STOCK TREND FORECASTING USING HYBRID PARTICLE SWARM OPTIMIZATION – SUPPORT VECTOR MACHINE (PSOSVM) TECHNIQUE

LEE ZHONG ZHEN

BORANG PENGESAHAN STATUS TESIS

JUDUL: STOCK TREND FORECASTING USING HYBRID PARTICLE SWARM OPTIMIZATION – SUPPORT VECTOR MACHINE (PSOSVM) TECHNIQUE

SESI PENGAJIAN: 2010/2011

Saya LEE ZHONG ZHEN
(HURUF BESAR)

mengaku membenarkan tesis (PSM/Sarjana/Doktor Falsafah) ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan FAkulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. ** Sila tandakan (/)

_____	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
_____	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
/	TIDAK TERHAD	

(TANDATANGAN PENULIS)

Alamat tetap: 803, Jalan Temiang,
Taman Nee Yan, 70200 Seremban,
Negeri Sembilan, Malaysia

Tarikh : 13/7/2011

(TANDATANGAN PENYELIA)

CHOO YUN HUOY
Nama Penyelia

Tarikh : 13/7/2011

CATATAN: *Tesis dimaksudkan sebagai Laporan Projek Sarjana Muda(PSM)
**Jika Tesis ini SULIT atau TERHAD, sila Lampirkan surat daripada pihak berkuasa.

STOCK TREND FORECASTING USING HYBRID PARTICLE SWARM OPTIMIZATION – SUPPORT VECTOR MACHINE (PSOSVM) TECHNIQUE

LEE ZHONG ZHEN

This report is submitted in partial fulfillment of the requirements for the Bachelor of Computer Science (Artificial Intelligence)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA
2011

DECLARATION

I hereby declare that this project report entitled

STOCK TREND FORECASTING USING HYBRID PARTICLE SWARM OPTIMIZATION – SUPPORT VECTOR MACHINE (PSOSVM) TECHNIQUE

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT : _____ Date: 13/7/2011
LEE ZHONG ZHEN

SUPERVISOR : _____ Date: 13/7/2011
DR. CHOO YUN HUOY

DEDICATION

To my beloved parents, your love and support are my greatest inspiration.

To my friends, for your motivation and support.

To my lecturer, for being receptive and critical, and challenging me to be a better student.

ACKNOWLEDGEMENTS

I would like to thank Dr. Choo Yun Huoy for giving assistance, guidance and encouragement to complete this project successfully.

I would also like to thank my beloved parents, Lee Yoke Paw and Lim Ming who have been giving me support and motivation throughout my project. I am also appreciative of my girlfriend, Tan Ai Ting for the non-stop encouragement throughout the project.

ABSTRACT

Stock trend forecasting is one of the important issues in stock market research. Accurate stock trend prediction by using a well suited algorithm is a tough challenge in financial industries because the distribution of stock data differs over time. In such case, Support Vector Machine (SVM) produces a fairly good result in stock trend forecasting. However, the accuracy of SVM can be affected if there exists too many input features or the data are noisy. Hence, a feature selection technique is hybridized with SVM in order to improve the forecasting accuracy in Malaysian stock market price trend. Then, the prediction results produced will be analysed. The Improvement Concept Selection Methodology (ICSM) is used as the methodology for this research project since ICSM is designed for R&D environments. A hybrid Particle Swarm Optimization – Support Vector Machine (PSOSVM) algorithm is implemented in stock trend analysis to improve the prediction results. This project implemented SVM with RBF kernel function and optimized two parameters, i.e. the γ and large margin parameter automatically using PSO method. The PSOSVM algorithm was tested on pre-sampled 17 years record of daily KLCI data. The experimental results show that PSOSVM has outperformed SVM technique significantly. Thus, it is concluded that the PSOSVM technique can be implemented in forecasting systems to improve prediction accuracy. It is suggested that future work extended from this project may focus on the selection of different particle in PSO as well as different kernel function in SVM.

ABSTRAK

Peramalan trend atau pergerakan harga saham merupakan salah satu isu penting dalam kajian mengenai pasaran saham. Ramalan trend saham secara tepat dengan menggunakan algoritma yang bersesuaian adalah sejenis cabaran yang amat sukar dalan industri kewangan kerana pengedaran data saham adalah tidak menentu. Dalam kes begini, Support Vector Machine (SVM) mampu menunjukkan keputusan yang agak baik dalam ramalan trend saham. Namun demikian, terlalu banyak input yang mengandungi data yang tidak lengkap akan mempengaruhi ketepatan algoritma tersebut. Justeru, sejenis teknik "feature selection" akan dihibridkan dengan SVM untuk meningkatkan ketepatan ramalan trend dalam pasaran saham Malaysia. Selepas itu, keputusan ramalan akan dianalisis. Improvement Concept Selection Methodology (ICSM) akan digunakan sebagain metodologi dalam projek ini kerana ia direka bentuk untuk tujuan penyelidikan dan pembangunan. Sejenis algoritma hibrid Particle Swarm Optimization – Support Vector Machine (PSOSVM) akan diaplikasikan dengan harapan keputusan ramalan yang lebih tepat dapat dihasilkan. SVM menggunakan fungsi kernel RBF dan setting parameternya, γ serta pmargin parameter akan dioptimumkan oleh PSO secara automatik. Algoritma PSOSVM diuji kaji dengan menggunakan 17 tahun dataset pra-sampel KLCI. Keputusan uji kaji menunjukkan PSOSVM adalah lebih baik daripada SVM. Oleh itu, PSOSVM boleh digunakan dalam sistem ramalan untuk miningkatkan keputusan ramalan. Pelabur saham boleh memakai model hybrid tersebut untuk mengaut keuntungan dengan mengenal pasti trend saham yang berpatutan. Bagi pembaikan lanjut, jumlah partikel yang untuk PSO dan fungsi kernel bagi SVM yang berbeza boleh digunakan untuk projek masa depan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATONS

Eq.	-	Equation
Fig.	-	Figure
Iss.	-	Issue
Pp.	-	Pages
UTeM	-	Universiti Teknikal Malaysia Melaka

CHAPTER I

INTRODUCTION

1.1 Project Background

Stock trend forecasting is one of the important issues in stock market research. Accurate stock prediction by using a well suited algorithm is a tough challenge in financial industries because the distribution of stock data differs over time. Over the years, various studies on the characteristics of stock market prices have been researched and published. Various literature reviews have shown that the stock market is highly stochastic, noisy, and non-stationary. The distribution of financial time series changes over time. There has been plenty of work done in this area but no real successful formula has been developed.

Unsupervised and supervised algorithm such as k-Nearest Neighbour and Support Vector Machine (SVM) respectively has proven to be able to generate good results. SVM is a set of related supervised learning methods able to be used for classification. Its classification power generally outperforms the classical Artificial Neural Network (ANN) algorithm since SVM is not prone to overfitting.

Even though Support Vector Machine has proven to be able to generate good prediction results, too many input features with noisy data may also affect the accuracy of the algorithm. Hence, a hybrid Particle Swarm Optimization – Support Vector Machine (PSOSVM) algorithm is proposed to overcome this problem so that more accurate predictions results can be produced.

Particle Swarm Optimization (PSO) is an optimization algorithm, thus it is able to select the optimal features for stock trend forecasting. The main focus is to select the better features or attributes in stock trend forecasting so that these selected features can lead to a better classification result for SVM.

1.2 Problem Statement

Support Vector Machine (SVM) as mentioned by Powell et al. [102] produces a fairly good result in stock trend forecasting. Results in [102] showed that SVM has better prediction results compared to unsupervised methods. However, the accuracy of SVM can be affected if there exists too many input features or the data are noisy. This is because redundant and correlated features confuse the classifier [103]. Hence, a feature selection technique is hybridized in hope to improve the performance of SVM. In this project, technical analysis is applied as an approach for stock trend prediction where technical indicators are used as input features for SVM. Technical indicators are measurement metrics used to determine future financial, stocks or economic trends. Indicators take price, time and volume as their input. These indicators serve as inputs for the prediction problem.

1.3 Objective

- To implement a feature selection technique in a prediction algorithm in order to improve the stock forecasting accuracy.
- To design and implement a hybrid Particle Swarm Optimization – Support Vector Machine (PSOSVM) technique to forecast Malaysian stock market price.
- To analyse the prediction results produced by PSOSVM as compared to SVM.

1.4 Scope

This project is focused mainly on Malaysian stock markets so that a feasible method can be used for predicting the trend of stock prices. Technical analysis will be the main focus since technical indicators are used as inputs to the PSOSVM for prediction. Then, the accuracy of the prediction result will base on the 'hit ratio' or in other words, prediction accuracy of both the methods.

1.5 Project Significance

The ability to predict the direction and not the exact value of the future stock prices is the most important factor in making money using financial prediction. All the investor needs to know to make a buying or selling decision is the expected direction of the stock. Hence, the proposed PSOSVM method is able to help technical analysts to better predict the trend of the stock.

1.6 Hypothesis

The proposed PSOSVM method is expected to outperform the pure SVM in prediction accuracy since Particle Swarm Optimization is a well suited optimization method for feature selection

1.7 Expected Output

The PSOSVM algorithm will be coded and run in Matlab. Besides that, the experimental results produced by Matlab will be illustrated using Microsoft Excel and SPSS.

1.8 Conclusion

In conclusion, a hybrid PSOSVM method, which is an AI technique for predicting the future trend of stock prices, is proposed. It will provide promising result in terms of prediction accuracy. By choosing an optimal set of technical indicators, PSO can help in improving the performance of the SVM system significantly.

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

Numerous studies and researches have been done to predict the future movement of stock market. Prediction the trend of the stock market is considered as a tough and challenging task in financial time-series forecasting. The main reason is because of the uncertainties involved in the movement of the stock market. Various factors affect the stock market and some of them include traders' expectations, events related to the government and economic conditions. Hence, predicting the movement of stock market price is quite a difficult job.

Over the years, many studies on the behaviour of stock market prices have been published. According to academic investigations, stock market prices movement are not random. It has been confirmed from many literature reviews that the stock market is highly stochastic, noisy, non-stationary and deterministically chaotic [1]-[2]. However, they also behave in a highly non-linear and dynamic manner [3]. In addition, the most important factor in earning profit using financial prediction is the capability to predict the trend and not the exact value of the future stock prices. Everything an investor needs to know to get a buy or sell signal is the just direction the stock.

A number of artificial intelligence (AI) and machine learning techniques have been used for decades to predict the stock market. Although many Artificial Neural Networks (ANN) techniques have been widely used, Support Vector Machine (SVM) has gained significant popularity over the past few years. SVM is a new assuring

non-parametric, non-linear classification technique, which already showed good results in the electric consumption forecasting, optical character recognition, medical diagnostics and other fields. Kim [4] has proposed a technique to predict the stock market direction by using technical analysis indicators as input features to SVM. A recent research on SVM and Back Propagation Neural Networks (BPNN) has been analysed. The experimental results showed that generally SVM outperformed BPNN albeit there are some stock markets for which BPNN have been found to perform better [5].

In order to obtain better prediction accuracy in stock trend forecasting, a feature selection process is needed. A suitable feature selection technique based on the number of features investigated for market price prediction is able to speed up the processing rate and predictive accuracy. In this review, different stock prediction techniques including Particle Swarm Optimization (PSO) and SVM will be studied to get a thorough view on the potency of hybridizing PSO and SVM.

2.2 Stock Market Prediction

A review on some of the theories describing the characteristics of the stock market is done to discuss the possibility whether the direction of the stock market prices can predicted or not. Analysts from the early centuries denied the probability of forecasting the stock market's future direction by giving various illustrations and example to disprove the existence of any forecasting techniques which are under research. However, technical analysts have also come out with proofs to show that there are actually many techniques which have been applied in the stock market field and these techniques have generated significant results which directly profits the stock market companies [2], [4], [5].

2.2.1 Random Walk Theory

Random Walk Theory (RWT) is a stock market theory that defines the past movement of the price of a stock cannot be used to forecast its future direction [6]. A random walk illustrates a walker who is drunk, he starts moving at some point, it is not possible to predict the direction of his next steps for he may even stop moving for a random amount of time. The step patterns he has taken cannot be used to predict his future steps. Hence, no strategy can be built to predict the future market price movement in short run.

This theory raised a lot of questions in 1973 when author Burton Malkiel wrote "A Random Walk Down Wall Street". The theory affirms that stock markets price movement will not follow any pattern and that the historical price movements cannot be used to predict future price movement. Summarising the previous sentences means whatever has come before is meaningless for predicting what is ahead.

As per the random walk theory, stock market prices follow a random path up and down; the chance of a stock's future price ascending is the same as it is for descending. Thus, shutting down the chance to predict with any accuracy which direction the market will move at any point of time.

In a nutshell, the idea is that stocks take a random and unpredictable path. A believer of the random walk theory tends to stand on the side that it is impossible to outperform the market supposing there is no additional risk. On the other hand, critics of the theory controvert that stocks prices do maintain trends as time passes – in short, that it is likely to predict the market by meticulously selecting entry and exit points for buying and holding of shares of stock on a stock market.

2.2.2 Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH) has been agreed by many as one of the capstones of financial economics. E.F Fama, in a paper in 1965 first explained the term "efficient market" as, on the average, competition will cause the entire

effects of new information on intrinsic values to be reflected "instantaneously" in actual prices [7]. The market is considered efficient only if the reaction of market prices to new information are instantaneous and without biased. The EMH shows the idea that information is efficiently assimilated into asset prices at any point in time that past information cannot be used to predict future price movements. Fama demonstrated the efficient markets model [7]. After it has been examined and supported by larger range of markets, it becomes popularly acknowledged by financial economists and applied in economics and finance communities. Therefore, three types of EMH are being categorized depending on the amount of information available.

a) Weak From EMH

The weak form EMH is established on the foundation through the use of fundamental analysis — stock values that are either overvalued or undervalued could be distinguished and thus allowing and agreeing that there could be too much profit generated from this fundamental analytical exercise, which is contradicting to the use of traditional investment strategies such as historical price values, financial strength ratios or stock market trends [8].

This form of efficiency also clarifies that the existing stock prices already reflect historical price and volume information. The information of a security containing historical prices is fully reflected in the current market price of that security. It is classified as weak form because security prices are the most easily and conveniently accessible pieces of information. It indicates that no one would be able to surpass the market using something that "everybody else knows". However, there are still a lot of financial researchers who are analysing the past stock price trends and stock trading information aiming on the goal to gain profit. This technique is called technical analysis that is claimed by EMH as ineffective or impractical for predicting future stock price movement.

The weak form efficiency test is strongly supported by evidence since its result is consistent and follows the random walk model. Although fundamental analysis has been statistically tested and shows significance in affecting stock prices, which is consistent with the expected return or 'fair game' efficient market model. the observation is still insufficient to prove the market is inefficient. Furthermore,

effects of new information on intrinsic values to be reflected "instantaneously" in actual prices [7]. The market is considered efficient only if the reaction of market prices to new information are instantaneous and without biased. The EMH shows the idea that information is efficiently assimilated into asset prices at any point in time that past information cannot be used to predict future price movements. Fama demonstrated the efficient markets model [7]. After it has been examined and supported by larger range of markets, it becomes popularly acknowledged by financial economists and applied in economics and finance communities. Therefore, three types of EMH are being categorized depending on the amount of information available.

a) Weak From EMH

The weak form EMH is established on the foundation through the use of fundamental analysis — stock values that are either overvalued or undervalued could be distinguished and thus allowing and agreeing that there could be too much profit generated from this fundamental analytical exercise, which is contradicting to the use of traditional investment strategies such as historical price values, financial strength ratios or stock market trends [8].

This form of efficiency also clarifies that the existing stock prices already reflect historical price and volume information. The information of a security containing historical prices is fully reflected in the current market price of that security. It is classified as weak form because security prices are the most easily and conveniently accessible pieces of information. It indicates that no one would be able to surpass the market using something that "everybody else knows". However, there are still a lot of financial researchers who are analysing the past stock price trends and stock trading information aiming on the goal to gain profit. This technique is called technical analysis that is claimed by EMH as ineffective or impractical for predicting future stock price movement.

The weak form efficiency test is strongly supported by evidence since its result is consistent and follows the random walk model. Although fundamental analysis has been statistically tested and shows significance in affecting stock prices, which is consistent with the expected return or 'fair game' efficient market model, the observation is still insufficient to prove the market is inefficient. Furthermore,

technical analysis such as chart reading requires high transaction cost which makes profit hard to gain. In fact, there is no real value to stock market investors. Yet, these may not be significant enough to reject the EMH. However, evidence can hardly be found to oppose that stock prices change in a long period or even for a single day. There is only ample evidence that provide insights to stock market which is irrelevant to model testing.

b) Semi-strong Form EMH

The semi-strong form EMH will modify ambiguous information by using rationalization and consistency based on real time data, which means there has been in some way different understanding on the information. In other words, share values may be slightly altered within a small range by investors' mind sets in random manners. It follows that there will also not be excessive gain generated as may be found under the fundamental analysis.

This form of efficiency also stipulates that all publicly available information is likewise already integrated into asset prices. In other words, all publicly available information is completely reflected in a current market price of that security. The public information does not only include historical prices, but also data documented in a company's financial statements, announcement, economic factors and etcetera. It also indicates that no one would be able to surpass the market using something that "everybody else knows". This implies that a company's financial reports are useless in predicting future price movements and ensuring high investment returns.

c) Strong Form EMH

The strong form EMH states that excessive profit could not be estimated in the long run. According to the normal distribution of profit gained, the stock value at any certain time should reflect the accurate position of all necessary information needed to determine the value of the stock price, which would not generate excessive gain for investors, provided that the revealing, disclosure and assessment of information are legally permitted.

This form of efficiency clarifies that private information is also already integrated by market prices and cannot be utilized to gain excessive trading profits. Therefore, the current market price of a security fully reflects all information,