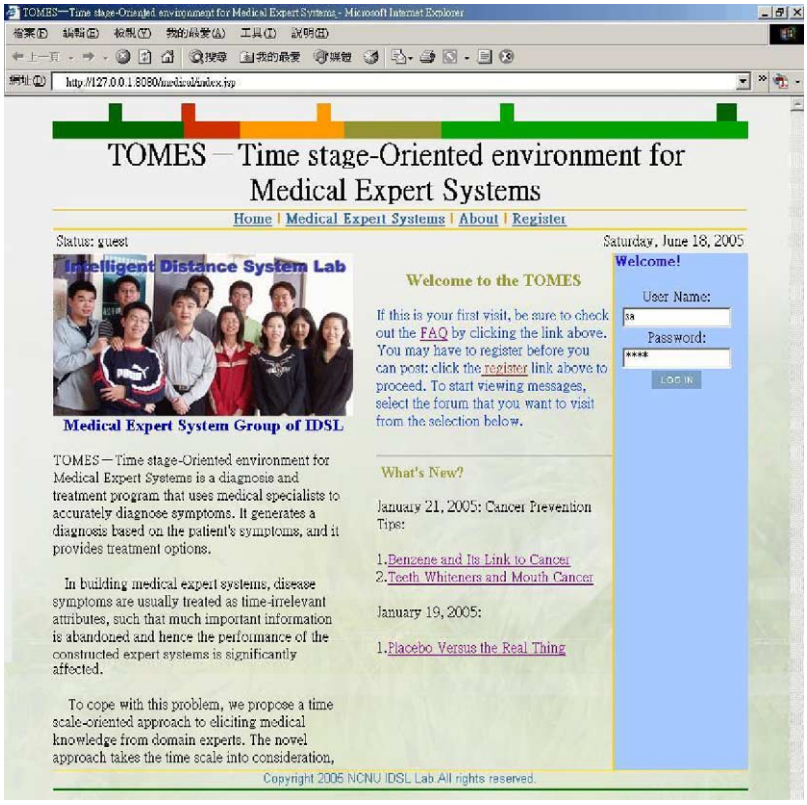

Fig. 1. Homepage of TOMES.

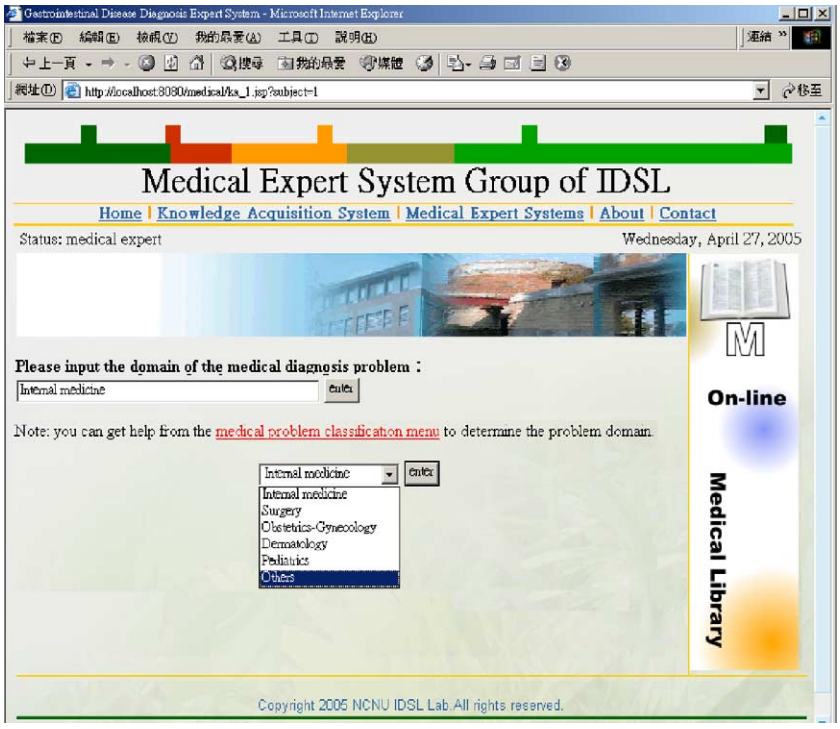

Fig. 2. User interface for defining problem domain.

Thirty-three cases given by the medical expert were used to test the performance of the constructed knowledge bases. Table 8 shows the diagnosis results given by the medical expert and the expert systems constructed by employing the repertory grid approach and the time scale-oriented approach, where an 'X' indicates that the expert system can not reach any conclusion, an integer number represent the disease number, and the '*D*–*T*' format represents the disease numbered *D* in *T*—the time scale. For example, '1–2' means 'Gastroenteritis in 2nd time scale'.

From the data given in Table 8, it can be seen that the expert system constructed with the time scale-oriented approach not

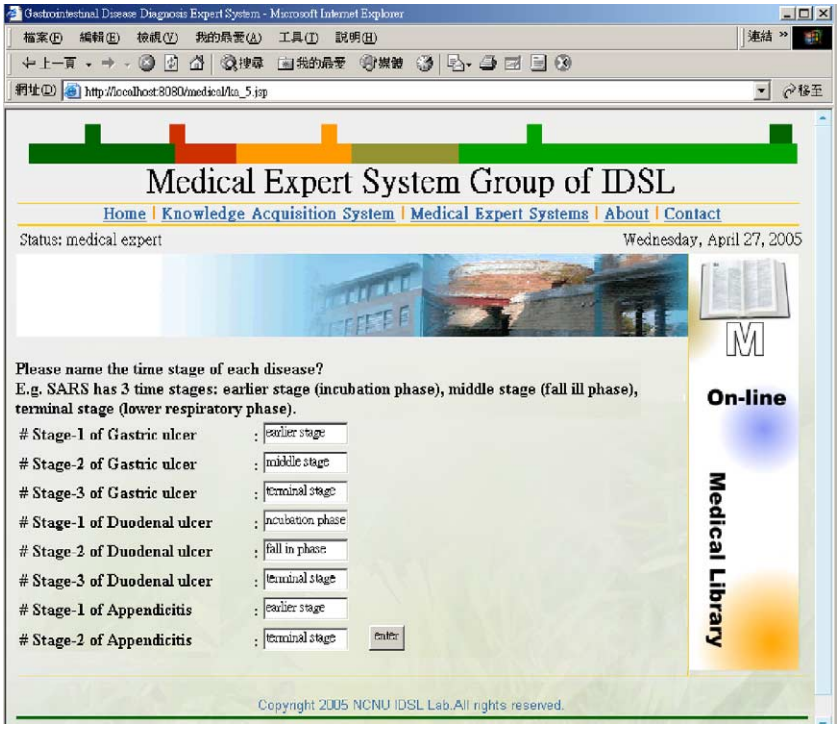

Fig. 3. User interface for entering the descriptions for each time stage.

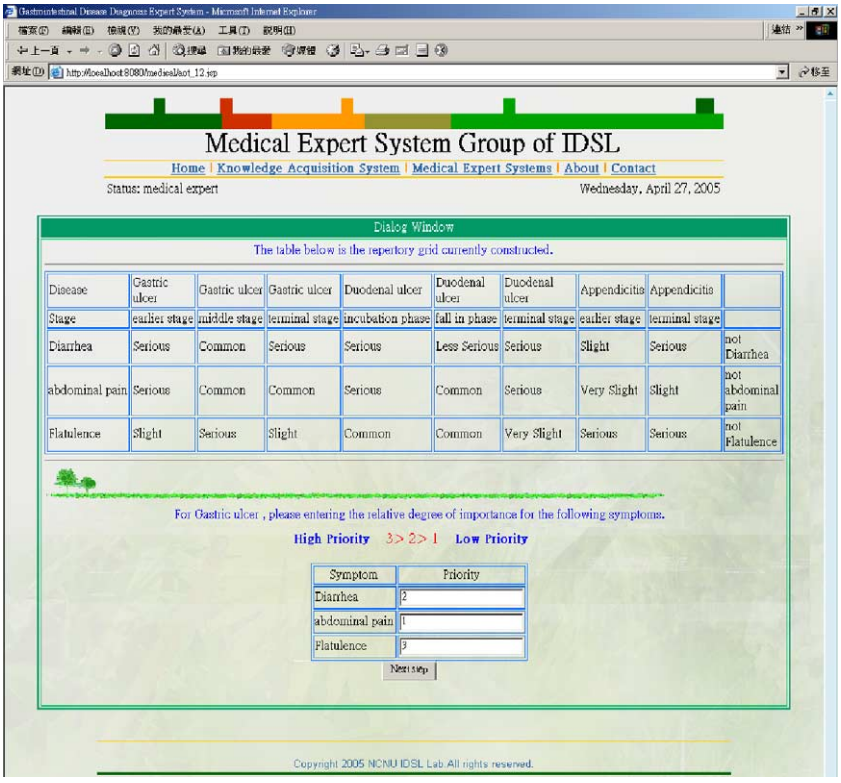

Fig. 4. User interface for comparing degrees of importance.

only can correctly diagnose the disease, but can also indicate the time scale of the disease. Nevertheless, the expert system with conventional repertory grid approach has provided several answers that are apparently different from those given by the medical expert.

Table 9 shows comparison for the correct-diagnosis rates of the two approaches. It can be seen that the performance of the time scale-oriented approach is significantly better than the traditional knowledge acquisition approach.

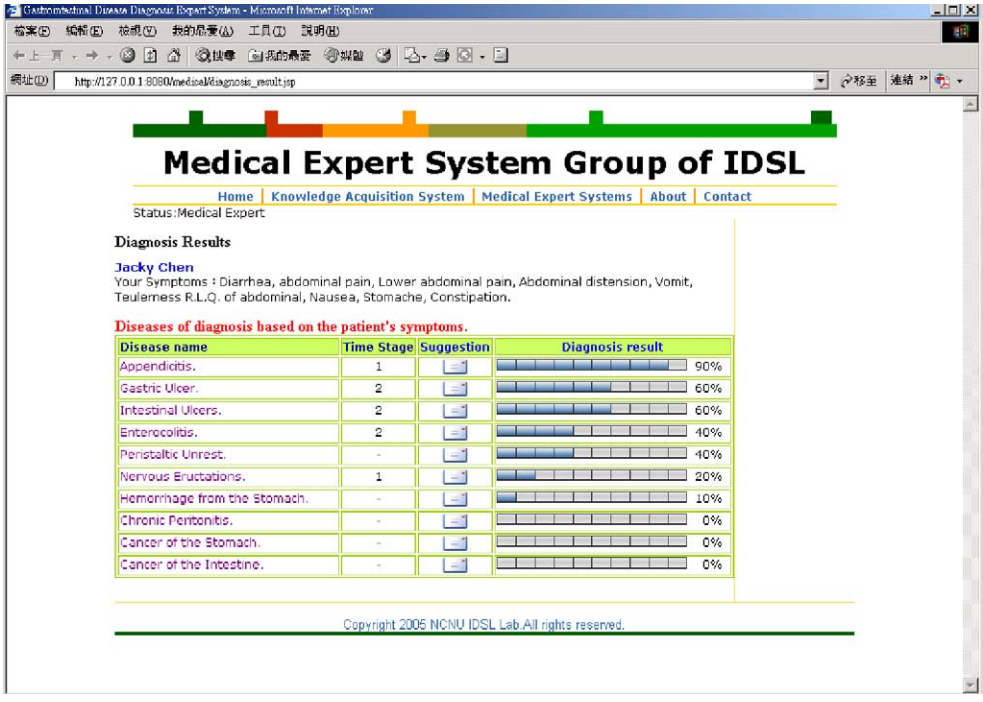

Fig. 5. User interface for presenting the diagnosis results.

Table 8
Testing results of the old and new prototypes.

Case number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
Medical expert	1	1	2	2	3	3	4	4	5	5	6	6	6	7	7	8	8	8	8	9	9	10	10	10	11	11	12	12	12	12	13	13	13
Repertory grid approach	1	10	2	X	3	7	4	X	1	5	1	6	5	4	7	1	8	X	8	9	10	11	10	9	13	11	12	12	X	12	X	9	13
Time scale-oriented approach	1-2	1-1	2-2	2-2	3-2	3-3	4-2	4-3	5-1	5-2	6-2	6-1	6-2	7-2	7-1	8-1	8-2	8-1	8-1	9-2	9-3	10-1	10-1	10-2	11-3	11-2	12-1	12-2	12-1	12-1	13-1	13-2	13-2

Table 9
Correct-diagnosis rate of each approach

	Correct-diagnosis rate (%)
Repertory grid approach	51.5
Time scale-oriented approach	100

6. Conclusion

To accurately eliciting medical diagnosis knowledge from domain experts, we propose a time scale-oriented approach, MDRG, in this paper. The novel approach attempts to take time scale into consideration while eliciting knowledge from medical experts, such that the variants of disease symptoms can be precisely expressed. In order to evaluate the performance of our approach, an experiment on a practical application was conducted. Based on the experimental results, it can be seen that MDRG can significantly improve the quality of the knowledge base. Currently, we are trying to develop and extend our approach to accumulate more clinical experiences in developing medical expert systems.

Acknowledgements

The authors are very grateful to Dr J.C. Huang for providing precious medical knowledge and conducting the experiment of this study. Part of this study is supported by the National Science Council of Taiwan, ROC, under contract numbers NSC-93-2524-S-260-001 and NSC-94-2524-S-024-003-EC3.

References

Barreiro, J. M., Caraca, J. P., Fernández, A., López-Illescas, A., Montes, C., & Olmo, J. (1997). *ISODEPOR: System for the Interpretation of Isoquinetics. Resúmenes de Comunicaciones del IV Congreso Internacional de Entrenamiento Deportivo*. León: Tecnología y Entrenamiento.

Boegl, K. (1997). *Design and implementation of a web-based knowledge acquisition toolkit for medical expert consultation systems*. Doctoral thesis, Technical University of Vienna, Austria.

Boose, J. H. (1984). Personal construct theory and the transfer of human expertise. *Proceedings of AAAI conference, California*.

Boose, J. H. (1985). A knowledge acquisition program for expert systems based on personal construct psychology. *International Journal of Man–Machine Studies*, *23*.

Boose, J. H., & Bradshaw, J. M. (1986). NeoETS: Capturing expert system knowledge in hierarchical rating grids. *IEEE expert system in government symposium*.

Boose, J. H., & Bradshaw, J. M. (1987). Expertise transfer and complex problems: Using AQUINAS as a knowledge-acquisition workbench for knowledge-based systems. *International Journal of Man–Machine Studies*, *26*, 3–27.

Buchanan, B., & Shortliffe, E. (Eds.). (1985). *Rule-based expert systems: The MYCIN experiments of the Stanford heuristic programming paper*. Reading, MA: Addison-Wesley.

Cairó, O. (1997). The KAMET methodology: A modeling approach for knowledge acquisition, EXPER SYS-97. In P. Smith (Subject Ed.) A. Niku-Lari (Series Ed.), *Technology transfer series*. IITT International.

Davis, R. (1979). Interactive transfer of expertise. *Artificial Intelligence*, *12*, 121–157.

de Dombal, F. T., Leaper, D. J., Staniland, J. R., McCann, A. P., & Horrocks, J. C. (1972). Computer-aided diagnosis of acute abdominal pain. *British Medical Journal*, *2*, 9–13.

Diederich, J., Ruhmann, I., & May, M. (1987). KRITON: A knowledge acquisition tool for expert systems. *International Journal of Man–Machine Studies*, *26*, 29–40.

Eshelman, L., Ehret, D., McDermott, J., & Tan, M. (1987). MOLE: A tenacious knowledge acquisition tool. *International Journal of Man–Machine Studies*, *26*, 41–54.

Feigenbaum, E. A. (1982). *Knowledge engineering in the 1980s*. Stanford, CA: Department of Computer Science, Stanford University.

Gaines, B. R. (1987). An overview of knowledge-acquisition and transfer. *International Journal of Man–Machine Studies*, *26*, 453–472.

Gale, W. A. (1987). Knowledge-based knowledge acquisition for a statistical consulting system. *International Journal of Man–Machine Studies*, *26*, 55–64.

Gonzalez, A. J., & Dankel, D. D. (1993). *The engineering of knowledge-based systems: Theory and practice*. Englewood Cliffs, NJ: Prentice-Hall.

Gruber, S. T. (1988). *The acquisition of strategic knowledge*. New York: Academic Press.

Hwang, G. J. (1995). Knowledge acquisition for fuzzy expert systems. *International Journal of Intelligent Systems*, *10*, 541–560.

Hwang, G. J., & Tseng, S. S. (1990). EMCUD: A knowledge acquisition method which captures embedded meanings under uncertainty. *International Journal of Man–Machine Studies*, *33*(4), 431–451.

Jantzen, J., Axer, H., & Keyserlingk, D. G. von (2002). Diagnosis of aphasia using neural and fuzzy techniques. In H.-J. Zimmermann, G. Tselentis, M. van Someren, & D. Dounias (Eds.), *Advances in computational intelligence and learning* (pp. 461–474). Massachussets: Kluwer Academic Publishers.

Jose, J., Nicholas, C. S., Jennings, R., Luo, X., & Shadbolt, N. R. (2003). Acquiring domain knowledge for negotiating agents: A case of study. *International of Human–Computer Studies*, *September*.

Kahn, G., Nowlan, S., & McDermott, J. (1985). MORE: An intelligent knowledge acquisition tool. *International Joint Conference of Artificial Intelligence* .

Kang, B. (1996). *Multiple classification ripple down rules*. PhD thesis, University of New South Wales.

Kelly, G. A. (1955). *The psychology of personal constructs* (Vol. 1). New York: W.W. Norton.

Kitto, C. M., & Boose, J. H. (1986). Heuristics for expertise transfer: The automatic management of complex knowledge acquisition dialogs. *IEEE Expert System in Government* .

Klinker, G., Bentolila, J., Genetet, S., Grimes, M., & McDermott, J. (1987). KNACK—report-driven knowledge acquisition. *International Journal of Man–Machine Studies*, *26*, 65–79.

Kolousek, G. (1997). *The system architecture of an integrated medical consultation system and its implementation based on fuzzy technology*. Doctoral thesis, Technical University of Vienna, Austria.

Leitich, H., Kiener, H. P., Kolarz, G., Schuh, C., Graninger, W., & Adlassnig, K.-P. (2001). A prospective evaluation of the medical consultation system CADIAG-II/RHEUMA in a rheumatological outpatient clinic. *Methods of Information in Medicine*, *40*, 213–220.

Liebowitz, J. (1997). Knowledge-based/expert systems technology in life support systems. *Kybernetes: The International Journal of Systems and Cybernetics*, *26*(5), 555–573.

Mahaman, B. D., Harizanis, P., Filis, I., Antonopoulou, E., Yialouris, C. P., & Sideridis, A. B. (2002). A diagnostic expert system for honeybee pests. *Computers and Electronics in Agriculture*, *36*, 17–31.

Mahaman, B. D., Passam, H. C., Sideridis, A. B., & Yialouris, C. P. (2003). DIARES-IPM: A diagnostic advisory rule-based expert system for integrated pest management in solanaceous crop systems. *Agricultural Systems*, *76*, 1119–1135.

Marcus, S. (1987). Taking backtracking with a grain of SALT. *International Journal of Man–Machine Studies*, *26*, 383–398.

Marcus, S., McDermott, J., & Wang, T. (1985). Knowledge acquisition for constructive systems. *Proceedings of the Ninth international joint conference on artificial intelligence, Los Angeles*.

Millar, A. (1990). WordNet: An on-line lexical resource. *Journal of Lexicography*, *3*(4).

Navigli, R., Velardi, P., & Gangemi, A. (2003). Ontology learning and its application to automated terminology translation. *IEEE Intelligent Systems*, *January/February*, 22–31.

O'Bannon, R. M. (1987). An intelligent aid to assist knowledge engineers with interviewing experts. *IEEE Western Conference on Expert Systems* .

Schreiber, G., Wielinga, B., & Breuker, J. (1993). *KADS: A principle approach to knowledge-based system development*. London: Academic Press.

Shaw, M. L. G., & Gaines, B. R. (1987). KITTEN: Knowledge initiation and transfer tools for experts and novices. *International Journal of Man–Machine Studies*, *27*, 251–280.

Shema, D. B., & Boose, J. H. (1988). Refining problem-solving knowledge in repertory grids using a consultation mechanism. *International Journal of Man–Machine Studies*, *29*, 447–460.

Steltzer, H., Trummer, B., HoÈltermann, W., Kolousek, G., Fridrich, P., Lewandowski, K., et al. (1999). Wissensbasierte Diagnostik und Therapieempfehlung mit Methoden der Fuzzy-Set-Theoriebei Patienten mit akutem Lungenversagen (ARDS). *AnaÈsthesiologie Intensivmedizin Notfallmedizin Schmerztherapie*, *34*, 218–223.

Weiss, S., Kulikowski, C. A., & Safir, A. (1978). Glaucoma consultation by computer. *Computers in Biology and Medicine*, *8*(1), 25–40.

Wielinga, B., Schreiber, A., & Breuker, J. (1992). KADS: A modeling approach to knowledge engineering. *Knowledge Acquisition*, *4*, 5–54.

Zhi-Hua, Zhou, Yuan, Jiang, Yu-Bin, Yang, & Shi-Fu, Chen (2002). Lung cancer cell identification based on artificial neural network ensembles. *Artificial Intelligence in Medicine*, *24*(1), 25–36.