Expert Systems with Applications 38 (2011) 2566–2576

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Dynamic financial distress prediction using instance selection for the disposal of concept drift

Jie Sun *, Hui Li

School of Economics and Management, Zhejiang Normal University, Jinhua City 321004, Zhejiang Province, PR China

ARTICLE INFO

Keywords:
Dynamic financial distress prediction
Concept drift
Instance selection
Multivariate discriminant analysis

ABSTRACT

Prior studies of financial distress prediction (FDP) all focus on static modeling and ignore whether the model is still suitable with time passing on. This paper devotes to the first investigation on what the concept of financial distress concept drift (FDCD) is, whether FDCD exists and how to dispose FDCD. We construct a dynamic FDP modeling based on instance selection for the disposal of FDCD. Dynamic FDP consists of instance selection, FDP modeling and future prediction. Instance selection methods including full memory window, no memory window, window of fixed size, window of adaptable size, and batch selection are used to tackle FDCD. For feature selection, we construct a wrapper by integrating forward and backward selections on Mahalanobis distance. Empirical results indicate that gradual and constant virtual concept drift does exist in FDP, and dynamic FDP models perform much better than static models. Meanwhile, window of fixed size and batch selection are more suitable for Chinese listed companies' dynamic FDP.

1. Introduction

Enterprise financial distress is a dangerous state that a firm faces when the firm suffers from serious outside frustration or inner financial activities are out of control. Such distress includes various conditions, such as low liquidity, inability to pay debts or dividend of preference stock, substantial and continual reduction in profitability, and bankruptcy. These conditions indicate financial distress from mild to serious in sequence. Financial distress is the synthetic reflection of deterioration of inner and outside risky factors of an enterprise. Even enterprise distress caused by non-financial factors tends to end up with financial distress. Therefore, financial distress prediction (FDP) is a very important tool for enterprise risk management. In current years, global financial crisis shows the fragility of international financial system, which further requires banks to strengthen risk management. Because most of bank capital flows to enterprises, banks care about the current and future financial states of their enterprise customers very much and FDP can directly serve banks' credit scoring. As a result, necessity of research on FDP gradually becomes the consensus of practitioners and academicians, and this area has been focused in recent years (Bose, 2006; Bose & Pal, 2006; Hu, 2008; Li & Sun, 2009, 2010; Premachandra, Bhabra, & Sueyoshi, 2009; Psillaki, Tsolas, & Margaritis, 2010; Sun & Li, 2009; Sun & Shenoy, 2007; West, Dellana, & Qian, 2005).

Fitzpartrick (1932) pioneered the research of FDP by analyzing financial ratios, which was followed by Beaver (1966) with univariate model. Altman (1968) firstly attempted to use multiple discriminant analysis (MDA) to construct a *Z*-score model, which was considered as a milestone in the field. After that, FDP research entered into multivariate stage, and various statistical models and intelligent models were applied to solve the problem of FDP. Ohlson (1980) broke through the linear separation and utilized the nonlinear Logit model to describe the nonlinear relationship between financial distress probability and financial ratios. In early 1990s, neural networks (NNs) began to be used in FDP and its performance was often compared with that of MDA and Logit. Odom and Sharda (1990) utilized back propagation NNs with the same financial ratios of Altman and compared it with MDA. Since then, lots of researchers devoted themselves to the study of FDP with NNs. Though minority of them made the conclusion that NNs model did not outperform statistical ones (Boritz & Kennedy, 1995; Etheridge & Sriram, 1997), the majority provided the supporting evidence that NNs model was superior to statistical ones in terms of prediction accuracy (Carlos, 1996; Fletcher & Goss, 1993; Leshno & Spector, 1996; Pendharkar, 2005; Yang, Platt, & Platt, 1999; Zhang, Hu, Patuwo, & Indro, 1999). However, NNs is a black-box method, and needs a lot of samples to train the structure in order to avoid over-fitting, and its training costs lots of time.

Besides NNs, other intelligent methods, such as decision tree (DT), genetic algorithm (GA), rough sets (RS), and case-based reasoning (CBR), were also applied in FDP. Frydman, Altman, and Kao (1985) used DT in forecasting financial distress. Shin and Lee

* Corresponding author. Tel.: +86 134 5494 2829.
E-mail address: sunjiehit@gmail.com (J. Sun).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.08.046

(2002) and Kim and Han (2003) respectively utilized GA to draw quantitative rules and qualitative rules for bankruptcy prediction. While, their rules' coverage were both relatively low. McKee (2000) developed a RS-based bankruptcy prediction model and applied it to forecast financial distress with USA data. CBR-based FDP methods usually use the k-nearest neighbor algorithm. Sun and Hui (2006) put forwards a FDP method on similarity weighted voting CBR with its application to Chinese listed companies' FDP. Support vector machine (SVM) is a relatively new intelligent model in pattern recognition. Both Shin, Lee, and Kim (2005) and Min and Lee (2005) applied SVM to Korean bankruptcy prediction and drew the conclusion that SVM outperforms MDA, Logit and NNs in predictive ability. Hui and Sun (2006) and Ding, Song, and Zen (2008) attempted to use SVM model for Chinese listed companies' FDP and obtained similar conclusion. Both Min, Lee, and Han (2006) and Wu, Tzeng, Goo, and Fang (2007) integrated GA with SVM to improve the predictive ability of SVM for FDP. The former uses GA to optimize both input features and model parameters, and the latter only uses GA to optimize model parameters. Hua, Wang, Xu, Zhang, and Liang (2007) interpreted and modified the outputs of SVM classifiers according to the result of logistic regression analysis, which was named as integrated binary discriminant rule by them.

Sun and Li (2008) put forward a FDP ensemble model by using weighted majority voting combination of diversified multiple classifiers, which was constructed by different classification algorithms on the same dataset. As a result, the ensemble got higher average accuracy and lower variance and coefficient of variation than any base classifier. Tsai and Wu (2008) compared the performance of the single NNs classifier with the (diversified) multiple NNs classifiers over three datasets for bankruptcy prediction and credit scoring problems. They found that NNs ensemble did not outperform the single best NNs classifier in many cases, which was attributed to too little training dataset for diversified single classifiers and the binary classification problem of FDP. Nanni and Lumini (2008) made a comparison among several ensembles for bankruptcy prediction and credit scoring, and showed that the method of Random Subspace outperformed the other ensemble methods. Hung & Chen (2009) proposed a selective ensemble for bankruptcy prediction, which selects some suitable classifiers on the base of an expected probability of each individual classifier to replace the voting policy.

These previous researches, including single classifier prediction and ensemble prediction, play an important role in forecasting financial distress, and almost all of them use MDA as a benchmark model for comparison. However, a common drawback of these researches is that they all focus on static modeling for prediction, that is, predictive models are constructed only with sample data of a certain period of time. With time passing on, static models can not effectively forecast financial distress in the changing economic environment or the changing enterprise operational environment. In the changing real world, new financially distressed enterprises gradually emerge to form sample data flow. Thus, the volume of sample data that are used to construct FDP model dynamically increases with the time passage. This characteristic of FDP in real world had never been focused on. Thus, in order to fit enterprises' dynamic operational environments with time passage, researches on dynamic FDP modeling should be conducted by breaking the assumption of all previous researches that sample volume never changes in FDP.

In this research, we define financial distress concept drift as the change in distribution of training data which dynamically increase with continuous emergence of new enterprises in financial distress, and attempt to empirically investigate whether this concept drift exists in FDP and to dispose it by using dynamic FDP modeling on the base of various sample selection methods. Since the research of MDA-based FDP is a milestone in this area and MDA is also the most classic and frequently used model in forecasting financial distress, we empirical test the feasibility and effectiveness of dynamic modeling for FDP by using the predictive model of MDA. This paper is organized as follows: Section 2 presents what financial distress concept drift is. Section 3 presents how to handle financial distress concept drift by using dynamic modeling with instance selection. Section 4 designs an experiment and discusses the results. Section 5 makes conclusion.

2. What is financial distress concept drift?

Concept drift is generally known as changes in the target concept, and these changes are induced by changes in the hidden context (Schlimmer & Granger, 1986). This definition refers to real concept drift. However, hidden changes in context may not directly or suddenly cause a change in the target concept, but just gradually cause a change in the underlying data distribution with the target concept remaining the same. For example, with time passage and increase of sample data volume for FDP, optimal features and useful knowledge for future FDP may also change. This kind of change belongs to change in data distribution. When such data distribution change happens, a model re-building process is necessary since the old model's error rate may no longer be acceptable. This phenomenon is called virtual concept drift (Delanya, Cunninghamb, Tsymbalb, & Coyle, 2005; Widmer & Kubat, 1993). Either real drift or virtual drift of the concept of financial distress may happen in real world.

For real concept drift of financial distress, the target concept of financial distress defined by a concrete enterprise will change when it evolves from one stage of enterprise life cycle to the other. For example, financial distress of enterprise in the starting-up period can be defined as deficiency of liquidity or cash flow difficulty, because an enterprises in this period is likely to have low profitability but still have an opportunity to grow up if the firm has relatively smooth cash flow. When an enterprise has entered the growing period, profitability quickly grows up and the enterprise has strong desire to expand capital. This objective may cause the firm to increase the amount of liability. Therefore, financial distress can be defined as excessive risk of financial leverage. An enterprise in maturing period will have stable and relatively high profitability. Thus financial distress can be defined as substantial or consecutive reduction in profitability. For an enterprise in recession period, it is rational to define financial distress as insolvency or bankruptcy. Such kind of change in financial distress concept due to the dynamic evolution of an enterprise can be considered as real concept drift of financial distress, which requires the company to re-build or re-choose FDP model whenever a new financial distress concept is defined.

For virtual concept drift, the underlying data distribution will gradually change with new sample data flowing in, although the target concept of financial distress does not change. The reason why data distribution changes, is that new characteristics of financial distress may appear with the variation of economic environment. Consider the condition that the target concept of financial distress is defined as negative profit in two successive years and a FDP model is constructed with sample data of a certain period. The model will gradually become unsuitable for the future prediction when the data starts drifting. Thus, it is imperative that the old FDP model should be revised or a new one be re-built.

Although both real and virtual financial distress concept drift will occur, this paper focuses on virtual concept drift for the following reason. When the target concept of financial distress defined by a company changes, namely real concept drift happens, it is of course necessary for the company to re-choose a FDP model suitable for the new financial distress concept or re-build a FDP model by re-selecting instances according to the new criteria. However, in the case that the target concept of financial distress remains unchanged, more and more instances, which accords with the unchanged target concept but shows some different characteristics, still keep emerging. Thus, it is worthy of study whether virtual concept drift of financial distress exists and how to re-build FDP model dynamically if it does exist.

3. Dynamic modeling for FDP on the base of instance selection

3.1. Framework

Instance selection is the most common technique in handling concept drift and this method only uses instances related to the current ones when constructing a predictive model. The most basic idea of handling concept drift by instance selection is to keep rebuilding model from a window that moves over recently emerging instances and use the learnt concepts for prediction only in the immediate future. This idea is on the assumption that new instances are always more important than old ones in making predictions and recent sample data will contain the information of current concept which needs to be learnt by predictive models (Klinkenberg, 2004).

New data for FDP will be collected in batches at time intervals, e.g., every one year, or every one season. More specifically, data of new companies in financial distress and corresponding healthy ones at each interval (assume the interval is a year in this research) are collected as a new batch of data. And these new data can be used to constitute the panel data flow for dynamic FDP modeling. Because the number of instances in each new batch at each interval is often not the same, we define the window on the base of the unit of year (batch) instead of instance number. Thus, the window size can be one year (batch), two years (batches), three years (batches), and so on, in place of some units of fixed number of instances. The framework of dynamic FDP modeling on the base of instance selection is illustrated as Fig. 1.

At each current year, instances that are used for modeling should first be selected from all available batches according to some instance selection mechanism. Those instances whose future financial state is to be predicted are called as current prediction instances. The near space of current prediction instances, namely the most recent batch, can be considered as criteria of instance selection by some instance selection methods. Meanwhile, a fast and effective feature selection method should be applied to the process of dynamic FDP modeling. Particularly, when the near space of current prediction instances is considered as criteria of instance selection, feature selection should be embedded into the process of instance selection. Based on the instances selected, a FDP model can be constructed by utilizing some kind of classification algorithm. The used classifying algorithm should be fast as well as effective, since the handling of concept drift will cost much time. In this research, we use Fisher discriminant analysis as the classification method for its simpleness, fastness as well as effectiveness. However, some other classification algorithms may be universally applicable in the proposed dynamic FDP system. After the dynamic model has been constructed, whether the current prediction instances will run into financial distress can be predicted. With the time passing by, the current year's prediction instances will gradually become new training instances when their financial state proves to be financial distress or not. Thus, a new batch of panel data can be formed. The process of model re-building, including instance selection, feature selection and model construction, should be carried out once again.

3.2. Methods tackling financial distress concept drift

Five windowing or instance selection methods, including full memory and no memory, fixed size and adaptable size, and batch selection, are to be used to for dynamic FDP modeling. And we also attempt to empirically test which one is more suitable for the domain of FDP when handling concept drift. These five methods are respectively presented as follows.

3.2.1. Full memory and no memory time window

Full memory time window method assumes that forgetting is not necessary in the dynamic FDP modeling. The learning machine generates its classification model from all previously seen instances and new instances are added to the window as they emerge at intervals. Meanwhile, no old ones are deleted from the window. As shown in Fig. 2(a), the window size becomes bigger and bigger with the current time point switches from year t to year $(t+1)$ and then from year $(t+1)$ to year $(t+2)$. This idea is very easy to be processed. However, the drawback of this method is that full memory time window cannot adapt to the new concept well if concept drift is strong because the new model inherits both old concept and new concept. Therefore, full memory time window is only suitable for very mild concept drift in which both old and new information is important. Another drawback of this method is that full memory time window will gradually become too large especially when the amount of data flowing in each year is big with time passage.

No memory time window refers to using a window of fixed size of one recent batch. This method assumes that data of former batches are unrelated with the current concept, and a new model should be built from the most current batch at each new time point by forgetting all old information. Fig. 2(b) shows the idea of no memory time window. Since only new data are used, the model can be quickly adapted to handle concept drift. This idea is entirely opposite to full memory time window. The drawback of this method is that models constructed from no memory time window often

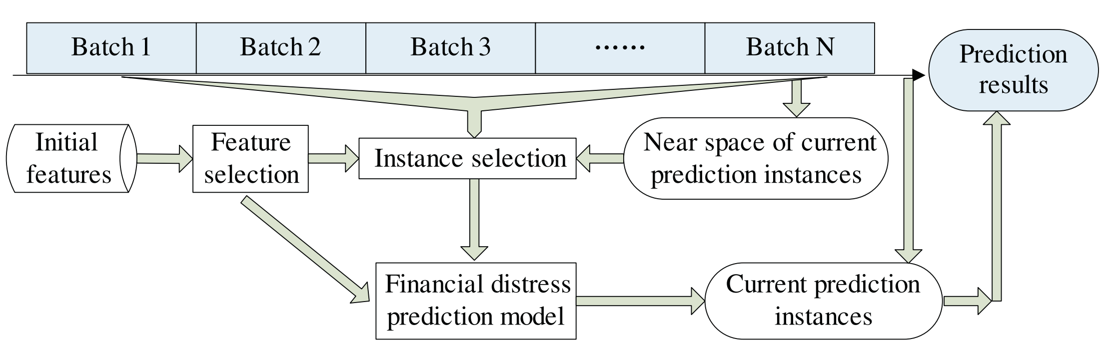


Fig. 1. Framework of dynamic FDP based on instance selection.

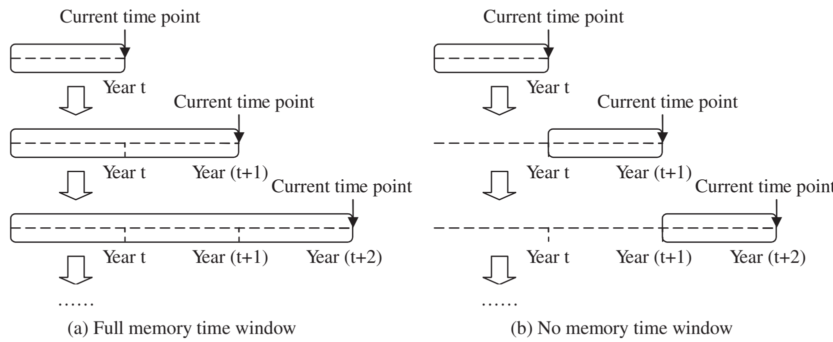


Fig. 2. Full memory time window and no memory time window.

lacks generalization for the too limited number of training number in the time period when concept keeps stable.

3.2.2. Time window of fixed size and adaptable size

For time window of fixed size, the key and difficult problem is how to choose an appropriate window size. Though small window has strong adaptability to the fast concept drift, yet it lacks generalization ability when the concept drift is mild because of the limited number of instances. In contrast, large window ensures good generalization in phases without concept drift, but it remembers too much old information that is not suitable for the new instances when concept drift happens (Klinkenberg, 2004). Taking the window size of two years for example, Fig. 3 illustrates dynamic FDP modeling on the base of time window of fixed size.

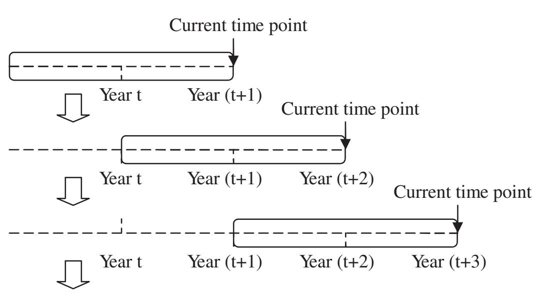


Fig. 3. Time window with fixed size of two years.

For time window of adaptive size, the window size is adjusted by some mechanism. Widmer and Kubat (1996) proposed the adaptive time window by heuristics, involving some parameters that are difficult to tune. Klinkenberg & Joachims (2000) presented an approach to select the window size so that the estimated generalization error on the newest batch is minimized. Supposing the current time point is year $(t+m)$, there are totally $(m+1)$ possible window sizes, as shown in Fig. 4. Assuming that the most recent batch of year $(t+m)$ is the most similar to the coming future, estimation error on the last batch of $(t+m)$ should be calculated for the FDP model constructed on each possible window size. And the window size with the minimum estimation error should be chosen. Therefore, feature selection and classification model construction should be carried out for each possible window size in the process of determining the adaptable window size.

3.2.3. Batch selection method

Klinkenberg (2004) proposed batch selection method to deal with concept drift, which differs from traditional windowing methods because selected instances do not cover several adjacent recent batches. Instead, this method selects batches that are similar to the most recent batch no matter when they happen. Basic idea of this method is shown as Fig. 5. Firstly, a model is learned on the most recent batch of $(t+m)$. Though this model is not good enough for future prediction in many cases, it is the most up-to-date one representing the current concept. Therefore, this model can be used to judge which former batches were generated from the similar concept to the most recent batch by comparing the

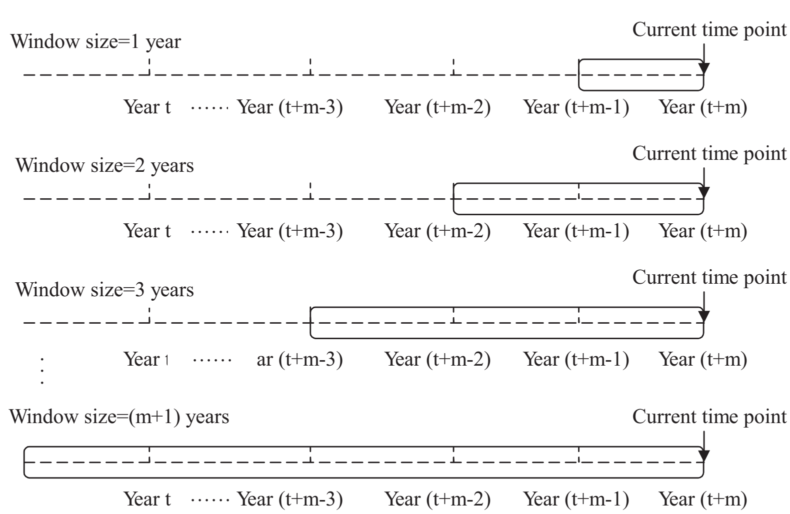


Fig. 4. Possible windows when the current time point is year $(t+m)$.

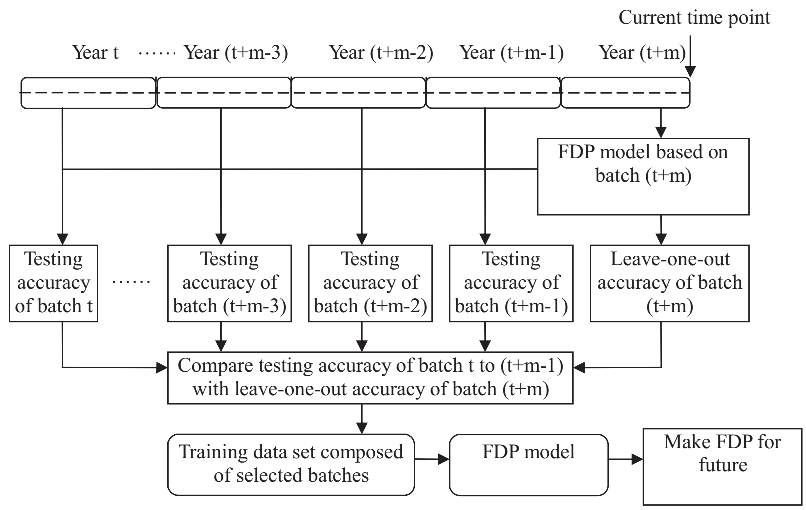


Fig. 5. Batch selection method.

model's testing accuracy on former batches with its leave-one-out accuracy on batch $(t+m)$. Suppose that the model's testing accuracy on batch i is denoted as TA_i $(i = t, t+1, \ldots, t+m-1)$, and leave-one-out accuracy on batch $(t+m)$ is denoted as $LA_{(t+m)}$. The batches that meet the condition: $1 - TA_i \leqslant p \times (1 - LA_{(t+m)})$ are selected to constitute training data set together with batch $(t+m)$. Here, p is the batch selection threshold parameter. The smaller p is, the less batches are selected and the more quickly the model constructed can adapt to the new concept after drifting. However, the drawback of this method is that only batches that are similar to the current one are selected with the possible result that lots of information is omitted. When concept keeps stable, too small value of the parameter p will lead to unsatisfying prediction performance because it omits too much useful information contained in former batches. In the process of batch selection, feature selection and classification model construction should also be carried out to build FDP model on the most current batch.

3.3. Feature selection

The initial features for dynamic FDP cover seven aspects, i.e. profitability, activity, solvency, growth, risk level, per share ratios, and cash flow ratios, which can provide comprehensive indication of firms' financial state. Each aspect is composed of several financial ratios, as listed in Table 1.

There are two categories of feature selection methods, that is, filter and wrapper. The former is independent of the classifier, while the latter uses the classifier's accuracies or criteria derived from the classifier to rank the discriminative power of the possible feature subsets (Ng, Yeung, Firth, Tsang, & Wang, 2008). In the dynamic FDP system, we use a wrapper, which integrates forward and backward selection on Mahalanobis distance in PRtools 4.1 (Heijden, Duin, Ridder, & Tax, 2004). Namely, whenever the classification model is to be constructed on certain data set in the process of dynamic FDP, the feature selection process should firstly be implemented and the model's accuracy is also used as

Table 1
Financial ratios used as initial features.

Category	Variable and feature	
Profitability	V01: Gross income to operating revenue	V02: Net profit to operating revenue
	V03: Earning before interest and tax to total asset	V04: Net profit to total assets
	V05: Net profit to current assets	V06: Net profit to fixed assets
	V07: Profit margin	V08: Net profit to equity
	V09: Return on invested capital	
Activity	V09: Account receivables turnover	V10: Inventory turnover
	V11: Account payable turnover	V12: Working capital turnover
	V13: Current assets turnover	V14: Fixed assets turnover
	V15: Long-term assets turnover	V16: Total assets turnover
	V17: Net assets turnover	
Solvency	V18: Current ratio	V19: quick ratio
	V20: Working capital ratio	V21: Asset-liability ratio Equity to debt ratio
	V22: Proportion of current assets to total assets	V23: Proportion of fixed assets to total assets
	V24: Proportion of equity to fixed assets	V25: Current liability to total liabilities
	V26: Debt to tangible assets ratio	V27: Ratio of liabilities to market value of equity
Growth ratios	V28: Growth rate of prime operating revenue	V29: Rate of capital preservation and appreciation
	V30: Growth rate of total assets	V31: Growth rate of net profit
Risk level	V32: Coefficient of financial leverage	V33: Coefficient of operating leverage
Per share ratios	V24: Operating revenue per share	V35: Earning per share
	V36: Net assets per share	
Cash flow ratios	V37: Cash flow to current liabilities ratio	V38: Cash rate of prime operating revenue
	V39: Net operating cash flow per share	V40: Net cash flow per share
	V41: Net operating cash flow to net profit ratio	

one of the criteria for feature selection. The algorithm is presented as follows.

Input. (1) Initial feature set, denoted as *Initial_feature*; (2) the number of features to be selected by forward selection method, denoted as *N*.

Output. Feature subset selected, denoted as *Selected_feature*.

```
Algorithm
//Use forward selection method based on Mahalanobis
   distance to select certain number of good features
Feature_num = N
Forw_feature = Forward (Initial_feature, Feature_num)
//Use backward selection method based on Mahalanobis
   distance to delete one feature from Forw_feat at each time
   and evaluate the remaining features set by model accuracy
Selected_feature = Forw_feature
Best_accuracy = Model_accuracy (Selected_feature)
Do the following
  Feature_num = Feature_num − 1
  Backw_feature = Backward
  (Selected_feature, Feature_num)
  Model_accuracy = Model_accuracy (Backw_feature)
  If Best_accuracy < Model_accuracy
    Best_accuracy = Model_accuracy
    Selected_feature = Backw_feature
  End
While Feature_num > 1
```

4. Experiments and discussion

4.1. Experimental design

The empirical experiment is designed to test whether this concept exists, to analyze the characteristic of financial distress concept drift, and to compare the performances of dynamic FDP methods based on instance selection with those of stationary FDP models. To simulate the time evolution process, we assumed that the current year gradually shift from 2002 to 2007. Thus, the sample companies stream in year after year in the form of batches. For a certain current year, FDP model is constructed according to the information of financial distress companies and their matching companies which are available in the current year. Since the model constructed in the current year has the ability of FDP two years in advance, the model's performance can be tested by the future two years' ST companies and their matching companies as Fig. 6 shows.

The current year is supposed to be 2002 at first. Thus, the performance of the newly constructed FDP model in 2002 can be tested by comparing the model's FDP results for 2003 and 2004 with the real financial distress results of 2004 and 2005. The same operation is conducted from 2003 to 2007 when they become the current year. Superior to former FDP experiment described in other researches, our experimental process not only utilizes panel data to build FDP model, but also tests the FDP model by the real future information. Hence, time passage and financial distress concept drift are considered, and different dynamic FDP methods based on instance selection to deal with financial distress concept drift problem can be analyzed.

4.2. Data collection

Because the experiment is carried out on real world information of Chinese listed companies, companies in financial distress are defined as those that are specially treated (ST) by Chinese Stock Exchange due to abnormal financial status. Chinese listed companies will be specially treated for the following two common reasons: (1) a company has had negative net profit in consecutive two years, or (2) its net capital per share is lower than the face value per share. Our study chooses financial distress samples according to these ST criteria. There are also another two reasons for ST: (1) companies purposely publish financial statements with serious false and misstatement, or (2) with other abnormal incidents described in

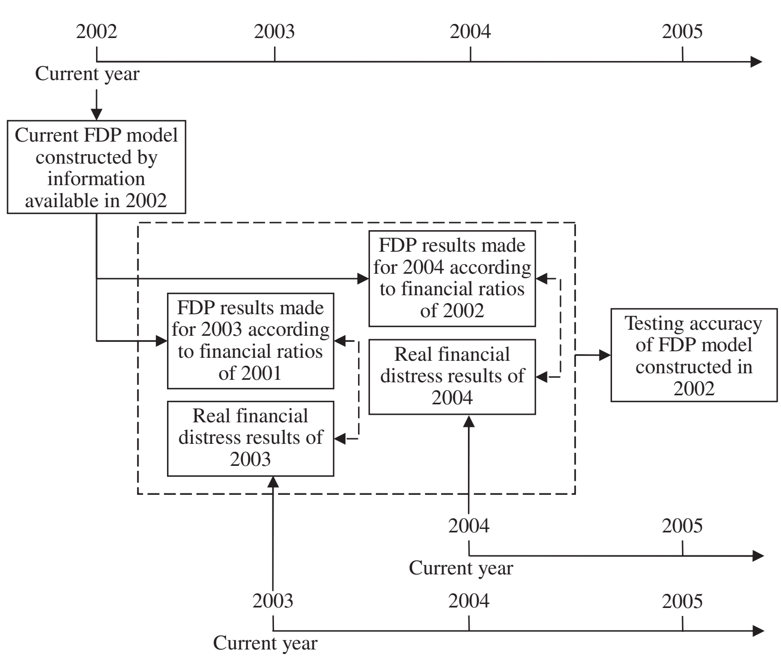


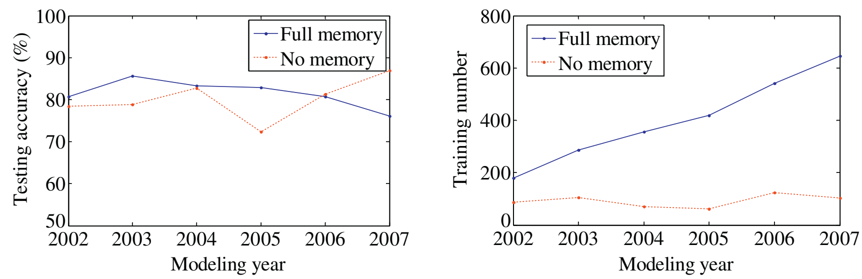
Fig. 6. Experimental design of dynamic FDP modeling and testing.

Table 2
Sample numbers of different years.

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008
Sample number	50	44	86	106	70	62	124	104	46

Table 3
Testing accuracies and training numbers of full memory and no memory windows.

Modeling year		2002	2003	2004	2005	2006	2007	Average
Full memory	Testing accuracy (%)	80.68	85.61	83.33	82.89	80.67	76.09	82.24
	Training number	180	286	356	418	542	646	
No memory	Testing accuracy (%)	78.41	78.79	82.80	72.37	81.33	86.96	78.76
	Training number	86	106	70	62	124	104	
Testing number		176	132	186	228	150	46	



(a) Testing accuracy of full memory and no memory (b) Training number of full memory and no memory

Fig. 7. Testing accuracies and training numbers of full memory window and no memory window.

Table 4
Testing accuracies and training numbers of fixed time window and adaptable time window.

Modeling year		2002	2003	2004	2005	2006	2007	Average
Window width = 2 years	Testing accuracy (%)	86.36	82.58	81.18	83.33	80.00	82.61	82.79
	Training number	130	192	176	132	186	228	
Window width = 3 years	Testing accuracy (%)	80.68	81.82	81.72	85.09	77.33	78.26	81.48
	Training number	180	236	262	238	256	290	
Window width = 4 years	Testing accuracy (%)	80.68	85.61	85.48	83.77	81.33	84.78	83.44
	Training number	180	286	306	324	362	360	
Window width = 5 years	Testing accuracy (%)	80.68	85.61	83.33	84.21	84.67	80.43	83.44
	Training number	180	286	356	368	448	466	
Window width = 6 years	Testing accuracy (%)	80.68	85.61	83.33	82.89	82.67	80.43	82.79
	Training number	180	286	356	418	492	552	
Adaptable time window	Testing accuracy (%)	80.68	85.61	81.18	82.89	82.67	82.61	82.46
	Training number	180	286	176	418	492	228	
Testing number		176	132	186	228	150	46	

Chinese Stock Listing Exchange Rule appear. These types of companies are excluded from our study to focus on truly financial distress.

In the experiment, FDP is made two years prior to distress. FDP models are constructed from financial ratios of the report year, which is two years before the ST year. It means that financial distress is attempted to be predicted according to financial ratios information two years in advance. Healthy companies are chosen from those that have never been specially treated by the matching method considering both industry and asset size. According to the information of Chinese listed companies from 2000 to 2008, the initial samples consist of totally 692 companies listed in Shenzhen Stock Exchange and Shanghai Stock Exchange. They divide into nine batches and one year corresponds to one batch. Table 2 lists the sample number of each year (batch).

4.3. Base predictive model

The study includes using Fisher discriminant analysis as the base predictive method.

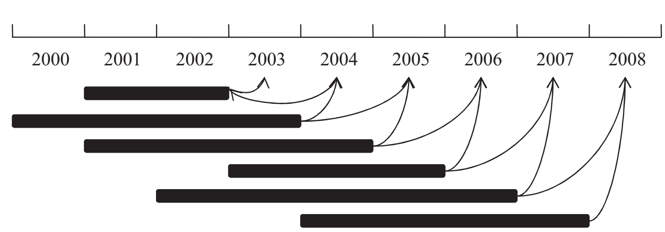


Fig. 8. Best window size at each modeling year.

Table 5
Testing accuracy and training number of batch selection.

Modeling year		2002	2003	2004	2005	2006	2007	Average
Threshold = 1	Testing accuracy (%)	78.41	78.79	82.80	72.37	81.33	86.96	78.76
	Training number	86	106	70	62	124	104	
Threshold = 1.5	Testing accuracy (%)	86.36%	82.58	82.80	72.37	81.33	82.61	80.61
	Training number	130	242	70	62	124	228	
Threshold = 2	Testing accuracy (%)	86.36	85.61	81.72	72.37	81.33	82.61	81.36
	Training number	130	286	262	62	124	228	
Threshold = 2.5	Testing accuracy (%)	80.68%	85.61%	81.72%	83.33%	80.00%	89.13%	82.57%
	Training number	180	286	262	298	230	404	
Threshold = 3	Testing accuracy (%)	80.68	85.61	83.33	84.21	82.00	84.78	83.22
	Training number	180	286	356	368	406	584	
Threshold = 3.5	Testing accuracy (%)	80.68	85.61	83.33	82.89	80.00	84.78	82.57
	Training number	180	286	356	418	456	584	
Threshold = 4	Testing accuracy (%)	80.68	85.61	83.33	82.89	80.67	76.09	82.24
	Training number	180	286	356	418	542	646	
Threshold = 4.5	Testing accuracy (%)	80.68	85.61	83.33	82.89	80.67	76.09	82.24
	Training number	180	286	356	418	542	646	
Testing number		176	132	186	228	150	46	

4.4. Experimental results and analysis

4.4.1. Tracking financial distress concept drift with full memory and no memory time window

Though full memory and no memory windowing methods usually do not perform very well in dealing with concept drift problems, they can be used to track whether financial distress concept drift exists and what type of concept drift it is. Experimental results of full memory and no memory are listed in Table 3 which includes information of testing accuracy, training number and testing number of the modeling years from 2002 to 2007. 2000, 2001 are excluded for the reason that only one batch or two batches of data are available and comparison among different instance selection methods for these two modeling years are of little sense. 2008 is also not included in the modeling years because the 2009 batch is not available yet and the FDP model constructed on the 2008 batch cannot be tested. The information of testing accuracy and training number is graphed in Fig. 7. With time going on, the training number of full memory will continually expand, and that of no memory often varies in a certain range which is determined by the available sample number of each year. Assuming that the concept of financial distress does not change at all, the FDP model's testing accuracy should theoretically keep at a certain level or even increase when being dynamically trained with more instances. However, as Table 3 and Fig. 7 shows, the testing accuracy of full memory increases from 80.68% at 2002s model to 85.61% at 2003s model, and then begins to continuously decrease to 76.09% until 2007s model. This indicates that more and more useless old information is contained in the newly updated full memory model with time going on and economic environment varying. Therefore, financial distress concept drift does exist and need to be considered in the domain of FDP. Furthermore, testing accuracy of no memory gradually increases from 78.41% at 2002s model to 82.80% at 2004s model, and then drops to 72.37% at 2005s model, after which it begins to increase to 86.96% at 2007s model. One financial distress concept gradually forms from 2002 to 2005 because 2004s no memory model can still predict future well. While, 2005s no memory model's prediction accuracy abruptly and significantly drops, representing that financial distress concept learned in 2005s no memory model is not suitable for future prediction, or another new financial distress concept begins to form after 2006.

4.4.2. Results of fixed time window and adaptable time window

Testing accuracies, training and testing numbers of fixed time window with different sizes and adaptable time window are listed in Table 4, in which the highest testing accuracy among different window sizes at each modeling year is in underlined. To illustrate it more clearly, the corresponding best window size at each modeling year is shown in Fig. 8. The best modeling window size for predicting 2003 and 2004 financial distress is two years' data which covers 2001 and 2002 data sets. Thereafter, the best window size for predicting future financial distress is all around four years, which means that the best modeling window size at 2003, 2004 and 2007 is four years although the best modeling window size at 2005 is three years and that at 2006 is five years. Though average testing accuracies of the fixed window size of four years and five years are both the highest among all window sizes, i.e. 83.44%, yet the window width of four years is most suitable for Chinese listed companies' dynamic FDP than other window sizes according to our empirical results.

As indicated in Table 4, the testing accuracy of adaptable time window at each modeling year does not perform better than time windows of fixed size. And its average testing accuracy of all six modeling years is 82.46%, which is only higher than that of fixed window of 3 years and is lower than those of all other fixed

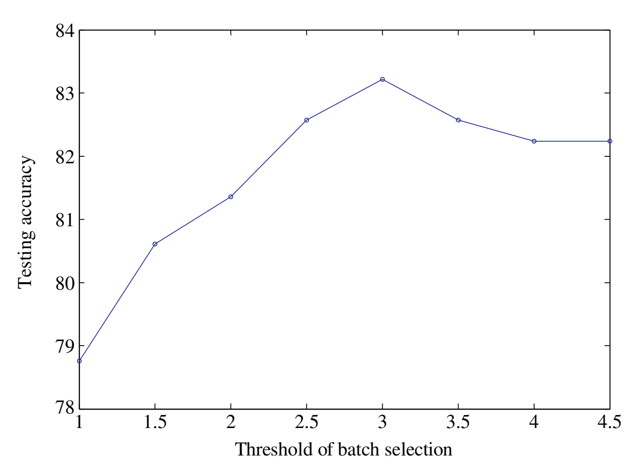


Fig. 9. Average testing accuracies with different batch selection thresholds.

Table 6
Testing accuracies and training numbers of stationary FDP models.

Dynamic modeling year		2002	2003	2004	2005	2006	2007	Average
Stationary model of 2000	Testing accuracy (%)	63.07	62.12	65.59	64.04	59.33	58.70	62.85
	Training number	50	50	50	50	50	50	
Stationary model of 2001	Testing accuracy (%)	65.91	62.88	73.12	73.68	70.67	69.57	70.63
	Training number	94	94	94	94	94	94	
Stationary model of 2002	Testing accuracy (%)	80.68	78.03	81.18	82.46	81.33	76.09	80.72
	Training number	180	180	180	180	180	180	

windows. This result indicates that the characteristics of financial distress are more frequently changeable than other concepts, e.g. text search preference, and historic characteristics of financial distress of the past few years are likely to reappear in the future. Consequently, the so-called adaptable window size which has the strict minimum estimation error for the current year's data set may often not really adapt to the coming future years. Furthermore, the adaptable time window costs much more time than fixed time window in the modeling stage because it carries out global search for the adaptable window size.

4.4.3. Results of batch selection method

Testing accuracies, training and testing numbers of batch selection methods with different threshold parameter values are listed in Table 5, and the average testing accuracies with different threshold parameter values are graphed in Fig. 9. The highest testing accuracies among different threshold parameter values at each modeling year are in underlined in Table 5, which shows dynamic FDP based on batch selection with the threshold parameter equaling to 3 gets the highest testing accuracy at four modeling years among all sixe modeling years. And it also has the highest average testing accuracy of 83.22%. In contrast, batch selection methods with other threshold parameter values get the highest testing accuracy at only two years or one year. As can be seen from Fig. 9, the average testing accuracy ascends from 78.76% to 83.22% when the batch selection threshold varies from 1 to 3 at interval of 0.5. While the average testing accuracy begins to descend when the batch selection threshold becomes further bigger. This result indicates that dynamic FDP for Chinese listed companies based on batch selection method is effective when the batch selection threshold is set around 3. When the batch selection threshold is too small, the instances selected for modeling are limited to the data whose financial distress characteristics are very similar to the most current batch. Thus, the model constructed from selected batches for future FDP will over-fit the current batch and will not be competent for future prediction, especially when the characteristics of future financial distress differ from the current batch's to some extent. Besides, the number of the selected instances will be too small to train a good enough FDP model when the batch selection threshold is too small.

As can be seen from Table 5, the training numbers are relatively small when the batch selection threshold equals to 1 or 1.5, and the average testing accuracies are relatively low too. On the other hand, when the batch selection threshold is too big, too much irrelevant information will be included in the selected batches, which also makes the model not suitable for future FDP. Therefore, how to set an appropriate batch selection threshold is the key problem of dynamic FDP based on batch selection, and our study provides some empirical support for solving this problem.

Table 7
Average testing accuracies of dynamic and stationary FDP methods.

Modeling methods of FDP	Average testing accuracy (%)
Full memory time window	82.24
No memory time window	78.76
Best fixed time window	83.44
Adaptable time window	82.46
Best batch selection	83.22
Stationary model of 2002	80.72
Stationary model of 2001	70.63
Stationary model of 2000	62.85

4.4.4. Comparison between dynamic and stationary FDP

To make clear whether dynamic FDP methods based on instance selection performs better than stationary FDP models, three stationary models are respectively constructed at the time point of 2000, 2001 and 2002. And the same testing instances as the above dynamic modeling methods are used to test their effectiveness in the process of time passage. As shown in Tables 6 and 7, the stationary model of 2000 has very poor prediction performance for future financial distress and the average testing accuracy is only 62.85%. On one hand, the financial distress concept after 2002 has drifted a lot from 2000, which makes the stationary model of 2000 unsuitable for FDP after 2002. On the other hand, the instance number of 2000 batch is too limited to build an effective discriminant model. The average testing accuracy of the stationary model of 2001 is much higher than that of 2000, because it further considers the financial distress characteristics of 2001 which is more similar to future financial distress concept, and at the same time the training number increases for the inclusion of 2001 batch. If the stationary model is built at 2002, the average prediction accuracy for future financial distress can be further enhanced to 80.72% for the similar reason of 2001. However, it is still much lower than the average testing accuracies of all other dynamic FDP modeling methods except no memory time window. Among all dynamic FDP modeling methods, fixed time window method gets the highest average testing accuracy of 83.44% with the window size properly set, and it is followed by batch selection method, whose average testing accuracy is 83.22% if the threshold parameter is appropriately set. Full memory time window method and adaptable time window method have similar medium performances, and no memory time window method is the worst one for dynamic FDP. As has been mentioned before, this phenomenon is due to the characteristics of financial distress concept drift, which belongs to gradual and constant drift instead of abrupt ones at intervals. Above all, dynamic FDP methods based on instance selection perform better than stationary ones, and fixed time window and batch selection methods are relatively more suitable for the domain of FDP than other dynamic modeling methods based on instance selection.

5. Conclusion and limitations

Effective enterprise FDP can help companies to prevent from deterioration of financial state in advance, so it is important for both inside management and outside interest parts. However, former researches on FDP ignore the phenomenon of financial distress concept drift with the passage of time, and only focus on how to construct static FDP models from sample data of certain period of time. Such static FDP models will gradually become unsuitable for future prediction when the concept of financial distress keeps drifting. This research focuses on how to dynamically re-building FDP model with time passage from a completely new view. A framework of dynamic FDP based on instance selection is designed and various instance selection methods, such as full memory window, no memory window, window with fixed size, window with adaptable size and batch selection, are applied to dynamic modeling of FDP. Among them, window with adaptable size and batch selection are relatively more complex than other instance selection methods, because they should integrate the process of feature selection and model construction into the process of instance selection to determine which batches should be selected. Initial features set composed of seven aspects of financial ratios are utilized, and a wrapper integrating forward and backward selection is proposed for the dynamic modeling of FDP. An empirical experiment based on Chinese listed companies is designed to simulate dynamic FDP with time passage by the rolling current year from 2002 to 2007. Experimental results of full memory window and no memory window indicate that financial distress concept drift does exist and need to be considered in the domain of FDP, although no memory window performs badly in proposing financial distress concept drift. Results of fixed time window and adaptable time window show that the fixed one is more suitable for dynamic FDP than the adaptable one, which is often the contrary in the domain like text search preference. We attribute the reason to the characteristic of financial distress concept drift which belongs to gradual and constant drift without truly stable stage. Consequently, the so-called adaptable window size which has the strict minimum estimation error for the most current batch may not really adapt to the coming future. Results of batch selection method indicate that it is applicable for Chinese listed companies' dynamic FDP when the batch selection parameter is set around 3. When the batch selection parameter is too small or too big, its effectiveness will be reduced. It is further proved in our experiment that dynamic FDP methods based on instance selection perform much better than traditional static models, which are not efficient enough for future FDP. Furthermore, window with fixed size and batch selection are more suitable for Chinese listed companies' dynamic FDP than other instance selection methods when their parameters are properly set.

This research also has some limitations. On one hand, the volume of data and the modeling years which simulate the evolution of time passage are limited. Thus the analysis on characteristics of financial distress concept drift may be not comprehensive enough. On the other hand, we only use Fisher discriminant analysis, the most classic and widely used FDP method, to construct the base predictive model. In future study, other classification techniques should be further integrated with the dynamic FDP modeling methods to handle financial distress concept drift.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (No. 70801054) and the Zhejiang Provincial Natural Science Foundation of China (No. Y6090392). The authors gratefully thank anonymous referees for their useful comments and editors for their work.

References

Altman, E. I. (1968). Financial ratios discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance, 23*(4), 589–609.

Beaver, W. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4*(Suppl.), 71–111.

Boritz, J. E., & Kennedy, D. B. (1995). Effectiveness of neural network types for prediction of business failure. *Expert Systems with Applications, 9*(4), 503–512.

Bose, I. (2006). Deciding the financial health of dot-coms using rough sets. *Information and Management, 43*(7), 835–846.

Bose, I., & Pal, R. (2006). Predicting the survival or failure of click-and-mortar corporations: A knowledge discovery approach. *European Journal of Operational Research, 174*(2), 959–982.

Carlos, S.-C. (1996). Self organizing neural networks for financial diagnosis. *Decision Support Systems, 17*, 227–238.

Delanya, S. J., Cunninghamb, P., Tsymbalb, A., & Coyle, L. (2005). A case-based technique for tracking concept drift in spam filtering. *Knowledge-Based Systems, 18*, 187–195.

Ding, Y., Song, X., & Zen, Y. (2008). Forecasting financial condition of Chinese listed companies based on support vector machine. *Expert Systems with Applications, 34*, 3081–3089.

Etheridge, H. L., & Sriram, R. S. (1997). A comparison of the relative costs of financial distress models: Artificial neural networks, logit and multivariate discriminant analysis. *Intelligent Systems in Accounting, Finance and Management, 6*, 235–248.

Fitzpartrick, P. J. (1932). A comparison of ratios of successful industrial enterprises with those of failed finns. *Certified Public Accountant, 10*, 598–605.

Fletcher, D., & Goss, E. (1993). Forecasting with neural networks: An application using bankruptcy data. *Information and Management, 24*, 159–167.

Frydman, H., Altman, E. I., & Kao, D. (1985). Introducing recursive partitioning for financial classification: The case of financial distress. *Journal of Finance, 40*(1), 269–291.

Heijden, F., Duin, R., Ridder, D., & Tax, D. (2004). *Classification, parameter estimation and state estimation: An engineering approach using MATLAB*. Wiley.

Hu, Y. C. (2008). Incorporating a non-additive decision making method into multi-layer neural networks and its application to financial distress analysis. *Knowledge-Based Systems, 21*(5), 383–390.

Hua, Z., Wang, Y., Xu, X., Zhang, B., & Liang, L. (2007). Predicting corporate financial distress based on integration of support vector machine and logistic regression. *Expert Systems with Applications, 33*, 434–440.

Hui, X.-F., & Sun, J. (2006). An application of support vector machine to companies' FDP. *Lecture Notes in Artificial Intelligence, 3885*, 274–282.

Hung, C., & Chen, J.-H. (2009). A selective ensemble based on expected probabilities for bankruptcy prediction. *Expert Systems with Applications, 36*, 5297–5303.

Kim, M-J., & Han, I. (2003). The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms. *Expert Systems with Applications, 25*, 637–646.

Klinkenberg, R. (2004). Learning drifting concepts: Example selection vs. example weighting. *Intelligent Data Analysis, 8*, 281–300.

Klinkenberg, R., & Joachims, T. (2000). Detecting concept drift with support vector machines. In *Proceedings of the seventh international conference on machine learning (ICML), USA* (pp. 487–494). Morgan Kaufmann.

Leshno, M., & Spector, Y. (1996). Neural network prediction analysis: The bankruptcy case. *Neurocomputing, 10*, 125–147.

Li, H., & Sun, J. (2009). Hybridizing principles of the Electre method with case-based reasoning for data mining: Electre-CBR-I and Electre-CBR-II. *European Journal of Operational Research, 197*(1), 214–224.

Li, H., & Sun, J. (2010). Business failure prediction using hybrid2 case-based reasoning (H2CBR). *Computers and Operations Research, 37*(1), 137–151.

McKee, T. E. (2000). Developing a bankruptcy prediction model via rough sets theory. *International Journal of Intelligent Systems in Accounting, Finance and Management, 9*, 159–173.

Min, J. H., & Lee, Y.-C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications, 28*(4), 128–134.

Min, S.-H., Lee, J., & Han, I. (2006). Hybrid genetic algorithms and support vector machines for bankruptcy prediction. *Expert Systems with Applications, 31*, 652–660.

Nanni, L., & Lumini, A. (2008). An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 36.

Ng, W. W. Y., Yeung, D. S., Firth, M., Tsang, E. C. C., & Wang, X.-Z. (2008). Feature selection using localized generalization error for supervised classification problems using RBFNN. *Pattern Recognition, 41*, 3706–3719.

Odom, M., & Sharda, R., (1990). A neural networks model for bankruptcy prediction. In *Proceedings of the IEEE international conference on neural network* (Vol. 2, pp. 163–168).

Ohlson, J. A. (1980). Financial ratios and probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*, 109–131.

Pendharkar, P. C. (2005). A threshold varying artificial neural network approach for classification and its application to bankruptcy prediction problem. *Computers and Operations Research, 32*, 2561–2582.

Premachandra, I. M., Bhabra, G., & Sueyoshi, T. (2009). DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique. *European Journal of Operational Research, 192*(2), 412–424.

Psillaki, M., Tsolas, I. E., & Margaritis, D. (2010). Evaluation of credit risk based on firm performance. *European Journal of Operational Research, 201*(3), 873–881.

Schlimmer, J. C., & Granger, R. H. (1986). Incremental learning from noisy data. *Machine Learning, 1*(3), 317–354.

Shin, K.-S., & Lee, Y.-J. (2002). A genetic algorithm application in bankruptcy prediction modeling. *Expert Systems with Applications, 23*, 321–328.

Shin, K.-S., Lee, T. S., & Kim, H.-J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications, 28*(1), 127–135.

Sun, J., & Hui, X.-F. (2006). FDP based on similarity weighted voting CBR. *Lecture Notes in Artificial Intelligence, 4093*, 947–958.

Sun, J., & Li, H. (2008). Listed companies' FDP based on weighted majority voting combination of multiple classifiers. *Expert Systems with Applications, 35*, 818–827.

Sun, J., & Li, H. (2009). Financial distress early warning based on group decision making. *Computers and Operations Research, 36*(3), 885–906.

Sun, L., & Shenoy, P. P. (2007). Using Bayesian networks for bankruptcy prediction: Some methodological issues. *European Journal of Operational Research, 180*(2), 738–753.

Tsai, C.-F., & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications, 34*, 2639–2649.

West, D., Dellana, S., & Qian, J. (2005). Neural network ensemble strategies for financial decision applications. *Computers and Operations Research, 32*(10), 2543–2559.

Widmer, G., & Kubat, M. (1993). Effective learning in dynamic environments by explicit context tracking. *Lecture Notes in Computer Science, 667*, 227–243.

Widmer, G., & Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. *Machine Learning, 23*(1), 69–101.

Wu, C.-H., Tzeng, G.-H., Goo, Y.-J., & Fang, W.-C. (2007). A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *Expert Systems with Applications, 32*, 397–408.

Yang, Z. R., Platt, M. B., & Platt, H. D. (1999). Probabilistic neural networks in bankruptcy prediction. *Journal of Business Research, 44*, 67–74.

Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research, 116*, 16–32.