Fuzzy Inference in RDB using Fuzzy Classification and Fuzzy Inference Rules

김 진 성
전주대학교 경영학부

Jin Sung Kim

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

Abstract

In this paper, a framework for implementing UFIS (Unified Fuzzy rule-based knowledge Inference System) is presented. First, fuzzy clustering and fuzzy rules deal with the presence of the knowledge in DB (DataBase) and its value is presented with a value between 0 and 1. Second, RDB (Relational DB) and SQL queries provide more flexible functionality for knowledge management than the conventional non-fuzzy knowledge management systems. Therefore, the obtained fuzzy rules offer the user additional information to be added to the query with the purpose of guiding the search and improving the retrieval in knowledge base and/or rule base. The framework can be used as DM (Data Mining) and ES (Expert Systems) development and easily integrated with conventional KMS (Knowledge Management Systems) and ES.

Keywords: Expert systems, Fuzzy clustering, Fuzzy rule, Knowledge management, Relational database, SQL.

1. Introduction

Click-and-mortar firms offer products, services, message boards, reference tools, search engines, and many other specialized customer values and can be an entry point to other sites in the Internet (Afuah & Tucci, 2001; Lake, 1998). Recently, many of the firms are interested in using the web mining techniques which refer to the use of data mining (DM) techniques to improve the customer value. It helps the forms to automatically retrieve, extract and evaluate (generalize/analyze) information for knowledge discovery from web documents, services and their customers (Arotaritei & Mitra, 2004; Martin-Bautista et al., 2004). However, early research in DM field concentrated on Boolean association rules, which are concerned only with whether an item is present in a transaction or not, without considering its quantity (Agrawal et al., 1993; Agrawal & Srikant, 1994). In addition, traditional DM technologies originally have some type of uncertainty, for instance, when the boundaries of a class of objects are not sharply defined (Borgodna et al., 2000; Kacprzyk & Zadrozny, 2000a, 2000b; Veryha, 2005). The most common, useful and widely accepted solution for this problem is the introduction of fuzzy sets (Bellma & Vojdani, 2000; Blanco et al., 2000; Bosc & Pivert, 2000; Dubois et al., 2000; Zadeh, 1989). Because of the fuzzy sets provide mathematical meanings to the natural language statements and become an effective solution for dealing with uncertainty. To overcome these limitations, we suppose UFIS (Unified Fuzzy rule-based knowledge Inference System).

2. Research Methodology

In this study, we propose a unified knowledge inference system framework UFIS as shown in Figure 1. UFIS mainly consists of three modules.

Web Library:
Web library contains reusable domain knowledge including domain ontology, product profiles and customer profiles. Web administrators transfer the data which are summarized and transformed raw data into web library.

Knowledge Generator:
Its main functions are selection & preprocessing of data, fuzzy clustering, fuzzy web mining, knowledge transformation, and interaction with RDBMS to manage the knowledge base. It extracts an executable knowledge, which corresponds to the transformed data generated above it. After the extraction of knowledge it interacts with RDMBS to restore and revise her knowledge bases.

Inference Engine:
Inference engine contains UI (User Interface), SQL-based inference, and justification. Conventional ES has text-oriented inference algorithm. However, in this study, UFIS use the RDB-based SQL inference engine. The main benefit of this inference engine is that there is no need to retransformation of text knowledge into a form of executable or inferable knowledge base.

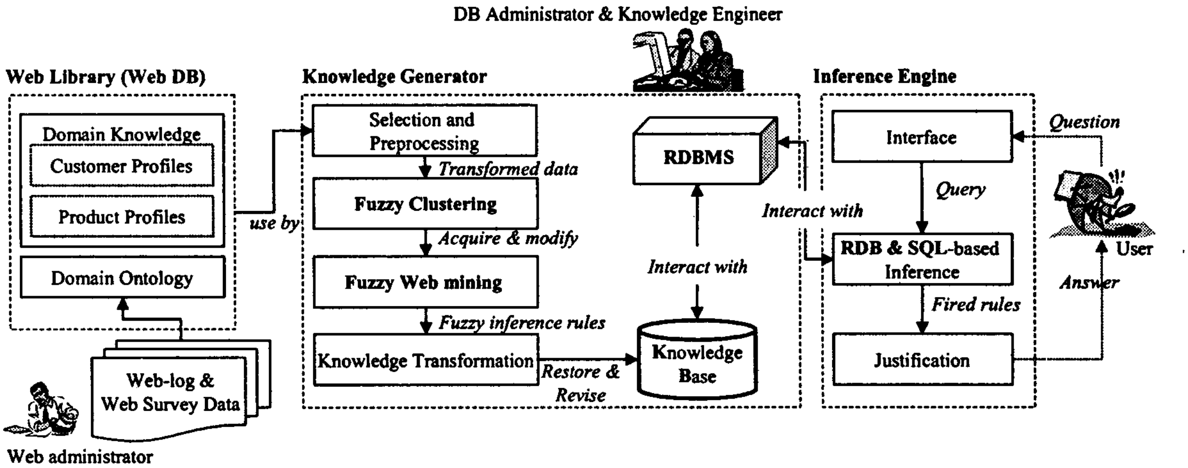


Figure 1 Research methodology

3. Implementation

3.1 Experimental data

Table 1 show the raw data used in this experiment which contains web log information, customers' profile, and web survey data expressing customers' purchasing behavior on cosmetic-related web site. Using the web resources, UFIS start to clustering and extracting the fuzzy association rules.

3.2 Fuzzy clustering

In this phase, we used FCM (Fuzzy C-Means) as a fuzzy clustering mechanism. Figure 2 shows the process and results of fuzzy clustering.

3.3 Fuzzy membership function

Traditional fuzzy membership values computed by fuzzy membership functions were divided into three categories, such as *numeric value, linguistic value, and hybrid (combination of numeric and linguistic) value*. In this study, the theory of fuzzy sets provides a mechanism for representing linguistic constructs such as '*Low*', '*Medium*', and '*High*'. Then, each linguistic construct was induced by the bell-shaped numeric fuzzy membership function π (Mitra & Pal, 1994). The fuzzy membership function π, lying in the range [0, 1], with F_j was defined as follows:

$$\pi(F_j;c,\lambda)=\begin{cases}2\left(1-\dfrac{|F_j-c|}{\lambda}\right)^2, & \text{for } \dfrac{\lambda}{2}\le |F_j-c|\le\lambda\\[2ex] 1-2\left(\dfrac{|F_j-c|}{\lambda}\right)^2, & \text{for } 0\le |F_j-c|\le\dfrac{\lambda}{2}\\[2ex] 0, & \text{otherwise}\end{cases}\qquad(1)$$

Table 1 Examples of web log & web survey data

(a) Examples of web log data

```
WebLogSample.log - 메모장
파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)
#Software: Microsoft Internet Information Services 6.0
#Version: 1.0
#Date: 2005-04-05 00:01:07
#Fields: date time s-ip cs-method cs-uri-stem cs-uri-query s-port cs-username c-ip
2005-04-05 00:01:07 202.31.241.??? GET /Default.asp - 80 - 211.63.11.60 Mozilla/4.(
2005-04-05 00:01:07 202.31.241.??? GET /blank.htm - 80 - 211.63.11.60 Mozilla/4.0+(
:
2005-04-05 01:07:10 202.31.241.??? GET /images/m_icon3.gif - 80 - 211.232.212.167 !
:
2005-04-05 02:04:34 202.31.241.??? GET /images/banner_bg.jpg - 80 - 211.230.171.36
:
2005-04-05 02:25:24 202.31.241.??? GET /data/board_skin_data/content/w_name_1.gif -
:
#Date: 2005-04-06 00:00:01
#Fields: date time s-ip cs-method cs-uri-stem cs-uri-query s-port cs-username c-ip
:
2005-04-06 00:00:10 202.31.241.??? GET /autoboard/board.asp code=qna&who=conti&dis·
:
2005-04-06 00:29:56 202.31.241.??? GET /data/banners/r_banner_20.gif - 80 - 202.31.
:
2005-04-06 00:30:01 202.31.241.??? GET /openwin_event/openwin_view.asp number=52 8(
:
2005-04-06 00:46:10 202.31.241.??? POST /autoboard/boardprocess.asp - 80 - 221.159.
Ln 1, Col 55
```

(b) Examples of web survey data

a1	a2	a3	a4	a5	b1	b2	b3	c1	c2	c3	c4	d1	d2
3	3	4	4	4	2	2	2	4	4	4	3	3	4
4	4	4	3	1	1	3	1	4	3	3	3	3	3
4	3	2	2	2	3	4	4	4	4	2	4	3	4
3	4	5	5	5	4	4	4	4	2	2	1	2	3
3	2	4	4	4	3	4	3	3	2	3	4	3	2
3	2	3	2	2	3	2	2	3	4	4	2	3	4
3	4	4	4	4	2	4	2	5	1	3	2	3	4
5	4	4	5	3	5	5	5	4	4	3	4	5	4
3	4	3	4	4	4	4	4	4	4	4	2	4	3
5	2	3	1	1	5	5	4	4	3	4	3	4	3
5	4	4	3	2	3	2	3	4	1	3	2	5	4
5	3	4	3	1	3	3	2	3	2	2	3	5	3
2	3	3	2	1	4	5	3	5	2	4	3	5	1
3	2	4	2	2	5	5	5	4	4	4	3	4	3
2	3	2	4	4	2	2	3	3	2	3	3	3	2
4	4	4	4	2	4	4	3	3	2	2	2	4	4
4	2	4	3	2	4	3	4	4	3	4	2	4	3
5	5	5	4	4	4	3	4	2	4	1	1	2	4
3	4	4	4	4	3	3	3	2	2	2	4	4	3
4	4	4	3	2	3	4	3	3	4	3	3	4	4
4	4	3	2	2	4	4	3	5	4	4	1	4	4
3	1	2	1	2	2	1	2	1	3	1	4	2	3
4	4	4	4	3	2	4	2	3	2	3	4	3	3
3	2	2	3	2	3	4	3	3	4	4	2	4	2
3	2	3	2	1	4	4	4	3	2	3	1	3	3
1	1	1	1	3	4	3	4	3	3	2	3	3	3
2	2	2	3	2	3	4	3	4	3	3	2	3	2
5	4	4	1	1	2	2	2	4	3	3	2	4	4
4	3	2	3	1	3	4	3	4	4	3	1	3	1

Where, λ > 0 is the radius of the π-function with c as the central point at which π(c; c, λ) = 1. Figure 3 shows the bell-shaped fuzzy membership functions π which has three linguistic variables *Low, Medium,* and *High*. Each factors and their values used to complete the fuzzy membership functions are shown in Table 2.

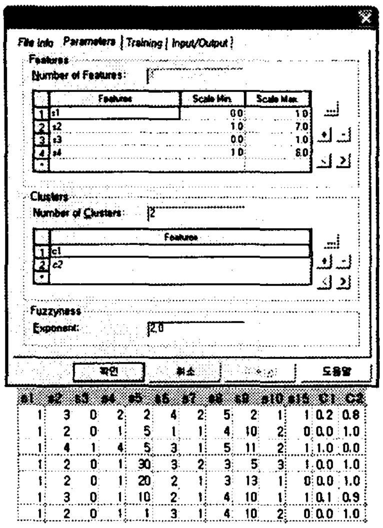
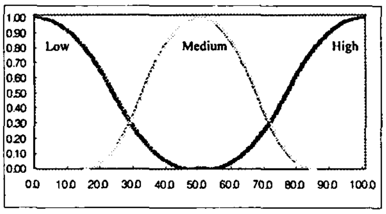
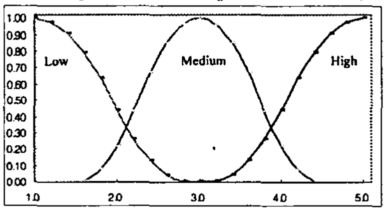


Figure 2 Result of FCM (C1: Class#1, C2: Class #2)



(a) Fuzzy membership function 1 (F_1)



(b) Fuzzy membership function 2 (F_2)

Figure 3 Fuzzy membership functions

Table 2 Value of factors used in fuzzy membership functions (F_1: function #1, F_2: function #2)

Q	L	M	H	Q	L	M	H
Center (c) or max pref.	0	50	100	Center (c) or max pref.	1	3	5
Min	0	15	50	Min	1	1.5	3
Max	50	85	100	Max	3	4.5	5
Lambda or width (λ)	47.1	35.0	47.1	Lambda or width (λ)	1.9	1.5	1.9
λ/2	23.5	17.5	23.5	λ/2	0.9	0.8	0.9

(Q: Quantity, L: Low, M: Medium, H: High)

3.4 Fuzzy Rule Extraction

In this phase, we used C5.0 which is one of well-known ML algorithms. Table 3 shows the result of rule extraction by using ML. The rules have a form as follows:

Rule *number* *predicted-value* (*Instance, Confidence*)
IF *antecedent_1*
AND *antecedent_2*
...
AND *antecedent_n*
THEN *predicted value*

Where, *Instance* means the number of records which contain the *antecedents* presented by the rule. The *Confidence* means the probability (%) and is computed as follows:

(1+*number of records where rule is correct*) / (2+*number of records for which the rule's antecedents are true*)

The *predicted-value* means the specific cosmetic product. In this study, we omitted the detailed name of which products.

Table 3 Example of fuzzy inference rules

Rule 1 CI (8, 0.20)
IF L6 = Low THEN CI
Rule 2 DF (2, 0.50)
IF L2 = Low AND L5 = Low AND L6 = Medium THEN DF
Rule 3 HR (1, 0.67)
IF L3 = Low AND L4 = High AND L5 = Medium AND L6 = Medium THEN HR
:
Rule 7 KR (5, 0.43)
IF L3 = High AND L4 = High AND L5 = Low AND L6 = Medium THEN KR
Rule 8 LG (2, 0.50)
IF L1 = High AND L4 = Low AND L6 = Medium THEN LG

(* L1: Good Advertisement, L2: Brand Image, L3: Good Design, L4: Skin Fitness, L5: Preference for Low-Price, L6: Fashion)

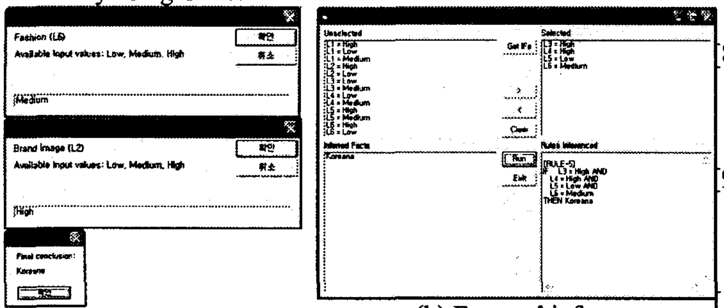
Using these fuzzy rules we examined our experimental data. As a result, which concerned to the CRM, we could find the *sustainability* and *changeability* of customers. The *changeability* means the probability of changing product from specific firm's product to another (competitive) firm's product. In contrast with *changeability*, *sustainability* means the capability of being maintained as a specific product. Table 4 shows the result of experiments.

Table 4 Result of inference (sustainability .vs. changeability)

Brand or products	Sustainability (%)		Changeability (%)	
CI	100.0		0.0	
DF	100.0		0.0	
HR	41.7		58.3	
KR	70.0	25.0	30.0	75.0
LG	50.0		50.0	
MI	66.7		33.3	
SL	33.3		66.7	

3.5 Inference in RDB

Figure 4 shows the backward inference (Figure 4(a)) and forward inference (Figure 4(b)) simultaneously by using UFIS.



(a) Backward inference (b) Forward inference

Figure 4 Inference by using RDB

4. Conclusion

In this study, we proposed unified fuzzy rule-based knowledge inference systems UFIS based on fuzzy clustering, machine learning inference rule, RDB, and SQL. The fuzzy classification and use of conventional SQL queries-based inference provide ease-to-use functionality for knowledge extraction and inference in ES. For the implementation of UFIS, the prototype based on Microsoft Visual Basic and MS-Access was developed. After the implementation and experiment with UFIS we found that the framework was effective to find the hidden knowledge from web DB and inference by using fuzzy rules, RDB and SQL.

References

[1] Agrawal, R. and Srikant, R. (1994), Fast algorithms for mining association rules, *Proceedings of the International Conference on Very Large Databases*, 487-499.

[2] Agrawal, R., Imielinski, T., & Swami, A. (1993), Mining association rules between sets of items in large databases, *Proceedings of ACM SIGMOD International Conference on Management of Data*, 207-216.

[3] Arotaritei, D. and Mitra, S. (2004), Web mining: A survey in the fuzzy framework, *Fuzzy Sets and Systems*, 148, 5-19.

[4] Bellma, M. and Vojdani, N. (2000), Fuzzy prototypes for fuzzy data mining, *Studies in Fuzziness and Soft Computing*, 39, 275-186.

[5] Blanco, I., Cubero, J., Pons, C., & Vila, A. (2000), An implementation for fuzzy deductive relational databases, *Studies in Fuzziness and Soft Computing*, 53, 183-208.

[6] Borgodna, G., Leporati, A., Lucarella, D., & Pasi, G. (2000), The fuzzy object-oriented database management system, *Studies in Fuzziness and Soft Computing*, 53, 209-236.

[7] Bosc, P. and Pivert, O. (2000), SQLf query functionality on top of a regular relational database management system, *Studies in Fuzziness and Soft Computing*, 39, 171-191.

[8] Dubois, D., Nakata, M., & Prade, H. (2000), Extended divisions for flexible queries in relational databases, *Studies in Fuzziness and Soft Computing*, 39, 105-121.

[9] Kacprzyk, J. and Zadrozny, S. (2000a), Data mining via fuzzy querying over the internet, *Studies in Fuzziness and Soft Computing*, 39, 211-233.

[10] Kacprzyk, J. and Zadrozny, S. (2000b), On combining intelligent querying and data mining using fuzzy logic concepts, *Studies in Fuzziness and Soft Computing*, 53, 67-84.

[11] Kaya, M. and Alhajj, R. (2005), Genetic algorithm based framework for mining fuzzy association rules, *Fuzzy Sets and Systems*, [online] available www.sicencedirect.com.

[12] Lake, M. (1998), *The new megasites: All-in-one Web superstores*, PCWorld.com.

[13] Martin-Bautista, M.J., Sanchez, D., Chamorro-Martinez, J., Serrano, J.M., & Vila, M.A. (2004), Mining web documents to find additional query terms using fuzzy association rules, *Fuzzy Sets and Systems*, 148, 85-104.

[14] Martin-Bautista, M.J., Sanchez, D., Chamorro-Martinez, J., Serrano, J.M., & Vila, M.A. (2004), Mining web documents to find additional query terms using fuzzy association rules, *Fuzzy Sets and Systems*, 148, 85-104.

[15] Mitra, S. & Pal, S.K. (1994), Logical Operation Based Fuzzy MLP for Classification and Rule Generation, *Neural Networks*, 7(2), 353-373.

[16] Veryha, Y. (2005), Implementation of fuzzy classification in relational databases using conventional SQL querying, *Information and Software Technology*, [online] available www.sicencedirect.com.

[17] Zadeh, L. (1989), Knowledge representation in fuzzy logic, *IEEE Transactions on Knowledge and Data Engineering*, 1, 89-100.