DL-$select: A Decision-List-Based Data-Mining System

Karl Weinmeister

Duke University
D101 Levine Science Research Center
Durham, NC 27708
karl.weinmeister@duke.edu

The application of machine-learning algorithms to the financial markets has been increasing in popularity in recent years. The majority of systems that have been created for the purpose of selecting stocks have utilized neural-network techniques. Our research has dealt with the feasibility of inductive logic approaches and the creation of a decision-list-based data-mining system, DL-Select.

Neural networks can model a variety of data distributions and handle inconsistent data well. But for complex problems such as financial analysis, the structure of a neural network can be difficult to interpret. Decision lists (Rivest 1987), however, are represented in an easily understood form: an extended “if-then-elseif-...else-” rule.

Iterative algorithms for decision lists append rules into a list and remove examples from the data set that are covered by these rules. Effective future learning depends on early rule selection, which, if made poorly, can reduce the accuracy of the entire decision list. The algorithm described by Rivest (Rivest 1987) avoids this obstacle by assuming 100% accurate rules in the training data, but consequently leaves open the problem of noisy data. The learning algorithm used in DL-Select, BruteDL (Segal and Etzioni 1994), addresses this and other issues by conducting a single search for homogenous rules—rules in which accuracy is independent of list position. Since homogenous rules need not be 100% accurate, BruteDL is better suited to handle the noise of financial-market data.

There has been significant discourse in the financial and academic community regarding the efficiency of financial markets. The efficient-markets hypothesis asserts that stock prices already reflect any available information, rendering markets. While the low risk of the portfolio is a fortuitous benefit, its high return confirms its potential.

The use of inductive logic and specifically decision lists are indeed useful for stock selection in the financial markets. We have developed a promising data-mining system, DL-Select, which can aid in the determination of a short-term portfolio. Future work on DL-Select includes further performance evaluations and the exploration of alternative knowledge-representation structures.

References


Figure 1: The DL-Select process