Advanced Science and Technology Letters
Vol.82 (SoftTech 2015), pp.41-47
http://dx.doi.org/10.14257/astl.2015.82.09

A New Forecast Model Based on Dempster-Shafer Theory and Support Vector Machine

Weidong Zhu, Yibo Sun, Yong Wu, Jingyu Liu

[1] School of Management, Hefei University of Technology, Hefei, Anhui, 230009, China;
zhuwd@hfut.edu.cn; Yibo2237@126.com,
wuyong@hfut.edu.cn; 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 add propensity scores to the results of the SVM forecasting model.

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

1 Introduction

DS evidence theory has been widely used in many 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 which the people analysis of proposition to get the basic credibility basis on, but also includes people's experience, knowledge, and the observation and study on the problem. 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 theory of evidence for decision-making. Shafer (1976) and Smets (2005) discussed how to get BPAs based on expert's experience and judgment [13, 14]; Yang Shan-lin(2005), Sikder(2007),Wang Jia-yang (2008, 2011) based on rough set theory from different perspective on information systems to get BPAs [15-18]; Deng(2011, 2012) used a similarity degrees of interval numbers to calculate BPAs from the historical data[19, 20]; Yang Lu-Jing(2005) studied the data sample by neural network, and then take the conditional probability as BPA[21].

SVM is a powerful classification and regression technique, it can maximize the prediction accuracy of the model, and deal with small sample very well, high dimension, non-normal data, and it is widely used in pattern recognition, function estimation, and time series prediction, etc.[22-24]. This paper introduced the SVM in the DS evidence theory, first of all, according to the specific classification problem to

ISSN: 2287-1233 ASTL

determine frame of discernment; Then it uses several different types of SVM's kernel function to deal with small sample, and get the results of BPA; Then using Dempster's Rule synthesis of BPA; Finally according to the judgment of DS - SVM fusion result, output the results of final classification prediction.

2 Basic Information

2.1. D-S Evidence Theory

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

Definition 1 (Frame of discernment, FOD). Assume: $\Theta = \{\theta_1, \theta_2 \ldots, \theta_n\}$ is a finite set of identifiable elements, called the frame of discernment. 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

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

The quantity *m* (*A*) is called *A*'s basic probability number. Any subset A of 2^{Θ} such that *m* (*A*)>0 is called a focal element.

Definition 3 (Dempster's rule of combination). Suppose m_1 and m_2 are basic probability assignments over the same frame *Θ*, with focal elements $A_1, \ldots A_K$ and $B_1, \ldots B_l$, respectively

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

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

$$K = \sum_{\substack{i,j \\ A_i \cap B_j = \phi}} m_1(A_i) m_2(B_j) < 1 \tag{6}$$

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.

2.2. SVM-Related Information

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 $\{x_i, y_i\}$ ($i = 1,2,\dots,l$, $y_i \in \{-1, 1\}$, $x_i \in R^D$, 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 is to choose and determine the kernel function of support vector machines, and then to study samples respectively by using support vector machine model, determined the BPA base on the output of the SVM model. Lastly, Using the DS evidence theory fused the BPAs, determine the output of DS - SVM model by analyses the fusion result.

When using SVM classification, propensity score can be calculated based on the classification results [25], 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 this article is based on the prediction results to build the BPA as follows:

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

4 Case Study

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

First we used SVM (SKF) and SVM (PKF) to classify 150 samples of Anderson's Iris data set, SVM (PKF) model of classification accuracy is 90%, accuracy of SVM (SKF) model is 95.3%. Using the formula 7 to deal with the results of the two models, changed the match scores of prediction results into BPAs, and Using the Dempster's Rule to compound the BPAs, the accuracy of predicted results is 98%, so the DS-SVM model can improve the accuracy of SVM model.

Especially, choose the two incorrect data for numerical demonstration' example. When the calyx length and width are 7 and 3.2 respectively, Petals of length and width are 4.7 and 4.7 respectively, SVM (SKF) put Versicolor judged to virginica wrongly, the match score is 0.683, SVM (PKF) makes correct judgment, the match score is 0.960, we can ensure BAP1 and BAP2 is SVM (SKF) model and SVM (PKF) model respectively by 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^*: $$m^*(A) = \begin{cases} 0.884 & A = \{Ve\} \\ 0.037 & A = \{Se\} \\ 0.079 & A = \{Vi\} \end{cases}$$

By the final synthesis results, the result of DS-SVM model is Iris Versicolor, It consistent with the original data set classification, fixed errors judgment of the SVM (SKF) model.

The propensity score can affect the creditability of the model by analysis model's result. Through the analysis of the sample, the propensity score of SVM (SKF), SVM (PKF) and DS-SVM as shown in the figure 1.

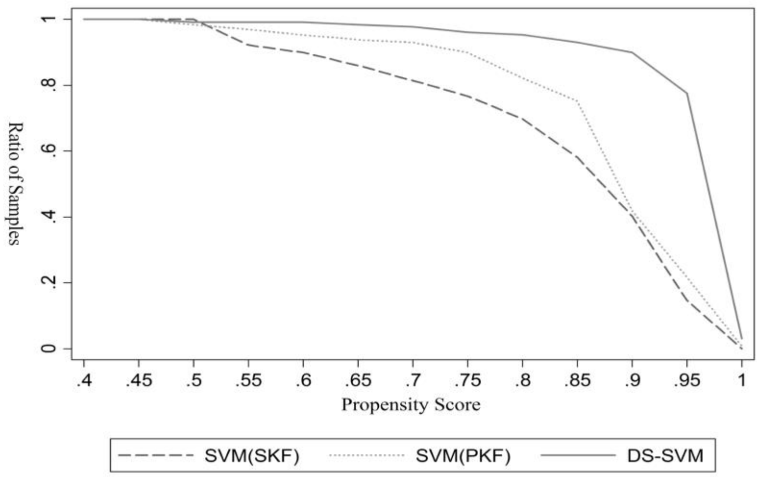

Fig. 1. Distribution of propensity score

5 Conclusion

To sum up, DS - SVM classification algorithm has the following two advantages.
(1) DS - SVM model to improve the reliability of the classification results of the SVM model. 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 to improves 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.

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, M. J.: 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, Y., WANG, J. Y.: 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. Platt J C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods[C] Advances in large margin classifiers. (1999).

