UTS

University of Technology, Sydney

IN-DEPTH OPTIMIZATION OF STOCK MARKET DATA MINING TECHNOLOGIES

A Thesis Submitted for the Degree of
Doctor of Philosophy

By

Li LIN

Supervisor: Prof. Chengqi Zhang

Co-Supervisor: Dr. Shichao Zhang

Faculty of Information Technology

University of Technology, Sydney

AUSTRALIA

29 July 2006

Certificate

Author: **Li Lin**

Title: **In-Depth Optimization of Stock Market Data Mining Technologies**

Degree: Philosopher Doctor

Date: 29 July 2006

I certify that this thesis has not already been submitted for any degree and is not being submitted as part of candidature for any other degree.

I also certify that the thesis has been written by me and any help that I have received in preparing this thesis and all sources used have been acknowledged in this thesis.

Signature of Student

Production Note:
Signature removed prior to publication.

i

Acknowledgements

I'd like to express my sincere thanks to those who once offered some help and concern to my study and life. My heart is filled with gratitude that words cannot express.

Contents

Abstract

Stock trading is a number of stocks to be exchanged from one trader to another trader. It consists of a trader selling a number of stocks at a price and a volume, and another trader buying the same stocks at the same price and the same volume. Most traders want to buy a stock at a low price and to sell the stock at a high price in order to make a profit. However, it is difficult to know whether the current trading (buy/sell) price is low or high. Some researchers have presented technical trading rules which are mathematical formulas with many parameters to solve this problem, such as moving average rules, filter rules, support and resistance, channel break-out rules, and so on. All these rules are based on historical data to generate the best parameters and use the same parameters in future trading to make a profit. When the parameters of a trading rule are set properly, the trading rule can help the traders to make a profit (buy/sell at a low/high price). Experiments have shown that technical trading rules are profitable.

However, there are still some disadvantages and limitations to the technical trading rules in real stock market trading. First, the technical trading rules do not integrate domain knowledge (expert experiences and domain constraints, etc). For example, some trading rules pattern maybe only generate three signals during one year trading to get the most profit. However, the pattern is unreasonable and it is unprofitable in future trading, because the pattern is only a mathematical maximum, but it is impracticable in stock trading. Second, the output of a parameter for a trading rule is only one single value. Sometimes, it may be a noise so the trading rule is inapplicable in future trading. Third, present algorithms to calculate parameters of trading rules are inefficient. Most trades are performed through internet such that they can buy and sell stocks in online and a trade is completed in a second. Real markets are dynamic such that trading rules have to be updated all the time depending on changing situations (new data come

in, new parameters will be recomputed). Current enumerate algorithms waste too much time to get new parameters. However, a one-second short delay in real stock trading will lose the best trading chance. Fourth, when we evaluate the performance of a stock, we need not only to consider its performance (profit and return), but also to compare it to other stocks performance. At present, trading rules do not compare to the other stocks performance when they are selected to generate a signal, so the selected stocks or rules may be not the best ones. Fifth, in stock markets, there are many stocks and many trading rules. The problem is how to match and rank stocks and rules to combine a profitable and applicable pair. However, trading rules do not solve this problem. Lastly, trading rule techniques do not consider the sizes of investments. However, in real market trading, different investments will result in a different performance of a pair.

We propose in-depth data mining methodologies based on technical trading rules to overcome these disadvantages and limitations mentioned above. In this thesis, we present the solutions to combat the existing six problems.

To address the first problem, we designed a domain knowledge database to store domain knowledge (expert experience and domain constraints). During the computing procedure, we integrated domain knowledge and constraints. We observed the output more reasonable as we considered domain knowledge.

To address the second problem, we optimized a sub-domain output instead of a single value, in the sub-domain all combinations of parameter can get a near-best result. Moreover, in the sub-domain, some experienced traders can also set or micro-tune parameters by themselves and a better performance is guaranteed.

To address the third problem, we adopted genetic algorithms and robust genetic algorithms to improve the efficiency. Genetic algorithms and robust genetic algorithms can get a near-optimal result in an endurable execution time, and the result is also near to the best one.

To address the fourth problem, we applied fuzzy sets and multiple fitness functions to evaluate stocks. Because many factors influence the performance of a stock, it is necessary to create a multiple fitness function for genetic algorithms and robust genetic algorithms.

To address the fifth problem, we built a stock-rule performance table to rank stock-rule pairs and find the best matching pairs. The stock-rule pair results showed that the ranked performance is better than that of randomly matched pairs.

Finally, to address the sixth problem, we drew a graph of the relationship between investments and number of stock-rule pairs to search maximal points, and to decide the number of pairs for different sizes of investments.

In summary, the purpose of this thesis is to identify optimal methodologies in stock market trading, to make more profit with less risk for investors. The experimental results showed that the methodologies are more profitable and predictable.