International Journal of u- and e- Service, Science and Technology
Vol.8, No.2 (2015), pp.395-404
http://dx.doi.org/10.14257/ijunesst.2015.8.2.38

A Study on DS-SVM-Based Forecast Model from the View of Information Fusion

Weidong Zhu[1], Yibo Sun[2], Yong Wu[3] and Jingyu Liu[4]

[1]School of Economics, Hefei University of Technology, Hefei, Anhui, 230009, China
Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei, Anhui, 230009, China
Institute of Industrial Information, Hefei University of Technology, Hefei, Anhui, 230009, China
[2]School of Management, Hefei University of Technology, Hefei, Anhui, 230009, China
[3]School of Management, Hefei University of Technology, Hefei, Anhui, 230009, China
Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei, Anhui, 230009, China
[4]School of Management, Hefei University of Technology, Hefei, Anhui, 230009, China
[1] zhuwd@hfut.edu.cn, [2] yibo2237@126.com,
[3] wuyong@hfut.edu.cn, [4] 834030756@qq.com

Abstract

Dempster-Shafer Theory is specially advantaged in information fusion, while Support Vector Machine (SVM) can well deal with high-dimensional limited sample data. This Article firstly forecasts the data samples by categories with multiple SVMs, and hence based thereon, fuses the resulting information from multiple SVM models by using DS theory. At the end, Anderson's Iris data set is used to simulate the system of the created DS-SVM model, which shows that the approaches proposed in this Article can not only increase the accuracy rate of the SVM forecasting model, but also improve the propensity scores of the SVM forecasting model.

Keywords: *DS evidence theory；SVM；information fusion；Dempster's Rule*

1. Introduction

Since first proposed by Dempster and then developed by Shafer, Dempster-Shafer theory of evidence, also called DS evidence theory or evidence theory has been widely used in numerous aspects as a kind of uncertain reasoning method. Such as group decision making, data fusion, and multiple attribute evaluation [1-7]. Many researchers have refined and expanded the evidence theory [3, 4, 8-12]. In the evidence theory, evidence includes not only attributes and objective environment when people analyze the problems, but also decision maker's experience, knowledge, and the observation and study on decision problems. To scientifically and reasonably process various types of evidences and get the accurate basic belief assignment (BPA) is a prerequisite for the practical application of the evidence theory for decision-making. Shafer and Smets discussed how to get BPAs based on expert's experience and judgment [13, 14]; Yang Shan-lin, Sikder, Wang Jia-yang obtained BPAs from different perspective on information systems based on rough set theory [15-18]; Deng used a similarity degree of interval numbers to

ISSN: 2005-4246 IJUNESST

calculate BPAs based on the historical data[19, 20]; Yang Lu-Jing studied the data sample by using the method of neural network, and took the conditional probability as BPA[21].

SVM is a powerful classification and regression technique while it can maximize the prediction accuracy of the model, and deal with small sample, high dimension, and non-normal data very well. It is widely used in pattern recognition, function estimation, regression analysis, time series predictive, text recognition, handwriting recognition, face image recognition, genetic classification and time series prediction, etc.[22-27]. This paper introduces the SVM in the DS evidence theory. To get the BPAs, we first determine the frame of discernment according to the specific classification problem and then use several different types of SVM's kernel function to deal with small sample. We get final classification prediction from the fusion of the results of DS-SVM model based on Dempster's Rule.

The rest of this paper is organized as follows. Section 2 reviews related concepts and formulas in DS evidence theory and SVM. Section 3 gives a detailed introduction about DS-SVM model including the determination of evidence identification framework and BPAs, the algorithm and process of DS-SVM. In Section 4, it shows a concrete example to help the readers understanding DS-SVM model, and analyzes the advantage of DS evidence theory for fusion. The fifth part summarizes the full text.

2. Basic Information

In this section, we briefly introduce DS evidence theory and Support Vector Machine.

2.1. DS evidence Theory

Evidence theory was originally investigated in the 1960's by Shafer, in this part, the basic concepts of DS theory are briefly discussed. It enables us to combine evidence from different sources and arrive at a degree of belief. It has been an important method for the study of information fusion.

Definition 1 (Frame of discernment, FOD). Assume: $\Theta = \{\theta_1, \theta_2 \dots, \theta_n\}$ is a finite set of identifiable elements, called the frame of discernment, Θ will acquire its meaning form what we know or we think we know. The set containing all subsets of Θ is named the power set and denoted by 2^{Θ}.

Definition 2 (Basic probability assignment, BPA). Θ is a frame of discernment, the function f: $2^{\Theta} \rightarrow [0, 1]$ is called basic probability assignment whenever

$$\begin{cases} m(\varnothing) = 0 \\ \sum_{A \subset \Theta} m(A) = 1 \end{cases} \quad (1)$$

The quantity $m(A)$ is called A's basic probability number, and it is understood to be the measure of the belief that is committed exactly to A. Any subset A of 2^{Θ} such that $m(A)>0$ is called a focal element.

Definition 3 (Belief function, BF).

$$Bel(A) = \sum_{B \subset A} m(B) \quad (2)$$

A function Bel: $2^{\Theta} \rightarrow [0, 1]$ is called a belief function over FOD Θ if it is given by (2) for some basic probability assignment f: $2^{\Theta} \rightarrow [0, 1]$.

Definition 4 (Plausibility Function, PF).

$$pl(A) = \sum_{A \cap B \neq \varnothing} m(B) \quad (3)$$

A function $pl: 2^{\Theta} \rightarrow [0,1]$ is called a plausibility function over FOD Θ if it is given by (3) for some basic probability assignment $f: 2^{\Theta} \rightarrow [0,1]$.

The relationship of belief function to plausibility function is:

$$pl(A) = 1 - Bel(\overline{A}) \tag{4}$$

Basic probability assignment, belief function and plausibility function can be calculated with each other, so if we know the one of them, other functions can be computed.

Definition 5 (Dempster's rule of combination). Suppose m_1 and m_2 are basic probability assignments over *FOD Θ*, with focal elements $A_1,\ldots A_K$ and $B_1,\ldots B_l$, respectively.

Suppose

$$\sum_{\substack{i,j \\ A_i \cap B_j = \phi}} m_1(A_i) m_2(B_j) < 1$$

Then the function *f*: $2^{\theta} \rightarrow [0,1]$ defined by

$$m(C) = m(A) \oplus m(B) = (1-K) \sum_{\substack{i,j \\ A_i \cap B_j = C}} m_1(A_i) m_2(B_j) \tag{5}$$

$$K = \sum_{\substack{i,j \\ A_i \cap B_j = \phi}} m_1(A_i) m_2(B_j) \tag{6}$$

For all non-empty $A \subset \theta$ is a basic probability assignment .The core of the belief function given by *m* is equal to the intersection of the cores of Bel_1 and Bel_2. Dempster's Rule of combination is the core of Theory of Evidence, it reflects the combined effects between multiple evidences.

2.2. SVM-Related Information

SVM classification mechanism can be summarized: Support vector machine (SVM) to find an optimal hyperplane that maximize margin between two classes, the schematic diagram of linearly separable optimal hyperplane as shown in the Figure 1.

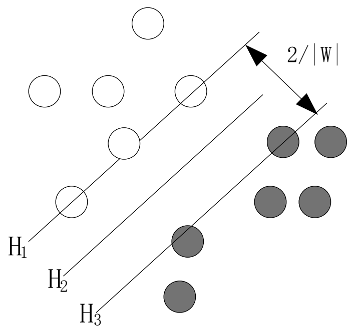


Figure 1. The Schematic Diagram of Linearly Separable Optimal Hyper Plane

Support vector machine show many unique advantages in tackling small sample, nonlinear and high dimensional space in the pattern recognition problems, the mathematical model are as follows: Input two type data in "*m*" space, SVM inside the space to construct a hyper-plane to distinguish between two types of data, The boundary of this hyper-plane distance of two classes of data is the largest. Specifically, assuming that a given training set $\{x_i, y_i\}$ ($i = 1,2,\ldots,l$, $y_i \in \{-1,1\}$, $x_i \in R^D$, the x_i contains D features), hyper-plane can use $w * x + b = 0$ signify, *w* is super-plane normal vector, all of the training samples meet the conditions:

$$y_i(x_i \cdot w + b) - 1 + \xi_i \geq 0,\ \xi_i \geq 0 \tag{7}$$

In order to solve the problem of linear inseparable introduction of non-negative slack variable ξ_i, $i = 1,2,\dots,l$, To maximize the hyper-plane of boundary can be converted into a convex quadratic optimization problem:

$$\begin{aligned} &\min \quad \|w\| \big/ 2w^2 + C \cdot \sum_{i=1}^{l} \xi_i \\ &s.t. \quad y_i(x_i \cdot w + b) - 1 + \xi_i \geq 0,\ \xi_i \geq 0 \end{aligned} \tag{8}$$

Where C is punishment coefficient, used to balance the size of the slack variables and classification boundary, by solving the above optimization problem, get:

$$f(x) = sign(\sum_{i=1}^{l} a_i y_i K(x_i, x) + b) \tag{9}$$

Where $K(x_i, x)$ is satisfy the Mercer kernel function.

3. DS-SVM-based Information Fusion Model

Firstly, DS-SVM prediction model has to choose and determine the kernel function of support vector machines, and then study samples respectively by using support vector machine model and get the BPAs from the results of the SVM model. Lastly, it gives the output from the fusion of BPAs based on the DS evidence theory.

3.1. Selection of Kernel Function

Kernel function (KF) can convert the inner product operation of high-dimensional space to low-dimensional input space operations, so solved the calculation in high dimensional feature space of the problem such as "dimension disaster" cleverly, at the same time can greatly reduce the amount of calculation. There are four main traditional kernel functions:

(1) Liner Kernel Function:

$$K(x, y) = x^T y \tag{10}$$

(2) Polynomial Kernel Function:

$$K(x, y) = \left[\left(x^T y \right) + c \right]^p,\ p \in N,\ c \geq 0 \tag{11}$$

(3) Radial Basis Function (RBF):

$$K(x, y) = exp\ (-\delta \left| x - y \right|^2),\delta > 0 \tag{12}$$

(4) Sigmoid Kernels Function (SKF):

$$K(x, y) = tanh(a(x, y) + v), a > 0,\ v > 0 \tag{13}$$

Kernel function is a key factor of SVM, the performance of different kernel functions and parameters of SVMs has a large difference. Selection of kernel function can be determined by the expert's experience of SVM in practical, can also be determined by trial of different kernel functions.

3.2. Method for Determination of SVM-Based BPA

From the evidence theory knowledge can be known, Evidence identification framework is built for the specific decision problem, Θ it is made for the information of decision makers " want to know " and " already know ". Assuming that the classification of the SVM model to predict problems have n kinds of classification results, so the identification frame Θ can be determined: $\Theta = \{\theta_1, \theta_2 \dots \theta_n\}$.

When using SVM classification, propensity score can be calculated based on the classification results [28], it signed the degree of accurate predicted, the degree of accurate predicted is higher, and the model forecast accuracy is higher. Assuming that the SVM prediction results is θ_i, the corresponding propensity score is λ; According to the meaning of BPA, 1-λ can be assigned to any element in the recognition framework except θ_i , but not sure the assigned to which elements specifically , thus the article is based on the prediction results to build the BPA as follows:

$$m(A)=\begin{cases}\lambda & A=\theta_i\\ 1-\lambda & A=C_\Theta\theta_i\\ 0 & A=\varnothing\end{cases} \tag{14}$$

Where $C_\Theta\theta_i$ means the complement of θ_i in frame of discernment Θ.

3.3. DS-SVM Model Calculation Flow

Step 1. Using different kernel function of SVM to classify data set;
Step 2. Identify framework of discernment, and convert all the study results of the SVM model into BPAs;
Step 3. Using the Dempster's Rule integrate BPAs into the final BPA*;
Step 4. Analyze the BPA*, determine the final classification results.

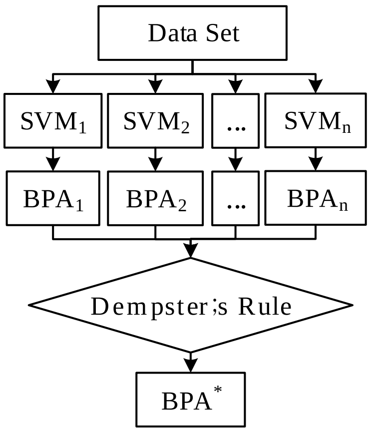


Figure 2. DS-SVM Calculation Flow

4. Case Study

The paper uses Anderson's iris data set to inspect DS_-SVM model. Anderson's Iris data set is a multivariate data set which introduced by Ronald Fisher in 1936. The data set consists of 50 samples from each of three species of Iris flowers: Iris setosa, Iris virginica and Iris versicolor (respectively as *Se*, *Ve* and *Vi*). Four attributes are used in quantitative analysis for each sample. These attributes are the length of sepal, the length of petal, the width of sepal and the width of petal.

Through the experiment, we choose sigmoid kernel function and polynomial kernel function to build a SVM forecasting model respectively. The SVM_1 classification algorithms uses sigmoid kernel function (denoted by SVM (SKF)), and the SVM_2 classification algorithms uses polynomial kernel function (denoted by SVM (PKF)), because the Anderson's Iris data set have three subgenus: *Se*, *Ve* and *Vi*, so the frame of discernment Θ has three elements: $\Theta=\{Se,\ Ve,\ Vi\}$.

We use SVM (SKF) and SVM (PKF) to classify 150 samples of Anderson's Iris data set, and the accuracy of the model has been shown in the figure 3. We can find that the accuracy of SVM (PKF) model is 90% and the accuracy of SVM (SKF) model is 95.3%. Using the formula 7 to deal with the results of the two models, we transform the match scores of prediction results into BPAs, and then we get the accuracy of 98% from the fusion of BPAs based on Dempster's Rule. Obviously the DS-SVM model can improve the accuracy of SVM model.

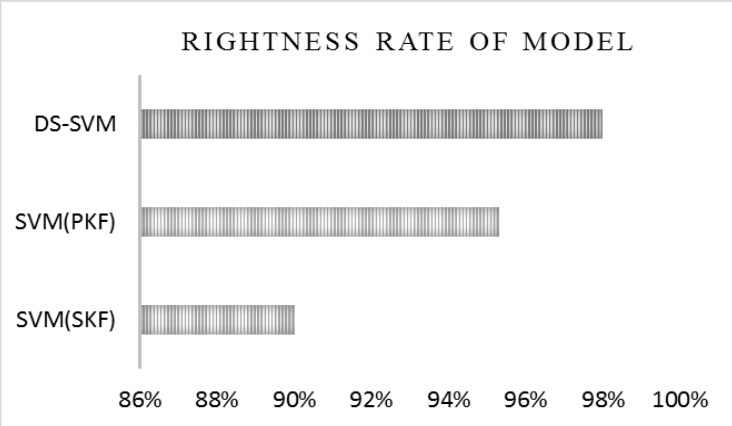


Figure 3. Rightness Rate of Model

Especially, we choose two incorrect data for numerical example. When the length and width of calyx is 7 and 3.2 and the length and width of petals is 4.7 and 4.7, SVM (SKF) has the wrong judgment "virginica" while the truth is "versicolor". The match score is 0.683. SVM (PKF) makes correct judgment, and the match score is 0.960. We can get BAP_1 of SVM (SKF) model and BAP_2 of SVM (PKF) model respectively from the formula (14).

$$BPA_1:\ m_1(A)=\begin{cases}0.683 & A=\{Vi\}\\ 0.317 & A=\{Se,Ve\}\end{cases}$$

$$BPA_2:\ m_2(A)=\begin{cases}0.960 & A=\{Ve\}\\ 0.040 & A=\{Se,Vi\}\end{cases}$$

Using Dempster's Rule to compound BPA_1 and BPA_2, we can get BPA^*:

$$BPA^*:\ m^*(A)=\begin{cases}0.884 & A=\{Ve\}\\ 0.037 & A=\{Se\}\\ 0.079 & A=\{Vi\}\end{cases}$$

From the final synthesis result, DS-SVM model gives the correct judgment "Iris Versicolor". It is consistent with the original data set classification, and is an amendment on the judgment of SVM (SKF) model.

Due to the propensity score of prediction can represent the accuracy of the prediction, the higher propensity score, and the higher accuracy. The propensity score can reflect the creditability of the model. Through the analysis of the sample, we get the propensity score of SVM (SKF), SVM (PKF) and DS-SVM in the Figure 4.

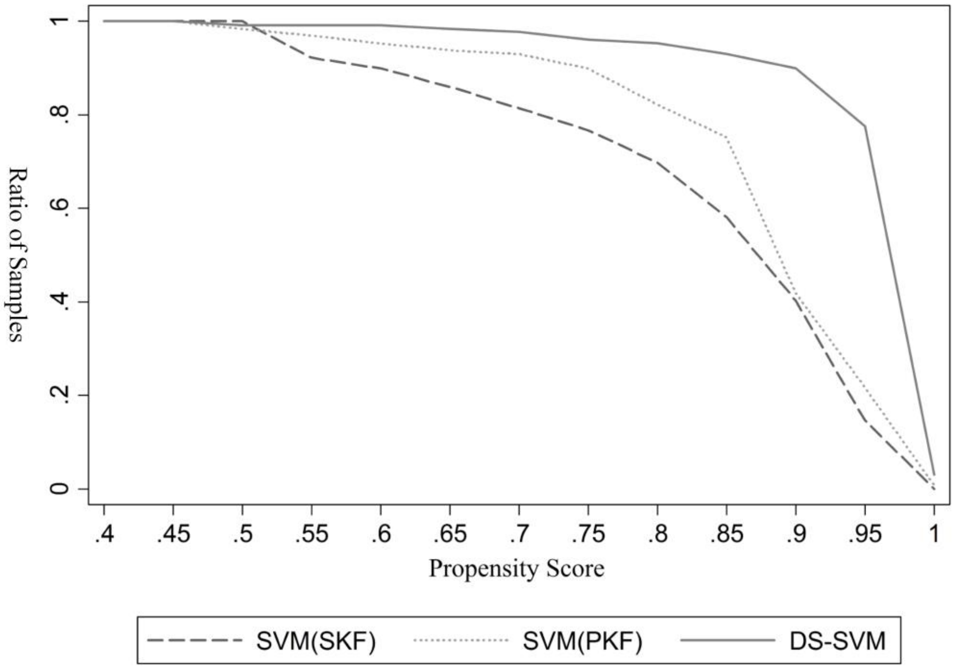


Figure 4. Distribution of Propensity Score

In the above diagram, when the propensity score is greater than 0.55, the sample proportion line of DS-SVM model is always higher than SVM (SKF) and SVM (PKF). It shows that DS-SVM model can improve the propensity score of the original model by using the DS evidence theory. For another, the line of SVM (SKF) and SVM (PKF) goes down quickly when the propensity score is greater than 0.75 and only about 40% of the samples have the propensity score greater than 0.9. However, the synthetic result based DS evidence theory show that over 90% of the samples have the propensity score greater than 0.9, over 79% of the samples have the propensity score greater than 0.95. This indicates that when DS evidence theory is introduced in SVM, the propensity score of the data set will be strengthened. In other words, evidence theory can improve the accuracy of prediction.

5. Conclusion

To sum up, DS-SVM classification algorithm has the following two advantages.

(1) DS-SVM model can enhance reliability of the classification results of the SVM model. DS evidence theory has a characteristic: If two evidences support a proposition at the same time, the support degree of the proposition get bigger after using the Dempster's Rule to fusion evidences. Due to the BPAs in the DS evidence theory is determined by the propensity score of each SVM model, so for the right data from SVM classification model, Dempster's Rule will make the propensity score of the model get bigger, and the reliability of the model is bigger.

(2) DS-SVM model can improve the original SVM model's accurate rate of the judgment. Through the analysis of section 4 can be found, for the error output of the model, synthesized by Dempster's Rule could change the judgment results of the model, thus can improve the accurate rate of the original SVM judgment.

Acknowledgements

This work is partially supported by National Natural Science Foundation of China, Grant Nos. 71071048, 71302063; the Soft Science Research Plan in Anhui Province, Grant No. 1402052006.

The authors would like to thank the referees, the editors and the anonymous reviewers for their helpful comments and constructive suggestions.

References

[1] Smets P. The combination of evidence in the transferable belief model. Pattern Analysis and Machine Intelligence, IEEE Transactions on 12(5): 447--458, **(1990)**.

[2] Beynon MJ. The Role of the DS/AHP in Identifying Inter-Group Alliances and Majority Rule Within Group Decision Making. Group Decision and Negotiation 15(1): 21-42, **(2006)**

[3] Feng T, Mi J, Zhang S. Belief functions on general intuitionistic fuzzy information systems. Information Sciences **(2014)**.

[4] Denoeux T. Maximum likelihood estimation from uncertain data in the belief function framework [J]. Knowledge and Data Engineering, IEEE Transactions on, 25(1): 119-130. **(2013).**

[5] Kang B, Deng Y, Sadiq R, Mahadevan S. Evidential cognitive maps. Knowledge-Based Systems 35:77-86, **(2012)**.

[6] Mercier D, Den Oe Ux T, Masson MELE. Belief function correction mechanisms. In: Foundations of Reasoning under Uncertainty, vol Springer, 203--222; **(2010)**.

[7] Smets P. Decision making in the TBM: the necessity of the pignistic transformation. International Journal of Approximate Reasoning 38(2): 133-147, **(2005)**.

[8] Leung Y, Ji N, Ma J. An integrated information fusion approach based on the theory of evidence and group decision-making. Information Fusion 14(4): 410-422, **(2013)**.

[9] Yang J, Xu D. Evidential reasoning rule for evidence combination. Artificial Intelligence 205:1-29, **(2013)**.

[10] Powell G, Roberts M. GRP1. A recursive fusion operator for the transferable belief model. In1--8; **(2011)**.

[11] Benferhat S, Saffiotti A, Smets P. Belief functions and default reasoning. In19--26; **(1995)**.

[12] Sentz K, Ferson S. Combination of evidence in Dempster-Shafer theory [M]. Sandia Report, **(2002)**.

[13] Shafer G. A mathematical theory of evidence. Princeton: Princeton university press, **(1976)**.

[14] Smets P. Belief functions on real numbers. International journal of approximate reasoning 40(3): 181--223, **(2005)**.

[15] Yang S L, Liu Y Z, Li Y F. Research on rough sets-based evidence acquirement and combination of DST [J]. Journal of management sciences in china, 8(5): 69-75, **(2005)**.

[16] Sikder IU, Gangopadhyay A. Managing uncertainty in location services using rough set and evidence theory. Expert Systems with Applications 32(2): 386--396, **(2007)**.

[17] Wu X，Wang J Y. Research on evidence acquirement and combination method based on rough set [J]. Computer Engineering and Applications 47(12): 34-35, 64, **(2011)**.

[18] ZHOU Yong, WANG Jia-yang. Study of evidence combination method based on rough set theory [J]. Computer Engineering and Applications, 44(7): 44-46, 50, **(2008)**.

[19] Kang B Y, Li Y, Deng Y. Determination of Basic Probability Assignment Based on Interval Numbers and Its Application [J]. Acta electronica sinica, (06): 1092-1096, **(2012)**.

[20] Xu P D, Hang D Q, Deng Y. An Optimal Transformation of Basic Probability Assignment to Probability [J]. Acta electronica sinica, (S1): 121-125, **(2011)**.

[21] Yang L J, Hao W. Multi-sensor Target Recognition Method Based on FMM Artificial Neural Network and D-S Evidence Reasoning [J]. Journal of Detection& Control, (06): 95-98, **(2005)**.

[22] Ha M, Wang C, Chen J. The support vector machine based on intuitionistic fuzzy number and kernel function. Soft Computing 17(4): 635-641, **(2013)**.

[23] Chen H, Yang B, Liu J, Liu D. A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. Expert Systems with Applications 38(7): 9014-9022, **(2011)**.

[24] Kisi O, Cimen M. A wavelet-support vector machine conjunction model for monthly stream flow forecasting. Journal of Hydrology 399(1): 132-140, **(2011)**.

[25] Wong WK, Xia M, Chu WC. Adaptive neural network model for time-series forecasting. European Journal of Operational Research 207(2): 807-816, **(2010)**.

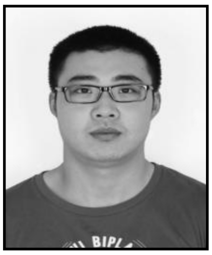
[26] Chapelle O, Vapnik V, Bousquet O, Mukherjee S. Choosing multiple parameters for support vector machines. Machine learning 46(1-3): 131-159, **(2002)**.

[27] Chapelle O, Vapnik V. Model Selection for Support Vector Machines[C]. Advances in Neural Information Processing Systems. 230-236, **(2000)**.

[28] Platt J C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods[C] Advances in large margin classifiers. **(1999)**.

Authors

Weidong Zhu, is a Doctor of Management, Hefei University of technology school of management professor, doctoral supervisor, Association for Information System of individual members. Mainly engaged in intelligent decision-making theory, accounting theory research and enterprise management information.



Yibo Sun, received his bachelor degree of engineering from Hefei University of Technology, Hefei, China in 2010. Now he is a doctoral candidate in School of Management, Hefei University of Technology. His current interests include cloud computing, intelligent decision-making theory.

Yong Wu, is a Doctor of Management, Hefei University of technology school of management professor, master tutor. Mainly engaged in intelligent decision-making theory, accounting theory research and Knowledge Management.

Jingyu Liu, received his Master degree of Economics from Hefei University of Technology, Hefei, China in 2007. Now he is a doctoral candidate in School of Management, Hefei University of Technology. His current interests is intelligent decision-making theory and financial econometrics.