Prediction of User Preferred Cosmetic Brand Based on Unified Fuzzy Rule Inference

Jin Sung Kim

School of Business Administration, Jeonju University
Hyoja-Dong 3-1200, Wansan-Ku, Jeonju, Jeonbuk 560-759, South Korea
Tel: +82-63-220-2932, Fax: +82-63-220-2787, E-mail: kimjs@jj.ac.kr

Abstract

In this research, we propose a Unified Fuzzy rule-based knowledge Inference Systems (UFIS) to help the expert in cosmetic brand detection. Users' preferred cosmetic product detection is very important in the level of CRM. To this purpose, many corporations trying to develop an efficient data mining tool. In this study, we develop a prototype fuzzy rule detection and inference system. The framework used in this development is mainly based on two different mechanisms such as fuzzy rule extraction and RDB (Relational DB)-based fuzzy rule inference. First, fuzzy clustering and fuzzy rule extraction deal with the presence of the knowledge in data base and its value is presented with a value between 0 ~1. Second, RDB and SQL (Structured Query Language)-based fuzzy rule inference mechanism provide more flexibility in knowledge management than conventional non-fuzzy value-based KMS (Knowledge Management Systems).

Keywords: Cosmetic, Data mining, Expert systems, Fuzzy clustering, Fuzzy rule, Knowledge management, RDB, SQL.

1. Introduction

Internet 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). Because of the fuzzy sets provide mathematical meanings to the natural language statements and become an effective solution for dealing with uncertainty (Zadeh, 1989).

Fuzzy rule-based knowledge management and/or inference models are often used to model systems in an input/output sense by means of IF-THEN rules. It is desirable that the rule base covers all the situations of the system that are of importance for appropriate decision making. In that case, the number of rules should be kept low to increase the generalizing ability of the system, and to ensure a compact and transparent model (Setnes, 2000). In some cases, to gain a compact and transparent model and to overcome these limitations, fuzzy classification mechanism is used (Kacprzyk & Zadrozny, 2000a).

Nevertheless, the limitation comes from the size of knowledge base (or rule base) is still remained as a tackling point of the development of knowledge management systems (KMS) and ES. To resolve this problem, Veryha (2005) suggested a framework for implementing fuzzy classification in information systems using conventional SQL querying. The first main contribution of Veryha (2005) is that the fuzzy classification and use of conventional DB-based SQL queries which provide easy-to-use functionality for data extraction. Second, the approach proposed a new mechanism can be used as and effective DM tool in large information systems and easily integrated with conventional relational databases (RDB). Third, the approach has several benefits including RDB-based flexible data combination/analysis and improvement of information presentation at the report generation phase because it is based on the RDB who presenting the relationships among knowledge sets.

To improve the effectiveness of DM, with these advantages, we suppose an UFIS (Unified Fuzzy rule-based knowledge Inference Systems) based on fuzzy rule extraction and RDB-based fuzzy rule inference. Figure 1 shows the structure and components of prototype UFIS.

Inference engine contains UI (User Interface), SQL-based inferrer, and Knowledge Justifier. Most of conventional ES have 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.

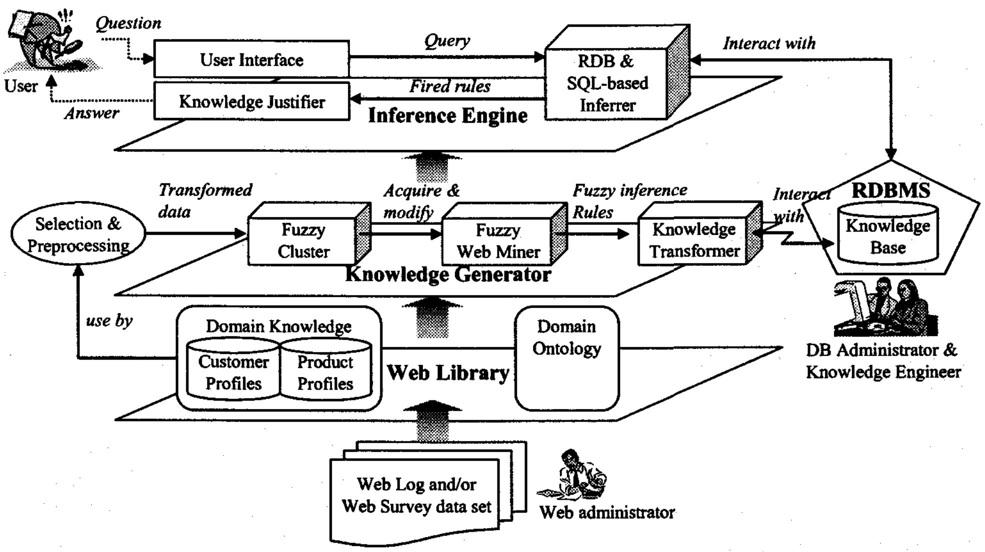


Figure 1 Prototype systems of UFIS

2. Research Methodology

Our proposed UFIS (*Unified Fuzzy rule-based knowledge Inference Systems*) mainly consists of three main modules *Web Library, Knowledge Generator,* and *Inference Engine.*

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

<u>*Knowledge Generator:*</u>
Main functions of knowledge generator are selection & preprocessing of data, fuzzy clustering, fuzzy web mining, knowledge transformation, and interaction with RDBMS to manage the knowledge base efficiently. Especially, as a fuzzy web miner, we will use the C5.0 machine learning (rule extraction) algorithm based on artificial intelligence. It can extract an executable knowledge set, 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.

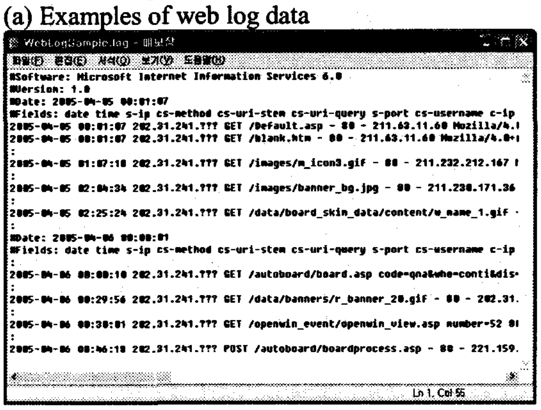
<u>*Inference Engine:*</u>

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 rules.

Table 1 Examples of web log & web survey data
(a) Examples of web log data



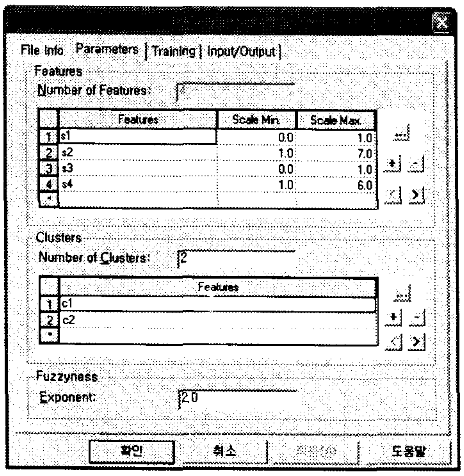
```
#Software: Microsoft Internet Information Services 6.0
#Version: 1.0
#Date: 2005-04-05 00:01:07
#Fields: date time s-ip cs-method cs-uri-stem cs-uri-query s-port cs-username c-ip
2005-04-05 00:01:07 202.31.241.??? GET /Default.asp - 80 - 211.63.11.60 Mozilla/4.
2005-04-05 00:01:07 202.31.241.??? 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&whe=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 80
:
2005-04-06 00:46:10 202.31.241.??? POST /autoboard/boardprocess.asp - 80 - 221.159.
```

(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

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.



s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s15	C1	C2
1	3	0	2	2	4	2	5	2	1	1	0.2	0.8
1	2	0	1	5	1	1	4	10	2	0	0.0	1.0
1	4	1	4	5	3	1	5	11	2	1	1.0	0.0
1	2	0	1	30	3	2	3	5	3	1	0.0	1.0
1	2	0	1	20	2	1	3	13	1	0	0.0	1.0
1	3	0	1	10	2	1	4	10	1	1	0.1	0.9
1	2	0	1	1	3	1	4	10	2	0	0.0	1.0

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

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:

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

Where, $\lambda > 0$ is the radius of the π-function with c as the central point at which $\pi(c;\, c,\, \lambda) = 1$. Each factors and their values used to complete the fuzzy membership functions are shown in Table 2.

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)

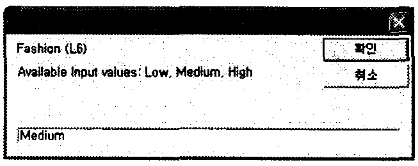
Using these fuzzy rules we examined our experimental data. As a result, which concerned to the CRM, we could find the *sustainability* and *changeability* of customers. The *changeability* means the probability of changing product from specific firm's product to another (competitive) firm's product.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

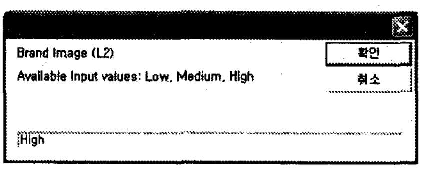
Table 4 Result of inference (Sustainability .vs. Changeability)

Brand (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	

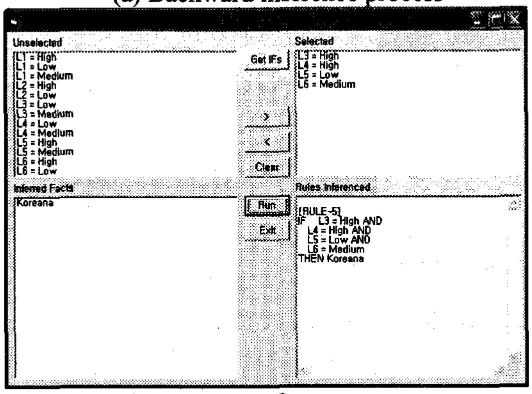
3.5 Inference in RDB

Figure 3 shows the backward inference (Figure 3(a)) and forward inference (Figure 3(b)) simultaneously by using UFIS.





(a) Backward inference process



(b) Forward inference process

Figure 3 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] Abdennadher, S. and Schuetz, H. (1998), Flexible query language, *Proceedings of Flexible Query Answering System Conference*, Roskilde, Denmark, 1-14.

[2] Afuah, A. and Tucci, C.L. (2001), *Internet business models and strategies: Text and cases*, McGraw-Hill, Inc.

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

[4] 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.

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

[6] Au, W.H. and Chang, K.C.C. (1998), An effective algorithm for discovering fuzzy rules in relational databases, *Proceedings of the IEEE International conference on Fuzzy systems*, II, 1314-1319.

[7] Bellma, M. and Vojdani, N. (2000), Fuzzy prototypes for fuzzy data mining, *Studies in Fuzziness and Soft Computing*, 39, 175-286.

[8] Bezdek, J.B., Keller, J., Krinapuram, R., & Pal, N.R. (1999), *Fuzzy models and algorithms for pattern recognition and image processing*, Kluwer Academic Publishers, Boston, MA.

[9] 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.

[10] 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.

[11] 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.

[12] Delgado, M., Marin, N., Sanchez, D., & Vila, M.A (2003), Fuzzy association rules: General model and applications, *IEEE Transactions on Fuzzy Systems*, 11, 214-225.

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

[14] Hong, T.P., Kuo, C.S., & Chi, S.C. (1999), Mining association rules from quantitative data, *Intelligent Data Analysis*, 3, 363-376.

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

[16] 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.

[17] Kaufmann, A. (1986), On the relevance of fuzzy sets for operations research, *European Journal of Operational Research*, 25(3), 330-335.

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

[19] Kraft, D.H., Martin-Bautista, M.J., Chen, J., & Vila, M.A. (2003), Rules and fuzzy rules in text: Concept, extraction and usage, *International Journal of Approximate Reasoning*, 34, 145-161.

[20] Kuok, C.M., Fu, A., & Wong, M.H. (1998), Mining fuzzy association rules in databases, *SIGMOD Record*, 27(1), 41-46.

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

[22] Lee, J.H. and Kwang, H.L. (1997), An extension of association rules using fuzzy sets, *Proceedings of the IFSA '97*, Prague, Czech Republic.

[23] 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.

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

[25] Ozer, M. (2005), Fuzzy *c*-means clustering and Internet portals: A case study, *European Journal of Operational Research*, 164, 696-714.

[26] Setnes, M. (2000), Supervised fuzzy clustering for rule extraction, *IEEE Transactions on Fuzzy Systems*, 8(4), 416-424.

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

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

[29] Zadrozny, S. and Kacprzyk, J. (1998), Implementing fuzzy querying via the internet/www: java applets active X controls and cookies, *Proceedings of Flexible Query Answering System Conference*, Roskilde, Denmark, 382-392.