Vehicle logo classification using support vector machine Ensemble

Wesal Abdallah and Li Hong

School of information science and technology, Central South University
Changsha, Hunan, 410083, P.R.China.

Abstract
The vehicle logo is a unique mark of vehicle make. It is classification is one of the vehicle identification methods. Vehicle Logos classification is widely used for security in many places such as government building, campus, roads. Most of the existing supervised classification methods are based on Support Vector Machines (SVM), which can yield ideal results. Although SVM can provide good generalization performance, but the classification results of the SVM in particular problem is often far from the theoretically expected level because its implementation in real problem is based on an approximated algorithm. To improve the limited classification performance of the SVM in vehicle logo classification, we propose to use SVM ensemble with bagging. In bagging each individual SVM is trained independently using randomly chosen training samples and then the results of the SVMs are aggregated to make a collective decision. The simulation results show that SVM ensemble outperforms the stand alone SVM method. We used two-dimensional principal component analysis (2DPCA) for logo's image feature extraction.

Keywords: *Vehicle logo, Classification, Support vector machines ensemble, bagging.*

1. Introduction

In the past decades, the issue of security and scientific management of vehicle have become more significant and the need for effective security and management systems has intensified [1]. Many areas were marked as restricted, since illegal access can have serious consequences for homeland security and can even result in the loss of lives in the case of an explosive armed vehicle [2]. When it comes to vehicle management, most of the current existent methods classify vehicle in broad categories, such as cars, buses, heavy goods vehicle etc. To increase security in access control applications for a vehicle that enters a restricted area or to have a good vehicles management system, vehicles' make need to be identified. Identifying vehicles' make can be accomplished through classification methods.

There are many methods that can be used for classification, such as logistic regression, Support Vector Machine (SVM), kernel logistic regression. The SVM is a new machine learning and classification method that developed rapidly and it has been widely used in many kinds of classification and pattern recognition problems. The basic idea of SVM is to transform the samples into a high-dimension space and to seek a separating hyperplane that separates the classes in this space. The separating hyperplane is chosen in such a way as to maximize its distance from the closest training samples. As a supervised machine learning technology, SVM is well-founded theoretically on statistical learning theory and it has been applied successfully in many fields of classification and pattern recognition. It usually outperforms other machine learning technologies, including Neural Networks (NN) and K-Nearest Neighbor (KNN) methods.

However, since SVM is originally built to accomplish binary classification tasks, a combination of SVMs should be used to accomplish the multi-class classification. There are several methods that can be used to enable the usability of SVM in multi-class classification [3, 4], but when it was applied to SVM the performance was not as good as binary classification. Also training SVM is time consuming for large scale problem. Some approximation algorithms were proposed to make SVM training more efficient, but using this approximation algorithm can degrade the SVM classification performance.

In this paper, we propose to use SVM ensemble for vehicle logo classification, to overcome the problems of SVM. In our proposed approach we used the two-dimensional principal component analysis (2DPCA), which is based on 2D image matrices for logo's image feature extraction. Bagging method is used to implement SVM ensemble. Bagging is a method for generating multiple classifiers using sampling with replacement. There are many strategies for aggregating these classifiers. Among these strategies is the average of the estimated probabilities, which is used in this paper. A variety of performance metrics have been used: accuracy, predicted

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 2, March 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

positive value, sensitivity, and the MCC to assess the performances of SVM and SVM ensemble.

2. Methods

2.1 Support vector machine (SVM)

The SVM is a new and promising classification and regression technique proposed by Vapnik and his group at AT&T Bell Laboratories [5]. It has emerged as a good classification technique and has achieved excellent generalization performance in a wide variety of applications [6], such as text recognition, web pages categorization, face detection and bioinformatics [7-11]. SVM initially dealt with two-class problems. There are several books for explaining the SVM. In brief given a training set $\{(x_i, y_i)\}, i=1,2,......N$, N is a total number of samples $y_i = \{1,-1\}, x_i \in R^P \subset R, i.e.$ x_i is a p dimension real vector. for the non-linear classification, SVM solves the following primal problem

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C\sum_{i=1}^{N} \xi_i \qquad (1)$$

$$\text{Subject to } \begin{cases} y_i(w^T \phi(x_i)+b) \geq 1-\xi i, i=1,2,3,....N \\ \xi i \geq 0, \ i=1,2,.....N \end{cases}$$

where ξi are slack variables, measuring the degree of misclassification of the sample x_i , C is the error penalty, penalizing the non-zero ξi . The bias b is a scalar, representing the bias of the hyperplane. w is the weight vector, defining a direction perpendicular to the hyperplane. $\phi(.)$ Is a nonlinear function (kernel function) which maps the input space into a higher dimensional space. Here the training vectors x_i are mapped into a higher (maybe infinite) dimensional space by the function $\phi(.)$. SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. The classification decision function in a general form in the linearly non-separable problem can be represented by

$$f(x) = sign(w^T \phi(x) + b) \qquad (3)$$

There are three typical kernel functions that can be used. These functions are as follows:
Polynomial: $K(x_i, x_j) = (x_i x_j + 1)^d$
Radial basis kernel function:
$K(x_i, x_j) = \exp(-g\|x_i - x_j\|^2)$
Sigmoid: $K(x_i, x_j) = \tanh(b(x_i x_j) + r)$

Where d,g, and r are the kernel parameters.

2.2 Support vector machine ensemble

An ensemble of classifiers is a collection of several classifiers whose individual decisions are combined in some way to classify the test examples [12, 13]. Hansen et al. [14] shows how the ensemble can have better performance than individual classifiers. However, The SVM has been known to show a good generalization performance and is easy to learn exact parameters for the global optimum [15]. Because of these advantages, their ensemble may not be considered as a method for improving the classification performance greatly. However, since in practical SVM has been implemented using the approximated algorithms in order to reduce the computation complexity of time and space, a single SVM may not learn exact parameters for the global optimum. Sometimes, the support vectors obtained from the learning is not sufficient to classify all unknown test examples completely. So, we cannot guarantee that a single SVM always provides the global optimal classification performance over all test examples.

To overcome this limitation, we propose to use an ensemble of support vector machines. There are different techniques that can be used to implement SVM ensemble. In this work, we suggest to use the SVM ensemble based on bagging technique, which is one of the most important recent developments in prediction and classification methodology. It was proposed by Leo Breiman [16]. Using bagging in many classification algorithms results in high improvement in performance and gives substantial gains in accuracy. Bagging works by sequentially applying a selected classification algorithm in respect to modifications of the training data set. This modification is accomplished by means of drawing random samples with replacement from the original training set. For each sample, a classifier should be created.

In our vehicle logo classification problem we will have different SVMs; each trained on the random drawn set and produces its own decision for the test set. The decisions of these different SVMs can be aggregated using different techniques such as majority voting, double-layer hierarchical combining, and the average of estimated probabilities strategy. The last technique can be used with the classifiers that produce probability estimates. In this paper we used bagging for the SVM ensemble and the average of estimated probabilities' strategy to aggregate

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 2, March 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

the decision of the different SVMs. The overall architecture of our SVM ensemble is shown in fig 1.

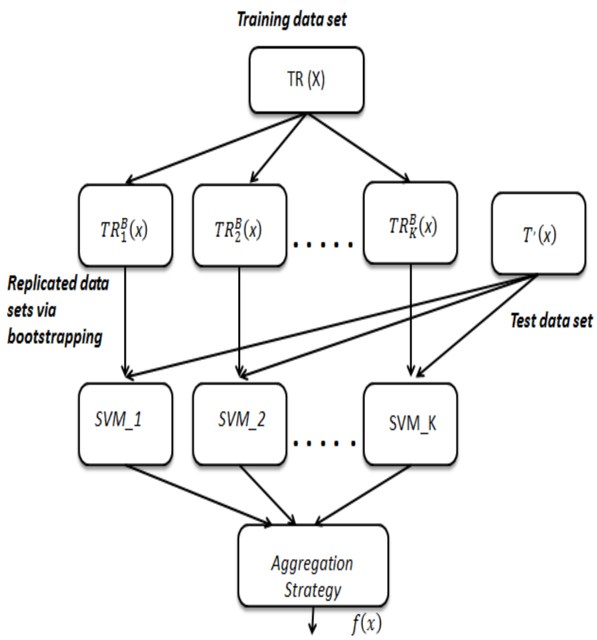

Fig. 1: A general architecture of the SVM ensemble.

Bagging procedure

In this work we used the following bagging procedure.

1. Initialization of the training data set T.
2. Draw a random samples (bootstraps) with replacement from the training set (some of the examples can be selected repeatedly and some may not be selected at all).
3. Train a SVM classifier using this sub training data set using radial bases function.
4. By repeating the previous steps *k* times, *k* classifiers will be obtained.
5. Aggregate the decision of the *k* classifiers using the average of estimated probabilities' strategy.

Any training instance in the training data set T has the probability $[1-(1-1/k)^k]$ of being selected, at least once in the *k* times randomly selected instances from the training data set. In this paper we set k=100. The probability above will approximately be equal to 0.63 which means that each sub training sample contains about 63% unique instances from the original training data set and this way the built classifiers' samples will not be identical.

2.3 The data set

The dataset contain vehicle logo with four classes .Each class refers to a vehicle logo name. The first class is Volkswagen logos, the second one is Hyundai logos, the third one is Nissan logos, and the last one is Toyota logos.

Fig 2: Example of vehicle logos image from dataset

2.4 Features selection

We used the Tow- Dimensional Principal Component (2DPCA) method for features extraction. 2DPCA is developed for logo's image feature extraction. As opposed to conventional PCA, 2DPCA is based on 2D matrices rather than 1D vector. That is, the image matrix does not need to be previously transformed into a vector. Instead, an image covariance matrix can be constructed directly using the original image matrices. In contrast to the covariance matrix of PCA, the size of the image covariance matrix using 2DPCA is much smaller. As a result, 2DPCA has two important advantages over PCA. First, it is easier to evaluate the covariance matrix accurately. Second, less time is required to determine the corresponding eigenvectors [17]. After a transformation by 2DPCA, a feature matrix is obtained for each vehicle logo image.

2.5 Performance measures

The question of method performance assessment is the central to constructing, deploying, and using classification method. The metric that we used in this work is the common metrics that are well known and have been widely used in machine learning. These metric are known as threshold metric. These metrics are: accuracy (ACC), the number of correct predictions on the test data is divided by the number of test data instances, sensitivity

(SN) or the fraction of the total positive examples that are correctly predicted, and the Predicted Positive Value (PPV), also called the precision. Beside the aforementioned measures we used Matthew's correlation coefficient (MCC), which is a robust and reliable performance measure that account for both over and under prediction. The result of the MCC is in the range of -1 and 1, where a value of 1 indicates a perfect positive correlation, a value of -1 indicates a perfect negative correlation, and a value of 0 indicates no correlation.

If we assume that our interest is to classify Nissan logo form others, then TP (true positives) will be the number of correctly classified Nissan, TN (true negatives) will be the number of correctly classified non-Nissan, FP (false positives) will be the number of non- Nissan incorrectly classified as Nissan and FN (false negatives) will be the number of Nissan incorrectly classified as non-Nissan. Using these quantities, the above mentioned metrics can be calculated as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

$$SN = \frac{TP}{TP + FN} \times 100$$

$$PPV = \frac{TP}{TP + FP} \times 100$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}}$$

2.5 Results and discussion

In this study we used LBSVM, which is a famous SVMs tools. All of the experiments are curried out using MATLAB version 2009 a. The LIBSVM parameters C and γ are optimized using the grid search. The optimal values of C and γ for the vehicle types logos are found to be $C = 2.0$ for Volkswagen and Nissan vehicle types, $C = 4.0$ for Hyundai and Toyota vehicle types. $\gamma = 0.03125$ for the all vehicle types. Then we implement SVMs on the datasets that contain each vehicle type logos using 5 cross validation approach. In 5 cross validation the dataset is divided into 5 samples. We adopted the leave-one-out cross-validation test, which is the most objective and rigorous cross-validation method compared with independent dataset test and sub-dataset test. In a full leave-one-out cross-validation test of 5 samples, one sample is removed from the set, the training is done on the remaining 4 samples and the test is done on the removed sample. This process is repeated 5 times by removing one sample in turn. The final prediction results are taken as the average of the results from the 5 testing samples. These results are shown in Table1.

Table 1: SVM results for vehicle logo classification

Vehicle logos	*ACC*	*PPV*	*SN*	*MCC*
Volkswagen	82.9	96.5	42.6	0.58
Hyundai	87.2	84.8	72.3	0.69
Nissan	87.8	1	18.5	0.68
Toyota	80.4.	85.3	38.9	0.49

We also implemented SVM ensemble using bagging on the vehicles logos. The experiment is constructed according the previous mentioned bagging procedure. A random sample was drawn with replacement from the original data set to form a training set. Each training set contains approximately 80% of the data from the original data set. Since LIBSVM support estimated probabilities, the average of estimated probabilities strategy has been used. Each training set generates estimated probabilities $\hat{p}\left(j/x\right)$, which is an object with prediction vector *x* belonging to class *j*. Then the class corresponding to *x* is estimated as argmaxj $\hat{p}(j/x)$. The results of SVM ensemble are obtained by averaging the $\hat{p}(j/x)$, and then this average is used as the final prediction. The results of the SVM ensemble are shown in Table 2.

Table 2: Bagging results for vehicle logo classification

Vehicle logos	*ACC*	*PPV*	*SN*	*MCC*
Volkswagen	83.2	98.8	44.3	0.63
Hyundai	90.6	90.6	81.2	0.79
Nissan	88.9	1	23.3	0.69
Toyota	81.2	88.5	41.7	0.56

From the results we can see that SVMs ensemble achieved results that are significantly better that stand alone SVM. These show that although SVMs is strong classifiers but their results can be improved using bagging for vehicle logo classification.

2.6 Conclusions

In this paper we presented SVM and SVM ensemble, which is a collection of several classifiers whose individual decisions are combined using the average of estimated probabilities strategy, for vehicle logo classification. We used 2DPCA, which is based on 2D matrices rather than 1D vector for logo’s image feature extraction. In 2DPCA the image matrix does not need to be previously transformed into a vector. Instead, an image covariance matrix can be constructed directly using the original image matrices. We first applied standalone SVM classifiers and then the results are compared with SVM ensemble. The results show that SVMs ensemble achieved results that are significantly better than the stand alone SVM. This indicates that, although SVMs are strong classifiers but in vehicle logo classification, their results can be improved significantly using ensemble with bagging.

References

[1] D. Robson, "Safe and secure in transit", Imaging and Machine Vision Europe, Dec. 2007 – Jan. 2008, Issue 24, p10.
[2] C. N. Anagnostopoulos, I. Giannoukos, T. Alexandropoulos, A. Psyllos, V. Loumos, and E. Kayafas, "Integrated vehicle recognition and inspection system to improve security in restricted access areas," in Intelligent Transportation Systems (ITSC)", 2010 13th International IEEE Conference on, 2010, pp. 1893-898.
[3] L. Bottou, C. Cortes , J. Denker, H. Drucker, I. Guyon, D. Lawrence, Y. LeCun, U. M¨uller, E. S¨ackinger, P. Simard, V. Vapnik: Comparison of classifier methods: "a case study in handwriting digit recognition" , Proceedings of the 13th International Conference on Pattern Recognition. IEEE computer Society Press, 1994, 77–87 397, 399.
[4] S. Knerrm, L. Personnaz, and G. Dreyfus.: "Single-layer learning revisited: a stepwise procedure for building and training a neural network". In: Fogelman, J. (eds.): Neurocomputing: Algorithms, Architechtures and Application, Springer-Verlag, 1990, 397, 400.
[5] C. Cortes, V. Vapnik, "Support vector network", Mach. Learn. 20, 1995, 273–297.
[6] L. Xi-Ying "A Study on Clustering Combination and Support Vector Machine Based on EELD Forecasting", IJCSI International Journal of Computer Science Issues, 2013, Vol. 10, Issue 1, No 3. 31-36.
[7] A. Bellili, M. Gilloux and P.Gallinari, "An MLP-SVM combination architecture for offline handwritten digit recognition: Reduction of recognition errors by Support Vector Machines rejection mechanisms", International Journal on Document Analysis and Recognition, Volume: 5, Issue: 4, 2003, p. 244 – 252.
[8] Q. Duan, D. Miao and K. Jin, "A Rough Set Approach to Classifying Web Page without Negative Examples", Advances in Knowledge Discovery and Data Mining, LNCS, Volume 4426, 2007, p.481-488.
[9] P. Shih, and C. Liu, "Face detection using discriminating feature analysis and Support Vector Machine ", Pattern Recogn., Vol 39, Issue 2, 2006, 260–276.
[10] M. E. Matheny, F. S. Resnic, N. Arora, and L.OhnoMachado , "Effects of SVM parameter optimization on discrimination and calibration for post-procedural PCI mortality", J. of Biomedical Informatics, Vol.40, No.6, 2007, 688-697.
[11] M .E .Mavroforakis and S .Theodoridis, "Ensemble methods for classification of patients for personalized medicine with high-dimensional data", Artif. Intell. Med., Vol.41, No.3, 2007, 197-207.
[12] T. Dietterich, "Machine learning research: four current directions", ArtiK. Intell. Mag, 18 (4) ,1998, 97–136
[13] T. Dietterich," Ensemble methods in machine learning", Multiple Classifier Systems, Lecture Notes in Computer Science, Calgari, Italy, Vol. 1857, 2000.
[14] L. Hansen, P. Salamon," Neural network ensembles", IEEE Trans, Pattern Anal. Mach. Intell. 12, 1990, 993–1001.
[15] C.J.C. Burges, "A tutorial on support vector machines for pattern recognition", Data Mining Knowledge Discovery, 2 (2), 1998, 121–167.
[16] L. Breiman, "Bagging predictors", Mach. Learn, 24 (2), 1996, 123–140.
[17] J. Yang, D. Zhang, A.F. Frangi, and J.Y. Yang. "Two-dimensional PCA: a new approach to appearance-based face representation and recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(1), 2004,131-137.