Development of a Knowledge Discovery System using Hierarchical Self-Organizing Map and Fuzzy Rule Generation

Taehoon Koo[a] and Jongtae Rhee[b]

[a] *E-net, R&D Center*
Sigma Tower, Fl. 5, 7-19 Shinchun-dong, Songpa-gu, Seoul 138-240 Korea
TEL : +82-2-6424-2177, FAX : 82-2-6424-2001, E-mail : robin@e-net.co.kr

[b] *Dongguk University*
3 ga 26, Pildong, Jung-gu, Seoul 100-717Korea
Tel: +82-2-2260-3378, Fax: +82-2-2260-3378, E-mail: jtrhee@dgu.ac.kr

Abstract

Knowledge discovery in databases(KDD) is the process for extracting valid, novel, potentially useful and understandable knowledge from real data. There are many academic and industrial activities with new technologies and application areas. Particularly, data mining is the core step in the KDD process, consisting of many algorithms to perform clustering, pattern recognition and rule induction functions. The main goal of these algorithms is prediction and description. Prediction means the assessment of unknown variables. Description is concerned with providing understandable results in a compatible format to human users. We introduce an efficient data mining algorithm considering predictive and descriptive capability. Reasonable pattern is derived from real world data by a revised neural network model and a proposed fuzzy rule extraction technique is applied to obtain understandable knowledge. The proposed neural network model is a hierarchical self-organizing system. The rule base is compatible to decision makers perception because the generated fuzzy rule set reflects the human information process. Results from real world application are analyzed to evaluate the system's performance.

Keywords:

Data mining; KDD(Knowledge Discovery in Database); Fuzzy rule generation

Introduction

Through recent developments of the computer and network technology, we obtain data from the enterprise and organization with various route. Therefore database size increase exceedingly in a short time. But the big size database should not help the decision making in strength the enterprise competitive power. Hens, it is important that acquisition information and knowledge from data more and more. There are many academic and industrial activities whit this new technology and application area[1][2].

Data Mining is techniques for obtain the hidden pattern, association rule, and functional relation from database. Knowledge Discovery in Database (KDD) is the process for extracting valid, novel potentially useful and understandable knowledge from real data[5][6].

The following figure explains the process knowledge acquisition from database through data mining. Database has the real world data for example phenomenon, fact case and so on. The purpose of KDD is an understanding and a prediction of the real world. And this process is divided into two classes, top down and bottom up, a point of view[18].

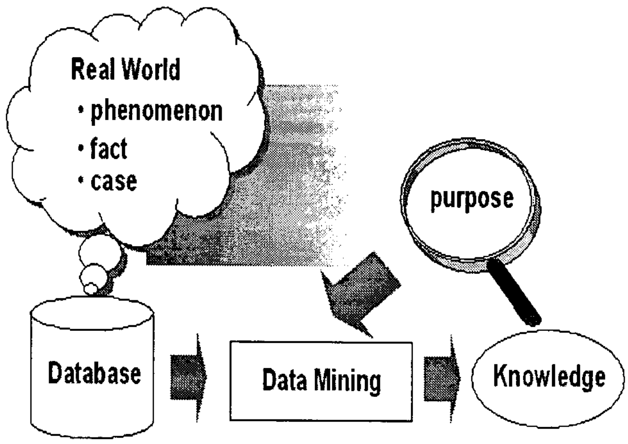

Figure 1. Concepts of Knowledge Discovery Process

Particularly, Data Mining is the core step in KDD process consisting of many algorithms to perform clustering, pattern recognition, and rule induction functions. Main goal in these algorithms is prediction and description. First thing means the assessment of unknown variables and other is concerned with providing understandable results descriptive capability. It is important that Data Mining techniques satisfy these two categories at the same time[5][9]. In the existing Data Mining techniques, Machine Learning methods is prominent concerning the description ability but has week performance of the generalization. Also, Neural Network method has opposite characteristics.

Proposed Method

In this study, we proposed the intelligent system satisfying these prediction and description goal in KDD process. Reasonable pattern is derived from real world data by a revised neural network model and a proposed fuzzy rule extraction technique is applied to obtain understandable knowledge. The proposed neural network model is a hierarchical self-organizing system. The rule base is compatible to decision makers perception because the generated fuzzy rule set reflects the human information process. Results from real world application are analyzed to evaluate the system's performance. Following figure shows the outline of the proposed method.

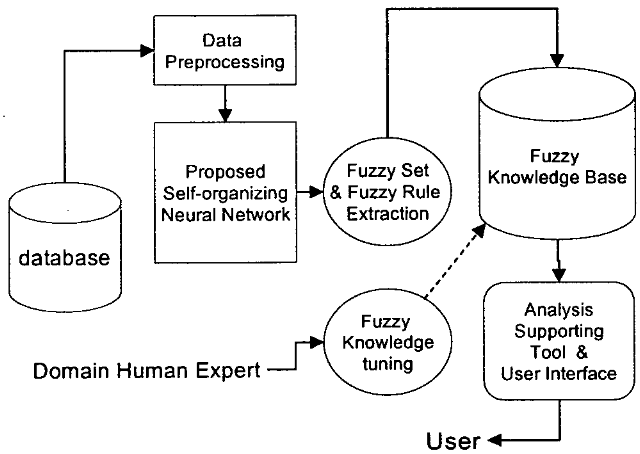

Figure 2. Outlines of the Proposed Method

Procedures of the proposed method stand for the following step.

Step 1. Selection Database

Select database and analyze their attributes so that detection dimension and target attributes.

Step 2. Preprocessing Database and Definition of Fuzzy Attributes

Analyze data in order to remove outlier, revise data dimension and ready to learning and control data set. As well as, define the membership function for construct the fuzzy knowledge base to the attribute in database. Then, a fuzzification is performed. Figure 3. is an example this process.

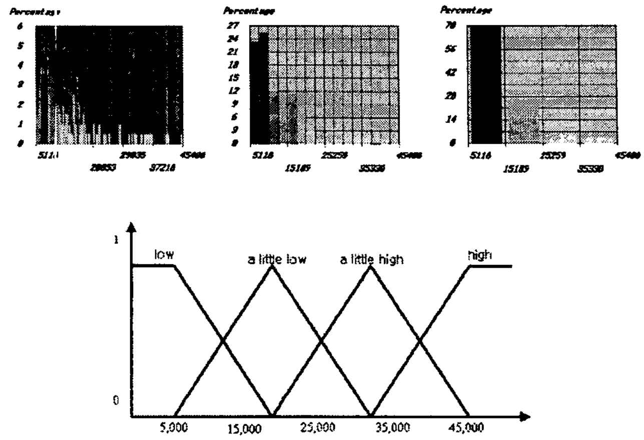

Figure 3. Define Fuzzy Attributes

Step 3. Rule Extraction with Proposed hierarchical Self-Organizing Map

Extract serious rules using proposed a hierarchical Self-Organizing Map from data set. Following Figure shows the architecture of the proposed neural network, which is revised Self-Organizing Map. Trough partial learning, this network obtain fuzzy rule from learning data set. The center cross mark in each data set express the partial winner during learning process and distance between learning data and winner stands for fuzzy membership value after learning. Followings explain the conceptual learning process.

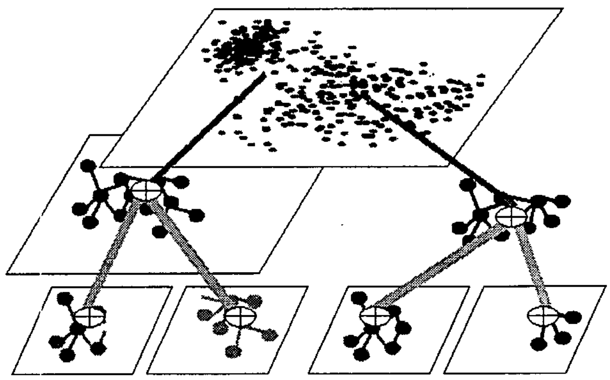

Figure 4. Conceptual Learning process

Step 4. Development of the Fuzzy Knowledge Base System

Construct a fuzzy knowledge base with fuzzy rule matrix and membership value, which is obtain through the step 3 process. Figure 4 shows an inference procedure in this fuzzy knowledge base

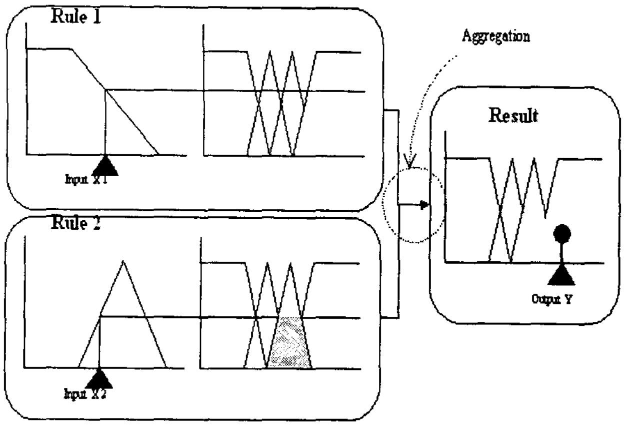

Figure 4. Inference Procedure in Fuzzy Knowledge Base

Basic Experiment

In order to demonstrate the proposed method, we applied data set, which collected by Jeffrey C. Schlimmer in 1987 and construct decision support system using proposed knowledge base. This data set consists of 22 attributes and these attribute type is various. For example, risk level is discrete type from 3 to –3, normalized-losses is continuous type from 65 to 256 and so on. The following figure display this data set.

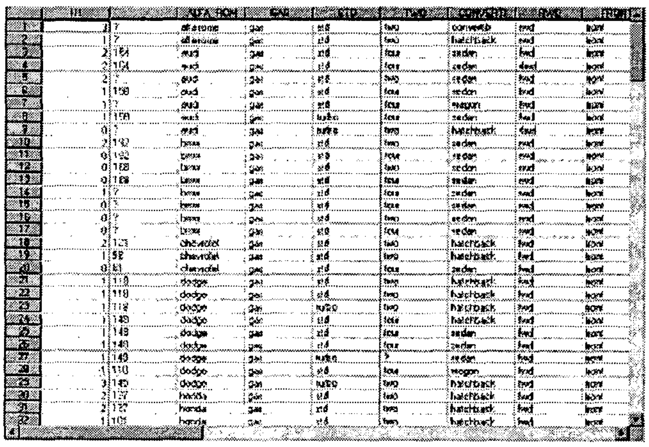

Figure 5. Experimental Data Set

Rule 1. Price of car is between $5,000 and $15,000

Rule 2. Price of car decrease from $45,000 to $5,0000

Rule 3 .If normalized losses is low then price is high

Rule 4. If Risk Level is low then price is high

Rule 5. If body style is wagon then price is low

Rule 6. If engine size is big then price is high

Rule 7. If length is small then price is low

...

Figure 6. Example the extracted fuzzy knowledge base

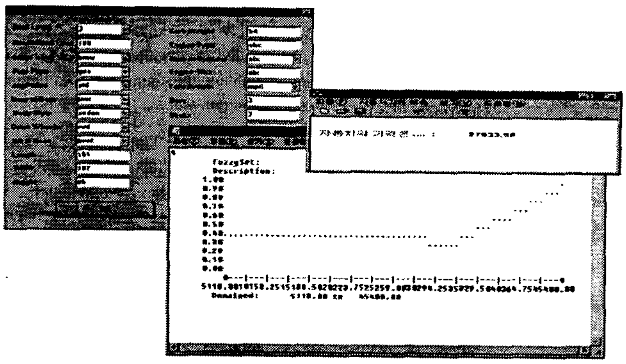

Figure 7. Example the extracted fuzzy knowledge base

Conclusion

In this paper, we proposed Knowledge Discovery System which is satisfy both prediction and description capabilities, and conduct an unsupervised competitive learning process, namely Self-Organizing Map, to detect fuzzy rule generation with hierarchical partial learning process. But this work is the initial research step so we test lots of different data and make validation and verification process.

Reference

[1] Agrawal, R., A. Arning, T.Bollinger, M.Mehta, J. Shafer and R. Srikant, "The Quest Data Mining System, KDD-96, 244-249, 1996.

[2] Brown, M., C. Harris, Nerofuzzy Adaptive Modelling and Control" Prentice Hall, 1994.

[3] Cox, E., Fuzzy Fundamentals, IEEE Spectrum, 58-61, 1992.

[4] Czarnecki, C., R. John, S. Bennett, The Appication of Fuzzy Logic to Real Time Multiple Robot Collision Avoidance, Proceeding of ICSC Fuzzy Logic Symposium 1996, 116-121, 1996.

[5] Fayyad, U., Piatetsky-Shapiro, P. Smyth, Knowledge Discovry and Data Mining : Toward a Unifying Framework, KDD-96, 1996.

[6] Fayyad, U., G. Piatetsky-Shapiro, P. Smyth, From Data Mining to Knowledge Discovery An Overview, Advances in Knowledge Discovery & Data mining, AAAI Press / The MIT Press, 1-34, 1996.

[7] Ishibuchi, H., K. Nozaki, H. Tanaka, Distributed Representation of fuzzy Rules and its application to pattern classification, Fuzzy Sets and Systems, Vol. 52, 21-32, 1992.

[8] Ivancic, F., A. Malaviya, L. Peters, An Automatic Rule Base Generation Method for Fuzzy Pattern Recognition with Multiphased Clustering, Proceedings of the 2nd International Conference of Knowledge-Based Intelligent Electronic Systems, 21-23, 1998.

[9] Jagielska, I., Linguistic Rule Extraction from Neural Networks for Descriptive Datamining, 1998 Second International Conference on Knowledge-Base Intelligent Electronic System, 21-23, 1998.

[10]John, R. I., Fuzzy Inferencing Systems -Problems and Some Solution, CMS Working Paper, No. 62, 1995.

[11]Karr, C., Design of an Adaptive Fuzzy Logic Controller Using Genetic Algorithm for Classification Problems, Second IEEE International Conference on Fuzzy Systems, Vol. 2, 1119-1124, 1993.

[12]Kasabov, N., K., Learning Fuzzy Rule through Neural Network, Proceeding of the First NewZealand International Conference on Artificial Neural Network and Expert System, 137-139, 1993.

[13]Kim, C. J., D. Russell, Automated Generation of Membership Function and Fuzzy Rule using Inductive Reasoning, IEEE, 1993.

[14]Lee, C. C., Fuzzy Logic in Cotrol System : Fuzzy Logic Cotroller, Part II. IEEE Transaction on Systems, Man and Cybernatics, Vol. 20, No. 2, 419-435, 1990.

[15]Lee, M. A., H. Takagi, Integrating Design Stages of Fuzzy Systems using Genetic Algorithms, Second IEEE international Conference on Fuzzy Systems, Vol. 1, 612-617, 1993.

[16]Malavia, A., H. Surmann, L. Peters, Automatic generation of fuzzy rule base for online Handwriting Recognition, in Proceedings of Second European Congress on Intelligent Techniques and Soft Computing, 1060-1065, 1994.

[17]Meredith, D. L., C. L. Karr, K. Kishna Kamur, The use of genetic algorithm in the design of fuzzy logic contorllers, 3rd Worshop on Neural Networks WINN'92, 549-545, 1992.

[18]Nakajima, H., Fuzzy Logic and Data Mining, Fuzzy Systems Symposium, 1996. Soft Computing in Intelligent Systems and Information Processing, 133-138, 1996.

[19]OMRON, Clearly Fuzzy, 1992.

[20]Schwarz, T. J., Fuzzy Systems in the Real World, AI Expert, 1990.

[21]Takagi, H., I. Hayashi, NN-Driven Fuzzy Reasoning, International Journal of Approximate Reasoning, Vol. 5, 191-212, 1991.

[22]Turksen, I. B., Measurement of membership function and their acquisition, Fuzzy Sets and Systems, Vol. 40, 5-38, 1991.

[23]Watanabe, N., Statistical Methods for Estimating Membership Function, Japanese Journal of Fuzzy Theory and Systems, Vol. 5, No. 4, 1979.

[24]Wang, S., Generating fuzzy membership functions : A monotonic neural network model, Fuzzy Sets and Sytems, Vol. 61, 71-81, 1994.

[25]Zadeh, L. A., Fuzzy Sets, Information and Control, Vol 8, 338-353, 1965.