Int. J. Mach. Learn. & Cyber.
DOI 10.1007/s13042-016-0545-8

ORIGINAL ARTICLE

Integration of incremental filter-wrapper selection strategy with artificial intelligence for enterprise risk management

Te-Min Chang[1] · **Ming-Fu Hsu**[1]

Received: 22 May 2015 / Accepted: 4 May 2016

Abstract The deterioration in enterprises' profitability not only threatens the interests of those firms, but also means related parties (investors, bankers, and stakeholders) could encounter tremendous financial losses, which could also impact the circulation of limited economic resources. Thus, an enterprise risk forecasting mechanism is urgently needed to assist decision-makers in adjusting their operating strategies so as to survive under any highly turbulent economic climate. This research introduces a novel hybrid model that incorporates an incremental filter-wrapper feature subset selection with the statistical examination and twin support vector machine (IFWTSVM) for enterprise operating performance forecasting. To promote a hybrid model's real-life application, the knowledge visualization extracted from IFWTSVM is represented in an easy-to-grasp style. The experimental results reveal that IFWTSVM's forecasting quality is very promising for financial risk mining, relative to other forecasting techniques examined in this study.

Keywords Risk management · Twin support vector machine · Feature subset selection

✉ Ming-Fu Hsu
hsumf0222@gmail.com

[1] Department of Information Management, National Sun Yat-sen University, 70 Lienhai Rd., Kaohsiung 80424, Taiwan, ROC

1 Introduction

Due to the radical changes that have taken place in global financial markets, modelling corporate survival and studying the reasons that lead to a failure/default have caught the considerable attention of both academics and practitioners over the last three decades [20, 27]. Financial failure/default often occurs when a corporate has suffered from chronic or serious losses and/or when it has become stuck in a financially insolvent problem. Most corporates facing financial failure/default not only have to deal with cash flow difficulties, but they also could disrupt the stability of financial markets as well as negatively influence investor confidence [13, 23]. One potential end result is the inefficient circulation of limited economic resources (like capital and money), which is one of many reasons why researchers place so much emphasis on building a financial pre-warning model.

In the existing research studies that have constructed financial failure/default warning models, numerous ratios derived from financial statements are taken as forecasting attributes, implying that these ratios represent a corporate's current status and a possible signal of upcoming financial troubles [2, 14, 45]. However, it has been widely recognized that the main cause of financial failure/default is poor management. According to Xu and Wang [58], operating efficiency reflects the strong or weak effectiveness of corporate management. Other works utilize accounting figures, such as return on assets (ROA) or return on equity (ROE), as proxies for measuring corporate operating efficiency, but those two specific measuring criteria only consider a single input and a single output variable. They actually improperly illustrate the whole aspect of a corporate's operating performance during highly volatile financial markets [33].

Published online: 12 May 2016

Data envelopment analysis (DEA) is a useful linear programming-based algorithm for measuring the relative efficiency of a set of decision-making units (DMUs) with multiple input and multiple output variables [4]. DEA does not need any pre-determination on the shape of the frontier surface and makes no assumptions concerning the internal operation of a DMU [9, 44]. Thus, DEA is a very suitable surrogate for assessing corporate operating efficiency.

Because Taiwan exhibited a large and stable growth rate of real GDP per capita that exceeded 8 % yearly on average over the 1950-2000 period [59], it has been admired as an economic miracle and has caught considerable researcher attention. Before the 1980s, the agriculture sector dominated Taiwan's economy, but after the 1980s the controlling power changed to labor-intensive manufacturing industries. In a traditional economy, physical assets (land, labor, equipment, and capital) are vital elements to determine the value of an enterprise and whether it can sustain or strengthen its competitive abilities. However, the Internet and the tremendous advancement of information and communication technologies have stirred these resources to circulate quickly all over the world, thus facilitating the rise of the current era of a knowledge-based economy [8, 35, 49]. In this knowledge-based economy, the core ingredient of any value-creating or value-added business activity has transformed from physical assets to knowledge-based assets [18, 43], which can help generate a corporate's knowledge resources and core competences, such as R&D innovation, employees' domain knowledge and professional skills, market share, and patents.

Compared with multinational or international firms, enterprises in Taiwan have scarce resources, such as technical staff and financial resources. It is thus essential for Taiwan managers or decision makers to realize what elements can enable their firm to be soundly competitive. Qiao et al. [39] stated that involvement in industry association networks can help corporates reach this goal by drawing on the networks to secure any missing resources. The network position of a corporate represents different opportunities to learn from other corporates and to access emerging ideas or knowledge that are important for developing innovative thinking and products [50, 51]. The network configuration also shapes the competitive priorities of a corporate and converts them into reputational and informational advantages and a competitive edge [21]. The literature has very rarely investigated industry association networks, despite the fact that supplier and buyer firms are embedded into such networks. Thus, this study implements a social network model to determine the prestige of a corporate embedded into industry association networks and examines how it affects corporate operating performance.

Most related studies in the literature put considerable effort on generating a warning mechanism to forecast financial failure/distress [28, 37], yet few works have researched corporate operating performance forecasting, which can be a prior stage enacted ahead of any potential financial failure/distress. If forecasting corporate operating performance is reliable, then managers or decision makers can initiate remedial actions to prevent further deterioration before the financial distress/failure bursts out, and investors can adjust their investment strategies to eliminate anticipated investment-related risks and allocate their personnel wealth toward suitable financial derivatives as a form of risk hedging.

This study's forecasting model construction can be chiefly divided into two main categories: one is based on a statistical technique, and the other is grounded on artificial intelligence [46, 48]. Although the forecasting model based on a statistical technique attains superior forecasting performance, it has to satisfy strict statistical assumptions that impede its practical applications. Support vector machine (SVM) is an effective and widely implemented form of artificial intelligence for classification and regression tasks in a variety of real-life problems that does not have to satisfy strict statistical assumptions [29, 54]. Although SVM has outstanding generalization ability compared with other artificial intelligence techniques (such as decision trees and neural networks), the model construction phase of SVM involves handling the quadratic programming problem (QPP), which is very time-consuming and restricts its utilization. Thus, Jayadeva et al. [24] proposed the twin support vector machine (TSVM) based on a non-parallel hyperplane. It seeks two non-parallel proximal hyperplanes whereby each hyperplane is closer to one of two classes and as far as possible from the other one [5, 42, 47]. The difference between traditional SVM and TSVM is that the former deals with a larger QPP, while the latter handles two reduced sized QPPs, which is the reason why TSVM calculates faster than the traditional SVM. Numerous research studies have demonstrated the effectiveness of TSVM over traditional SVM. Thus, this research conducts TSVM to handle the forecasting task.

Our major contribution to the literature is three-fold. First, for the performance measuring field, this study implements DEA to measure efficiency through multiple input and multiple output variables, thus appropriately illustrating the full aspect of a corporate's operating performance versus just a limited consideration with a single input and a single output variable. Second, for industry association networks, this study implements a social network technique to identify the critical position of an enterprise embedded into the network that is affecting its operating performance. Third, in our experimental study we utilize a novel hybrid model (called IFWTSVM) that incorporates the incremental filter-wrapper subset selection technique with twin SVM so as to strengthen the forecasting quality.

The reminder of this study is organized as follows. Section 2 presents the experimental methodologies and proposes the forecasting model. Section 3 shows the experimental outcomes. The final section concludes.

2 Methodologies

2.1 Data envelopment analysis: DEA

Data envelopment analysis (DEA), proposed by Charnes et al. [10], is a decision support approach grounded on linear programming for efficiency measurement. The fundamental idea of DEA is to compute and calculate an efficiency frontier, assessed by the relative performance of homogeneous entities called decision-making units (DMUs) in terms of the distance of per units to the ideal frontier. DEA is constructed by utilizing observed multiple input and multiple output variables [15].

Suppose there are pDMU$_s$, each with q input variables and r output variables. The performance ranking of the relative efficiency of DMU$_s$ is determined by the following equations.

$$Max \sum_{k=1}^{r} a_k y_{ks} \Bigg/ \sum_{j=1}^{q} b_j x_{js} \tag{1}$$

s.t.

$$\sum_{k=1}^{r} a_k y_{kj} \Bigg/ \sum_{j=1}^{q} b_j x_{ji} \tag{2}$$

$a_k \geq 0, b_j \geq 0 \ \forall k,$

where $k = 1, \ldots, r$; $j = 1 \ldots, q$; $i = 1, \ldots, p$. The amounts of output k and input j generated by DMU$_i$ are expressed as y_{ki} and x_{ji}, separately. Here, a_k denotes the weight given to output k, and b_j denotes the weight given to input j.

Each DMU decides the weights of input and output variables so as to maximize the relative performance score. A score of 1 means a DMU is efficient (that is, the DMU is located on the efficiency frontier), while a score less than 1 means that it is inefficient (that is, the DMU is located below the efficiency frontier). In addition, the frontier yields the parameters needed to evaluate what an inefficient DMU should do to become efficient [19]. Thus, it has become a widely implemented technique to handle the task of performance ranking.

2.2 Incremental filter-wrapper subset selection: IFWSS

Feature selection is a procedure of identifying a subset of features from the original set of features that are relevant to a particular learning problem or task. The subset should be sufficient and appropriate to illustrate target concepts, preserving a superior quality in expressing the original features [57]. This procedure helps to facilitate and strengthen the forecasting performance of the learned mechanism by: (a) eliminating the problem due to higher dimensionality; (b) improving the mechanism's generalization ability; (c) decreasing the mechanism's computational complexity; and (d) increasing the learning and inference speed [17, 57].

The feature subset selection algorithms can be divided into two main categories: (a) filter algorithms and (b) wrapper algorithms. The former assess the goodness of a feature by merely utilizing intrinsic properties of the data, and the latter implement the classifier embedded into the algorithm to assess the effectiveness of the features. Wrapper algorithms achieve better performance than filter algorithms, but have the disadvantage of being very time-consuming. In order to preserve the advantages of a wrapper algorithm, but at the same time preventing being stuck by its heavy computational burdens, Ruiz et al. [40] introduced the hybrid filter-wrapper algorithm. The basic concept of the hybrid algorithm is to rank the feature's relevance with respect to the class through filter algorithms. Sequentially, we implement the following algorithm to go over the ranking by incrementally feeding those features that are effective into the classification procedure, where the effectiveness of including a new feature is evaluated in a wrapper manner [6].

Grounded on the concept of the hybrid filter-wrapper algorithm, this study introduces an emerging feature subset selection algorithm—namely, incremental filter-wrapper subset selection (IFWSS) approach. IFWSS can be divided into two main stages: (a) filter stage and (b) wrapper stage. In the filter stage, we use symmetrical uncertainty (SU), a kind of mutual information technique, to rank the predictive features. Equation (3) presents the mathematical format.

$$SU(F_i, C) = 2\left(\frac{E(C) - E(C|F_i)}{E(C) + E(F_i)}\right), \tag{3}$$

where C denotes the class, and $E()$ denotes the Shannon entropy. Features are arranged in increasing SU order—that is, the higher the SU order, the more important it is.

The first feature in the ranking is sequentially fed into rough set theory (RST) due to its outstanding ability to handle imprecision, uncertainty, and vagueness (Pawlak and Skowron 2007). A five-fold cross-validation is then executed to eliminate the problem of over-fitting. To get a more compact feature subset, we adjust the process of feature subset selection. The concept is that when a new feature is being evaluated, we examine not only the effectiveness of adding it to the selected feature subset, but also swapping it with any of the already included features. Furthermore, the evaluation process undergoes statistical testing—that is, if the added feature can improve the model's forecasting performance based on statistical testing (e.g., t test with a significance level of 5 %), then the added feature is preserved and vice versa. The relevant criterion grounded on statistical testing prevents the approach from adding new features due to noise or outliers. Thus, it is clear that IFWSS obtains feature subsets with sufficient information and superior quality [7, 40].

2.3 Twin support vector machine: TSVM

Vapnik [53] proposed an emerging neural network-based technique—namely, support vector machine (SVM) - that is based on the structural risk minimization (SRM) principle of maximum margin, dual theory, and kernel trick that come from statistical learning theory. Due to its excellent generalization ability for classification and regression tasks, SVM has caught a great deal of attention in the past decades. Though SVM achieves outstanding generalization ability compared with numerous artificial intelligence techniques, its model construction procedure involves the complex solution of a quadratic programming problem (QPP). This can be very time-consuming for datasets with a large amount of features.

Twin support vector machine (TSVM), introduced by Jayadeva et al. [24], was proposed to handle the aforementioned problem that traditional SVM encounters. TSVM discriminates the pattern of classes by implementing two non-parallel hyperplanes. It transforms a single complex QPP in traditional SVM into two reduced QPPs. In contrast to SVM, TSVM has been proven to be very competitive in terms of forecasting quality as it is around four times faster [41].

We offer an illustration of TSVM as follows [24, 41, 47]. Let the instances of positive and negative class labels be expressed by f and g, respectively. The matrix X_1 - $\Re^{f\times h}$ denotes the positive class data instance, matrix X_2 - $\Re^{g\times h}$ denotes the negative class data instance, and $\Re$ expresses the h-dimensional real space. Equation (4) presents the two non-parallel hyperplanes in h-dimensional real space $\Re^h$ [27].

$$x^T w_1 + c_1 = 0 \quad and \quad x^T w_2 + c_2 = 0, \tag{4}$$

where w_1 and w_2 represent the hyperplane's normal vector, and c_1 and c_2 are error terms.

The linear TSVM is formulated in Eqs. (5)-(6).

$$\begin{gathered}\min_{w_1,c_1,\zeta}\frac{1}{2}\|X_1w_1+b_1c_1\|^2+d_1b_2^T\zeta\\ \text{subject to the constraints}\\ -(X_1w_1+b_1c_1)+\zeta\geq b_2,\quad \zeta\geq 0\end{gathered} \tag{5}$$

$$\begin{gathered}\min_{w_2,c_2,\gamma}\frac{1}{2}\|X_2w_2+b_2c_2\|^2+d_2b_1^T\gamma\\ \text{subject to the constraints}\\ -(X_1w_2+b_1c_2)+\gamma\geq b_1,\quad \gamma\geq 0,\end{gathered} \tag{6}$$

where ζ, γ denote the slack variables, and d_1, d_2 represent penalty parameters. Two vectors of appropriate dimension and possessing all values of 1 are expressed by b_1 and b_2. Equations (5) and (6) advance the implementation to establish the TSVM classification mechanism. The Lagrange multiplier executed in Eq. (5) can be obtained as follows:

$$\begin{aligned}L(w_1,c_1,\zeta,\alpha,\beta)=\frac{1}{2}\|X_1w_1+b_1c_1\|^2+d_1b_2^T\zeta+\alpha^T((X_2w_1\\ +b_2c_1)-\zeta+b_2)-\beta^T\zeta\end{aligned} \tag{7}$$

The two vectors α and β denote the Lagrange multipliers. Equation (7) under the Karush–Kuhn–Tucker (KKT) conditions is represented as follows:

$$\frac{\partial L}{\partial w_1}=X_1^T(X_1w_1+b_1c_1)+X_2^T\alpha=0 \tag{8}$$

$$\frac{\partial L}{\partial c_1}=b_1^T(X_1w_1+b_1c_1)+b_2^T\alpha=0 \tag{9}$$

$$\frac{\partial L}{\partial \zeta}=d_1b_2^T-\beta^T-\alpha^T=0 \tag{10}$$

$$-(X_2w_1+b_2c_1)+\zeta\geq b_2,\ \zeta\geq 0 \tag{11}$$

$$\alpha^T((X_2w_1+b_2c_1)-\zeta+b_2)=0,\ \beta^T\zeta=0 \tag{12}$$

$$\alpha\geq 0,\quad \beta\geq 0 \tag{13}$$

After combining Eqs. (8) and (9), we obtain Eq. (14).

$$\begin{bmatrix}X_1^T\\ b_1^T\end{bmatrix}[X_1\ b_1]\begin{bmatrix}w_1\\ c_1\end{bmatrix}+\begin{bmatrix}X_2^T\\ b_2^T\end{bmatrix}\alpha=0 \tag{14}$$

We reconstruct Eq. (14) with $M=[X_1b_1], N=[X_2b_2]$, and $U_1=\begin{bmatrix}w_1\\ c_1\end{bmatrix}$ as follows:

$$M^TMU_1+N^T\alpha=0 \tag{15}$$

$$U_1=-(M^TM)^{-1}N^T\alpha \tag{16}$$

However, it is complicated to calculate the inverse of M^TM. To handle the aforementioned task, we add a regularization term ηI in Eq. (16), and I denotes the identity matrix, which are represented as follows:

$$U_1=-(M^TM+\eta I)^{-1}N^T\alpha \tag{17}$$

We get the normal vector and error term from the dissimilar class labels under a similar process.

$$U_2=-(N^TN+\eta I)^{-1}M^T\iota \tag{18}$$

We feed the normal vectors and error terms into Eq. (4) to construct the non-parallel hyperplane. Sequentially, the hyperplane for each dissimilar class label can be decided by TSVM, and a new data instance is assigned to a class label i by Eq. (19).

$$\mathrm{Classlabel\ i} = \min\left|\mathrm{x}^{\mathrm{T}}\mathrm{w_i} + \mathrm{c_i}\right| \quad \text{for} \quad \mathrm{i} = 1, 2 \tag{19}$$

The test instance class label is determined by calculating the perpendicular distance from each hyperplane and taking the shortest distance. TSVM embedded with a kernel trick can be extended to handle data with a non-linear structure. The mathematical formulation of non-linear TSVM is expressed in Eqs. (20)-(21).

$$\begin{gathered}\min_{(u_1,c_1,\zeta)} \frac{1}{2}\left\|\Theta(X_1,G^T)U_1 + b_1c_1\right\|^2 + d_1b_2^T\zeta \\ \text{subject to the constraints} \\ -(\Theta(X_2,G^T)U_1 + b_2c_1) + \zeta \geq b_2,\ \zeta \geq 0\end{gathered} \tag{20}$$

and

$$\begin{gathered}\min_{(u_2,c_2,\gamma)} \frac{1}{2}\left\|\Theta(X_2,G^T)U_2 + b_2c_2\right\|^2 + d_2b_1^T\gamma \\ \text{subject to the constraints} \\ (\Theta(X_1,G^T)U_2 + b_1c_2) + \gamma \geq b_1,\ \gamma \geq 0\end{gathered} \tag{21}$$

Here, $G = [X_1X_2]^T$ and Θ denotes the kernel function.

Equation (22) expresses the discriminating hyperplane with a kernel function.

$$\Theta(x^T, G^T)U_1 + c_1 = 0 \quad \text{and} \quad \Theta(x^T, G^T)U_2 + c_2 = 0. \tag{22}$$

The Lagrange multiplier injected into Eq. (20) can be represented as follows:

$$\begin{aligned}L(U_1, c_1, \zeta, \alpha, \beta) = {} & \frac{1}{2}\left\|\Theta(X_1,G^T)U_1 + b_1c_1\right\|^2 + d_1b_2^T\zeta \\ & + \alpha^T((\Theta(X_2,G^T)U_1 + b_2c_1) - \zeta + b_2) - \beta^T\zeta\end{aligned} \tag{23}$$

Equations (24) and (25) are now implemented to calculate the normal vectors and error terms for two non-parallel hyperplanes.

$$Z_1 = \begin{bmatrix} U_1 \\ c_1 \end{bmatrix} = -(Q^TQ)^{-1}Q^T\alpha \tag{24}$$

$$Z_2 = \begin{bmatrix} U_2 \\ c_2 \end{bmatrix} = (R^TR)^{-1}Q^T\alpha, \tag{25}$$

where $Q = [\Theta(X_1, G^T)b_1]$ and $Q = [\Theta(X_2, G^T)b_2]$. Equation (26) is executed to determine the class label.

$$\text{Class } i = \min\left|\Theta(x^T, G^T)U_i + c_i\right| \quad \textit{for} \quad i = 1, 2 \tag{26}$$

The distance of the new data instance is measured from both kernel discriminating surfaces, and it is used to determine the class label from which its distance is shorter (Chen et al. 2013; [52].

2.4 IFWTSVM model

This study provides a novel hybrid model - namely, IFWTSVM - that combines incremental filter-wrapper technique and twin support vector machine for enterprise risk management. A large amount of prior works on enterprise risk management only takes return on assets (ROA) or return on equity (ROE) into consideration, but both of these two assessing criteria fall into the one input and one output variable category, which is not appropriate for representing the full aspect of an enterprise's operating performance. With the advantage of handling multiple input and output variables simultaneously, DEA can rank and more suitably depict an enterprise's operating performance. The performance rankings generated by DEA in the highest quintile (20 %) and in the lowest quintile (20 %) are designated as efficient and inefficient operating performances, respectively.

After going through the performance measuring stage, this study collects the related information to construct the enterprise risk forecasting model. However, as too much information will mislead or disturb the judgement made by decision makers, a feature subset selection in thus an essential and inevitable stage. We conduct the incremental filter-wrapper subset selection (IFWSS) technique to handle the task of feature selection. IFWSS in this study can be divided into two stages: filter stage and wrapper stage. In the filter stage, we implement a filter measure (that is, symmetrical uncertainty) to obtain the prestige of the feature's relevance in regard to the class. A sequential approach is then executed to go over the prestige of the feature by incrementally adding those features that are relevant to the classification process, where the relevance of including a new feature is assessed in a wrapper manner [7]. The merits of the hybrid filter-wrapper mechanism are that it preserves the highest proportion of wrapper advantages and eliminates computational burden or complexity to an acceptable level, instead of allowing it to occur under the original wrapper algorithm [40]. We take RST as a surrogate to assess the effectiveness of the added feature in the wrapper stage.

The informative features determined by IFWSS are used to construct the enterprise risk forecasting model. The forecasting model (that is, TSVM) belongs to a data-driven mechanism. We divide the original dataset into two sets: training dataset and testing dataset. The former establishes the forecasting model, and the latter examines the forecasting performance of a well-trained model. We implement five-fold cross-validation to alleviate the impact of over-fitting. The decision maker can view IFWTSVM as a guideline to modify or adjust an investment strategy to survive in a highly competitive environment, and investors can use it to allocate their personnel wealth toward suitable derivatives. Figure 1 depicts the flowchart of IFWTSVM.

3 Experimental results

3.1 The data and features

This research takes publicly-listed electronics firms in Taiwan from 2012 to 2014 as the sample due to their considerable impact on local investors' personnel wealth. In order to depict their operating performance or relationship in their respective supply chain, information about the enterprises was gathered from the public websites of Taiwan Economic Journal Data Bank (TEJ), Taiwan Stock Exchange Corporation (TSEC), and Taipei Exchange (TE).

Enterprises running a business in this current knowledge economy are quite different from those that are from a labor-intensive economy. The core competitiveness of an enterprise has changed from physical assets to knowledge-based assets, which encompass its dynamic ability as generated through its core competences and knowledge resources [18, 35]. In contrast to multinational or global firms, enterprises in Taiwan have limited capital as well as scarce knowledge and human resources. Furthermore, with global trade liberalization, enterprises in Taiwan have been encountering numerous challenges from all over the world.

How to survive in a highly competitive environment is an essential task for managers and decision makers. Qiao et al. [39] stated that involvement in industry association networks can help enterprises reach or find implicit resources, by drawing on such networks for any missing resources. Gnyawali and Madhavan [21] noted that the network configuration also shapes the competitive priorities of an enterprise, transforming them into reputational and informational advantages and a competitive edge. A few research studies have taken industry association networks into consideration to construct an enterprise risk forecasting mechanism. Thus, this study utilizes a social network technique to identify the prestige of an enterprise embedded in industry association networks and tests how this prestige contributes to the enterprise's operating performance.

Measuring the network prestige of an enterprise is about determining the centrality of a node that has an essential structural impact on a firm's leadership, satisfaction, and efficiency [16]. Betweenness centrality can evaluate the number of times a particular node lies "between" the various other nodes in the network. It can be viewed as a broker or gatekeeper that controls the information flow in the network [16]. The betweenness of node p_k goes as follows:

$$C_{Between}(p_k) = \sum_{i<}^{n} \sum_{j}^{n} \frac{g_{ij}(p_k)}{g_{ij}}, \quad i \neq j \neq k, \tag{27}$$

where the number of geodesics (that is, the shortest paths) linking p_i and p_j is expressed as g_{ij}, and the number of geodesics linking p_i and p_j that includes node p_k is represented as $g_{ij}(p_k)$.

To execute the forecasting model for enterprise risk management, which is normally based on financial ratio analysis, predictor features should first be determined. The predictor features include three main parts in this study: one is the efficiency measurement by DEA, another consists of numerous features picked up from financial statements, and the other is the enterprises' prestige evaluation through the social network technique. Table 1 presents the input and output features used for efficiency measurement, and Table 2 lists the features used for constructing the enterprise risk management forecasting model.

3.2 Experimental design

We carry out the experiment in accordance with the following stages.

Input: An enterprise risk management dataset.

Output: The forecasting outcome.

Stage 1: Prepare the dataset with financial and social network features and pre-decided performance ranks by DEA.

Stage 2: Feed the dataset into IFW to get the informative features for constructing the forecasting model.

Stage 3: Divide the dataset into two parts: the training dataset, which is used to construct the forecasting model; and the testing dataset, which is implemented to examine the performance of the constructed model. Execute five-fold cross-validation to overcome the problem of over-fitting.

Stage 4: Calculate the model's performance measurement under four evaluation criteria.

Stage 5: To ensure the outcome does not happen by chance, perform a statistical test.

Stage 6: To measure the robustness of the introduced model, further consider another performance ranking by ratio analysis (that is, ROA).

End: The model can be taken as a roadmap to adjust managerial planning and control activities in order to survive in highly viotile business environment.

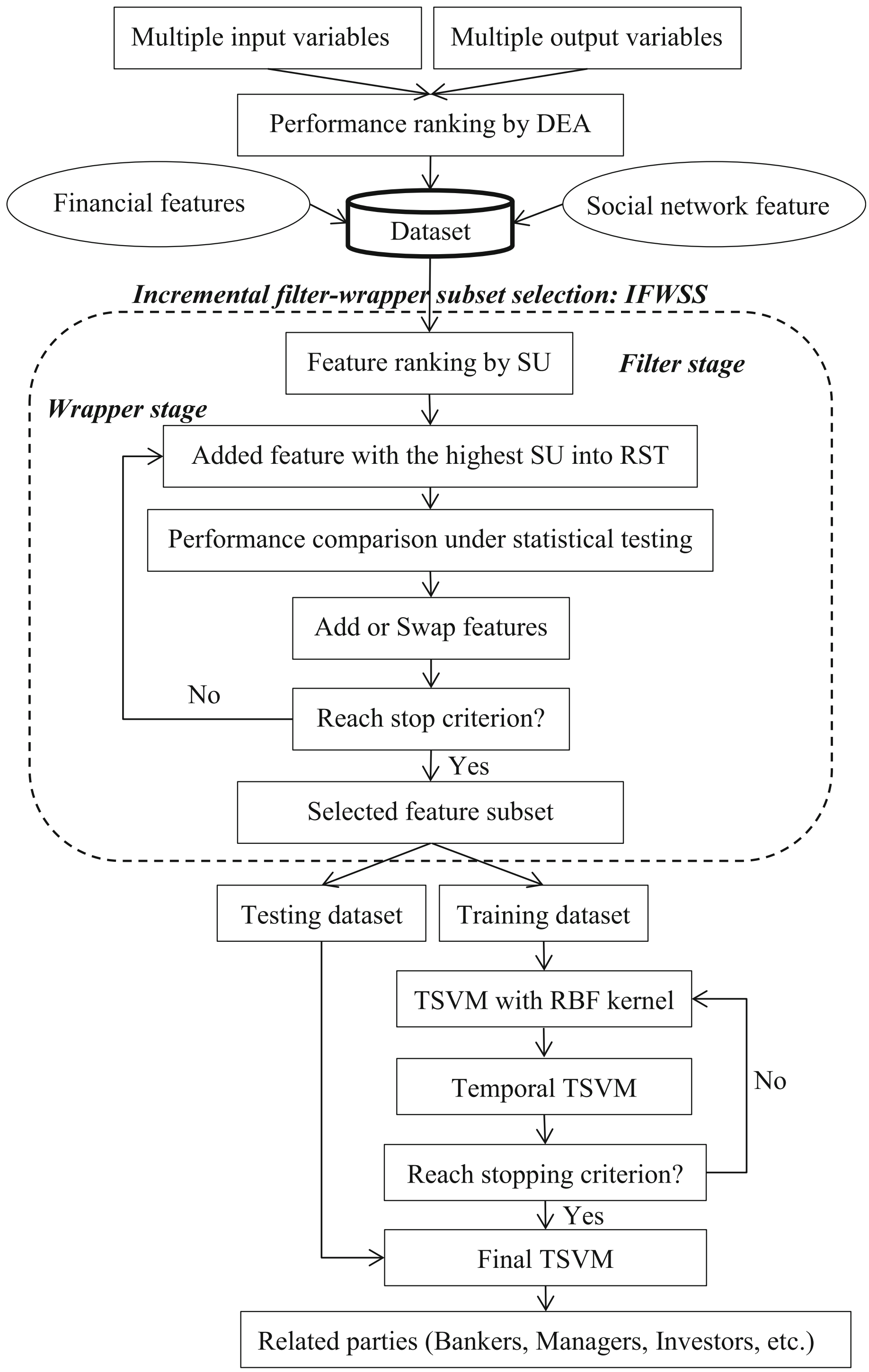

Fig. 1 The IFWTSVM model

3.3 Performance measures

There are plenty of performance measures for forecasting tasks. Commonly implemented performance measures in financial forecasting domains include accuracy, precision, recall, and F-measure. We offer brief illustrations of these measures as follows [30, 38].

Accuracy: It is the percentage of precisely discriminated modules, and the mathematical format is depicted as:

$$Accuracy = (TN + TP)/(TP + FP + FN + TN) \quad (28)$$

True positive (TP): It is the number of precisely discriminated fault-prone modules, and it is also named as a sensitivity measure.

Table 1 The input and output variables for DEA

Measuring technique: DEA	
Input oriented variables	Output oriented variables
IOV 1:Total assets	OOV 1: Total sales;
IOV 2: Total liabilities	OOV 2: Net income
IOV 3: Cost of sales	

Table 2 The used features

Symbols	Illustration
Financial property	
X1: TL/TA	OOV 2: Net income
X2: CA/CL	Total liability to total assets
X3: EBIT/IE	Current assets to current liabilities
X4: NI/S	Earnings before interest and tax (EBIT) to interest expense
X5: NI/TA	Net income to sales
X6: NI/FA	Net income to total assets
X7: S/FA	Net income to fixed assets
X8: WC/TA	Sales to fixed assets
X9: C/TA	Working capital to total assets
X10: I/S	Cash to total assets
X11: CA/S	Inventory to sales
X12: CF/TL	Current assets to sales
X13: NI/(TA-TL)	Cash flow to total liabilities
X14: LTD/TA	Net income to (total assets – total liabilities)
X15: C/S	Long-term liabilities to total assets
X16: ITR	Cash to sales
X17: S/MP	Insolvency turnover rate
X18: OI/MC	Sales to market capitalization
X19: RTR	Operating income to market capitalization
Network prestige property	Receivable turnover rate
X20: NP	Network prestige

$True\_Positive = (TP)/(TP + FN)$.

False positive (FP): It is the number of non-fault-prone modules that are misclassified as being fault-prone.

$False\_Positive = (FP)/(FP + TN)$.

True negative: It is the number of precisely discriminated non-fault-prone modules.

$True\_Negative = (TN)/(TN + FP)$.

False negative (FN): It is the number of fault-prone module that are misclassified as being non-fault-prone.

$False\_Negative = (FN)/(FN + TP)$.

The precision, recall, and F-measure are expressed as follows:

$$\Pr ecision = (TP)/(TP + FP), \tag{29}$$

Table 3 The selected features by IFW

Symbols	Illustration
X1: TL/TA	Total liability to total assets
X4: NI/S	Net income to sales
X7: S/FA	Sales to fixed assets
X10: I/S	Inventory to sales
X19: RTR	Receivable turnover rate
X20: NP	Network prestige

$$\mathrm{Re}call = (TP)/(TP + FN) \tag{30}$$

$$F - measure = (2 * \Pr ecision * \mathrm{Re}call)/(\Pr ecision + \mathrm{Re}call) \tag{31}$$

3.4 Results and discussion

Feature selection eliminates the cost of time and space requirement and improves the model's performance by removing irrelevant or redundant features. Thus, this process can be viewed as an inevitable stage in machine learning application. We execute incremental filter-wrapper subset selection (IFW), providing the results in Table 3. To examine the usefulness of IFW, we take it as a baseline and compare it with the other feature selection techniques usually implemented in the financial domain (i.e., stepwise regression: SR; and correlation matrix: CM), the singular filter technique (i.e., SU), and the singular wrapper technique (i.e., RST), with the results in Table 4a. To examine that the results do not happen by chance, we conduct a statistical examination. Table 5 lists the results, which reveal that IFW outperforms the other four feature selection techniques.

One interesting finding is that IFWSS picks up the network prestige property (X20: NP). To examine the effectiveness of NP (X20), this study implements a with-and-without strategy and presents the results in Table 6. Sequentially, the statistical test is performed to eliminate the result that happens by chance, and the outcomes are in Table 7. This table indicates that NP (X20) not only can strengthen the forecasting performance, but can also increase those assessing criteria. The rationale is, in fact, that enterprise association networks enable members to participate in negotiations, mobilize for regional economic cooperation, and facilitate their innovation ability, which is a vital element in the knowledge-based economy [1, 26]. In addition, to examine the effectiveness of the introduced IFWTSVM, this study compares it with other three classifiers: support vector machine (SVM), random forest (RF), and neural networks (NN). The result reveals that the proposed IFWTSVM is a promising alternative model for enterprise risk forecasting (see Tables 8, 9).

Table 4 The experimental results

Model	IFWTSVM	SWTSVM	CMTSVM	SUTSVM	RSTTSVM
ACC	91.24	75.16	76.48	82.08	82.08
PRE	92.08	75.68	77.36	82.4	82.96
REC	90.59	74.90	76.00	81.90	81.53
F-M	91.32	75.29	76.67	82.14	82.24

ACC accuracy, *PRE* precision, *REC* recall, *F-M* F-measure

Table 5 The statistical results

Model	Baseline: IFWTSVM			
	SWTSVM	CMTSVM	SUTSVM	RSTTSVM
Hypothesis	$H_0 : \mu_{IFWTSVM} = \mu_{SWTSVM,CMTSVM,SUTSVM,RSTTSVM}$ $H_1 : \mu_{IFWTSVM} \neq \mu_{SWTSVM,CMTSVM,SUTSVM,RSTTSVM}$			
Assessing criteria: ACC				
Results	*P*-value = 0.000 (***)	P-value = 0.000 (***)	P-value = 0.000 (***)	P-value = 0.001 (***)
Assessing criteria: PRE				
Results	P-value = 0.000 (***)	P-value = 0.000 (***)	P-value = 0.000 (***)	P-value = 0.001 (***)
Assessing criteria: REC				
Results	P-value = 0.000 (***)	P-value = 0.000 (***)	P-value = 0.000 (***)	P-value = 0.004 (***)
Assessing criteria: F-M				
Results	P-value = 0.000 (***)	P-value = 0.000 (***)	P-value = 0.000 (***)	P-value = 0.001 (***)

ACC accuracy, *PRE* precision, *REC* recall, *F-M* F-measure

* P-value < 0.1; ** P-value < 0.05; *** P-value < 0.01

Table 6 The results under with-and without strategy (Average performance)

Model	IFWTSVM		SVM		RF		NN	
Criteria	With	Without	With	Without	With	Without	With	Without
ACC	91.24	82.92	84.80	79.08	81.84	77.92	79.32	74.28
PRE	92.08	83.52	85.84	79.28	83.52	77.92	79.92	75.04
REC	90.59	82.61	84.09	78.98	80.79	77.93	78.99	73.90
F-M	91.32	83.03	84.95	79.13	82.13	77.92	79.44	74.46

ACC accuracy, *PRE* precision, *REC* recall, *F-M* F-measure

Table 7 The statistical results

Model	IFWTSVM	SVM	RF	NN
Hypothesis	$H_0 : \mu_{with} = \mu_{without}$ $H_1 : \mu_{with} \neq \mu_{without}$			
ACC	P-value = 0.000 (***)	P-value = 0.003 (***)	P-value = 0.005 (***)	P-value = 0.001 (***)
PRE	P-value = 0.000 (***)	P-value = 0.001 (***)	P-value = 0.001 (***)	P-value = 0.024 (***)
REC	P-value = 0.003 (***)	P-value = 0.011 (**)	P-value = 0.028 (**)	P-value = 0.000 (***)
F-M	P-value = 0.000 (***)	P-value = 0.002 (***)	P-value = 0.003 (***)	P-value = 0.003 (***)

ACC accuracy, *PRE* precision, *REC* recall, *F-M* F-measure

* P-value < 0.1; ** P-value < 0.05; *** P-value < 0.01

The outstanding forecasting performance of IFWTSVM comes with a critical weakness—it lacks interpretability, which impedes its real-life application [31]. To improve and get more insight into the decision-making process of IFWTSVM, we implement the swarm intelligence technique (namely, particle swarm optimization: PSO) to yield

Table 8 The forecasting results of each model

Model	IFWTSVM				SVM			
Condition	ACC	PRE	REC	F-M	ACC	PRE	REC	F-M
CV-1	91.00	92.80	89.58	91.16	87.20	89.60	85.50	87.50
CV-2	90.80	92.40	89.53	90.94	85.20	86.00	84.65	85.32
CV-3	92.20	92.00	92.37	92.18	84.40	85.20	83.86	84.52
CV-4	92.00	91.20	92.68	91.94	83.80	83.60	83.94	83.77
CV-5	90.20	92.00	88.80	90.37	83.40	84.80	82.49	83.63
AVG.	91.24	92.08	90.59	91.32	84.80	85.84	84.09	84.95
Model	RF				NN			
Condition	ACC	PRE	REC	F-M	ACC	PRE	REC	F-M
CV-1	84.60	87.20	82.89	84.99	80.80	81.20	80.56	80.88
CV-2	81.80	83.20	80.93	82.05	79.00	80.00	78.43	79.21
CV-3	80.80	82.00	80.08	81.03	78.40	79.60	77.73	78.66
CV-4	82.40	84.80	80.92	82.81	79.20	77.60	80.17	78.86
CV-5	79.60	80.40	79.13	79.76	79.20	81.20	78.08	79.61
AVG.	81.84	83.52	80.79	82.13	79.32	79.92	78.99	79.44

ACC accuracy, *PRE* precision, *REC* recall, *F-M* F-measure

Table 9 The statistical outcomes of each model

Base model Compared model	IFWTSVM SVM	IFWTSVM RF	IFWTSVM NN
Hypothesis	$H_0 : \mu_{IFWTSVM} = \mu_{SVM,RF,NN}$ $H_1 : \mu_{IFWTSVM} \neq \mu_{SVM,RF,NN}$		
ACC	P-value = 0.001 (***)	P-value = 0.000 (***)	P-value = 0.000 (***)
PRE	P-value = 0.001 (***)	P-value = 0.002 (***)	P-value = 0.000 (***)
REC	P-value = 0.002 (***)	P-value = 0.001 (***)	P-value = 0.000 (***)
F-M	P-value = 0.001 (***)	P-value = 0.000 (***)	P-value = 0.000 (***)

ACC accuracy, *PRE* precision, *REC* recall, *F-M* F-measure

* P-value < 0.1; ** P-value < 0.05; *** P-value < 0.01

the knowledge represented in an "If–Then" style. This optimization technique was introduced by Eberhart and Kennedy (1995) and is grounded on the analogy of the behavior of flocks of birds or schools of fish. PSO has demonstrated that it can reach faster convergence in comparison with a genetic algorithm (GA). Furthermore, from the viewpoint of application, PSO scales well and is not sensitive to population size [36].

The selection of the most essential rule sets from IFWTSVM can be transformed into optimization tasks that can be handled by PSO. The principles of rule selection are based on two assessing criteria: (1) Coverage and (2) Precision.

Coverage criterion: $Coverage = NCP(S)/|S|, \quad (32)$

where $NCP(S)$ denotes the number of instances that have been correctly discriminated in dataset S, and $|S|$ is the number of instances in dataset S.

Precision criterion: $Precision = (1 + TP)/(1 + TP + FP), \quad (33)$

where TP denotes the number of True Positive, and FP denotes the number of False Positive. This assessing criterion is not only performed to evaluate the quality of the Rules, but also is implemented to prune the rules when the value of Pr *ecision* is smaller than 0.1. For a more detailed illustration of the rule generation procedure, one can refer to Chen and Ludwig [12]. Table 10 lists the rules derived from IFWTSVM.

3.5 Robust test

Measuring an enterprise's operating efficiency can also be related to financing and investing strategies. Thus, how to appropriately evaluate its operating performance is a critical task under a turbulent and competitive business environment. Many researchers have employed ROA or ROE

Table 10 The knowledge visualized from IFWTSVM

If 'condition'	*Then* 'decision'
X1: TL/TA is between 0.54 and 0.63, X4: NI/S is between -0.03 and 0.04, and X20:NP is between 0.05 and 0.01	Inefficiency
X7: S/FA is between 1.32 and 2.13, X10: I/S is between 0.02 and 0.08 and X20:NP is between 0.64 and 0.73	Efficiency
X4: NI/S is between 0.17 and 0.28, and X19: RTR is between 2.12 and 3.12	Efficiency
X1: TL/TA is between is between 0.27 and 0.39, X7: S/FA is between 0.03 and 0.11	Inefficiency
X7: S/FA is between 1.32 and 2.13 and X20: NP is between 0.64 and 0.73	Efficiency

Table 11 The forecasting results of each model

Model	IFWTSVM				SVM			
Condition	ACC	PRE	REC	F-M	ACC	PRE	REC	F-M
CV-1	86.80	84.40	88.66	86.48	84.00	83.60	84.27	83.94
CV-2	87.00	86.40	87.45	86.92	81.80	82.00	81.67	81.84
CV-3	89.40	89.60	89.24	89.42	86.40	84.40	87.92	86.12
CV-4	90.20	88.00	92.05	89.98	81.80	79.20	83.54	81.31
CV-5	88.00	86.00	89.58	87.76	80.40	80.40	80.40	80.40
AVG.	88.28	86.88	89.40	88.11	82.88	81.92	83.56	82.72
Model	RF				NN			
Condition	ACC	PRE	REC	F-M	ACC	PRE	REC	F-M
CV-1	78.40	80.00	77.52	78.74	79.40	80.40	78.82	79.60
CV-2	79.40	79.20	79.52	79.36	78.60	79.60	78.04	78.81
CV-3	77.20	74.00	79.06	76.45	80.80	82.00	80.08	81.03
CV-4	79.60	79.20	79.84	79.52	82.60	84.40	81.47	82.91
CV-5	80.00	80.80	79.53	80.16	79.00	80.80	77.99	79.37
AVG.	78.92	78.64	79.09	78.85	80.08	81.44	79.28	80.34

ACC accuracy, *PRE* precision, *REC* recall, *F-M* F-measure

as a surrogate to determine a firm's operating performance, but different criteria lead to quite different outcomes. To robust the usefulness of the introduced model, this study changes the assessing criteria from DEA to ROA, with the results in Tables 11, 12. The research findings state that the proposed IFWTSVM still achieves optimal forecasting quality.

4 Conclusions

The proper forecasting of an enterprise's potential financial distress/crisis is of considerable importance to managers, decision makers, stakeholders, and other parties. Numerous researchers in prior works have put forth considerable attention on the financial ratios of enterprises when constructing a financial distress/crisis forecasting model. In comparison with previous financial crisis or distress forecasting studies, research on enterprise performance measurement is quite rare. Because the main reason an enterprise encounters financial trouble is due to poor management, its operating efficiency can be used to reflect the effectiveness of firm management. Thus, with its ability to handle multiple input and output variables, this study has used DEA to evaluate the operating performance of enterprises.

In the current knowledge-based economy, an enterprise's core competitiveness has changed from physical assets to knowledge-based assets, which encompass the operational performance it generates through knowledge resources. The structural position of an enterprise embedded into industry association networks offers various opportunities for it to learn from other enterprises and to have access to new knowledge that is essential for generating innovative ideas and attractive products. In other words, an enterprise with a suitable understanding of its structural position can become a gatekeeper of reliable or novel information concerning the know-how and market trends that can be spread by networks. In fact, the public sectors can allocate much more resources towards specific enterprises that have a superior network position so as to strengthen their operating performances

Table 12 The statistical outcomes of each model

Base model Compared model	IFWTSVM SVM	IFWTSVM RF	IFWTSVM NN
Hypothesis	$H_0 : \mu_{IFWTSVM} = \mu_{SVM,RF,NN}$ $H_1 : \mu_{IFWTSVM} \neq \mu_{SVM,RF,NN}$		
ACC	P-value = 0.009 (***)	P-value = 0.000 (***)	P-value = 0.000 (***)
PRE	P-value = 0.018 (**)	P-value = 0.014 (**)	P-value = 0.002 (***)
REC	P-value = 0.015 (**)	P-value = 0.000 (***)	P-value = 0.000 (***)
F-M	P-value = 0.010 (**)	P-value = 0.001 (**)	P-value = 0.000 (***)

ACC accuracy, *PRE* precision, *REC* recall, *F-M* F-measure

* P-value < 0.1; ** P-value < 0.05; *** P-value < 0.01

as well as to upgrade a country's overall industrial level. The literature has very infrequently considered industry association networks when constructing an enterprise operating performance forecasting model. To fill this gap, we have herein set up the structural position model of an enterprise.

Inspired by the hybrid mechanism, this study has proposed a novel architecture, the IFWTSVM model, for enterprise operating performance forecasting. IFWTSVM can be divided into two main parts: (1) essential feature subset selection and (2) forecasting model construction. We execute the essential feature subset selection by an incremental filter-wrapper strategy and a statistical examination. With its superior forecasting quality and efficiency, we select TSVM to construct the forecasting model. The result can be taken as a pre-warning model for investors to modify their investing strategies and to more suitably allocate their personnel wealth into financial derivatives for risk hedging purposes.

To facilitate the application of the black-box model (the TSVM-based model), this study has implemented PSO to extract the inherent knowledge from IFWTSVM. The knowledge is represented in a more intuitive and human readable way. Managers of enterprises can thus utilize the rule base (i.e., knowledge base) as a tutorial to adjust their operational direction as well as to decrease their firm's exposure to risk.

Several issues remain that could be considered in future research. First, the forecasting model can be extended into a more sophisticated hybrid mechanism, such as a parallelized learning strategy and a fuzziness category-based divide-and-conquer strategy, so as to strengthen forecasting performance [3, 34, 55]. Second, in addition to data dimensionality reduction techniques (i.e., feature selection and feature extraction), other data preprocessing techniques (i.e., data uncertainty elimination by the fuzzy set theory) can be implemented for comparison [22, 56].

References

1. Abbasi A, Altmann J, Hossain L (2011) Identifying the effects of co-authorship networks on the performance of scholars: a correlation and regression analysis of performance measures and social network analysis measures. J Informetr 5:594–607
2. Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J Financ 23:589–609
3. Ashfaq RAR, Wang XZ, Huang JZX, Abbas H, He YL (2016) Fuzziness based semi-supervised learning approach for intrusion detection system (IDS). Inform Sci. doi:10.1016/j.ins.2016.04.019
4. Avkiran NK (2015) An illustration of dynamic network DEA in commercial banking including robustness tests. Omega 55:141–150
5. Bai L, Wang Z, Shao YH, Deng NY (2014) A novel feature selection method for twin support vector machine. Knowl-Based Syst 59:1–8
6. Bermejo P, Gámez JA, Puerta JM (2011) A GRASP algorithm for fast hybrid (filter-wrapper) feature subset selection in high-dimensional datasets. Pattern Recognit Lett 32:701–711
7. Bermejo P, Ossa L, Gámez JA, Puerta JM (2012) Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking. Knowl-Based Syst 25:35–44
8. Boj JJ, Rodriguez-Rodriguez R, Alfaro-Saiz JJ (2014) An ANP-multi-criteria-based methodology to link intangible assets and organizational performance in a balanced scorecard context. Decis Support Syst 68:98–110
9. Bray S, Caggiani L, Ottomanelli M (2015) Measuring transport systems efficiency under uncertainty by fuzzy sets theory based data envelopment analysis: theoretical and practical comparison with traditional DEA model. Transp Res Proced 5:186–200
10. Charnes A, Cooper W, Rhodes E (1978) Measuring the efficiency of decision-making units. Eur J Oper Res 2:429–444
11. Chen WJ, Shao YH, Hong N (2014) Laplacian smooth twin support vector machine for semi-supervised classification. Int J Mach Learn Cybern 5:459–468
12. Chen M, Ludwig SA (2014) Particle swarm optimization based fuzzy clustering approach to identify optimal number of clusters. J Artif Intell Soft Comput Res 4:43–56
13. Dagher J, Kazimov K (2015) Banks' liability structure and mortgage lending during the financial crisis. J Financ Econ 116:565–582
14. Dimitras AI, Zanakis SH, Zopounidis C (1996) A survey of business failures with an emphasis on prediction methods and industrial applications. Eur J Oper Res 90:487–513
15. Fancello G, Uccheddu B, Fadda P (2014) Data envelopment analysis (D.E.A.) for urban road system performance assessment. Proced—Soc Behav Sci 111:780–789

16. Freeman L (1979) Centrality in social networks: conceptual clarification. Soc Netw 1:215–239
17. Gan Q, Hasan BAS, Tsui CS (2014) A filter-dominating hybrid sequential forward floating search method for feature subset selection in high-dimensional space. Int J Mach Learn Cybern 5:413–423
18. Guthrie J, Ricceri F, Dumay J (2012) Reflections and projections: a decade of intellectual capital accounting research. Br Acc Rev 44:68–82
19. Guimarães VA, Leal Junior IC, Almada Garcia PA (2014) Environmental performance of Brazilian container terminals: a data envelopment analysis approach. Proced—Soc Behav Sci 160:178–187
20. Geng R, Bose I, Chen X (2015) Prediction of financial distress: an empirical study of listed Chinese companies using data mining. Eur J Oper Res 241:236–247
21. Gnyawali DR, Madhavan R (2001) Cooperative networks and competitive dynamics: a structural embeddedness perspective. Acad Manage Rev 26:431–445
22. He YL, Wang XZ, Huang JZX (2016) Fuzzy nonlinear regression analysis using a random weight network. Inform Sci. doi:10.1016/j.ins.2016.01.037
23. Huber C, Scheytt T (2013) The dispositif of risk management: reconstructing risk management after the financial crisis. Manage Acc Res 24:88–99
24. Jayadeva, Khemchandani R, Chandra S (2007) Twin support vector machines for pattern classification. IEEE Trans Pattern Anal Mach Intell 29:905–910
25. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of the fourth IEEE international conference on neural networks, Perth, Australia. IEEE Service Center 1942–1948
26. Kim DY (2014) Understanding supplier structural embeddedness: a social network perspective. J Oper Manage 32:219–231
27. Kim KJ (2003) Financial time series forecasting using support vector machines. Neurocomputing 55:307–319
28. Kim HS, Sohn SY (2010) Support vector machines for default prediction of SMEs based on technology credit. Eur J Oper Res 201:838–846
29. Kumar MA, Gopal M (2010) Fast multiclass SVM classification using decision tree based one-against-all method. Neural Process Lett 32:311–323
30. Kou G, Peng Y, Wang G (2014) Evaluation of clustering algorithms for financial risk analysis using MCDM methods. Inform Sci 275:1–12
31. Lin SJ, Hsu MF (2016) Incorporated risk-adjusted balanced scorecards and hybrid AI techniques for risk management. Neural Comput Appl. doi:10.1007/s00521-016-2253-4
32. Lin SJ (2016) Hybrid kernelized fuzzy clustering and multiple attributes decision analysis for corporate risk management. Int J Fuzzy Syst. doi:10.1007/s40815-016-0196-7
33. Lin SJ, Chen TF (2016) Multi-agent architecture for corporate operating performance assessment. Neural Process Lett 43:115–132
34. Lu SX, Wang XZ, Zhang GQ, Zhou X (2015) Effective algorithms of the Moore-Penrose inverse matrices for extreme learning machine. Intell Data Anal 19:743–760
35. Oliveira L, Rodrigues LL, Craig R (2010) Intangible assets and value relevance: evidence from the Portuguese stock exchange. Br Acc Rev 42:241–252
36. Prado RP, Galan SG, Exposito JE, Yuste AJ (2010) Knowledge acquisition in fuzzy-rule-based systems with particle-swarm optimization. IEEE T Fuzzy Syst 18:1083–1097
37. Peng Y, Wang G, Kou G, Shi Y (2011) An empirical study of classification algorithm evaluation for financial risk prediction. Appl Soft Comput 11:2906–2915
38. Peng Y, Wang G, Wang H (2012) User preferences based software defect detection algorithms selection using MCDM. Inf Sci 191:3–13
39. Qiao P, Ju X, Fung HG (2014) Industry association networks, innovations, and firm performance in Chinese small and medium-sized enterprises. China Econ Rev 29:213–228
40. Ruiz R, Riquelme JC, Aguilar-Ruiz JS (2006) Incremental wrapper-based gene selection from microarray data for cancer classification. Pattern Recognit 39:2383–2392
41. Shao YH, Chun CH, Wang XB, Deng NY (2011) Improvements on twin support vector machines. IEEE T Neural Netw 22:962–968
42. Shao YH, Chen WJ, Zhang J, Wang Z, Deng NY (2014) An efficient weighted Lagrangian twin support vector machine for imbalanced data classification. Pattern Recognit 47:3158–3167
43. Striukova L, Unerman J, Guthrie J (2008) Corporate reporting of intellectual capital: evidence from UK companies. Br Acc Rev 40:297–313
44. Sueyoshi T, Wang D (2014) Sustainability development for supply chain management in US petroleum industry by DEA environmental assessment. Energy Econ 46:360–374
45. Sun J, Li H (2008) Data mining method for listed companies' financial distress prediction. Knowl-Based Syst 21:1–5
46. Sun J, He KY, Li H (2011) SFFS-PC-NN optimized by genetic algorithm for dynamic prediction of financial distress with longitudinal data streams. Knowl-Based Syst 24:1013–1023
47. Soman S, Jayadeva (2015) High performance EEG signal classification using classifiability and the Twin SVM. Appl Soft Comput 30:305–318
48. Tinoco MH, Wilson N (2013) Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. Int Rev Financ Anal 30:394–419
49. Tsai CF, Lu YH, David Yen C (2012) Determinants of intangible assets value: the data mining approach. Knowl-Based Syst 31:67–77
50. Tsai W (2000) Social capital, strategic relatedness and the formation of intraorganizational linkages. Strat Manage J 21:925–939
51. Tsai W (2001) Knowledge transfer in intraorganizational networks: effects of network position and absorptive capacity on business unit innovation and performance. Acad Manage J 44:996–1004
52. Tomar D, Agarwal S (2015) Twin support vector machine: a review from 2007 to 2014. Egypt Inform J 16:55–69
53. Vapnik V (1995) The nature of statistical learning theory. Springer, New York
54. Wang XY, Yang HY, Zhang Y, Fu ZK (2013) Image denoising using SVM classification in nonsubsampled contourlet transform domain. Inf Sci 246:155–176
55. Wang XZ, Ashfaq RAR, Fu AM (2015) Fuzziness based sample categorization for classifier performance improvement. J Intell Fuzzy Syst 29:1185–1196
56. Wang XZ (2015) Uncertainty in learning from big data-editorial. J Intell Fuzzy Syst 28:2329–2330
57. Xiang J, Han XH, Duan F, Qiang Y, Xiong XY, Lan Y, Chai H (2015) A novel hybrid system for feature selection based on an improved gravitational search algorithm and k-NN method. Appl Soft Comput 31:293–307
58. Xu X, Wang Y (2009) Financial failure prediction using efficiency as a predictor. Expert Syst Appl 36:366–373
59. Yang CH, Motohashi K, Chen JR (2009) Are new technology-based firms located on science parks really more innovative?: evidence from Taiwan. Res Policy 38:77–85