Decision Support Systems 50 (2010) 258–269

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches

Chih-Fong Tsai *, Yu-Chieh Hsiao

Department of Information Management, National Central University, Taiwan

ARTICLE INFO

Article history:
Received 10 May 2009
Received in revised form 4 August 2010
Accepted 17 August 2010
Available online 21 August 2010

Keywords:
Stock prediction
Feature selection
Data mining
Principal Component Analysis
Genetic algorithm
Decision trees

ABSTRACT

To effectively predict stock price for investors is a very important research problem. In literature, data mining techniques have been applied to stock (market) prediction. Feature selection, a pre-processing step of data mining, aims at filtering out unrepresentative variables from a given dataset for effective prediction. As using different feature selection methods will lead to different features selected and thus affect the prediction performance, the purpose of this paper is to combine multiple feature selection methods to identify more representative variables for better prediction. In particular, three well-known feature selection methods, which are Principal Component Analysis (PCA), Genetic Algorithms (GA) and decision trees (CART), are used. The combination methods to filter out unrepresentative variables are based on union, intersection, and multi-intersection strategies. For the prediction model, the back-propagation neural network is developed. Experimental results show that the intersection between PCA and GA and the multi-intersection of PCA, GA, and CART perform the best, which are of 79% and 78.98% accuracy respectively. In addition, these two combined feature selection methods filter out near 80% unrepresentative features from 85 original variables, resulting in 14 and 17 important features respectively. These variables are the important factors for stock prediction and can be used for future investment decisions.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Stock investments are a very popular investment activity around the world. However, the stock market is always difficult to accurately predict due to many reasons, such as the political situation (for some specific countries), the global economy, etc. Without the good ability of predicting stock price, successful investments are very difficult to make.

In literature, some basic important factors, such as financial ratios, technical indexes, and macroeconomic indexes have been proved as the important factors of affecting stocks' rise and fall. However, different studies select their factors (i.e. input variables) differently for their prediction models [3]. That is, the opinion of the important factors for stock prediction is somewhat different in related work since there is no exact answer to the question of what are the most representative variables. On the other hand, it is the fact that using different input variables can make the same prediction model performs differently. Therefore, constructing the optimal stock prediction model for investors is very challenging.

In general, some related work considers a feature selection step to examine the usefulness of their chosen variables for effective stock prediction, e.g. [5,15,53]. This is because not all of the pre-chosen features are informative or can provide high discrimination power. This can be called as the curse of dimensionality problem [33]. As a result, feature selection can be used to filter out redundant and/or irrelevant features from a chosen dataset resulting in more representative features for better prediction performances [50].

Related work which considers feature selection is usually based on one chosen method only (c.f. Table 1). That is, the chosen feature selection method is supposed to select usable features for stock prediction. However, using different feature selection methods is likely to produce different results (i.e. different variables selected). Therefore, if we could apply a number of different feature selection methods and then combine the selection results, we can not only understand the most important and representative variables that all the feature selection methods 'agree', but also further improve prediction performances over using one single feature selection methods.

The idea of combining multiple feature selection methods is derived from classifier ensembles (or multiple classifiers) [26]. The aim of classifier ensembles is to obtain highly accurate classifiers by combining less accurate ones. They are intended to improve the classification performance of a single classifier. That is, the combination is able to complement the errors made by the individual classifiers on different parts of the input space. Therefore, the performance of classifier ensembles is likely better than one of the best single classifiers used in isolation.

* Corresponding author.
E-mail address: cftsai@mgt.ncu.edu.tw (C.-F. Tsai).

0167-9236/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2010.08.028

Table 1
Comparisons of related work.

Work	Dataset	Prediction model	Input variables	Feature selection
Huang and Tsai (2009) [15]	Taiwan index futures (FITX)	A hybrid SOM[a]-SVR[b] model	13 Technical indexes	Filter-based feature selection
Lai et al., (2009) [28]	Taiwan Stock Exchange Corporation	K-means, GA-based fuzzy decision tree	7 Technical indexes	Step-wise regression
Lin et al., (2009) [36]	S&P 500	ESN[c], BPNN[d], RNN[e]	Technical indexes	PCA
Zarandi et al., (2009) [54]	An automotive manufactory in Asia	A type-2 fuzzy logic system	Fundamental & Technical indexes	Regularity Criterion (RC)
Li and Kuo (2008) [34]	Taiwan Weighted Stock Index	SOM + BPNN	Technical indexes	Discrete wavelet transform (DWT)
Chang and Liu (2008) [5]	TSE index and MediaTek	A TSK type fuzzy rule based system	8 Technical indexes	Step-wise regression
Yu et al., (2005) [53]	S&P 500 index data	GA-based SVM[f]	18 Technical indexes	GA
Enke and Thawornwong (2005) [10]	S&P 500 stock index	Linear regression model, BPNN, GRNN[g], PNN[h]	31 Financial and economic variables	Information gain
Ince and Trafalis (2004) [19]	NASDAQ	BPNN and SVM	Technical indexes	PCA and FA[i]
Lam (2004) [29]	364 S&P companies	BPNN	16 Financial & 11 macroeconomic indexes	–
Abraham et al., (2001) [1]	NASDAQ	BPNN, neuro-fuzzy system	Fundamental indexes	PCA
Hulme and Xu (2001) [17]	Australian Stock Exchange (ASX)	GA-based NN	Fundamental indexes	–
Kim and Han (2000) [24]	Daily Korea stock price index (KOSPI)	A hybrid model of BPNN and GA	12 Technical indexes	GA

[a] SOFM Self-organizing feature map.
[b] SVR Support vector regression.
[c] ESN: Echo state network.
[d] BPNN: Back-propagation neural network.
[e] RNN: Recurrent neural network.
[f] SVM: Support vector machine.
[g] GRNN: Generalized regression neural network.
[h] PNN: Probabilistic neural network.
[i] FA: Factor analysis.

As a result, the major research objective of this paper is to examine whether the prediction model using the selected features (i.e. variables) by the combination of multiple feature selection methods can provide better performances (higher accuracy and lower errors) than using single feature selection methods. In particular, three combination strategies to combine multiple selection results are assessed, which are the union, intersection, and multi-intersection approaches. Moreover, the combination of multiple feature selection methods is able to allow us to identify much better representative variables for stock prediction.

The rest of this paper is organized as follows. Section 2 reviews related literature, including stock price theory and analysis methods and feature selection methods used in this paper which are Principal Component Analysis, genetic algorithm, and decision trees. In addition, related work is compared in terms of their datasets used, prediction models constructed, feature selection methods applied, etc. Section 3 presents the experimental setup, including the chosen dataset, the combination approaches to combine multiple feature selection methods, the process of constructing the prediction model based on artificial neural networks, and the evaluation methods. Section 4 shows the experimental results and a conclusion is provided in Section 5.

2. Literature review

2.1. Stock price theory and analysis methods

2.1.1. Stock price theory

Stock prices mean the actual transaction price through the buyers and sellers in the market. Stock prices are determined by the laws of supply and demand [6]. In theory, whether the price of a stock is high or low, it is decided by the buyers and sellers' transactions in the open market. When supply and demand change, the stock price must be changed. That is, if the supply exceeds the demand, the stock price must fall; if the demand exceeds the supply, the stock price must rise. Therefore, we can see that the supply and demand factors directly affect stock prices.

However, there are many other factors affecting stock prices. In past decades, the academic community has developed many related theories about stock prices. The most common one is the Efficient Market Hypothesis (EMH) proposed by Fama [11].

Fama's Efficient Market Hypothesis supposes that the investment activity is a “Fair-Game Market”. It means all information has disclosed in the stock market, and reflects on stock prices. According to the difference of disclosed information, there are three kinds of Efficient Market Hypothesis as follows.

- The Weak Form Efficient Market: the variations of price, volume of trade and other historical information have fully reflected in stock prices. Hence, using past information to analyze stock situations cannot get excess returns. Because of this reason, technical analysis is not applied under this situation.
- The Semi-strong Form Efficient Market: all readily-available public information including the variation of price, volume of trade, financial statements and other information have fully reflected in stock prices. Therefore, it cannot acquire excess returns by using the information that everyone knows. For this reason, fundamental analysis is not applied under this situation.

Input: set of training examples T
Begin:
  initialize concept description $C = \varnothing$
while (there are still positive examples in T)
{
  initialize the GA with random disjuncts $\{d_i \mid i = 1 \ldots N\}$ where N is the population size;
  repeat
    compute F(d_i), the fitness function value for each disjunctd_i ;
    reproduce a new population by applying the genetic operators;
  until (the stopping criteria for the GA is met);
  $C = C \vee d_{best}$ (add the best disjunct to the concept);
  Remove all the positive examples from T that are covered by d_{best} ;
}
Output: concept description C learned by the GA

Fig. 1. The procedure of searching the instance space by GA.

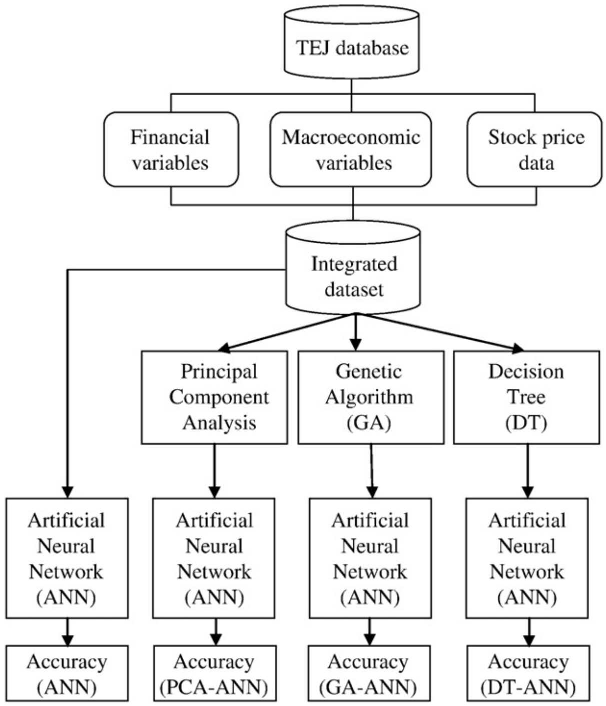

Fig. 2. The first stage experiment.

- The Strong Form Efficient Market: all information includes public and privileged information is fully reflected in prices. Privileged information involves knowledge available to a market marker, insider information available to corporate managers, etc. Therefore, both of public and privileged information cannot predict the market situation.

2.1.2. Stock price analysis methods

In literature, the most common analytical approaches are fundamental analysis and technical analysis described below.

- Fundamental analysis. Fundamental analysis believes that every stock has its intrinsic value. If the share prices lower than the intrinsic value, it means the stock is undervalued. In this case, we should buy this stock, and vice versa. Hence, a fundamental analysis is the process of analyzing information contained in financial statements, such as the company's annual report, balance sheets, and income statements [41]. Some commonly used financial ratios for stock price forecasting are current ratio, return on assets, liabilities ratio, etc.
  In addition, economic factors also belong to this category. It depends on the statistics of the macroeconomics data and they have a significant influence on the returns of individual stocks as well as stock index in general as they possess a significant impact on the growth and earnings' prospects of the underlying companies. Moreover, economic variables also affect the liquidity of the stock market. Some examples of the economic variables are inflation rates, employment figures and producers' price index, etc. After taking all these factors into account, the analyst can make a decision about whether to sell or buy a stock [29].
- Technical analysis. Technical analysis, also known as "charting", has been a part of financial practice for many decades [32,37]. It studies the historical price and volume movements of a stock by using charts as the primary tool to forecast future price movements [39]. This theory believes that the trends and patterns of an investment instrument's price, volume, breadth, and the trading activities reflect most of the relevant market information that a decision maker can utilize to determine its value [29]. Other technical indexes, which have been used for stock price prediction are such as moving average (MA) [29,37], moving average convergence and divergence (MACD) [9], psychological line (PSY) [9], relative strength index (RSI) [29], commodity channel index (CCI) [21], etc. For detailed descriptions, please refer to Achelis [2] and Jobman [21].

2.2. Feature selection

In many research problems, such as pattern recognition, it is important to choose a group of set of attributions with more prediction information. That is, if the number of irrelevant or redundant features is reduced drastically, the running time of a learning algorithm is also reduced. Moreover, a more general concept can be yielded. Performing feature selection can lead to many potential benefits, which are facilitating data visualization and data understanding, reducing the measurement and storage requirements, reducing training and utilization times, defying the curse of dimensionality to improve prediction performances, etc. [13,25,38].

The following describe three well-known feature selection methods, which are Principal Component Analysis, genetic algorithm, and decision trees.

2.2.1. Principal Component Analysis

Principal Component Analysis (PCA) is a multivariate statistical technique. It aims at reducing the dimensionality of a dataset with a large number of interrelated variables. In particular, it extracts a small set of factors or components that are constituted of highly correlated elements, while retaining their original characters. After performing PCA, the uncorrelated variables which are called components, will replace the original variables. The total variability of a dataset produced by the complete set of m variables can often be accounted for primarily by a smaller set of k components of these variables ($k<m$). Therefore, the new dataset consists of n records on k components rather than n records on m variables as the original one. Specifically, eigenvalues and eigenvetors of the principal components are computed in order to find a linear combination of the original variables that makes the greatest variance. The first principal component accounts for as much of the variability in the data, and the second principal component accounts for the remaining variability and so on. Particularly, the level of the variability for each feature lies in the range [0,1], in which the feature with 1 represents the highest variability. Therefore, if we need the components (i.e. features) which can explain 90% (i.e. 0.9) of the variability, features with 90% of the variability or higher can be selected [22].

2.2.2. Genetic algorithms

The main idea of Genetic Algorithms (GA) is from Darwin's theory of evolution from natural selection in the survival of the fittest. GA attempts to computationally mimic the processes by which natural selection operates. It works with a set of candidate solutions called population and generates successive populations of alternate solutions that are represented by a chromosome [14]. Associated with the characteristics of exploitation and exploration search, GA can deal with large search spaces efficiently, and hence has less chance to get a local optimal solution than other algorithms [16].

In Siedlecki and Sklansky [46], a given feature subset is represented as a binary string (a 'chromosome') of length n (the total number of features), with a zero or one in position i denoting the absence ('0') or presence ('1') of feature i in the set. Then, each chromosome is evaluated to determine its fitness, which determines how likely the chromosome is to survive and breed into the next generation. New chromosomes are created from old chromosomes by the process of crossover and mutation. In addition, doing these operators over and over again until some termination criterion is satisfied, we can find the evolution of the optimal solution in a complex space.

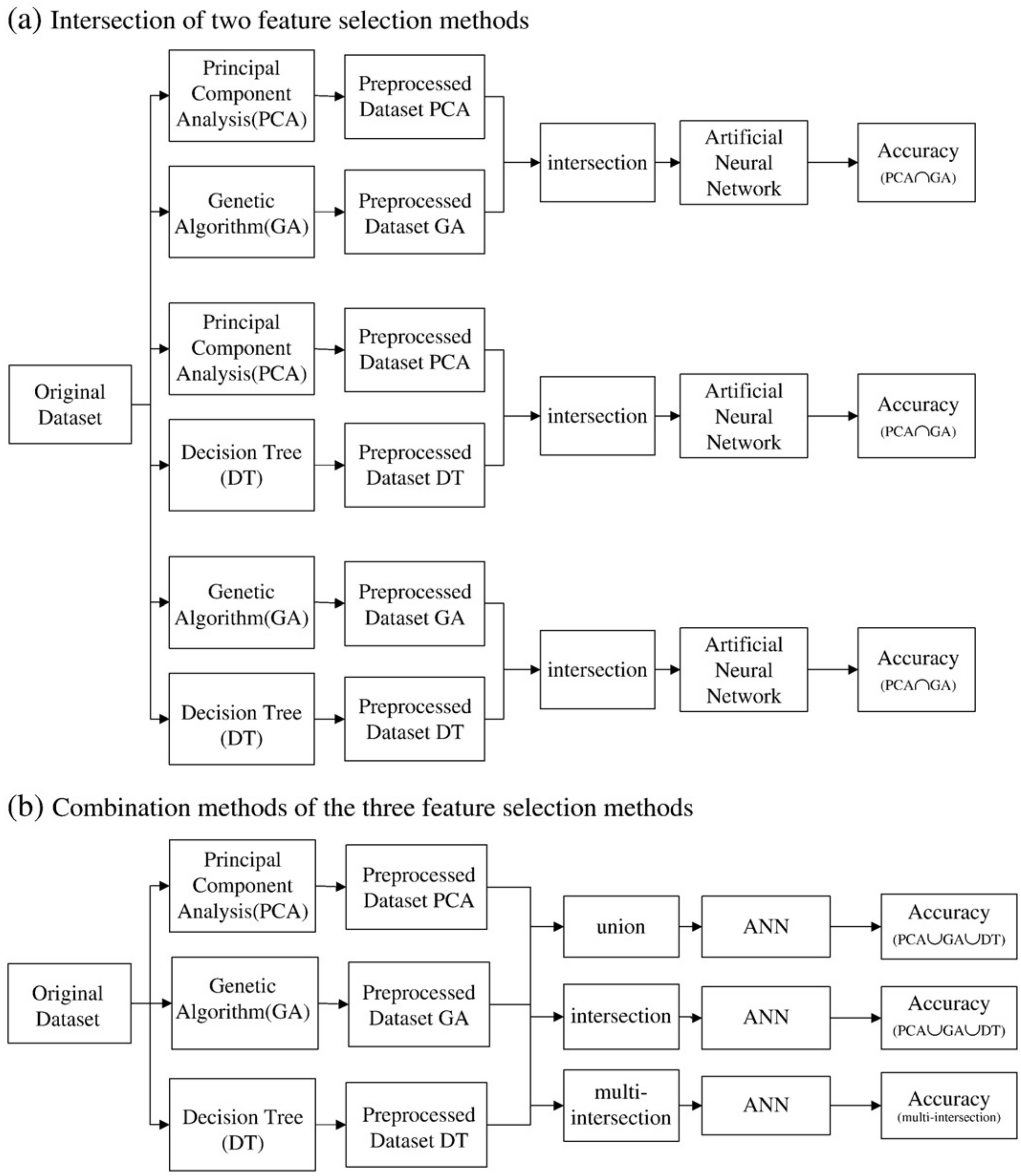

Fig. 3. The second stage experiment.

Fig. 1 shows the procedure for searching the instance space by GA [47]. That is, each member of the population in the GA is a single disjunct and the GA tries to find the best possible disjunct at each generation. Then, the best disjunct replaces the rest through the operators. After GA converges, the best disjunct found is retained and the positive examples it covers are removed. This process is repeated until all the positive instances are covered. The final rule or concept is then the disjunct of all the disjuncts found.

Note that the fitness function looks at the number of positive and negative examples covered by the rule, and it also assigns partial credit for the number of attribute intervals on that rule that match the corresponding attribute values on a positive training example. For

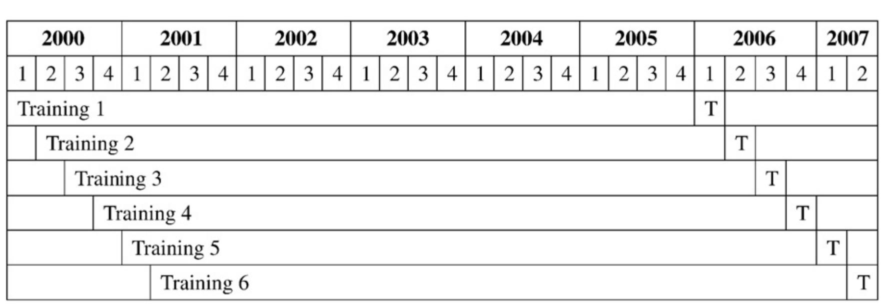

Fig. 4. Sliding window by one quarter based testing data.

<table>
<tr><th colspan="4">2000</th><th colspan="4">2001</th><th colspan="4">2002</th><th colspan="4">2003</th><th colspan="4">2004</th><th colspan="4">2005</th><th colspan="4">2006</th><th colspan="2">2007</th></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>1</td><td>2</td><td>3</td><td>4</td><td>1</td><td>2</td><td>3</td><td>4</td><td>1</td><td>2</td><td>3</td><td>4</td><td>1</td><td>2</td><td>3</td><td>4</td><td>1</td><td>2</td><td>3</td><td>4</td><td>1</td><td>2</td><td>3</td><td>4</td><td>1</td><td>2</td></tr>
<tr><td colspan="24">Training 1</td><td colspan="6">T</td></tr>
<tr><td colspan="1"></td><td colspan="24">Training 2</td><td colspan="5">T</td></tr>
<tr><td colspan="2"></td><td colspan="24">Training 3</td><td colspan="4">T</td></tr>
<tr><td colspan="3"></td><td colspan="24">Training 4</td><td colspan="3">T</td></tr>
<tr><td colspan="4"></td><td colspan="24">Training 5</td><td colspan="2">T</td></tr>
</table>

Fig.5. Sliding window by other-quarter based testing data.

instance, a chromosome is an n dimensional binary vector, where n is the total number of features. If the i-th bit of the vector is 1, then the i-th feature is included in the subset. On the contrary, if the i-th is 0, the feature is not included. The fitness function is determined for each chromosome in the population.

2.2.3. Decision trees

The construction of a decision tree involves a collection of decision nodes, connected by branches, extending downward from the root node until terminating in leaf nodes. The leaf nodes of a decision tree means a result and the path from the root node to the leaf nodes constitute the required combination of conditions [30].

The Classification and Regression Trees (CART) [4] is a statistical technique that can select from a large number of explanatory variables those that are most important in determining the response variable to be explained [13]. The decision trees produced by CART are strictly binary, containing exactly two branches for each decision tree. The root node t is separated into two samples based on some condition. The samples that fit the condition will be separated into the left nodes (t_l), and the others will be separated into the right nodes (t_r). P_L and P_R are the path that the node t goes through t_l and t_r. In particular, a decision tree is based on the entropy theory that the attribute (or feature) with the highest information gain (or greatest entropy reduction) is chosen as the test attribute for the non-leaf node. As a result, the decision nodes (t, t_l, and t_r) can be regarded as representative features over a given dataset [7].

2.3. Related work

This section compares related work in terms of their datasets used, prediction models constructed, feature selection methods considered, etc. Table 1 shows the comparative result.

Regarding Table 1, we can see that much related work only considers one specific index, i.e. either technical indexes or fundamental indexes (including economic factors). However, only Zarandi et al., [54] use both fundamental and technical indexes for stock prediction. Even related work uses the same index; the number of input variables used in these studies is different [3]. Therefore, currently there are no generally agreed representative variables for stock prediction. In addition, in the current stage there is no 'best' feature selection method for stock prediction. Consequently, related work only applies one chosen feature selection method to filter out irrelevant variables. This motivates us to collect all relevant variables used for stock prediction in literature and then combining multiple

Table 2
The fundamental and macroeconomic indexes.

Fundamental indexes		
ROA(A): EBI%	Gross margin growth	Quick ratio
Gross margin%	Operation income growth	Liabilities ratio
Operating income%	Net income growth	Total asset turnover
Net income%	Ordinary income growth	Account receivable turnover
Continued net income%	Continued income growth	Inventory turnover
Cash flow ratio	Total asset growth	Fixed asset turnover
Sales Growth ratio	Return on total asset	Days payables outstanding
Current ratio		
Macroeconomic indexes		
US gross national product	Monitoring indicator	Export foreign exchange volume
US gross domestic product	Leading indicators	Government purchase
US unemployment rate	WPI increase rate	Government revenue
US Industrial Production	CCI Increase Rate	Taiwan Consumer Price Index (CPI)
US export trade amount	Import price index increase rate	Taiwan wholesale price index WPI
US import trade amount	Export price index increase rate	GNP deflator
US consumer price index CPI	US lagging indicator	Industrial production
US producer price index PPI	Foreign investment approval	Electric product export order
US real GDP	Taiwan unemployment rate	Machinery product export order
US real economic growth rate	Narrow monetary supply M1A	Electric machinery product export order
US CCI increase rate	Narrow monetary supply M1B	Information and communication product export order
US customer confident index CCI	Monetary supply M1B increase ratio	Taiwan total trading volume
US personal expenditure	Broad monetary supply M2	US total trading volume
US personal income (Quarter)	Broad monetary supply M2 increase rate	Import volume in dollar
US monetary amount (M1)	Narrow monetary supply M1A Increase Rate	Export amount to US
US monetary supply (M2)	Taiwan rediscount rate	Import amount from US
US industrial production increase rate	Foreign exchange rate	Export volume index
US current account of GDP in ratio	Foreign exchange reserves	Import volume index
Taiwan export volume in NT	Merchandise trade volume	Export growth rate
Taiwan import volume in NT	Merchandise export (F.O.B)	Import growth rate
Total import volume change rate in NT	Merchandise import (F.O.B)	Quasi money

Table 3
Parameter settings of GA.

Work	Population size	Crossover rate	Mutation rate
De Jong and Spears [8]	50	0.6	0.001
Grefenstette [12]	30	0.9	0.01
Kim and Han [24]	20	0.6	0.033

feature selection methods to identify more representative variables for improving prediction performances.

3. Experimental design

3.1. The experimental process

3.1.1. The first experimental stage

The experiment contains two stages. For the first stage, this paper considers fundamental indexes as the input variables including financial and macroeconomic variables from the Taiwan Economic Journal (TEJ) database. That is, financial and macroeconomic variables are concatenated. In addition, the stock price information (i.e. the output variable) corresponding to the fundamental indexes is collected to be the dataset for later experiments. In particular, this is the original dataset without feature selection for training and testing the artificial neural network (ANN) as the prediction model (c.f. Section 3.5).

Next, the original dataset is processed by Principal Component Analysis (PCA), Genetic Algorithms (GA), and decision trees (CART) respectively, in order to filter out unrepresentative variables. As a result, three processed datasets from the three feature selection methods can be obtained respectively. Then, each of the three processed datasets is divided into the training and testing datasets to construct the prediction model based on Artificial Neural Networks (ANN) for stock prediction. Therefore, the aim of the first stage is to find out whether using one of these three feature selection methods can allow ANN to provide better performances than the model without feature selection. Fig. 2 shows the first stage experiment.

3.1.2. The second experimental stage

For the second stage, the three feature selection methods are combined by the union, intersection, and multi-intersection methods (c.f. Section 3.6) in order to predict stock prices more effectively and find out more representative variables. Fig. 3 shows the second stage experiment.

3.2. The dataset

The data source of this paper is based on the Taiwan Economic Journal (TEJ) database. In addition, the listed electronic corporations which are published by the Taiwan Stock Exchange (TSE)[1] are considered. This is because the government has invested a lot of efforts and resources in the electronic industry and many investors invest much money on this industry. Therefore, the electronic industry has become the mainstream in the stock market and it is the most competitive industry in Taiwan. In particular, its transactions contain over 70% of the Taiwan stock market.

This study chooses the data from the first quarter of 2000 to the second quarter of 2007. Seasonal data are considered because of the volatility of stock prices. Moreover, there will be insufficient samples if the research adopts the annual financial report. Therefore, in order to cooperate with the data of financial reports in seasons, the selected index will mainly be based on the months of 3, 6, 9, and 12. In total, there are 4140 data samples (i.e. case companies) composed of 2117 and 2023 samples for stocks' rise and fall respectively. Therefore, on average each quarter contains 159 data samples.

Furthermore, the sliding window method [34,41] is used to divide the sample data into different groups of training and testing data. The sliding window strategy is widely used in many frequent data mining, including stock market prediction [49]. In this paper, there are two testing strategies based on the sliding window. The first one is to predict the single quarter of the stock price shown in Fig. 4. That is, for example, the training data of the first group is from the first quarter of 2000 to the fourth quarter of 2005. Then, the testing data (T) is based on the next quarter (i.e. the first quarter of 2006). Therefore, the model is trained and tested for six times. As a result, there are six different rates of accuracy of the prediction model. Particularly, the proportion of training and testing data is 24:1.

The second strategy of using the sliding window is to predict the other quarters except the training ones shown in Fig. 5. That is, the

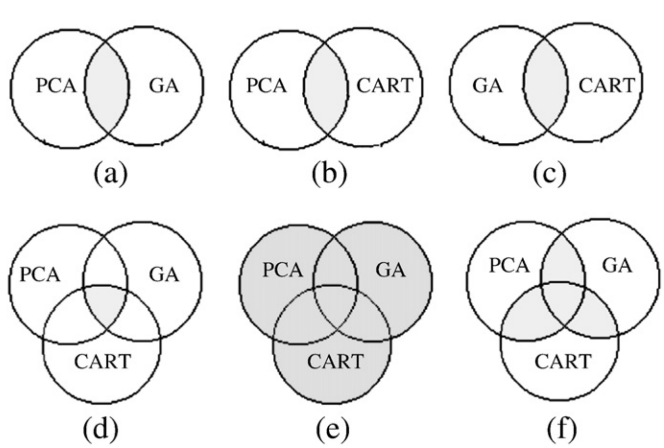

Fig. 6. The combination methods.

Table 4
Confusion matrix.

↓Actual\predicted→	Rise	Fall
Rise	a	b
Fall	c	d

Average accuracy $= \frac{a+d}{a+b+c+d}$.

Error rates for stocks' rise $= \frac{b}{a+b}$.

Error rates for stocks' fall $= \frac{c}{c+d}$.

Table 5
Prediction accuracy of by one quarter based testing data.

	MLP (%)	PCA + MLP (%)	CART + MLP (%)	GA + MLP (%)
TEST1	72.46	73.16	71.74	73.19
TEST2	74.64	75.36	75.36	73.19
TEST3	74.64	73.91	75.36	74.64
TEST4	93.48	93.48	93.48	93.48
TEST5	64.49	62.32	62.32	62.32
TEST6	51.45	93.48	94.2	93.48
Avg. accuracy	71.86	78.62	78.74	78.38

Table 6
Prediction accuracy by other-quarter based testing data.

	MLP (%)	PCA + MLP (%)	CART + MLP (%)	GA + MLP (%)
TEST1	73.19	78.14	78.26	77.78
TEST2	79.86	78.7	79.28	72.61
TEST3	79.89	57.61	79.35	80.98
TEST4	56.76	82.61	54.11	79.71
TEST5	53.26	77.9	77.9	77.9
Avg. accuracy	68.59	74.99	73.78	77.8

[1] http://www.tse.com.tw.

Table 7
Prediction accuracy of multiple feature selections by one quarter based testing data.

	Union (%)	Multi-intersection (%)	PCA∩GA (%)	PCA∩CART (%)	GA∩CART (%)
TEST1	71.74	73.91	71.74	27.54	27.54
TEST2	71.01	71.74	75.36	75.36	24.64
TEST3	75.36	74.64	74.64	74.64	74.64
TEST4	93.48	93.48	93.48	93.48	93.48
TEST5	62.32	62.32	62.32	62.32	62.32
TEST6	43.48	93.48	93.48	93.48	93.48
Avg. accuracy	69.57	78.262	78.50	71.14	62.68

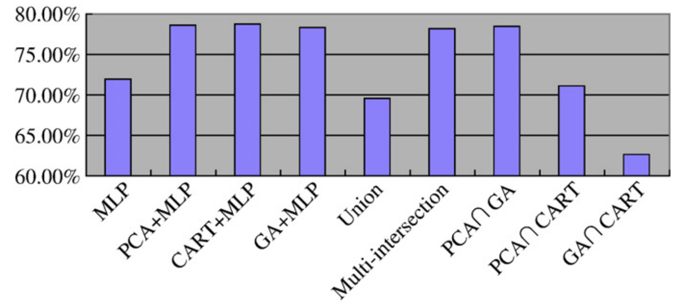

Fig. 7. Prediction accuracy of the MLP models by the one quarter based testing dataset.

training data is the same as the first strategy. However, the testing data (T) are based on the other quarters except the training ones. For example, the first group of the training data is based on the first quarter of 2000 to the fourth quarter of 2005. For the testing data, it is from the first quarter of 2006 to the second quarter of 2007. This is the situation when one only uses a model trained by 'Training 1' for stock prediction in any periods from 2006 to 2007. As a result, there are five different models developed which provide five different prediction results respectively. Specifically, the proportions of training and testing data are 4:1 (24:6), 24:5, 6:1 (24:4), 8:1 (24:3), and 12:1 (24:2) respectively.

3.3. Variables

As Huang and Tsai [15] and Kim [23] pointed out that technical indexes are applied to daily price change in the stock price, this paper considers fundamental indexes and macroeconomic indexes as the input variables except technical indexes for predicting the quarter based dataset. This is because the Taiwan stock market is the Weak Form Efficient Market, which does not reflect all public information in stock prices [35].

Regarding literature review, all of the fundamental and macroeconomic indexes considered in related work are selected. In total, there are 85 variables selected for each data sample which are listed in Table 2.

Note that as the United States is an important trade partner of Taiwan, the economy of United States greatly influences the Taiwan stock market. For example, for the year 2003, the share of Taiwan export to US was 18% and the share of Taiwan imports from US was 13.2%. Therefore, a great deal of United States macroeconomic indexes are considered in this paper. In addition, the experimental results show that most of the representative features selected by combining multiple feature selection methods (which can provide the highest accuracy rate and are important factors for Taiwan stock prediction) are United States macroeconomic indexes (c.f. Section 4.3.3).

For the output variables (i.e. class labels for the prediction model), since it is hard to define the degree of stocks' rise and decline, i.e. different investors may have different definitions about stock price rising and declining, the first attempt of this paper is to simply define two class labels, which are "1" and "−1". For the output class labels of "1", it means the stock price is higher than the previous quarter and "−1" means that the stock price is lower than the previous quarter. That is, the output (or classification) variable for each data sample is based on comparing its stock price between the $i+1$-th quarter and the i-th quarter. For example, for a specific case company if its stock price of the second quarter in 2006 is higher than the one of the first quarter in 2006, then the output (or classification) variable of the second quarter in 2006 is "1".

Note that it can cause the problem of 'predicting' the known stock price movement since the quarterly data are only available after that quarter is over. Therefore, to test the prediction model over a specific quarter, the input variables are based on its previous quarter. For example, given a constructed prediction model trained by 'Training 1' shown in Fig. 4, the input variables of the first quarter in 2006 (T) are based on the fourth quarter in 2005 and so on.

3.4. Feature selection

3.4.1. Principal Component Analysis

To perform PCA, the factors accounting for greater than 10% of the variance (eiqenvalues>1) are kept in the analysis and the factor loading 0.5 are used as informative variables [45]. Specifically, we set the factor loading equals to or greater than 0.5 to extract the important variables from the dataset. To enhance these factors' interpretability, we consider the varimax factor rotation method to minimize the number of variables that have high loading on a factor. That is, varimax rotation maximizes the sum of the variance of the squared loadings. Specifically, for each factor, high loadings (i.e. correlations) will result in a few variables, and the rest will be near zero [22].

In addition, the selection of the important principal component is based on the requirement that the percentage of the total variance is 95% [52]. Note that after the factor loading which is lower than 0.5 is deleted from the original dataset, the total variance of the processed dataset has attained to 95.45%.

Table 8
Prediction accuracy of multiple feature selections by other-quarter based testing data.

	Union (%)	Multi-intersection (%)	PCA∩GA (%)	PCA∩CART (%)	GA∩CART (%)
TEST1	77.17	77.54	75.72	62.68	58.57
TEST2	77.39	78.99	79.57	79.86	69.71
TEST3	59.64	80.98	81.16	59.24	59.24
TEST4	55.56	83.09	83.09	74.88	83.09
TEST5	51.09	77.9	77.9	34.42	77.9
Avg. accuracy	64.17	79.7	79.49	62.22	69.7

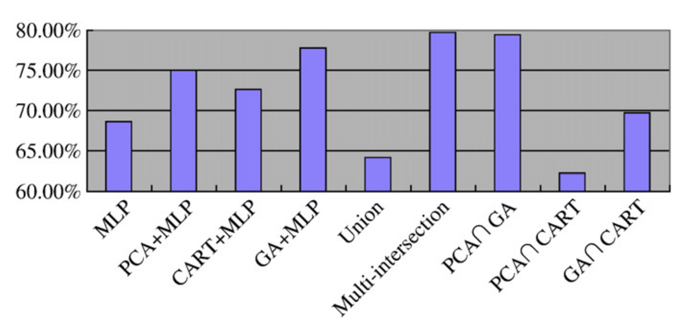

Fig. 8. Prediction accuracy of the MLP models by the other-quarter based testing dataset.

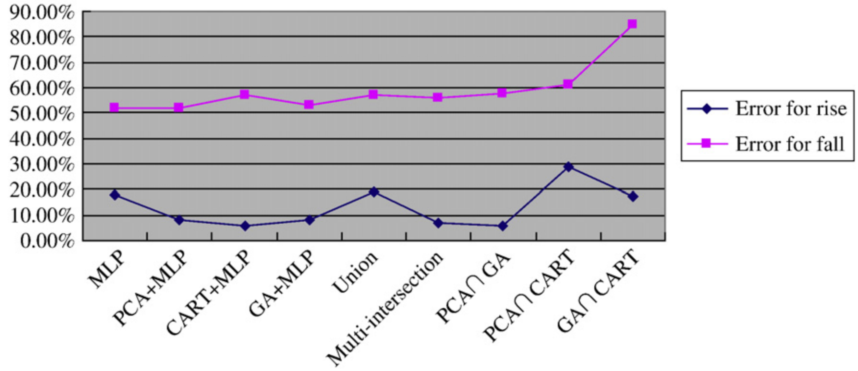

Fig. 9. Error rates of the MLP models by one quarter based testing dataset.

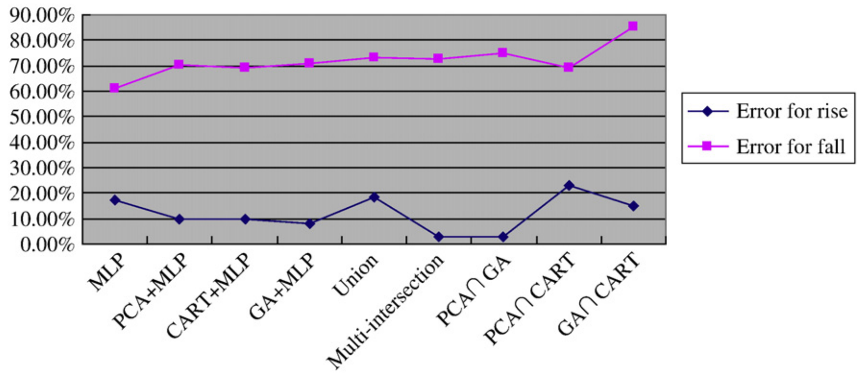

Fig. 10. Error rates of the MLP models by the other-quarter based testing dataset.

3.4.2. Genetic Algorithm

In this study, the second feature selection method we used is based on Genetic Algorithms (GA). There are several different parameter settings for GA shown in Table 3. According to the prediction performance, the parameters used in this paper for later comparisons are as follows: the population size is set to 20, the crossover rate is set to 0.6 and the mutation rate is set to 0.033.

3.4.3. Decision trees

The CART (Classification and Regression Trees) is used as the third feature selection method. The default value[2] is used to establish the initial decision tree (based on a given training set) and then pruning the least related variables in order to select the explanatory variables and split point with the highest reduction of impurity.

To prune the initial decision tree, the minimum support and the score method are considered to create the tree branches. In particular, we set the minimum support for 100 (i.e. to delete the rules which contain less than 100), and the result does not make the prediction performance different. In addition, entropy and Bayesian methods can be used to create the tree branches, and we found that the entropy method can provide the best prediction performance over the given dataset.

3.5. Artificial neural network

In this paper, we used multi-layer perceptron (MLP) artificial neural networks with the back-propagation learning algorithm as the baseline prediction model. This is because approximately 95% of business application studies utilize MLP [48]. In addition, the most popular learning method is back-propagation [18,40]. Since the focus of this paper is not on developing a novel prediction model, it is feasible to construct the widely applied model, i.e. MLP, as the baseline prediction model for comparisons. The following parameters of constructing a MLP network are as follows:

- Learning rate. The learning rate is the parameter in the learning rule that aids the convergence of errors. In the general case, a learning rate of 0.9 is recommended. However, if the learning rate is too high, it will cause the error to oscillate and thus prevent the converging process [43].
- Hidden layer. Regarding prior studies [33,42,51], it is found that using one hidden layer of MLP in the area of stock price prediction can have better performances. Therefore, in this paper, we consider one hidden layer to construct the MLP model.
- Hidden layer node. In literature [42,51], there is no precise number to the node of the hidden layer. If there are too few nodes, the network cannot reflect the relationship between input variables, which may result in the under-fitting problem. On the other hand, too many nodes will cause the over-fitting problem easily. Hence, we use 6, 12, and 18 respectively in order to find the optimal number of the hidden layer node.
- Training epoch. The linking value will gradually be adjusted when the MLP model is trained continuously. In order to make the error of the target value and the output of the neural network become closer, it will become convergent when two of the values do not change. In this paper, we consider the training epoch of 100, 300, 500, and 1000 respectively to find the best training epoch.

As a result, there will be 72 and 60 models for each feature selection method over the one quarter and other-quarter based testing datasets respectively. That is, for the example of the one quarter based testing dataset, it contains six testing subsets based on the sliding window and every sample is used for training for 12 times (i.e. three different hidden layer nodes and four different training epochs).

[2] The toolbox is based on Weka (http://www.cs.waikato.ac.nz/ml/weka/).

Table 10
The selected variables by PCA∩GA and the multi-intersection approach.

PCA∩GA		The multi-intersection approach	
1	US gross national income	1	US gross national income
2	US producer price index	2	US Producer Price Index
3	US annual changes in consumer price index	3	US annual changes in consumer price index
4	US personal consumption expenditures	4	US personal consumption expenditures
5	US annual changes in industrial production index	5	US annual changes in industrial production index
6	US current account to GDP ratio	6	US current account to GDP ratio
7	Taiwan unemployment rate	7	Taiwan unemployment rate
8	Quasi money	8	Quasi money
9	Export amount to US	9	Export amount to US
10	US merchandise trade volume	10	US merchandise trade volume
11	The export order for electric products	11	The export order for electric products
12	GNP deflator	12	GNP deflator
13	US monetary supply	13	US monetary supply
14	Narrow monetary supply	14	Narrow monetary supply
		15	Import quantum index
		16	Annual changes in export price index
		17	Industrial production index

3.6. Combination methods

Regarding Fig. 3, there are six different methods of combining the chosen three feature selection methods. Fig. 6 shows the concept diagram of these combination methods. That is, Fig. 6(a), (b), and (c) are the intersections of two feature selection methods and the result of the intersection method is based on the repeated variables selected by two of the combined feature selection methods. Fig. 6(d) is the intersection of the three feature selection methods.

On the other hand, the result of using the union combination method is based on all variables that have been selected by each of the three feature selection methods shown in Fig. 6(e).

Finally, for the multi-intersection method, the repeated variables of PCA and GA, PCA and CART, GA and CART are selected as shown in Fig. 6 (f).

3.7. Evaluation strategies

To assess the performance of the developed prediction models, accuracy and error rates are examined. They can be measured by a confusion matrix shown in Table 4.

4. Results

4.1. Single feature selection methods

Tables 5 and 6 show the rate of prediction accuracy of the four different MLP models based on the one quarter and other-quarter based testing datasets respectively. As we can see, the results are slightly different if different testing datasets are considered. In particular, larger testing dataset could degrade the prediction performance of the models. Note that the prediction accuracy rate for each test set is based on the best parameter setting of MLP (c.f. Section 3.5). That is, for each test set ('T' in Figs. 4 and 5) there are 12 MLP models constructed and only the best MLP, which provides the highest rate of accuracy, is listed here.

Based on the one quarter based testing dataset, the MLP models followed by PCA, CART, and GA performs similarly, which can provide about 78% accuracy. This may be because the testing data are only based on one quarter, i.e. the testing data size is relatively small, which cannot make these MLP models perform significantly different. On the other hand, for the other-quarter based testing dataset, GA + MLP performs the best (77.8% on average). Moreover, only the model of GA + MLP degrades the least accuracy rate from the one quarter to other-quarter based testing datasets. This implies that feature selection using GA could make the prediction model more stable than the other feature selection methods.

It is interesting that for 'TEST5' of one quarter based testing data (Table 5), all of the models do not perform well, which means that the first quarter of 2007 is difficult to forecast. However, for 'TEST6' the baseline MLP model performs even worse than 'TEST5', but the other three models followed by feature selection provide relative good performances. This is similar to 'TEST4' and 'TEST5' of other-quarter based testing data (Table 6). Therefore, it implies that the baseline MLP model is not suitable and unstable for predicting newer testing data.

4.2. Multiple feature selection methods

Tables 7 and 8 show the prediction performances of the MLP models by combining multiple feature selection methods over the one quarter and other-quarter based testing datasets respectively. Note that the feature selection result of GA∩CART is the same as the intersection between PCA, GA, and CART (PCA∩GA∩CART).

The results indicate that the intersection between PCA and GA outperforms the other combination approaches over the one quarter based testing dataset. On the other hand, combining multiple feature

Table 11
T-test of prediction accuracy (*p* value) by the one quarter based testing dataset.

	Baseline	PCA	CART	GA	Union	Multi-intersection	PCA∩GA	PCA∩CART	GA∩CART
Baseline		0.139	0.193	0.001	0.311	0.000	0.000	0.973	0.009
PCA			0.792	0.375	0.040	0.005	0.001	0.143	0.000
CART				0.287	0.081	0.012	0.000	0.300	0.004
GA					0.000	0.015	0.001	0.004	0.000
Union						0.000	0.000	0.609	0.026
Multi-intersection							0.110	0.000	0.000
PCA∩GA								0.000	0.000
PCA∩CART									0.001
GA∩CART									

Table 9
Numbers of features selected vs. accuracy rates.

	PCA	CART	GA	Union	Multi-intersection	PCA∩GA	PCA∩CART	GA∩CART
No. features selected	64	11	17	72	17	14	5	2
Accuracy (one quarter)	78.66% (2)	78.743% (1)	78.38% (4)	69.57% (7)	78.262% (5)	78.50% (3)	71.14% (6)	62.68% (8)
Accuracy (other quarters)	74.99% (4)	72.62% (5)	77.8% (3)	64.17% (7)	79.7% (1)	79.49% (2)	62.22% (8)	69.7% (6)
Avg. accuracy	76.83% (4)	75.68% (5)	78.09% (3)	66.87% (6)	78.98% (2)	79.00% (1)	66.68% (7)	66.19% (8)

selection methods by the multi-intersection approach performs the best based on the other-quarter based testing dataset. However, the rates of prediction accuracy by both combination approaches over the two testing datasets do not have a big difference, i.e. less than 0.3%.

4.3. Further comparisons

4.3.1. Prediction accuracy

Figs. 7 and 8 further compare average prediction accuracy of the MLP models using different feature selection methods over the one quarter and other-quarter based testing datasets respectively.

For prediction accuracy, as we can see that the MLP model followed by each of the three single feature selection methods performs better than the baseline MLP model over the two different testing datasets. On the other hand, combining multiple feature selection methods by the PCA∩GA and multi-intersection approaches also perform better than the baseline MLP model.

Although the prediction performances of using single feature selection methods (i.e. PCA, CART, and GA) and combining multiple feature selection methods (i.e. PCA∩GA and multi-intersection) do not have a big difference, the later methods can provide much higher accuracy than the single feature selection methods over the other-quarter based testing dataset.

4.3.2. Prediction errors

Figs. 9 and 10 show the error rates of the MLP models using different feature selection methods over the one quarter and other-quarter based testing datasets respectively. It is interesting that all of these prediction models do not perform well for predicting stocks' fall. We believe that this is because we did not exclude the data which may be difficult to forecast, such as the president election in the first quarters of 2000 and 2004 and the 9/11 and SARS events from 2000 to the first quarter of 2001.

However, the error rate of predicting stocks' rise is relatively lower. Particularly, the MLP models by PCA∩GA and the multi-intersection approach outperform the others. This implies that given a new stock, investors who would like to make successful investments can only rely on the decision of the prediction model for the stock rises. If investors follow the output of the prediction model for the case of stocks' fall, then investors are very likely to make incorrect decisions. In other words, these models can help investors make decisions for buying 'rising' stocks, rather than selling 'falling' stocks if they have held.

4.3.3. Selected features vs. prediction accuracy

Table 9 compares these feature selection methods in terms of the number of features selected and their corresponding accuracy rates.

Regarding Table 9, very few input variables still have a high discriminate power for stock prediction. For example, the 11 features (out of 85) selected by CART allow the MLP model to produce 78.743% accuracy over the one quarter based testing dataset and 17 features selected by the multi-intersection approach for 79.7% accuracy over the other-quarter based testing dataset. In other words, these variables can be regarded as the important factors of affecting stocks' rise and fall.

On average, combining PCA and GA by the intersection approach provides the highest accuracy rate (79%) and the multi-intersection approach performs the second (78.98%). For single feature selection methods, GA performs the best (78.09%). On the other hand, the union combination approach, PCA∩CART and GA∩CART performs the worst (i.e. below 70%). Therefore, we can conclude that combining multiple specific feature selection methods is able to allow the stock prediction model to perform better than using single feature selection methods. However, the combination methods used need to be carefully considered.

Table 10 lists the variables selected by PCA∩GA and the multi-intersection approach. Both approaches select the same 14 variables, in which the later one selects three more variables. This indicates that the U.S. stock market has a leading effect to the Taiwan stock market. Therefore, for future stock prediction and investments, these 14 variables can be considered.

4.3.4. Statistical analysis

To analyze the level of significant difference of prediction accuracy by using different feature selection methods, *t*-test is used. Tables 11 and 12 show the *t*-test result over the one and other-quarter based testing datasets respectively.

Regarding above analyses, we can see that considering single feature selection methods, PCA, CART, and GA do not make MLP perform significantly different over the two testing datasets. However, only the model of GA + MLP provides a high level of significant difference from the baseline MLP. For combining multiple feature selection methods, the prediction results of the MLP models using the PCA∩GA and multi-intersection approaches over the two testing datasets are significantly different from the ones using single and other combined multiple feature selection methods.

5. Conclusion

In stock prediction, fundamental and technical indexes composed of different variables have been widely used in literature. As feature selection aiming at selecting more representative features for better prediction results, most of the related studies only use one chosen feature selection method for stock prediction. This paper compares three different feature selection methods, i.e. Principal Component Analysis (PCA), Genetic Algorithms (GA), and decision trees (CART) and combines them based on union, intersection, and multi-intersection approaches to examine their prediction accuracy and errors.

The experimental results show that combining multiple feature selection methods can provide better prediction performances than using single feature selection methods. In particular, the intersection between PCA and GA and the multi-intersection of PCA, GA, and CART perform the best, which provide the highest rate of prediction accuracy and the lowest error rate of predicting stocks' rise. This finding directly corresponds to the success of classifier ensembles, which is based on the diversity of individual classifiers [26]. That is, the ways of selecting features by PCA, GA, and CART individually are different, which can make the selected features by these three methods much diversified (see Table 9)[3]. Therefore, the multi-

[3] The numbers of features selected by GA and CART are 17 and 11 respectively, but there are only three features, which are selected by both GA and CART.

Table 12
T-test of prediction accuracy (*p* value) by the other-quarter based testing dataset.

	Baseline	PCA	CART	GA	Union	Multi-intersection	PCA∩GA	PCA∩CART	GA∩CART
Baseline		0.859	0.068	0.015	0.384	0.000	0.000	0.356	0.800
PCA			0.359	0.031	0.573	0.000	0.000	0.612	0.664
CART				0.391	0.036	0.001	0.000	0.022	0.198
GA					0.001	0.000	0.000	0.003	0.020
Union						0.000	0.000	0.967	0.841
Multi-intersection							0.007	0.000	0.000
PCA∩GA								0.000	0.000
PCA∩CART									0.830
GA∩CART									

intersection of PCA, GA, and CART can provide the best performance. For the intersection between PCA and GA, they select 14 features, which are the same as the 14 features out of 17 by the multi-intersection approach. This approach of course also performs very well, which is similar to multi-intersection of PCA, GA, and CART.

Moreover, these two combined approaches select 14 and 17 important variables respectively from the 85 original variables, which filter out many unrepresentative variables. These variables can be used not only for practical investment decisions, but also for future research as the 'standard' input variables to construct novel prediction models for comparisons.

It should be noted that although this paper considers three popular feature selection methods, there are other methods available in literature, for example, information gain [31], independent component analysis [27], and other variants of PCA, such as kernel PCA [44], asymmetric PCA [20], etc. However, from the practical standpoint, it is difficult to conduct a comprehensive study on all existing feature selection methods. In addition, currently it is hard to define the most representative method in the stock prediction domain, and there is no comparative study based on these methods, which can be regarded as one of the future research issues.

References

[1] A. Abraham, N. Baikunth, P.K. Mahanti, Hybrid intelligent systems for stock market analysis, Lecture Notes in Computer Science 2074 (2001) 337–345.
[2] S.B. Achelis, Technical Analysis from A to Z, McGraw-Hill, New York, 2000.
[3] G.S. Atsalakis, K.P. Valavanis, Surveying stock market forecasting techniques — part II: soft computing methods, Expert Systems with Applications 36 (3) (2009) 5932–5941.
[4] L. Breiman, J. Friedman, R. Olshen, S. Stone, Classification and Regression Trees, Chapman & Hall/CRC Press, Florida, 1984.
[5] P.C. Chang, C.H. Liu, A TSK type fuzzy rule based system for stock price prediction, Expert Systems with Application 34 (1) (2008) 135–144.
[6] G. Cordinly, Guide to the Stock Exchange, 2nd EdRichard D., Irwin, Inc, 1907.
[7] M. Dash, H. Liu, Feature selection for classification, Intelligent Data Analysis 1 (1997) 131–156.
[8] K.A. De Jong, W.M. Spears, An analysis of the interacting roles of population size and crossover in genetic algorithms, Proceedings of the First Workshop on Parallel Problem Solving from Nature, 1990, pp. 38–47.
[9] J.L. Du, Know-how of Applications of Technical Indices in Taiwan Stock Market, Wealth Press, 2003.
[10] D. Enke, S. Thawornwong, The use of data mining and neural networks for forecasting stock market returns, Expert System with Applications 29 (4) (2005) 927–940.
[11] E.F. Fama, Random walks in stock market prices, Financial Analysis Journal 21 (1965) 55–59.
[12] J.J. Grefenstette, Optimization of control parameters for genetic algorithms, IEEE Transactions on Systems, Man, and Cybernetics 16 (1) (1986) 122–128.
[13] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, Journal of Machine Learning Research 3 (2003) 1157–1182.
[14] J. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, 1975.
[15] C.L. Huang, C.Y. Tsai, A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting, Expert System with Applications 36 (2) (2009) 1529–1539.
[16] C.L. Huang, C.J. Wang, A GA-based feature selection and parameters optimization for support vector machines, Expert Systems with Applications 31 (2006) 231–240.
[17] D. Hulme, S. Xu, Application of genetic algorithm to the optimisation of neural network configuration for stock market forecasting, Lecture Notes in Artificial Intelligence 2256 (2001) 285–296.
[18] H. Ince, T.B. Tradalis, A hybrid model for exchange rate prediction, Decision Support Systems 42 (2) (2006) 1054–1062.
[19] H. Ince, T.B. Tradalis, Kernel principal component analysis and support vector machines for stock price prediction, IEEE International Joint Conference on Neural Networks, 2004, pp. 2053–2058.
[20] X. Jiang, Asymmetric principal component and discriminant analyses for pattern classification, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (5) (2009) 931–937.
[21] D.R. Jobman, The Handbook of Technical Analysis: A Comprehensive Guide to Analytical Methods, Trading Systems and Technical Indicators, McGraw-Hill, New York, 1994.
[22] I.T. Jolliffe, Principal Component Analysis, Springer Verlag, New York, 1986.
[23] K.J. Kim, Financial time series forecasting using support vector machines, Neurocomputing 55 (2003) 307–319.
[24] K.J. Kim, I. Han, Genetic algorithm approach to feature discretization in artificial neural network for the prediction of stock price index, Expert Systems with Applications 19 (2) (2000) 125–132.
[25] Y. Kim, Toward a successful CRM: variable selection, sampling, and ensemble, Decision Support Systems 41 (2) (2006) 542–553.
[26] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, J. On, Combining classifiers, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (3) (1998) 226–239.
[27] N. Kwak, C. Kim, H. Kim, H. Dimensionality, Reduction based on ICA for regression problems, Neurocomputing 71 (2008) 2596–2603.
[28] R.K. Lai, C.Y. Fan, W.H. Huang, P.C. Chang, Evolving and clustering fuzzy decision tree for financial time series data forecasting, Expert Systems with Applications 36 (2) (2009) 3761–3773.
[29] M. Lam, Neural network techniques for financial performance prediction: integrating fundamental and technical analysis, Decision Support System 37 (4) (2004) 567–581.
[30] D.T. Larose, Data Mining Method and Models, John Wiley & Sons, Inc., New Jersey, 2006.
[31] C. Lee, G.G. Lee, Information gain and divergence-based feature selection for machine learning-based text categorization, Information Processing and Management 42 (1) (2006) 155–165.
[32] W. Leigh, N. Modani, R. Hightower, A computational implementation of stock charting: abrupt volume increase as signal for movement in New York stock exchange composite index, Decision Support Systems 37 (4) (2004) 515–530.
[33] J. Li, M.T. Manry, P.L. Narasimha, C. Yu, Feature selection using a piecewise linear network, IEEE Transactions on Neural Networks 17 (5) (2006) 1101–1115.
[34] S.T. Li, S.C. Kuo, Knowledge discovery in financial investment for forecasting and trading strategy through wavelet-based SOM networks, Expert Systems with Applications 34 (2) (2008) 935–951.
[35] K.P. Lim, R.D. Brooks, M.J. Hinich, Nonlinear serial dependence and the weak-form efficiency of Asian emerging stock markets, Journal of International Financial Markets, Institutions and Money 18 (5) (2008) 527–544.
[36] X. Lin, Z. Yang, Y. Song, Short-term stock price prediction based on echo state networks, Expert Systems with Applications 36 (3) (2009) 7313–7317.
[37] A.W. Lo, H. Mamaysky, J. Wang, Foundation of technical analysis: computations, algorithm, statistical inference, and empirical implementation, Journal of Finance 55 (4) (2000) 1705–1765.
[38] D. Mladenic´, M. Grobelnik, Feature selection on hierarchy of web documents, Decision Support Systems 35 (1) (2003) 45–87.
[39] J.J. Murphy, Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications, New York Institute of Finance, 1999.
[40] S. Olafsson, X. Li, S. Wu, Operations research and data mining, European Journal of Operational Research 187 (3) (2008) 1429–1448.
[41] D. Olson, C. Mossman, Neural network forecasts of Canadian stock returns using accounting ratios, International Journal of Forecasting 19 (3) (2003) 453–465.
[42] M. Paliwal, U.A. Kumar, Neural networks and statistical techniques: a review of applications, Expert Systems with Applications 36 (1) (2009) 2–17.
[43] T.S. Quah, B. Srinvasan, Improving returns on stock investment through neural network selection, Expert Systems with Applications 17 (4) (1999) 295–301.
[44] B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation 10 (5) (1998) 1299–1319.
[45] V. Shih, Q. Zhang, M. Liu, Comparing the performance of Chinese banks: a principal component approach, China Economic Review 18 (2007) 15–34.
[46] W. Siedlecki, J. Sklansky, A note on genetic algorithms for large-scale feature selection, Pattern Recognition Letters 10 (5) (1989) 335–347.

[47] R. Sikora, S. Piramuthu, Framework for efficient feature selection in genetic algorithm based data mining, European Journal of Operational Research 180 (2) (2007) 723–737.
[48] K.A. Smith, J.N.D. Gupta, Neural networks in business: techniques and applications for the operations researcher, Computers & Operations Research 27 (11–12) (2000) 1023–1044.
[49] S.K. Tanbeer, C.F. Ahmed, B.-S. Jeong, Y.-K. Lee, Sliding window-based frequent pattern mining over data streams, Information Sciences 179 (22) (2009) 3843–3865.
[50] C.-F. Tsai, Feature selection in bankruptcy prediction, Knowledge-Based Systems 22 (2) (2009) 120–127.
[51] P.M. Tsang, P. Kwok, S.O. Choy, R. Kwan, S.C. Ng., J. Mak, J. Tsang, K. Koong, T.L. Wong, Design and implementation of NN5 for Hong Kong stock price forecasting, Engineering Applications of Artificial Intelligence 20 (4) (2007) 453–461.
[52] N.K. Vitanov, K. Sakai, Z.I. Dimitrova, SSA, PCA, TDPSC, ACFA: useful combination of methods for analysis of short and nonstationary time series, Chaos, Solitions and Fractals 37 (1) (2008) 187–202.
[53] L. Yu, S. Wang, K.K. Lai, Mining stock market tendency using GA-based support vector machines, Lecture Notes in Computer Science 3828 (2005) 336–345.
[54] F.M.H. Zarandi, B. Rezaee, I.B. Turksen, E. Neshat, A type-2 fuzzy rule-based expert system model for stock price analysis, Expert Systems with Applications 36 (1) (2009) 139–154.

Dr. Chih-Fong Tsai obtained a PhD at School of Computing and Technology from the University of Sunderland, UK in 2005 for the thesis entitled “Automatically Annotating Images with Keywords”. He is now an associate professor at the Department of Information Management, National Central University, Taiwan. He has published over 20 refereed journal papers including *ACM Transactions on Information Systems*, *Pattern Recognition*, *Information Processing & Management*, *Applied Soft Computing*, *Neurocomputing*, *Knowledge-Based Systems*, *Expert Systems with Applications*, *Expert Systems*, *Online Information Review*, *International Journal on Artificial Intelligence Tools*, *Journal of Systems and Software*, etc. In 2008, he received the ‘Highly Commended Award’ (Emerald Literati Network 2008 Awards for Excellence) for a paper published in *Online Information Review* (“A Review of Image Retrieval Methods for Digital Cultural Heritage Resources”). His current research focuses on multimedia information retrieval and data mining applications.

Miss Yu-Chieh Hsiao received the Master's degree from the Department of Accounting and Information Technology, National Chung Cheng University, Taiwan. Her research interest focuses on data mining applications.