2003 International Symposium on Advanced Intelligent Systems
September 25-28, 2003, Jeju, Korea

Development of Expert Systems using Automatic Knowledge Acquisition and Composite Knowledge Expression Mechanism

Jin Sung Kim

School of Business Administration, Jeonju University
Hyoja-Dong 3-1200, Wansan-Ku, Jeonju, Jeonbuk, 560-759, Korea
E-mail: kimjs@jeonju.ac.kr

Abstract

In this research, we propose an automatic knowledge acquisition and composite knowledge expression mechanism based on machine learning and relational database. Most of traditional approaches to develop a knowledge base and inference engine of expert systems were based on IF-THEN rules, AND-OR graph, Semantic networks, and Frame separately. However, there are some limitations such as automatic knowledge acquisition, complicate knowledge expression, expansibility of knowledge base, speed of inference, and hierarchies among rules. To overcome these limitations, many of researchers tried to develop an automatic knowledge acquisition, composite knowledge expression, and fast inference method. As a result, the adaptability of the expert systems was improved rapidly. Nonetheless, they didn't suggest a hybrid and generalized solution to support the entire process of development of expert systems.

Our proposed mechanism has five advantages empirically. First, it could extract the specific domain knowledge from incomplete database based on machine learning algorithm. Second, this mechanism could reduce the number of rules efficiently according to the rule extraction mechanism used in machine learning. Third, our proposed mechanism could expand the knowledge base unlimitedly by using relational database. Fourth, the backward inference engine developed in this study, could manipulate the knowledge base stored in relational database rapidly. Therefore, the speed of inference is faster than traditional text -oriented inference mechanism. Fifth, our composite knowledge expression mechanism could reflect the traditional knowledge expression method such as IF-THEN rules, AND-OR graph, and Relationship matrix simu ltaneously. To validate the inference ability of our system, a real data set was adopted from a clinical diagnosis classifying the dermatology disease.

Keywords: Expert systems, Composite knowledge expression, Relational database, Dermatology, Inference mechanism, Query.

1. Introduction

The purpose of this study is to develop an automatic knowledge base construction and maintaining mechanism. To this purpose, we combined automatic knowledge acquisition, relational database, and SQL-based backward inference engine. Automatic knowledge acquisition is commonly regarded as a major obstacle and bottleneck in the process of designing and implementing expert systems. Failure to acquire and encode appropriate amounts of relevant knowledge lead to limited consultation performance of the expert systems (Eriksson, 1991; Hong et al., 2002).

Knowledge used in inference could be acquired through one of two ways: either manual or automatic. Through the former researches automated knowledge acquisition tools are developed to help the knowledge engineer or even the expert himself to build and maintain the required knowledge systems (Gruber, 1987; Rafea et al., 2003). Unfortunately, however, expandability and reusability degree are still limited since no one of this tool integrates between task and domain (Allsopp et al., 2002). This issue was one of our motivations in developing our mechanism. Our proposed mechanism has six advantages as follows.

- Automatic rule extraction from the specific and incomplete database.
- Efficient rule reduction.
- Unlimited knowledge base expansion.
- Speed up of the backward inference based on SQL.
- Reflection of the traditional knowledge expression methods.
- Grow up the reusability in knowledge base.

Therefore, our proposed mechanism could cover different classification problems and applied for different empirical diagnosis. To validate the performance of our proposed mechanism, we developed the prototype expert system shell.

Then, the dermatology data set (medical diagnosis) was referred from the UCI machine learning data repository (UCI ML Grup, 2003). Because this type of medical data was very difficult to predict (Brasil et al., 2001), and they all share the clinical features of erythema and scaling with very little differencesm, the differential diagnosis of dermatopathy is a difficult problem (Rafea et al., 2003).

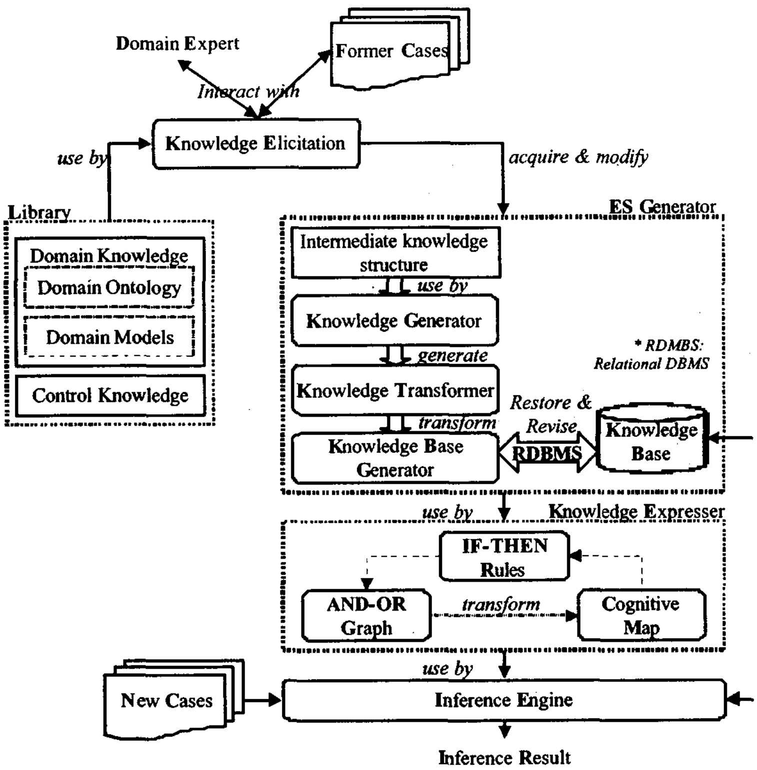


Figure 1. Research Methodology

2. Methodology

The architecture of the research methodology is graphically shown in Figure 1. The mechanism includes five main components namely: knowledge elicitation, library, ES (expert systems) generator, knowledge expresser, and inference engine. These components are similar with the research architecture of Rafea et al. (2003). In this study, however, we expanded Rafes et al. (2003)'s research architecture with other components as shown in Figure 1.

- *Library*: Contains both reusable domain knowledge and control knowledge.
- *Knowledge Elicitation*: Its main functions are to create, maintain, and restore knowledge elicited from the external input, fetch the relevant knowledge components from the library, and transform this knowledge into appropriate knowledge structure.
- *ES Generator*: Automatically generates an executable knowledge, which corresponds to the intermediate knowledge generated above. It contains knowledge generator, knowledge transformer, and knowledge base generator. During the knowledge transformation, ES Generator uses the RDBMS to restore and revise her knowledge bases.
- *Knowledge Expresser*: Support the three knowledge expression methods such as, IF-THEN rules, AND-OR graph, and Relationship matrix. It could help users to understand the knowledge base efficiently.
- *Inference Engine*: In this study, we developed SQL-based backward inference engine. Therefore, its inference speed is faster than other text-oriented inference.

3. Implementation and application

We developed the prototype system SEES (Self Evolving Expert Systems) using the Visual Basic and Microsoft Access in a Windows-XP environment. SEES was composed of five components 1) Knowledge Elicitor, 2) Library, 3) ES Generator, 4) Knowledge Expresser, and 5) Inference Engine.

3.1 Library

Dermatology database contains 34 attributes, 33 of which are linear valued and one of them is nominal. The diseases in this group are psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis, and pityriasis rubra pilaris (UCI ML Group, 2003).

Table 1. Domain knowledge for dermatopathy
(a) Clinical attributes

Clinical Attributes (takes values 0,1,2,3, unless otherwise indicated)	
Attribute #	Descriptions
v1	erythema
v2	scaling
v3	definite borders
v4	itching
v5	koebner phenomenon
v6	polygonal papules
v7	follicular papules
v8	oral mucosal involvement
v9	knee and elbow involvement
v10	scalp involvement
v11	family history *(0 or 1)*
v34	Age *(linear)*

(b) Histopathological attributes

Histopathological Attributes (take values 0,1,2,3)	
v12	melanin incontinence
v13	eosinophils in the infiltrate
v14	PNL infiltrate
v15	fibrosis of the papillary dermis
v16	Exocytosis
v17	acanthosis
v18	hyperkeratosis

v19	parakeratosis
v20	clubbing of the rete ridges
v21	elongation of the rete ridges
v22	thinning of the suprapapillary epidermis
v23	spongiform pustule
v24	munro microabcess
v25	focal hypergranulosis
v26	disappearance of the granular layer
v27	vacuolisation and damage of basal layer
v28	spongiosis
v29	saw-tooth appearance of retes
v30	follicular horn plug
v31	perifollicular parakeratosis
v32	inflammatory monoluclear inflitrate
v33	band-like infiltrate

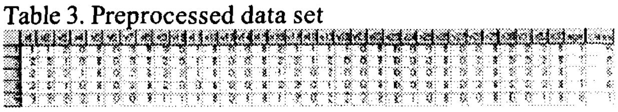
Table 2. Number of instances and classes

Value	Class	Number of instances
1	psoriasis	112
2	seboreic dermatitis	61
3	lichen planus	72
4	pityriasis rosea	49
5	cronic dermatitis	52
6	pityriasis rubra pilaris	20
	Total	366

3.2 Knowledge Elicitor

The first step to be done is to cleanse the original data so that the preprocessed data may become more traceable (Lee et al., 2002). Table 3 shows the preprocessed data set.

Table 3. Preprocessed data set



3.3 ES Generator

Machine learning algorithm such as C4.5 (Quinlan, 1988) was known to yield a set of rules. In this module, we regard a knowledge base as a domain expert system. Table 4 shows an excerpt of the derived rules. In this module we used Clementine 6.0.1 (SPSS, 2002) as a knowledge generator. The generated knowledge is encoded as production rules of the kind in Table 4. The rules are stored internally in the LISP form shown, from which the English version is generated. The tightly structured database form makes it possible for SEES to be designed to execute them in a form of SQL-based inference. The rules translated into a readable English format, as in Table 4.

Table 4. SEES production rule

PREMISE: (AND (LESS_THAN_OR_EQUAL_TO CNTXT BAND-LIKE-INFILTRATE 1)
(GREATER_THAN CNTXT FIBROSIS-OF-THE-PAPILLARY-DERMIS 1))
ACTION: (CONCLUDE CNTXT CLASS CHRONIC-DERMATITIS)

SEES' s English translation:
IF 1) the band-like infiltrate is less than or equal to 1, and
2) fibrosis of the papillary dermis is greater than 1
THEN diagnosis is chronic dermatitis.

MYCIN produces an explanation by printing out an English version of the chain of rules used. TEIRESIAS, suggested by Davis (1977, 1980), can provide more complex explanation facilities. Each of rules is constructed from a predicate function with an associative triple OAV- (object, attribute, value)- as its argument. Each premise clause typically has the following four components:

<Predicate function> / <Object> / <Attribute> / <Value>
(GREATER_THAN/CNTXT/FIBROSIS-OF-THE-PAPILLARY-DERMIS/1)

Figure 2 shows the knowledge base (rules) restored in relational database.

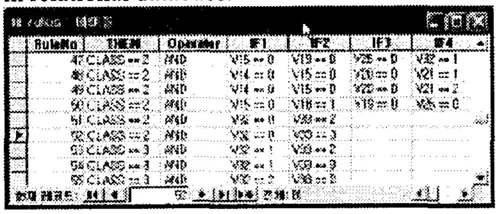


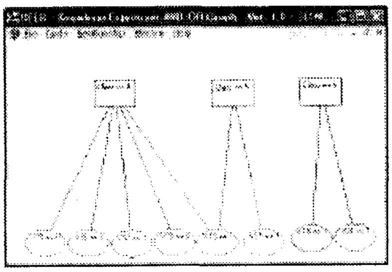
Figure 2. Knowledge base restored in relational database

3.4 Knowledge Expresser

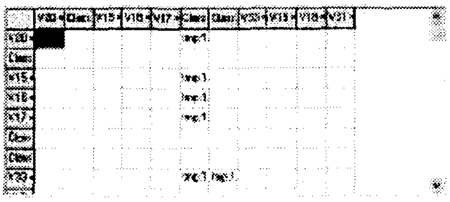
Totally 81 rules were extracted by using APRIORI and C5.0 algorithms which were developed by Agrawal (1993) and Quinlan (1993). Table 5 shows the example of IF-THEN rules. Figure 3(a) and (b) shows the AND-OR graph and Relationship matrix separately.

Table 5. SEES' s OAV type production rule (IF-THEN rules)

Rule #71
IF 1) the fibrosis of the papillary dermis is equal to 3, AND
2) band-like infiltrate is equal to 1
THEN diagnosis is cronic dermatitis
Rule #80
IF 1) the eosinophils in the infiltrate is equal to 2, AND
2) perifollicular parakeratosis is equal to 3
THEN diagnosis is pityriasis rubra pilaris



(a). AND-OR graph for dermatology



(b) Relationship matrix

Figure 3. AND-OR graph and relationship matrix for knowledge base

3.5 Inference Engine

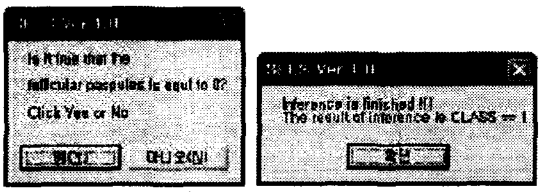
SQL-based backward inference algorithm was developed by Visual Basic. Therefore, rule consistency check and incompleteness check was easier than other traditional text-driven works.

After the construction of knowledge base, SEES ready to execute inference. In this sense, the SEES has backward inference engine, which uses SQL-based query. Table 6 shows the randomly selected patients' brief clinical data and histopathological data to validate the ability of SEES's inference engine.

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27	V28	V29	V30	V31	V32	V33	Age	Class
1	1	0	1	0	0	3	0	1	0	0	0	0	1	0	0	1	1	1	0	0	0	0	0	1	0	0	1	0	2	2	1	0	10	6
3	2	1	2	2	0	0	0	2	2	0	0	0	1	0	0	2	3	2	2	2	3	0	3	0	0	0	0	0	0	0	3	1	36	1
1	1	1	0	0	0	1	0	0	0	0	0	1	1	0	2	1	1	1	0	0	0	0	0	0	0	0	3	0	0	0	1	0	50	2

Table 6. Patients' clinical data set for validation

Figure 4 shows the whole inference process and final inference result for the 2nd patient's data set.



(a) Dialogue window

(b) Inference result (CLASS==1; psoriasis)

Figure 4. Inference result of SEES by using a knowledge base for dermatology

The dialogue window shows the variable (IF conditions) and their translated sentence in the same time to elicit the user's correct response. This dialog window works until the final inference result was fired (Figure 4(b)).

4. Conclusion

In this study, we suggested an automatic expert systems construction and maintaining mechanism The proposed mechanism consisted of the five main components Library, Knowledge Elicitation, ES Generator, Knowledge Expresser, and Inference Engine. Our mechanism was based on machine learning, relational database, and traditional backward inference algorithm, which were mainly aimed at expand the adaptability, expandability, and reusability of knowledge base. It is expected that our proposed expert system construction mechanism and expert system shell will have a significant impact on the research domain related to self-evolving expert system. Then, further research topics still remains. First, this expert system shell should be improved as an Internet-based system to support the Web-based user's decision-making. Second, other AI (artificial intelligence) technologies (such as fuzzy logic, neural networks, rough set, and etc.) may improve the inference ability of our expert system shell.

Reference

Agrawal, R., Imielinski, T., & Swami, A. (1993), Database mining: A performance perspective, *IEEE Transactions on Knowledge and Data Engineering*, 5(6), 914-925.

Allsopp, D.J., harrison, A., & Sheppard, C. (2002), A database architecture for reusable CommonKADS agents specification components, *Knowledge-Based Systems*, 15, 275-283.

Brasil, L.M., Azevedo, F.M., & Barreto, J.M. (2001), Hybrid expert system for decision supporting in the medical area: complexity and cognitive computing, *International Journal of Medical Informatics*, 63, 19-30.

Davis, R. (1980), Meta-rules: Reasoning about control, *AI Journal*, 15, 179-222.

Davis, R., & Buchanan, B. (1977), Meta-level knowledge: Overview and applications, *International Joint Conference on Artificial Intelligence (IJCAI 1977)*, 5, 920-928.

Eriksson, E. (1991), *Meta-tool support for knowledge acquisition*, PhD Thesis, Linköping University, Sweden.

Gruber, T.R. (1987), Acquiring strategic knowledge from experts, In J.H. Boose, & Gaines, B.R. (Eds.), *The foundations of knowledge acquisition*, New York: Academic Press.

Hong, T.P., Tseng, L.H., & Wang, S.L. (2002), Learning rules from incomplete training examples by rough sets, *Expert Systems with Applications*, 22, 285-293.

Lee, K.C., Kim, J.S., Chung, N.H., & Kwon, S.J. (2002), Fuzzy cognitive map approach to Web-mining inference amplification, *Expert Systems with Applications*, 22, 197-211.

Quinlan, J.R. (1993), *C4.5 - Programs for Machine Learning*, The Morgan Kaufmann series in machine learning, Morgan Kaufman Publishers, San Mateo, CA, USA.

Rafea, A., Hassen, H., & Hazman, M. (2003), Automatic knowledge acquisition tool for irrigation and fertilization expert systems, *Expert Systems with Applications*, 24, 49-57.

UC Irvine Machine Learning Group, http://www.ics.uci.edu/~mlearn/MLSummary.html, 2003.