한국경영과학회/대한산업공학회 2003 춘계공동학술대회
2003 년 5 월 16 일-17 일 한동대학교(포항)

연관규칙과 퍼지 인공신경망에 기반한 하이브리드 데이터마이닝 메커니즘에 관한 연구

A Study on the Hybrid Data Mining Mechanism Based on Association Rules and Fuzzy Neural Networks

김 진 성
전주대학교 경영학부 교수

Jin Sung Kim
School of Business Administration, Jeonju University
Hyoja-Dong 3-1200, Wansan-Ku, Jeonju, Jeonbuk, 560-759
kimjs@jeonju.ac.kr

Abstract

In this paper, we introduce the hybrid data mining mechanism based on association rule and fuzzy neural networks (FNN). Most of data mining mechanisms are depended on the association rule extraction algorithm. However, the basic association rule-based data mining has not the learning ability. In addition, sequential patterns of association rules could not represent the complicate fuzzy logic. To resolve these problems, we suggest the hybrid mechanism using association rule-based data mining, and fuzzy neural networks. Our hybrid data mining mechanism was consisted of four phases. First, we used general association rule mining mechanism to develop the initial rule-base. Then, in the second phase, we used the fuzzy neural networks to learn the past historical patterns embedded in the database. Third, fuzzy rule extraction algorithm was used to extract the implicit knowledge from the FNN. Fourth, we combine the association knowledge base and fuzzy rules. Our proposed hybrid data mining mechanism can reflect both association rule-based logical inference and complicate fuzzy logic.

Keywords: fuzzy neural networks, association rules, data mining, rule extraction.

1. Introduction

Data mining is one of interested topics in the field of knowledge discovery in database (Bonchi, et al., 2001; Chakrabarti et al., 1999; Changchien & Lu, 2001; Hui & Jha, 2000; Lee et al., 2002; Song et al., 2001), and has been recognized as a new area for database research. The area can be defined as efficiently discovering interesting rules from large collections of data.

The association rule extraction mechanism, which was proposed by Agrawal et al.(1993), was a most popular tools to execute the data mining. Given a set of transactions, where each transaction is a set of item, an association rule is an expression of the form X → Y. X and Y means the sets of items. An example of an association rule is: "20% of transactions that contain beer also contain diapers; 10% of all transactions contain both these items." Here 20% is called the *confidence* of the rule, and 10% the *support* of the rule.

However, one of the most critical problems with basic data mining mechanism is the lack of learning ability. In addition, it couldn't represent the fuzzy logic embedded in real world database. Combine the learning ability and the fuzzy logic with the association rule mining is very difficult for general decision makers because they require high expertise in data mining, artificial intelligence and fuzzy logic (Lee et al., 2002). In this sense, we propose a hybrid data mining mechanism based on association rule mining, fuzzy neural network, and fuzzy rule extraction algorithm. Fuzzy neural networks and fuzzy rule extraction algorithm were used to extract the implicit fuzzy knowledge from database.

2. Methodology

Our proposed hybrid data mining mechanism was based on fuzzy membership function, association rule mining and fuzzy neural networks. Which was aimed at enriching the adaptability of knowledge based to real-world case and reasoning ability. The proposed mechanism consists of the four phases-association rule extraction, fuzzy neural networks, and fuzzy rule extractions. Figure 1 shows our hybrid data mining mechanism and methodology.

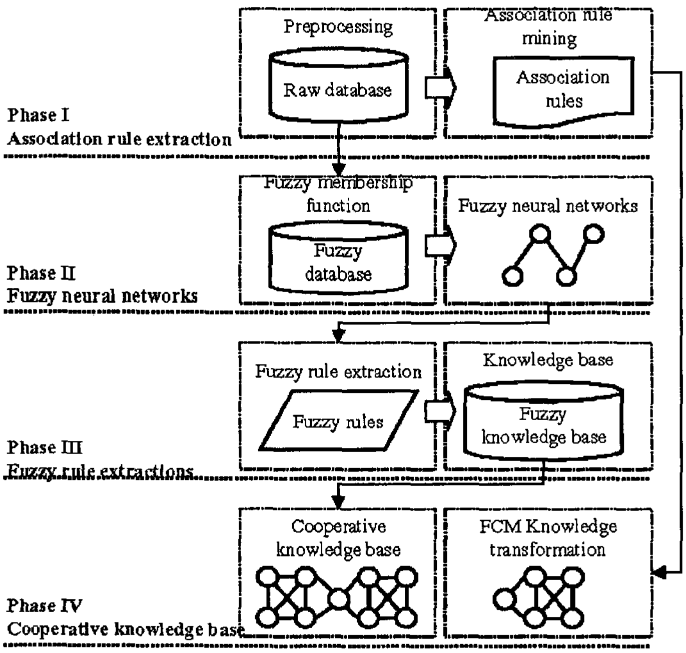

Figure 1. Research methodology

2.1 Phase I: Association rule extractions

The first phase is to preprocess the raw database and association rule mining. In this phase, we adopted the association rules mining technique to extract the relationships among items and attributes.

2.2 Phase II: Fuzzy neural networks

The second phase is to adapt the fuzzy membership function to traditional databases. As a result, raw database was transformed into fuzzy database. Then, we used the fuzzy neural networks to learn the implicit knowledge from the fuzzy database.

2.3 Phase III: Fuzzy rule extractions

The fourth stage of the proposed hybrid data mining mechanism is to apply the fuzzy rule extraction algorithm the fuzzy neural networks. Then, initial knowledge base was extended by these fuzzy rules.

2.4 Phase IV: Cooperative knowledge base

The final stage of our proposed mechanism starts with the transformation of association rules into knowledge base. Then, association rule-based knowledge base was combined and with fuzzy rules extracted from fuzzy neural networks.

3. Implementation

To prove the quality of hybrid causal knowledge base construction mechanism, we used hepatitis data stored in University of California Irvine's machine learning data repository. First, totally 155 data was selected. After the pre-processing such as missing data elimination, however, totally 80 data was used for validation. Which was composed of 19 input variables and 1 output variable (two classes 1:die, 2:live). The prototype system was implemented by using the Excel and VBA language in a Windows XP environment. In addition, SPSS and Clementine 6.0.1 was also used to preprocess the raw-data and extract the association rules. We call this prototype system as AFC (Association rule and Fuzzy neural network-based Cooperative knowledge base). Figure 2 shows the raw database for hepatitis check.

No	Class	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
1	2	30	2	1	2	2	2	2	1	2	2	2	2	2	1	85	18	4	?	1
2	2	50	1	1	2	1	2	2	1	2	2	2	2	2	0.9	135	42	3.5	?	1
3	2	78	1	2	2	1	2	2	2	2	2	2	2	2	0.7	96	32	4	?	1
4	2	31	1	?	1	2	2	2	2	2	2	2	2	2	0.7	46	52	4	80	1
5	2	34	1	2	2	2	2	2	2	2	2	2	2	2	1	?	200	4	?	1
6	2	34	1	2	2	2	2	2	2	2	2	2	2	2	0.9	95	28	4	75	1
7	1	51	1	1	2	1	2	1	2	2	1	1	2	2	?	?	?	?	?	1
8	2	23	1	2	2	2	2	2	2	2	2	2	2	2	1	?	?	?	?	1
9	2	39	1	2	2	1	2	2	2	1	2	2	2	2	0.7	?	48	4.4	?	1
10	2	30	1	2	2	2	2	2	2	2	2	2	2	2	1	?	120	3.9	?	1
11	2	39	1	1	1	2	2	2	1	1	2	2	2	2	1.3	78	30	4.4	85	1
12	2	32	1	2	1	1	2	2	2	1	2	1	2	2	1	59	249	3.7	54	1
13	2	41	1	2	1	1	2	2	2	1	2	2	2	2	0.9	81	60	3.9	52	1
14	2	30	1	2	2	1	2	2	2	1	2	2	2	2	2.2	57	144	4.9	78	1
15	2	47	1	1	1	2	2	2	2	2	2	2	2	2	?	?	60	?	?	1
16	2	38	1	1	2	1	1	1	2	2	2	2	1	2	2	72	89	2.9	46	1
17	2	66	1	2	2	1	2	2	2	2	2	2	2	2	1.2	102	53	4.3	?	1
18	2	40	1	1	2	1	2	2	2	1	2	2	2	2	0.6	62	166	4	63	1
19	2	38	1	2	2	2	2	2	2	2	2	2	2	2	0.7	53	42	4.1	85	2
20	2	38	1	1	1	2	2	2	1	1	2	2	2	2	0.7	70	28	4.2	62	1

Figure 2. Raw database for hepatitis check

3.1 Phase I: Association rule extraction

The association rule mining algorithm we adopted here is an APRIORI algorithm (Agrawal et al., 1993), which was known to yield a set of association rules. Based on the hepatitis data in Figure 2, the corresponding association rules were extracted with a threshold of 80% confidence. Figure 3 shows the association rule extraction process using Clementine.

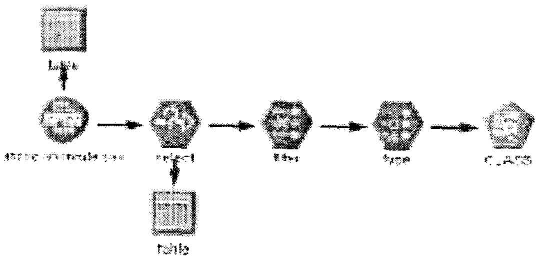

Figure 3. Association rule extraction process using Clementine

Table 1 shows an excerpt of the derived association rules. The association rules shown in Table 1 are straightforward and easy to understand and interpret.

Table 1. Example of association rules from the database

CLASS = 1 <= V9 = 2 & V8 = 2 & V3 = 1 & V11 = 2 & V20 = 2 (9:22.5%, 0.889)
CLASS = 1 <= V9 = 2 & V8 = 2 & V3 = 1 & V5 = 2 & V20 = 2 (13:32.5%, 0.846)
CLASS = 1 <= V9 = 2 & V8 = 2 & V3 = 1 & V6 = 1 & V20 = 2 (11:27.5%, 0.909)
CLASS = 1 <= V9 = 2 & V8 = 2 & V3 = 1 & V4 = 1 & V20 = 2 (7:17.5%, 1.0)
CLASS = 1 <= V9 = 2 & V8 = 2 & V3 = 1 & V7 = 1 & V20 = 2 (9:22.5%, 0.889)
CLASS = 1 <= V9 = 2 & V8 = 2 & V3 = 1 & V20 = 2 & V12 = 1 (9:22.5%, 0.889)
CLASS = 2 <= V9 = 2 & V8 = 2 & V14 = 2 & V20 = 1 & V7 = 2 (16:40.0%, 1.0)
CLASS = 2 <= V9 = 2 & V8 = 2 & V11 = 2 & V20 = 1 & V7 = 2 (16:40.0%, 1.0)
CLASS = 2 <= V9 = 2 & V8 = 2 & V13 = 2 & V20 = 1 & V7 = 2 (16:40.0%, 1.0)

3.2 Phase II: Fuzzy neural networks

In the first phase, we adapted the fuzzy membership functions to transform the real data into fuzzy sets. Fuzzy membership functions used in this phase was as follows (Mitra & Pal, 1994):

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

$$\lambda_{medium} = \frac{1}{2}(F_{\max} - F_{\min})$$

$$C_{medium} = F_{\min} + \lambda_{medium}$$

$$\lambda_{low} = \frac{1}{fdenom}(c_{medium} - F_{\min})$$

$$C_{low} = C_{medium} + 0.5 * \lambda_{low}$$

$$\lambda_{high} = \frac{1}{fdenom}(F_{\max} - C_{medium})$$

$$C_{high} = C_{medium} + 0.5 * \lambda_{high}$$

Figure 4 shows the fuzzified database transformed by fuzzy membership functions.

No	V2_L	V2_M	V2_H	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15_L	V15_M	V15_H		Class	Class
1	0.96	0.61	0.04	0.10	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.98	0.14	0.00		0.90	0.10
2	0.78	0.91	0.22	0.10	0.10	0.10	0.90	0.90	0.90	0.10	0.10	0.90	0.90	0.90	0.90	0.99	0.40	0.01		0.90	0.10
3	0.99	0.43	0.01	0.10	0.90	0.10	0.10	0.90	0.90	0.90	0.10	0.90	0.10	0.90	0.90	0.99	0.19	0.00		0.90	0.10
4	0.67	0.97	0.33	0.10	0.90	0.10	0.10	0.90	0.90	0.90	0.10	0.90	0.90	0.90	0.90	0.93	0.14	0.00		0.90	0.10
5	1.00	[illegible]	0.00	0.10	0.90	0.90	0.10	0.90	0.90	0.90	0.10	0.90	0.90	0.90	0.90	[illegible]	[illegible]	[illegible]		0.90	0.10
6	0.83	0.87	0.17	0.10	0.10	0.90	0.10	0.10	0.10	0.90	0.90	0.90	0.90	0.10	0.90	0.91	0.88	0.13		0.90	0.10
7	0.73	0.95	0.27	0.10	0.10	0.90	0.10	0.90	0.90	0.90	0.10	0.90	0.90	0.90	0.90	0.93	0.04	0.00		0.90	0.10
8	0.83	0.87	0.17	0.10	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.95	0.06	0.00		0.90	0.10
9	0.83	0.97	0.17	0.10	0.10	0.10	0.90	0.90	0.90	0.10	0.10	0.90	0.90	0.90	0.90	0.95	0.06	0.00		0.90	0.10
10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.10	0.10	[illegible]	[illegible]	[illegible]	0.90	0.90	0.90	0.90	0.90	0.90	0.16	0.00	...	0.90	0.10

Figure 4. Fuzzified database

3.3 Phase III: Fuzzy rule extractions

After the learning of fuzzy neural networks, we adopted the fuzzy rule extraction algorithm proposed by Mitra & Pal (1994) to fuzzy neural network. Fuzzy rule extraction algorithm was shown in Table 2.

Table 2. Fuzzy rule extraction algorithm

Step I: Path generation by backtracking

Step 1.1: Find the intermediate node i which has a positive effect on output node j in H(output) layer. If $w_{ji}^{H-1} > 0$, Then select node i in H-l layer

Step 1.2: Select the connection weights between i and j.

Step 1.3: Select the input node, which has an output value more than 0.5. Then, find the connection weight from the lower layer until there's no connection weight.

Step 1.4: Sort the selected connection weight list.

Step 2: Sentence generation

Adapt two conditions as follows:

Condition 1: Define the conditions for sorting. Then, generate the If-Then rules.

Condition 2: Select the linguistic hedge or real values.

Table 3 shows the fuzzy rules extracted from fuzzy neural networks. Where, each value means the fuzzy membership value.

Table 3. Sample of fuzzy rules extracted from fuzzy neural networks

CLASSS = 1<=V2_L=0.96 & V2_M=0.61 & V3=0.90 & V4=0.90 & V5=0.90 & V6=0.90 & V7=0.90 & V8=0.90 & V9=0.90 & V10=0.90 & V11=0.90 & V12=0.90 & V13=0.90 & V14=0.90 & V15_L=0.98 & V15_M=0.68 & V16_L=1.00 & V16_H=1.00 & V17_=0.92 & V17_H=0.82 & V18_L=0.98 & V18_M=0.81 & V18_H=0.87 & V19_L=0.62 & V19_M=0.73 & V19_H=0.91 (95%)

CLASSS = 1<=V2_L=0.78 & V2_M=0.91 & V5=0.90 & V6=0.90 & V7=0.90 & V8=0.90 & V9=0.90 & V10=0.90 & V11=0.90 & V12=0.90 & V13=0.90 & V14=0.90 & V15_L=0.99 & V16_L=0.99 & V16_M=0.64 & V17_=0.93 & V18_H=1.00 & V19_H=1.00 (95%)

CLASSS = 2<=V2_M=0.95 & V2_H=0.99 & V4=0.90 & V5=0.90 & V6=0.90 & V7=0.90 & V8=0.90 & V9=0.90 & V10=0.90 & V11=0.90 & V12=0.90 & V13=0.90 & V14=0.90 & V15_L=0.97 & V15_M=0.56 & V16_L=0.97 & V16_M=0.55 & V16_H=0.66 & V17_=1.00 & V17_M=0.91 & V17_H=0.79 & V18_M=1.00 & V18_H=0.55 & V19_L=1.00 & V19_M=0.88 & V19_H=0.81 & V20=0.90 (90%)

CLASSS = 2<=V2_M=0.95 & V2_H=0.73 & V4=0.90 & V5=0.90 & V6=0.90 & V7=0.90 & V8=0.90 & V9=0.90 & V10=0.90 & V11=0.90 & V13=0.90 & V15_L=0.81 & V15_M=0.88 & V16_L=1.00 & V17_=0.89 & V18_M=0.60 & V18_H=0.96 & V19_M=0.96 & V19_H=0.70 & V20=0.90 (90%)

3.4 Phase IV: Cooperative knowledge base

After the extraction of association rules and fuzzy rules, we combined two different kinds of knowledge bases into cooperative knowledge base. Table 4 shows the cooperative knowledge base.

Table 4. Example of cooperative knowledge base

CLASS = 1 <= V9 = 2 & V8 = 2 & V3 = 1 & V11 = 2 & V20 = 2 (9:22.5%, 0.889)

CLASS = 1 <= V9 = 2 & V8 = 2 & V3 = 1 & V5 = 2 & V20 = 2 (13:32.5%, 0.846)

CLASS = 1 <= V9 = 2 & V8 = 2 & V3 = 1 & V6 = 1 & V20 = 2 (11:27.5%, 0.909)

CLASS = 2 <= V9 = 2 & V8 = 2 & V14 = 2 & V20 = 1 & V7 = 2 (16:40.0%, 1.0)

CLASS = 2 <= V9 = 2 & V8 = 2 & V11 = 2 & V20 = 1 & V7 = 2 (16:40.0%, 1.0)

CLASS = 2 <= V9 = 2 & V8 = 2 & V13 = 2 & V20 = 1 & V7 = 2 (16:40.0%, 1.0)

CLASSS = 1<=V2_L=0.96 & V2_M=0.61 & V3=0.90 & V4=0.90 & V5=0.90 & V6=0.90 & V7=0.90 & V8=0.90 & V9=0.90 & V10=0.90 & V11=0.90 & V12=0.90 & V13=0.90 & V14=0.90 & V15_L=0.98 & V15_M=0.68 & V16_L=1.00 & V16_H=1.00 & V17_=0.92 & V17_H=0.82 & V18_L=0.98 & V18_M=0.81 & V18_H=0.87 & V19_L=0.62 & V19_M=0.73 & V19_H=0.91 (95%)

CLASSS = 1<=V2_L=0.78 & V2_M=0.91 & V5=0.90 & V6=0.90 & V7=0.90 & V8=0.90 & V9=0.90 & V10=0.90

& V11=0.90 & V12=0.90 & V13=0.90 & V14=0.90 & V15_L=0.99 & V16_L=0.99 & V16_M=0.64 & V17_=0.93 & V18_H=1.00 & V19_H=1.00 (95%)

CLASSS = 2<=V2_M=0.95 & V2_H=0.99 & V4=0.90 & V5=0.90 & V6=0.90 & V7=0.90 & V8=0.90 & V9=0.90 & V10=0.90 & V11=0.90 & V12=0.90 & V13=0.90 & V14=0.90 & V15_L=0.97 & V15_M=0.56 & V16_L=0.97 & V16_M=0.55 & V16_H=0.66 & V17_=1.00 & V17_M=0.91 & V17_H=0.79 & V18_M=1.00 & V18_H=0.55 & V19_L=1.00 & V19_M=0.88 & V19_H=0.81 & V20=0.90 (90%)

CLASSS = 2<=V2_M=0.95 & V2_H=0.73 & V4=0.90 & V5=0.90 & V6=0.90 & V7=0.90 & V8=0.90 & V9=0.90 & V10=0.90 & V11=0.90 & V13=0.90 & V15_L=0.81 & V15_M=0.88 & V16_L=1.00 & V17_=0.89 & V18_M=0.60 & V18_H=0.96 & V19_M=0.96 & V19_H=0.70 & V20=0.90 (90%)

4. Conclusions

We introduced the problem of mining generalized association rule. Earlier work on association rules did not consider the learning ability of the mining systems, and restricted the items in the association rules to the leaf-level items in the taxonomy. An obvious solution to the problem is to replace each former knowledge base with an "extended fuzzy rules" that contains the implicit knowledge of fuzzy neural networks. We presented four phased new hybrid mining mechanism. Empirical evaluation showed that this mechanism extract more flexible, and the result of experiment with a hepatitis database proved to be valid and robust. In conclusion, this study has shown how the association rules and fuzzy neural networks can be brought together to create cooperative knowledge base. It is expected that the proposed hybrid knowledge based construction mechanism will have a significant impact on the research domain related to the human perception and knowledge management. However, this "basic and hybrid" data mining approach is complicate and not very fast. Further research topics still remaining are as follows:

(1) The basic technology of association rule mining used for this study needs to be improved so that more fuzzy knowledge can be analyzed.
(2) Fuzzy membership functions need to be integrated with other rule refining and reasoning mechanism.
(3) Complicate FNN construction processes and fuzzy rule refinement algorithm was need to be improved with other useful knowledge management mechanisms.

References

[1] Agrawal, R., Imielinski T., and Swami, A., "Mining Association Rules between Sets of Items in large Databases," *In Proc. of the ACM SIGMOD Conference on Management of Data*, Washington, D.C., pp.207-216, 1993.

[2] Bonchi, F., Giannotti, F., Gozzi, C., Manco, G., Nanni, M., Pedreschi, D., Renso, C., and Ruggieri, S., "Web Log Data Warehousing and Mining for Intelligent Web Caching," *Data & Knowledge Engineering*, 39, pp.165-189, 2001.

[3] Chakrabarti, S., Dom, B.E., Kumar, S. R., Raghavan, P., Rajagopalan, S., Tomkins, A., Gibson, D., and Kleinberg, J.M., "Mining the Web's Link Structure," *Computer*, 32, pp.60-67, 1999.

[4] Changchien, S.W., and Lu, T.C., "Mining Association Rule Procedure to Support On-Line Recommendation by Customers and Products Fragmentation," *Expert Systems with Applications*, 20, pp.325-335, 2001.

[5] Hui, S.C. and Jha, G.., "Data Mining for Customer Service Support," *Information & Management*, 38, pp.1-13, 2000.

[6] Lee, K.C., Kim, J.S., Chung, N.H., and Kwon, S.J., "Fuzzy Cognitive Map Approach to Web-mining Inference Amplification," *Expert Systems with Applications*, 22, pp.197-211, 2002.

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

[8] Song, H.S., Kim, J.K., and Kim, S.H., "Mining the Change of Customer Behavior in an Internet Shopping Mall," *Expert Systems with Applications*, 21, pp.157-168, 2001.