Rough Neuro Voting System for Data Mining: Application to Stock Price Prediction

Hiromitsu Watanabe[1], Basabi Chakraborty[2], and Goutam Chakraborty[2]

[1] Graduate School of Software and Information Science
Iwate Prefectural University, Japan
g236d012@edu.soft.iwate-pu.ac.jp
[2] Faculty of Software and Information Science
Iwate Prefectural University, Japan
basabi@soft.iwate-pu.ac.jp, goutam@soft.iwate-pu.ac.jp

Abstract. This work proposes a rough neuro voting system with modified definitions of rough set approximations for knowledge discovery from complex high dimensional data. Proposed modification of rough set concepts has been used for attribute subset selection. Ensemble of neural networks are used for analysing subspaces of data in parallel and a voting system is used for final decision. The rough neuro voting system is used for stock price prediction with considering other influencing factors in addition to day-to-day stock data. The proposed approach shows effective in predicting increment or decrement of the nextday's stock price from simulation experiment.

Keywords: Data Mining, Stock Price Prediction, Rough Neuro Voting System, Neural Network Ensemble.

1 Introduction

Classical data anylysis techniques based mainly on statistics and mathematics are no longer adequate for analyzing increasingly huge collection of data in variety of domains. New intelligent data analysis methodologies are evolving for discovery of knowledge from complex data bases. Though many successful applications of the above tools are reported in the literature [1], advanced hybrid techniques are necessary for better understanding of the inherent knowledge in the form of simple rules for high dimensional complex data. Soft computing methodologies are widely used for solving real life problems as they provide a flexible information processing capability for handling ambiguity, uncertainty and incompleteness prevalent in real life data. Neural networks and rough set theory are useful tools for classification and rule generation from raw data.

The basic task behind discovery of knowledge from raw data is to divide the data set in the attribute space into different classes, and define the class boundaries by simple yet accurate rules. Most of the real world data with large number of attributes comes with noises some of which are irrelevant to the decision under consideration. In general, one major step is to find which attributes are important. Depending on the method used for rule extraction the real valued data are