

The Research on Volleyball Trajectory Simulation Based on Cost-Sensitive Support Vector Machine

Yan Binghong

Northwestern Polytechnical University, Xi'an, China

Abstract: Aim of study of the study is to effectively solve the imbalanced data sets classification problem, the field of machine learning proposed many effective algorithms. On Seabrook's view, now the classification methods for class imbalance problem can be broadly divided into two categories, one class is to create new methods or improve the existing methods based on the characteristics of class imbalance. Another is to reduce class imbalance effect re-use of existing methods by resembling technique. But there are also some drawbacks in resembling method. Sample set sampling may lead to excessive learning; the next sample may result in the training set. Therefore, by dividing the training set, it does not increase the number of training samples; also it is without loss of useful information in the sample to obtain a certain degree of balance in a subset of the problem.

Keywords: Artificial Neural Network, autonomous hybrid power system static var compensator

INTRODUCTION

Support vector machine theory development has undergone a process of continuous improvement, the first study of statistical learning theory in the 1960s by V. Vapnik began to study and he can be called the founder of SVM. In 1971, "The Necessary and Sufficient Conditions for the Uniforms Convergence of Averages to Expected Values" in an article, V. Vapnik and A. Chervonenkis proposed VC dimension theory, it is an important theoretical basis of the SVM. V. Vapnik 1982 Estimation of Dependences Based on Empirical Data "a book made of structural risk minimum theory, this theory put forward is the cornerstone of the epoch-making significance, SVM algorithm. Boser, Guyon and Vapnik in 1992 proposed the optimal classifier. Cortes and Vapnik in 1993, further discussion of the nonlinear optimal boundary classification problem. V. Vapnik published in 1995, "The nature of statistical Learning Theory", a book, a complete proposed SVM theory.

The core of the support vector machine 1992 1995 SVM algorithm, SVM is so far the most successful statistical learning theory to achieve is still in the stage of continuous development. Support vector machine development process lasted a very short, but it has a solid theoretical foundation, because it is based on statistical learning theory and in recent years has emerged a lot of theoretical research, but also laid a solid foundation for applied research (Maloof, 2003; Barandela *et al.*, 2003).

BAYESIAN DECISION THEORY AND INSPIRATION

Bayesian decision theory is an important part of the subjective theory Bayeux Spirax summarized. It was founded by the British mathematician Bayesian.

Bayesian decision, in the case of incomplete data information, can estimate an unknown part with the subjective probability and finally uses Bayesian probability formula to fix the probability of occurrence. At last, the same expectations and correction of the final decision-making optimal probability is made.

Bayesian decision theory's basic idea can be described as follows: First class conditional probability density parameter expression and a priori probability of the premise, using Bayesian probability formula to convert it to a posteriori probability already know, the final utilization the posterior probability Size to optimal prediction decisions (Chawla *et al.*, 2002; Japkowicz and Stephen, 2002; Barandela *et al.*, 2004).

Bayesian inverse probability and using it as a universal method of reasoning is a significant contribution to the statistical reasoning. Bayesian formula is expressed on Bayes' theorem, a mathematical formula.

Assuming that B_1, B_2, \dots is a prerequisite of a process, $P(B_i)$ is the a priori probability, estimate the likelihood of each prerequisite. This process has been a result of A the Bayesian formula provided the preconditions to make a new evaluation method based on the appearance of a re-estimate of the probability of occurrence of B in the premise of a $P(B_i | A)$ posterior probability (Akbai *et al.*, 2004; Han *et al.*, 2005; Tang and Chen, 2008).

A set of methods and theory based on Bayesian formula has a very wide range of applications in real life. The following is a mathematical formula of the Bayesian formula expression:

Bayesian formula: Let D_1, D_2, \dots, D_n be a partition of the sample space S, said that the probability of the event

D_i to $P(D_i)$ and $P(D_i) > 0, i = 1, 2, \dots, n$. For any event x $P(x) > 0$:

$$P(D_j | x) = \frac{p(x | D_j)p(D_j)}{\sum_{i=1}^n p(x | D_i)p(D_i)}$$

Bayesian decision theory analysis: Bayesian decision theory is analyzed from the following aspects:

- The training sample collection known as classification categories probability distribution and has labeled category premise, this time to estimate the probability distribution parameters need to be estimated from the training sample collection.
- If there is no relevant information for any related probability distribution of the classification categories, but already know that the training sample set and the discriminant function in the form of labeled category, when you want to estimate the discriminant function of the parameters necessary, you need to estimate from the sample collection.
- Do not know any information about the probability distribution of the classification categories farewell function of the form, but does not know, only the training sample set of the category has been marked, then the probability distribution function of the parameters necessary to estimate from the training sample set.
- If other information are not only not being labeled set of samples of the class, this time in order to estimate the parameters of the probability distribution, the need for sample collection clustering operation.
- One of the best, already know the probability distribution of the classification categories, this case does not require training sample set directly using Bayesian decision theory can be optimal classifier.

Inspiration: Set any of the same this x belongs to class j of probability $P(j | x)$ Bayesian decision theory, the risk of sample classification i need to minimize conditions:

$$P(i | x) = \sum_j P(j | x)C(i, j) \tag{1}$$

The conditions minimize risks known as Bayesian risk. $i, j \in \{c_1, c_2, \dots, c_m\}$, m said the number of categories; $C(i, j)$ is a class sample j classification for i the risk, $i = j$ correct classification, $i \neq j$ wrong classification. Based on the accuracy 0-1 loss classifier IF $C(i, j) = 0, i = j, C(i, j) = 1, i \neq j$, the classification task is to find the maximum a posteriori probability.

Solve the problem of cost-sensitive data mining, $i \neq j, C(i, j) \neq C(j, i)$, relying only on the posterior probability of x to determine the category of the sample is not desirable. Given the cost of misclassification of a class of samples, the cost matrix can be re-constructed. Based on the formula (4-1) since you can solve the problem of data mining and make the minimum cost of misclassification, Bayesian decision theory can be used based on the above analysis shows that the cost of the different cost of misclassification embedded in the different categories of samples sensitive data mining (Wang *et al.*, 2007; Zhou and Liu, 2006; Rehan *et al.*, 2004).

CONSIDERATION MINIMIZE

Domingo proposed a new method of a classification model into consideration sensitive model, called Element consideration; it is a through meta-learning process, has been estimated $P(j | x)$ after class probability of the sample and then by a minimum desired consideration to modify the sample category tag.

Meta-learning process in the global scope, after knowledge received local learning secondary learning. Use appropriate learning program for each individual dispersed data meta-learning objectives and these procedures are performed together last relatively independent classifiers who classifiers are merged together to form a new data set, then the use of a new learning program to operate the new data set and finally generate meta-knowledge. Meta-learning biggest feature is any suitable algorithm can be used in the training phase to produce relatively independent classifier. Because of the use of meta-learning method, using a variety of integration methods in the initial stages, the last generated meta-classifier higher accuracy of prediction. Meta-learning flow chart shown in

Yuan cost of a classification learning algorithm based on Bayes decision theory. The algorithm process for the first of several from the training set sampling multiple models from multiple models in the training set for each sample belonging to the posterior probability of each class $P(j | x)$ then calculated for each sample belonging to each category expectations Consideration $R(j | x)$, last modified minimum expected cost category tag, get a new set of data, resulting in a new model, the minimum expected cost from

$$R(i | x) = \sum_j P(j | x)cost(i, j) \tag{2}$$

Each sample in the training set x , first the posterior probability $P(j | x)$ and then calculated according to the formula (1) which belongs to each category i consideration Further Reconstruction x class is marked:

$$\hat{y} = \begin{cases} +1, \\ -1, \text{ If } R(+1|x) \leq R(-1|x) \end{cases}$$

Misclassification class is marked with the sample information, is called the "true" class label sample.

COST-SENSITIVE SUPPORT VECTOR MACHINES

Different samples with different misclassification cost, cost-sensitive support vector machines (Weiss, 2004; Kubat and Matwin, 1997; Yin *et al.*, 2011; Zhao *et al.*, 2012) (CS-SVM) is the different samples misclassification integrated into the design of SVM, considering each sample has a different misclassification cost original sample set:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), x_i \in R, \quad (3)$$

$$y_i \in \{+1, -1\}, i = 1, 2, \dots, n$$

Reconstructed as follows:

$$(x_1, y_1, c_{o1}), (x_2, y_2, c_{o2}), \dots, (x_i, y_i, c_{oi}), \dots, (x_n, y_n, c_{on}), \quad (4)$$

$$x_i \in R, y_i \in \{+1, -1\}, c_{oi} \geq 0, i = 1, 2, \dots, n$$

Where in c_{oi} first i samples misclassification for the normal number, it depends on the x or the y_i . Let the sample set can be the hyperplane the $(w \cdot x) + b = 0$ classification and then the SVM-based data imbalance minimizes the objective function:

$$R(w, \xi) = \frac{1}{2} \|w\|^2 + C \left(\sum_{i=1}^n c_{oi} \xi_i \right) \quad (5)$$

s.t.

$$y_i (w \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, \dots, n \quad (6)$$

Including: $\|w\|^2$ for the structure of the consideration; for the experience of a consideration; C is the relaxation factor, the role is a balance between the consideration for the control structure and the experience of the consideration, for solving optimization problems is constructed as follows Lagrange equation:

$$L_p = \frac{1}{2} w \cdot w + C \sum_{i=1}^n c_{oi} \xi_i - \sum_i \hat{\alpha}_i \{y_i (x_i \cdot w + b) - 1 + \xi_i\} - \sum_i \hat{u}_i \xi_i \quad (7)$$

Where in: and the Lagrange coefficients minimize formula (7).

CS-SVM classification-oriented class offset the relatively small cost of misclassification, thus making it costly misclassification of samples can be correctly classified reducing overall misclassification.

COST-SENSITIVE SUPPORT VECTOR MACHINE BASED ON THE DATA SET DECOMPOSITION

Imbalance data set to solve classification problems in use sampling methods for data processing, there may exist because the training samples when the sampling of data sets on the training samples increased lead to excessive learning and performs down-sampling the reduction in the number and loss of the classification information of the sample. In order to avoid the above situation, a method is to divide the training set, the training set is divided into a plurality of subset have a certain degree of balance, the use of machine learning methods to train in each subset and then integrated learning, this method neither increasing the number of training samples, you will not lose a sample of classified information.

Solve the cost-sensitive data mining problem, when $i \neq j C(i, j) \neq C(j, i)$, if only rely on x maximum a posteriori probability does not determine the category of the sample, when given sample misclassification cost matrix can be re-constructed Bayeux misclassification costs embedded in different misclassification cost price sensitive issues, the Adams decision-making theory provides an implementation framework, based on Bayesian decision theory can to achieve the cost-sensitive, so that the global minimum.

Using Bayesian decision theory to deal with the cost-sensitive data set is divided into subcategories i , Class i subclass i minimum expected cost relative to other subclass misclassification Category i samples costly, then it will be the original does not belong to the subclass i on minimum expected cost part of the sample is allocated to the sub-class i , which also is a change in the sample class tag to reconstruct the sample.

The foregoing analysis, presents a cost-sensitive support vector machine (KCS-SVM) based on the data set decomposition.

It has the

$$L = \{(x_1, y_1), \dots, (x_m, y_m), y_i \in \{-1, +1\}, 1 \leq i \leq m\} \quad L = L_+ \cup L_-$$

of the training set, where in: positive class sample sets, negative class sample set $L_- = \{(x_{n+1}, -1), \dots, (x_m, -1)\}$, $x_i \in X \subseteq R^n$, represents a sample, n represents the training set the positive class number of samples, $m - n$ represents the number of the training set samples negative class.

In this algorithm, it is first decomposed into sample sets arbitrary negative class subsets. Then will decompose each negative subset L_- the positive class sample collection L_+ merged together, the combined k training set $L_i, 1 \leq i \leq k$. The proportion of the number of samples of the training set of positive and negative class can be controlled by adjusting the support vector

Table 1: Positive and negative sample size of data set

Number of samples	Data sets	The total number of samples	Is the number of samples	Number of negative samples	Unbalance rate
	Hepatitis	155	32	123	3.84
	Soybean	683	44	639	14.52

Table 2: Cost matrix

The true class category	Predicted Positive class samples	Negative class samples
Positive class samples	cost(p,p) = 0	cost(N,p) = 1
Negative class samples	cost(p,N) = 5	cost(N,N) = 0

machine trained on each subset can output posterior probability.

Take the training sample concentration of each sample $x_i \subseteq L$ were obtained in each sub-classifier posterior probability $P_i(+1|x)$, $P_i(-1|x)$ matrix according to the set in consideration of the use of meta-learning, misclassification cost of training samples

$$R_i(+1|x) = \sum_i P_i(-1|x) \cdot \text{cost}(-1,+1)$$

$$R_i(-1|x) = \sum_i P_i(+1|x) \cdot \text{cost}(+1,-1)$$

take minimum misclassification the $\min(R)$, according to the conditions to determine true class label of the sample L', thus making the sample integrated consideration misclassification, above the sample concentration of each sample to regain class label and then reconstructed sample set, because the reconstruction of the samples the Chichi become misclassification cost, so you can take advantage of the cost-sensitive support vector machine, to get a decision function with misclassification, making the classification of the overall misclassification minimum.

SIMULATIONS

Experiment data: Algorithm used in the experimental data set is used in the study of the unbalance data classification 2 discloses data sets are from <http://www.ics.uci.edu/mllearn/MLRepository.htm> standard data set obtained, respectively the hepatitis, soybean data set, these two data sets hepatitis for two types of problems, soybean original multi-class problem, consider the convenience of calculation, the experiments in this soybean data set is first converted to the two types of questions to the class label positive class and all kinds of combined negative class. The number of various samples of each data set, such as shown in Table 1.

Assuming the Consideration known cost matrix as shown in Table 2.

Experimental results and analysis: Data set randomly split into a training set and a test set, cross-validation method, all samples were randomly divided into 8 parts and each sample imbalance rate remains equal

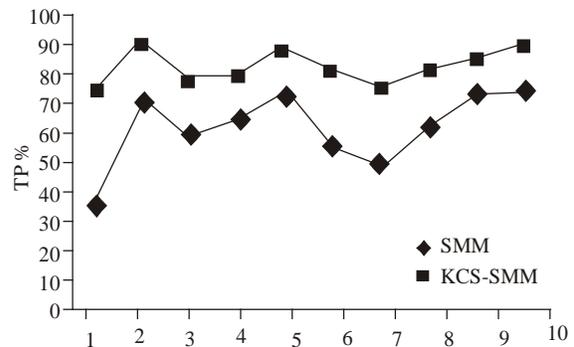


Fig. 1: Data set hepatitis is the correct class sample classification accuracy

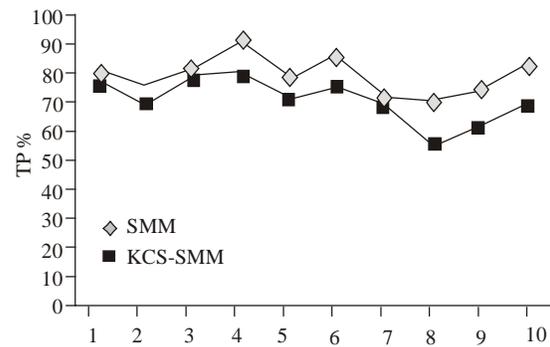


Fig. 2: Data sets hepatitis negative class samples correctly classified accuracy

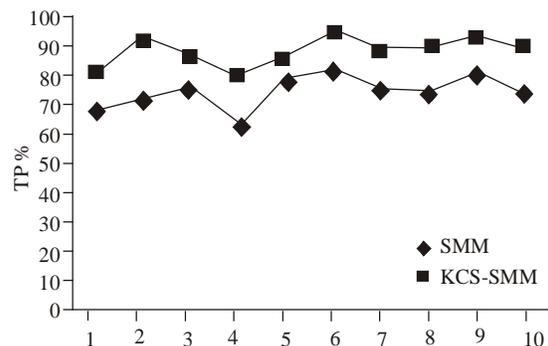


Fig. 3: Data set the soybean is class sample correct classification accuracy

to the rate of overall imbalance. Then every seven randomly selected as the training set, the rest as a test set.

Use standard SVM, KCS-SVM; two methods were compared as shown in Table 3 and 4.

Table 3: Data sets of hepatitis two types of sample classification accuracy (%)

The number of experiments	1		2		3		4		5	
Index/Classifier	TP	TN								
SVM	36	80	71	75	60	81	65	90	73	78
KCS-SVM	75	75	91	69	78	78	80	79	89	71
The number of experiments	6		7		8		9		10	
Index/Classifier	TP	TN								
SVM	55	85	44	71	61	69	73	73	72	82
KCS-SVM	81	75	75	68	90	55	85	61	79	68

Table 4: Soybean sample two types of correct classification accuracy (%)

The number of experiments	1		2		3		4		5	
Index/Classifier	TP	TN								
SVM	67	95	71	80	75	87	62	79	78	81
KCS-SVM	80	85	92	73	85	80	79	67	85	69
The number of experiments	6		7		8		9		10	
Index/Classifier	TP	TN								
SVM	81	67	75	85	73	91	80	82	73	73
KCS-SVM	93	63	87	79	89	83	92	92	89	84

Table 5: Two data sets misclassification

Data sets/Classifier	Hepatitis	Soybean
SVM	5.1	4.8
KCS-SVM	3.7	3.1

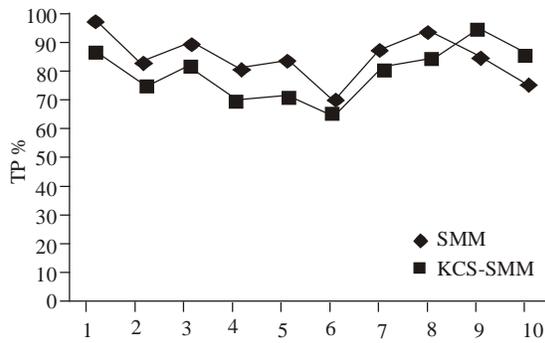


Fig. 4: Data sets negative class soybean sample correct classification accuracy

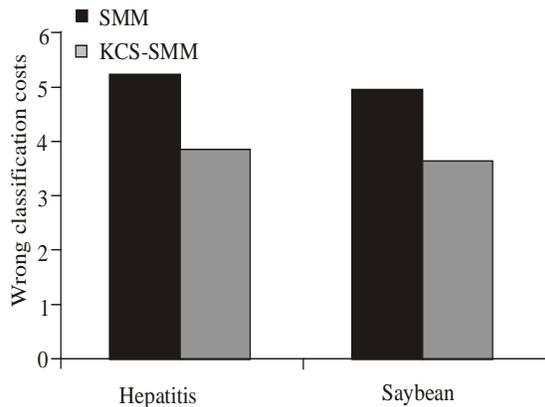


Fig. 5: SVM KCS-SVM misclassification costs in different data sets to compare

Since cost-sensitive support vector machine and known cost matrix evaluation criteria for cost-sensitive learning, the cost-sensitive evaluation indicators are: total classification error cost (*Total Costs*), the average

misclassification cost (*Average Costs*) as shown in Fig. 1 to 4.

The two classification confusion matrix and cost matrix, *Total Costs* and *Average Costs*, respectively, as follows: $Total\ Costs = FP \times cost(N, P) + FN \times cost(P, N)$
Average

$$Costs = \frac{FP \times cost(N, P) + FN \times cost(P, N)}{TP + TN + FP + FN}$$

Misclassification costs calculated for each data set, taking 10 times to calculate the average of the results, the results as shown in Table 5.

As shown in Fig. 5, the experimental results from the two data sets and SVM, KCS-SVM in the positive class samples obtained classification accuracy but also the positive class samples misclassification, while the negative class samples lower on classification accuracy and lower consideration misclassification. The positive class sample classification accuracy greater than average misclassification role negative samples classification accuracy, thus reducing the average misclassification cost

According to the experiment can be seen KCS-SVM method effectively reduce the sample misclassification cost.

CONCLUSION

This study based on Bayesian decision theory and risk minimization revelation by the decomposition of the training set a new consideration based on the data set decomposition sensitive support vector machine. Firstly, the sample set according to a certain data set of decomposition rules be broken down into several subsets, training support vector machine that can output posterior probability for each subset, according to Platt's proposed use sigmoid function mapping posteriori output probability of the sample and the minimum

classification consideration, learning theory and the costs to get the samples to get a sample of the "true" class label, after the training is completed for each training sample, the reconstruction of a data set, a new set of data is the integration of the misclassification data set, then training the new sample set using cost-sensitive support vector machine, a new classifier, the classifier can make minimum misclassification. Simulation results show that the new cost-sensitive support vector function effectively reduce misclassification cost, to achieve good results.

REFERENCES

- Akbai, R., S. Kweks and N. Japkowicz, 2004. Applying support vector machines to imbalanced datasets. Proceeding of the 5th European Conference on Machines Learning, Springer-Verlag, Berlin, 2004: 39-50.
- Barandela, R., J.S. Sanchez, V. Garcia and E. Rangel, 2003. Strategies for learning in class imbalance problems. *Pattern Recogn.*, 36(3): 849-851.
- Barandela, R., R.M. Valdovinos, J.S. Sanchez and F.J. Ferri, 2004. The Imbalanced Training Sample Problem: Under or Over Sampling. In: Fred, A., T.M. Caelli, R.P.W. Duin, A.C. Campilho and D. de Ridder (Eds.), *SSPR and SPR. LNCS*, Springer-Verlag, Heidelberg, Berlin, 3138: 806-814.
- Chawla, N.V., K.W. Bowyer, L.O. Hall and W.P. Kegelmeyer, 2002. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.*, 16(3): 321-357.
- Han, H., W.Y. Wang and B.H. Mao, 2005. Borderline-smote: A new over-sampling method in imbalanced data sets learning. Proceeding of the International Conference on Intelligent Computing (ICIC05), *Lecture Notes in Computer Science*. 3644: 878-887.
- Japkowicz, N. and S. Stephen, 2002. The class imbalanced problem: A systematic study. *Intelligent Data Analysis. Intell. Data Anal.*, 6(5): 429-449.
- Kubat, M. and S. Matwin, 1997. Addressing the curse of imbalanced datasets. One-sided Sampling Proceedings of the 14th International Conference on Machine Learning, Nashville, Tennessee, pp: 178-186.
- Maloof, M.A., 2003. Learning when data sets are imbalanced and when costs are unequal and unknown. Proceeding of the ICML-2003 Workshop on Learning from Imbalanced Data Sets II, AAA I Press, Washington DC, 2003: 531-542.
- Rehan, A., K. Stephen and J. Nathalie, 2004. Applying support vector machines to imbalanced datasets. Proceeding of the 15th European Conference on Machines Learning, Springer-Verlag, Berlin, pp: 39-50.
- Tang, S. and S.P. Chen, 2008. The generation mechanism of synthetic minority class examples. Proceeding of the International Conference on Information Technology and Applications in Biomedicine, pp: 444-447.
- Wang, J.J., M.T. Xu, H. Wang and J.W. Zhang, 2007. Classification of imbalanced data by using the SMOTE algorithm and locally linear embedding. Proceeding of the 8th International Conference on Signal Processing.
- Weiss, G.M., 2004. Mining with rarity: A unifying framework. *ACM SIGKDD Explorations*, 6(1): 7-19.
- Yin, Y., X.D. Zhu and L. Xu, 2011. A media sensitive cache replacement algorithm. *IEIT J. Adapt. Dyn. Comput.*, 2011(1): 7-11.
- Zhao, Z.L., B. Liu and W. Li, 2012. Image classification based on extreme learning machine. *IEIT J. Adapt. Dyn. Comput.*, 2012(1): 5-11.
- Zhou, Z.H. and X.Y. Liu, 2006. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE T. Knowl. Data Eng.*, 18(1): 63-77.