A Development of Forward Inference Engine and Expert Systems based on Relational Database and SQL

김진성
전주대학교 경영학부

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 a mechanism to develop an inference engine and expert systems based on relational database and SQL (structured query language). Generally, former researchers had tried to develop an expert systems based on text-oriented knowledge base and backward/forward (chaining) inference engine. In these researches, however, the speed of inference was remained as a tackling point in the development of agile expert systems. Especially, the forward inference needs more times than backward inference. In addition, the size of knowledge base, complicate knowledge expression method, expansibility of knowledge base, and hierarchies among rules are the critical limitations to develop an expert systems. To overcome the limitations in speed of inference and expansibility of knowledge base, we proposed a relational database-oriented knowledge base and forward inference engine. Therefore, our proposed mechanism could manipulate the huge size of knowledge base efficiently, and inference with the large scaled knowledge base in a short time. To this purpose, we designed and developed an SQL-based forward inference engine using relational database. In the implementation process, we also developed a prototype expert system and presented a real-world validation data set collected from medical diagnosis field.

Keywords: Database, SQL, Inference Engine, Forward Inference, Expert System.

1. Introduction

Over the past 20 years, expert systems had been widely used in domains where mathematical models could not be easily built, human experts were not available or the cost of querying an expert was high. Especially, expert systems were commonly used when an inadequate algorithm or no algorithmic solution exists. Although a wide variety of expert systems have been built, knowledge acquisition and development of knowledge base are remains a development bottleneck. Generally, a knowledge engineer was needed to produce a dialog with a human expert and to explicit (encode) the knowledge elicited into a knowledge base to develop a domain expert system. However, the whole development process was very time-consuming (Buchanan & Shortliffe, 1984; Giarratano & Riley, 1989). Therefore, shortening the time in developing was then the most important factor for the successful development of an expert system (Hong et al., 2002). To this purpose, machine-learning mechanism have been adopted to ease the knowledge acquisition bottleneck. Among proposed approaches, deriving rules from training examples was the most common (Hong et al., 2000, 2001; Kodratoff & Michalski, 1983; Tsumoto, 1998).

The standard rule structure used in constructing an expert system or knowledge base was OAV (object-attribute-value) type. The structure of standard rule was ((IF condition THEN action). As an extension of standard production rule, Michalski and Winston (1986) proposed the censored production rule (CPR) of the form (<**IF** condition **THEN** action **UNLESS** censor>) as an underlying representational and computational mechanism to enable logic based systems to exhibit variable precision logic (VPL) in which certainty varies, while specificity stays constant. The form of CPR is as follows:

IF <premise> **THEN** <decision> **UNLESS** <censors>
And can be written **P → DLC**

Where P is the premise, D is the decision, and C is the censor. The premise is a conjunction of literals; the decision is a single literal; and the censor is a disjunction of literal. CPRs embody both object level and control level information (Hong et al., 2002). In addition, the method of reasoning or inference was particularly important in expert systems because inference was the basic technology by which expert systems solve problems. As a result of inference, expert system could offers several alternatives or a specific solution.

Forward inference was well known method in the area of rule-based expert systems. Which begins with a set of known facts, derives new facts using rules whose premises match the known facts, and continues this process until a goal state is reached or until no further rules have premises that match the known or derived facts (Durkin, 1994). Figure 1 shows the whole forward inference process.

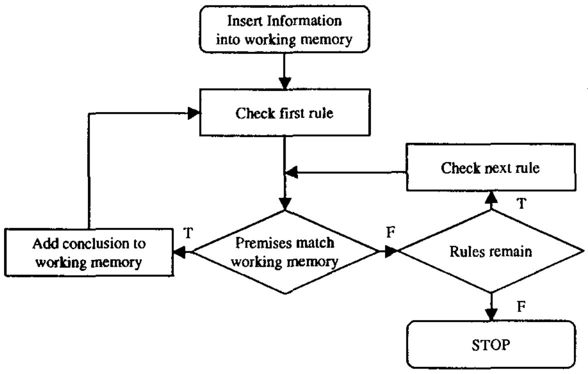


Figure 1. Forward inference process

However, the speed of inference and expandability of knowledge base were still remained as a tackling point to develop an expert system. To overcome these limitations, in this study, we developed a RDB (relational database)-based expandable knowledge base. Then, SQL-based high-speed forward inference engine (process) was implemented to inference that knowledge base effectively.

The remainder of this paper is organized as follows. The research methodology was briefly proposed in Section 2. Our prototype system was presented in Section 3. Conclusion and future work are finally given in Section 4.

2. Methodology

Our research methodology was graphically presented in Figure 2. This methodology 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 Rafea et al. (2003)'s research architecture with other components as shown in Figure 2.

- *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 forward/backward inference engine. Therefore, its inference speed is faster than other text-oriented inference.

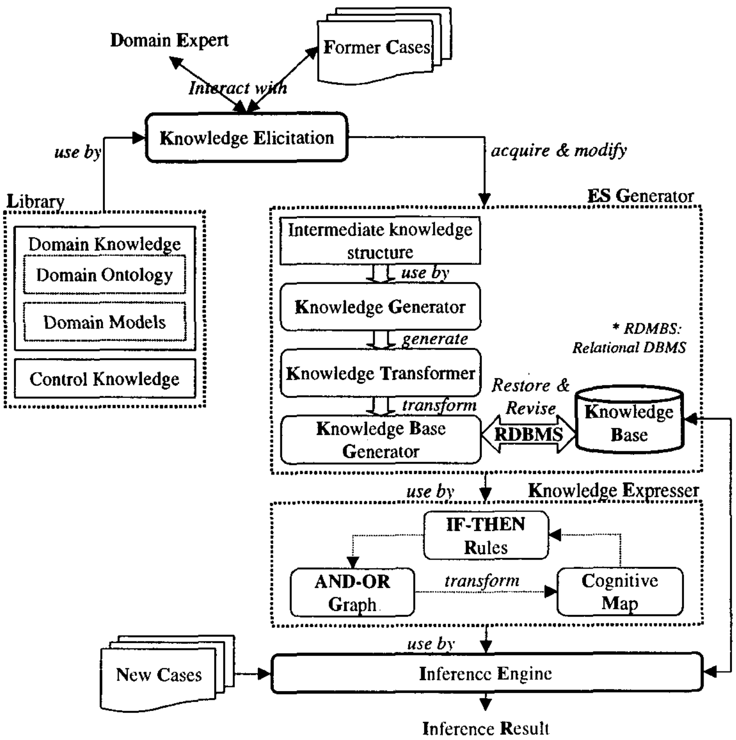


Figure 2. Research methodology

Figure 3 shows our proposed high-speed forward inference mechanism.Generally, most of traditional expert systems' inference engine use text-oriented pattern-matching mechanism to find a specific rule. When the more rules we get, therefore, the more time was spent. In the first phase, we removed all the rules, which didn't contain any facts matched with pre-inserted information.

We called this phase as *Elimination of garbage rules*. In the second phase, the forward inference was executed. In contrast with traditional forward inference mechanism, we deleted matched rules and unmatched rules during inference because these rules has no more information available after the check. Therefore, we could reduce the forward inference time effectively.

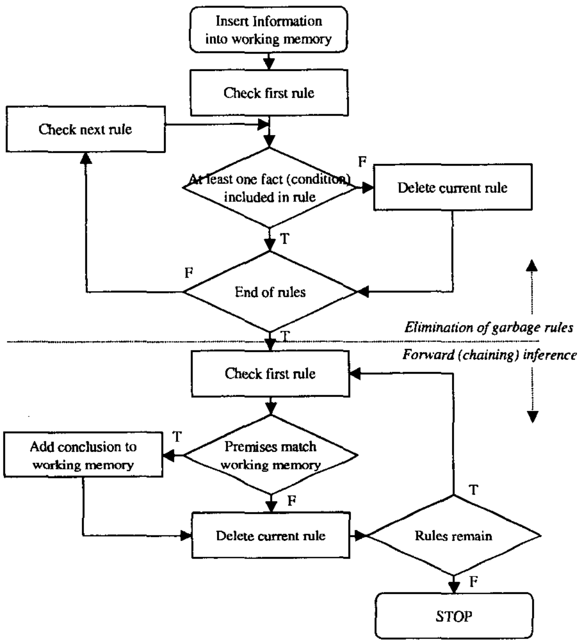


Figure 3. High-speed forward inference process

3. Implementation

To validate our proposed mechanism, we developed the prototype expert system shell SEES (Self Evolving Expert Systems) using the Visual Basic and Microsoft Access in a Windows-XP environment. Experimental data was collected from UCL Machine Learning group (2003). Which contains 366 dermatology analysis data. 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.

To extract the rules from this data set we used preprocessed data set. Table 1 shows the preprocessed data set.

Table 1. Preprocessed data set

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17
1	1	1	2	0	0	0	3	0	3	0	1	0	0	0	0	2	1
2	2	2	1	0	0	0	3	0	2	0	1	0	0	0	0	2	2
3	2	2	2	1	0	0	2	0	2	2	1	0	0	0	0	1	2
4	2	2	1	0	0	0	2	0	2	0	0	0	0	0	0	3	2
5	3	3	3	2	1	0	0	0	1	1	1	0	0	1	0	1	2

	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28	v29	v30	v31	v32	v33	age	class
1	1	1	1	0	0	0	0	0	0	0	0	0	2	2	2	0	7	6
2	1	2	0	0	0	0	0	0	0	0	2	0	3	3	2	0	7	6
3	0	2	0	0	0	0	0	0	0	0	2	0	2	2	2	0	7	6
4	0	1	0	0	0	0	0	0	0	0	3	0	2	2	2	0	7	6
5	0	2	2	2	2	2	1	0	0	0	0	0	0	0	1	0	8	1

We used the Clementine 6.0.1 (SPSS, 2002) as a knowledge generator. The generated knowledge is encoded as production rules (Table 2). 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 shown in Table 2.

Table 2. 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)

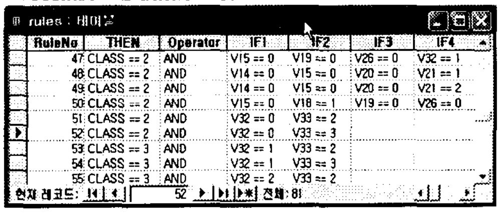
SEES's English translation:
IF 1) the band-like infiltrate is less than or equal to 1, and
2) fibrosis of the papillary dermis is greater than 1
THEN diagnosis is chronic dermatitis.

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



RuleNo	THEN	Operator	IF1	IF2	IF3	IF4
47	CLASS == 2	AND	V15 == 0	V19 == 0	V26 == 0	V32 == 1
48	CLASS == 2	AND	V14 == 0	V15 == 0	V20 == 0	V21 == 1
49	CLASS == 2	AND	V14 == 0	V15 == 0	V20 == 0	V21 == 2
50	CLASS == 2	AND	V15 == 0	V18 == 1	V19 == 0	V26 == 0
51	CLASS == 2	AND	V32 == 0	V33 == 2		
52	CLASS == 2	AND	V32 == 0	V33 == 3		
53	CLASS == 3	AND	V32 == 1	V33 == 2		
54	CLASS == 3	AND	V32 == 1	V33 == 3		
55	CLASS == 3	AND	V32 == 2	V33 == 2		

Figure 4. Knowledge base restored in relational database

Totally 81 rules were extracted by using APRIORI and C5.0 algorithms which were developed by Agrawal (1993) and Quinlan (1993). Table 3 shows the example of IF-THEN rules.

Table 3. OAV type production rule (IF-THEN rules)

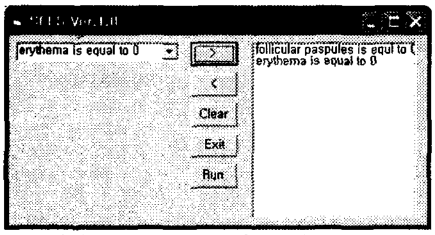
Rule #71
IF 1) the fibrosis of the papillary dermis is equal to 3, AND
2) band-like infiltrate is equal to 1
THEN diagnosis is cronic dermatitis
Rule #80
IF 1) the eosinophils in the infiltrate is equal to 2, AND
2) perifollicular parakeratosis is equal to 3
THEN diagnosis is pityriasis rubra pilaris

SQL-based forward 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 use forward inference mechanism. Table 4 shows the randomly selected patients' brief clinical data and histopathological data to validate the ability of SEES's inference engine.

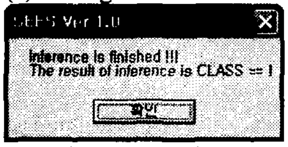
Table 4. Patients' clinical data set for validation

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	51	2

Figure 5 shows the dialogue window for information insert, and the final inference result of the 2nd patient's data.



(a) Dialogue window for information insert



(b) Inference result (CLASS==1; psoriasis)
Figure 5. Inference result of SEES

The dialogue window shows the every information possible then the user could select the information for specific patient. As a result, the system showed the final inference result using dialog window (Figure 5(b)).

4. Conclusion

In this study, we proposed an RDB-based knowledge base construction mechanism and SQL-based forward inference algorithm. The proposed mechanism was 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 high-speed forward inference algorithm, which were mainly aimed at the expandability of knowledge base, and shorten the inference time. It is expected that our proposed mechanism will have a significant impact on the research domain related to intelligent expert systems. Then, further research topics still remains. First, this expert system shell should be improved as a Web-based multiple decision-making system to support the Internet user's decision. Second, other intelligent decision support mechanism such as approximate reasoning, case-based reasoning, rough set, fuzzy logic, and etc. may improve the reasoning ability and adaptability of expert system dramatically.

References

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

[2] Durkin, J. (1994), *Expert systems: Design and development*, Prentice Hall International, Inc.

[3] 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.

[4] Hong, T.P., Wang, T.T., & Chien, B.C. (2001), Mining approximate coverage rules, *International Journal of Fuzzy Systems*, 3(2), 409-414.

[5] Hong, T.P., Wang, T.T., & Wang, S.L. (2000), Knowledge acquisition from quantitative data using the rough-set theory, *Intelligent Data Analysis*, 4, 289-304.

[6] Kodratoff, Y., & Michalski, R.S. (1983), *Machine learning: An artificial intelligence approach*, Vol.3, California: Kaufmann Publishers.

[7] Michalski, R., & Winston, P. (1986), Variable precision logic, *Artificial Intelligence*, 29, 121-145.

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

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

[10] Tsumoto, S. (1998), Extraction of experts' decision rules from clinical databases using rough set model, *Intelligent Data Analysis*, 2, 215-227.

[11] UC Irvine Machine Learning Group (2003), http://www.ics.uci.edu/~mlearn/MLSummary.html.