

An Evaluation of Divide-and-Combine Strategies for Image Categorization by Multi-Class Support Vector Machines

Can Demirkesen, Hocine Cherifi

▶ **To cite this version:**

Can Demirkesen, Hocine Cherifi. An Evaluation of Divide-and-Combine Strategies for Image Categorization by Multi-Class Support Vector Machines. 23rd International Symposium on Computer and Information Sciences, 2008. ISCIS '08, Oct 2008, Istanbul, Turkey. pp.1 - 6, 10.1109/ISCIS.2008.4717904 . hal-00612239

HAL Id: hal-00612239

https://hal.archives-ouvertes.fr/hal-00612239

Submitted on 28 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An Evaluation of Divide-and-Combine Strategies for Image Categorization by Multi-Class Support Vector Machines

C. Demirkesen[1] and H. Cherifi[1,2]

1: Institue of Science and Engineering
Galatasaray University
Ortakoy, Istanbul, 34257, Turkey
Phone: (90) 212 227 44 80 E-mail: candemirkesen@gmail.com

2: Faculté des Sciences Mirande
Université de Bourgogne
BP 47870 21078 DIJON FRANCE
E-mail : hocine.cherifi@u-bourgogne.fr

Abstract

Categorization of real world images without human intervention is a challenging ongoing research. The nature of this problem requires usage of multiclass classification techniques. In divide-and-combine approach, a multiclass problem is divided into a set of binary classification problems and then the binary classifications are combined to obtain multi-class classification. Our objective in this work is to compare several divide-and-combine multiclass SVM classification strategies for real world image classification. Our results show that One-against-all and One-against-one MaxWins are the most efficient methods.

Key Words- image categorization, multiclass classification, support vector machines,

1. Introduction

Increasing need for image organization and retrieval triggers new research of multiclass image classification methods. From model-based methods to learning algorithms, there are many choices for an appropriate classifier. Among these, support vector machines (SVMs) appear to be a good candidate because of their ability to generalize in high-dimensional spaces without the need to add a prior knowledge. However, achieving high image classification accuracy is quite challenging.

SVM is originally designed for binary classification [1]. The idea is to separate two classes by calculating the maximum margin hyper plane between the training examples. Several methods have been proposed to extend SVM in order to classify more than two classes because classification problems are mostly multi class. Image classification is naturally a multi class problem as well. Basically there are two major approaches for extending SVM to multiclass classification: (1) considering all data in a single optimization. (2) Combining several binary SVM classifiers; generally the first approach is called 'all-in-one' (AIO) [2], and the second 'divide-and-combine'. Due to several complexity issues the all-in-one method is generally avoided in favor of divide-and-combine approach. The main methods for divide-and-combine are One-Against-All (OAA) also known as 'winner-takes-all' [3], One-Against-One (OAO) implemented by four different strategies: (1) MaxWins majority voting [4], (2) pairwise coupling [5], (3) error correcting codes [6], (4) directed acyclic graph (DAG) [7], Half-Against-Half (HAH) [8]. There is some work in the literature comparing these methods for classical datasets like iris, wine, glass, letter etc. In [9], OAA, OAO MaxWins and DAG are compared. The authors show that there is not one method that performs best for every dataset but that OAO MaxWins and DAG perform better with large number of classes. In [7], OAO MaxWins, OAA, DAG and Neural Networks are compared. The authors show that the methods have comparable performance on accuracy and error rate but that OAO and DAG need less time for training phases. HAH, OAO MaxWins, OAA and DAG are compared in [8]. It is shown that HAH is competitive with the others in accuracy. This conclusions point out that one should compare these methods for a specific classification problem, in this case image classification, because the best method can depend on the problem at hand. In [10], OAO MaxWins, OAO-Pairwise Coupling, DAG and Neural Networks are compared for natural texture images like grass, leaves, brick etc. using a mixed color and texture representation. The authors conclude that OAO MaxWins and DAG have almost the same performance and they are both better then neural networks in terms of accuracy. In [11], OAO MaxWins, OAA, DAG, maximum likelihood and back propagation neural networks are compared for satellite images like water, construction, wood, bare soil etc. using topographical raster data for image representation. According to their results OAO with MaxWins majority voting is the most performing in terms of accuracy. Despite all these studies there is not a fully complete comparison of multi-class SVM classification methods. The comparisons in [7] and [11] do not cover OAO pairwise coupling. In [10], OAA and ECOC are missing. Even in the most complete comparison [2], OAO-Pairwise Coupling method is missing. In addition to our knowledge there is not one comparison of these methods for classification of real world natural scenes like forest,

978-1-4244-2881-6/08/$25.00 ©2008 IEEE

Authorized licensed use limited to: ULAKBIM UASL - Galatasaray Universitesi. Downloaded on February 26, 2009 at 04:44 from IEEE Xplore. Restrictions apply.

coast, mountain or city view categories and previous results show that performances are greatly influenced by nature of the data.

In this paper, we evaluate and compare all of the divide-and-combine multiclass SVM methods mentioned above on a challenging image database by following an experimental approach. We compare performance of the methods for natural image categorization task using global and local image representations. The rest of the paper is organized as follows: Section 2 introduces multiclass methods. In section 3 features used for image representation are briefly presented. Experimental results are given in section 5 and finally conclusion in section 6.

2. Binary Classification by SVM

Support vector machines have strong theoretical foundations and excellent empirical successes. We consider SVMs in the binary classification setting. In their simplest form, SVMs are hyper planes that separate the training data by a maximal margin.

We are given the training data {x1 ... xn} in some space X and their labels {y1 … yn} where yi in {-1, 1}.The objective is to find a hyperplane which divides the set of instances such that all the points with the same label are on the same side of the hyperplane [8]. This amounts to finding *w* and *b* so that

$$y_i\left(w\cdot x_i+b\right)>0,\quad i=1,\ldots,N \tag{1}$$

If a hyperplane satisfying (1) exists, it is possible to write

$$\min_{1\le i\le N} y_i\left(w\cdot x_i+b\right)\ge 1,\quad i=1,\ldots,N \tag{2}$$

Once *w* and *b* are calculated the decision function is:

$$f(x)=sign(w\cdot x+b) \tag{3}$$

3. K-Classification by Divide-and-Combine

Strategies described below can be applied to build N-class classifiers using binary SVM classifiers. They can be decomposed in two main steps: (1) classification, (2) fusion. In the classification step an instance x is classified by all of the binary classifiers. In the fusion step, classification outputs are combined together to provide a decision.

3.1 One-Against-One SVM classifiers

One-Against-One (OAO) method involves N(N-1)/2 binary SVM classifiers. Each classifier is trained to separate each pair of classes. There are different strategies used to combine these binary classifiers.

3.1.1 MaxWins Majority Voting

Each of the *N*(*N*-1)/2 SVMs casts one vote for its favored class, and finally the class with most votes wins. This strategy must address the case of equality between the votes of the classes. This weakness can be rectified by adding a very simple rule to the original algorithm. For example, [9] propose to decide the class with smaller index if two classes are equally voted.

3.1.2 Pairwise Coupling

In pairwise coupling [5], a pairwise probability p_{ij} is obtained from each binary SVM output noted as $f_{ij}(x)$.

$$p_{ij}=\frac{1}{2}f_{ij}(x)+0.5 \tag{1}$$

These pairwise probabilities are coupled into a common set of posterior probabilities p_i

$$p_i=\frac{2}{N(N-1)}\sum_{j\neq i}\sigma(p_{i,j}) \tag{2}$$

Where σ (.) is one of the functions presented in Table 1.

σ	$\sigma(p_{i,j})$
PWC1	$\sigma(p_{ij})=p_{ij}$
PWC2	$\sigma(p_{ij})=\begin{cases}1, if p_{ij}\ge 0.5\\ x, otherwise\end{cases}$
PWC3	$\sigma(p_{ij})=\begin{cases}1, p_{ij}\ge 0.5\\ 0, otherwise\end{cases}$
PWC4	$\sigma(p_{ij})=\frac{1}{1+e^{-1/2\,p_{ij}-0.5}}$

Table 1. Examples of pairwise coupling functions

The decision function is given by:

Authorized licensed use limited to: ULAKBIM UASL - Galatasaray Universitesi. Downloaded on February 26, 2009 at 04:44 from IEEE Xplore. Restrictions apply.

$$c(x) = \arg\max_{1 \le i \le N}(p_i) \quad (3)$$

3.1.3 DAG SVM classifiers

Directed Acyclic Graph (DAG) SVM is proposed by Platt et al [7]. Training is the same as the OAO using N(N-1)/2 binary SVMs. However, in the testing phase, a directed acyclic graph with N(N-1)/2 internal nodes and N leaves is used. Testing a sample starts at the root node and it moves to either left or right depending on the output value. Therefore, we go through a path before reaching a leaf node, which indicates the predicted class. An example DAG for three classes is given in Figure 1.

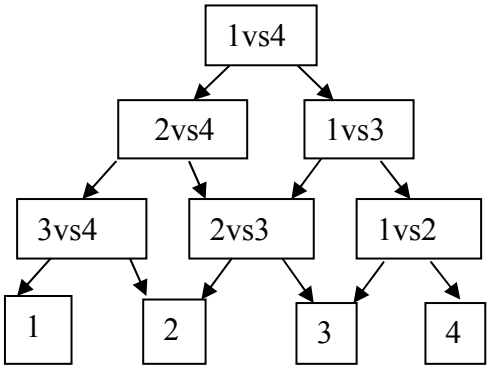

Fig. 1. Decision DAG for four classes. Each node is a binary classifier

3.1.4 Error Correction Output Coding

Error-correcting output coding (ECOC) is another one-against-one combination strategy [6]. A unique binary string of length L called codeword is assigned to each class. L binary classifiers are built by learning L function. A set of 6-bit codewords assigned to four classes is shown in Table 2, where the columns $f_1, \ldots, f_6$ represent the classifiers to be trained. All columns are distinct and each classifier learned is a disjunction of the classes. For example, $f_2(x) = 1$ if x is in class 1 or 4.

Class	f_1	f_2	f_3	f_4	f_5	f_6
1	1	1	1	1	1	1
2	0	0	0	1	1	1
3	0	0	1	1	0	0
4	0	1	0	1	0	1

Table 2. Error-correcting output codes for four-class problem with minimum Hamming distance of two.

To classify a test example x, the L SVM outputs are combined to generate a codeword $C = (c_1, \ldots, c_L)$ as follows:

$$c_i = \begin{cases} 1 \ if & f_i \ge 0 \\ 0 & otherwise \end{cases}$$

The generated codeword C is compared to the pre-assigned codewords for each class; the closest class in terms of Hamming distance between the codewords gives the classification of x.

3.2 One-Against-All SVM classifiers

One-Against-All (OAA) is the most common approach [3]. It involves N binary SVM classifiers, one for each class. Each binary SVM is trained to separate one class from the rest. The winning class is the one that corresponds to the SVM with highest output value i.e. the largest decision function value. This approach may suffer from error caused by imbalanced training sets.

3.3 Half-Against-Half multi-class SVM

In Half-Against-Half (HAH), similar classes are grouped together as a bigger category. Two categories are obtained from the dataset and then matched against each other. The structure of the classifier is the same as a decision tree with each node as binary SVM. The classification procedure goes from root to the leaf guided by the binary SVMs. Given that similar or close classes are grouped together for training phase the biggest challenge is to determine the optimum divisions [8]. A sample decision tree for classifying four classes is given in Figure 2.

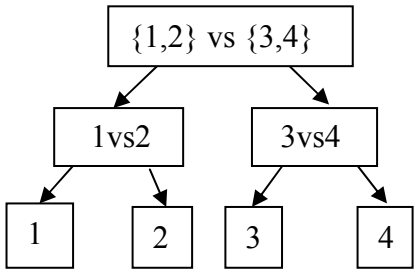

Fig. 2. Decision tree for classifying four classes. Each node is a binary classifier

4. Image Representation

Image representation can be generally thought of in two ways: local representation and global representation. In local representation, an image is either segmented in roughly homogeneous regions in color or texture, or it is divided into regular sub-blocks for further processing. Either way characteristics are extracted from regions of the image. In global representation, a characteristic is

Authorized licensed use limited to: ULAKBIM UASL - Galatasaray Universitesi. Downloaded on February 26, 2009 at 04:44 from IEEE Xplore. Restrictions apply.

extracted from the whole image without segmenting or dividing it into sub-blocks.

These two approaches have both advantages and disadvantages. For example when global image representation is used, an individual object in the image may have due care or attention. Small details can be neglected. When local representation is used global semantic of the image is not taken into account. Spatial relationship between regions or objects is not considered.

When the role of local and global representation has been studied by experiments on humans, it is noted that local and global information are separately processed by human visual system then integrated together. In a previous work we evaluated a collection of local and global features used to represent color, texture, edge and spectral information for binary classification [13]. We have shown that texture leads to the highest classification accuracy as a local representation, while gist is the most performing global representation. We used co-occurrence matrices to characterize texture information. Gist is a low dimensional representation of the scene structure based on the output of filters tuned to different orientations and scales. It is calculated using power spectrum of the image and a series of Gabor filters [14]. To sum up the previous work, holistic spatial scene properties may be best estimated by gist and local information is best described by texture information. The combination of texture and gist improves classification performance in binary classification. The same representations can be adopted for multi class image classification.

5 Experimental Results

5.1 Image Database

Our image database contains 8 categories of natural scenes: highway(260), streets(292), forest(328), open country(410), inside of cities(308), tall buildings(356), coast(360) and mountain(374) images (Numbers in brackets represent the size of each categories). The database provided by Oliva and Torralba was collected from a mixture of COREL images as well as personal photographs [11]. All images are colored and sized of 256x256 pixels. For each classification experiment 100 images of each category are reserved for test purpose and the remaining images are used as training set. Sample images for the 8 categories are given in Figure 3.



Fig. 3. Sample images of the database. From top left to bottom right: Forest, Highway, Coast, Street, Inside of city, Street, Mountain, Open Country, Tall building

5.2 Choice of Modalities



Fig. 4. Sample images from the least similar classes: Forest, highway, coast, street



Fig. 5. Sample images from the most similar classes: Inside of city, street, tall building, mountain

We use our image database to generate two groups of images that contain both four classes. These groups are arranged in such a way that one group contains the four most similar classes and the other the four least similar ones. We use these two groups in the remaining experimentations to compare multiclass classification methods. We suppose that if two classes are similar then the binary classification performance for these classes is low and vice versa. In other words, similarity of two classes varies in the opposite way with binary classification accuracy of these classes. In order to obtain the groups of classes mentioned above we performed binary classifications between every possible pair of classes in our image database (Combination (8, 2) =28) based on a local texture feature that is obtained by extracting four attributes namely energy, entropy, homogeneity and inertia from gray level co-occurrence matrix. This feature is extracted from block of 64x64 pixels. We sorted the binary classification results by accuracy. Keeping in mind that 6 classifiers are needed to

Authorized licensed use limited to: ULAKBIM UASL - Galatasaray Universitesi. Downloaded on February 26, 2009 at 04:44 from IEEE Xplore. Restrictions apply.

build a 4-class classifier the 6 best performing classifiers sufficient to construct a 4-class classifier are selected; these four classes are the most similar ones according to the feature that is used. Following the same procedure the four least similar classes are found. The four most similar classes are Inside of city, Street, Tall building and Mountain and the four least similar classes are Forest, Highway, Coast, Street (Figure 4 and Figure 5). This result is in accordance with ordering based on spectral signature as presented in [11]. For each category the spectral signature is obtained by averaging the power spectra of a few hundred images that belong to this category. The authors showed that spectral signature is very appropriate to discriminate the categories. Categories very close to each other exhibit similar spectral signatures while for more distant categories the shape of the spectral signatures is less similar.

5.3 Comparison of Pairwise Coupling Functions

The goal of this paragraph is to compare the pairwise coupling functions defined in 3.1.2. Once the best of these functions is found the further comparison of the methods will cover only the pairwise coupling method with the best performing function only. This will increase the readability and clarity of further results and help us to analyze the results in a more appropriate manner. Results are given in terms of accuracy which is the most widely used performance measure in image classification domain. Accuracy is the simplest way to compare two confusion matrices because it is a measure that represents the whole classification with a single value. It varies in interval [0,1], 1 for perfect classification and 0 for a classifier that makes wrong decision systematically.

Table 3 shows the classification results of the least similar classes using local and global representations. Texture representation as a local representation used in this experiment is defined in 5.2. Gist used as a global representation is described in 3.

	Local	Global
PWC1	***0.775***	***0.892***
PWC2	*0.760*	*0.883*
PWC3	*0.710*	*0.835*
PWC4	*0.692*	*0.822*

Table 3. Classification results for the least similar classes using pairwise coupling

According to the results in Table 3, PWC1 is the most performing pairwise coupling function among the four different functions. Functions are ranked as PWC1 to PWC4.

The same experiment is repeated for the classification of the most similar classes. Results are given in Table 4. The ranking of the functions remain the same except for PWC3 and PWC4 for global representation only. The winner is PWC1 again for this experiment. According to these result we can conclude that the most performing pairwise coupling function is PWC1.

	Local	Global
PWC1	***0.590***	***0.817***
PWC2	*0.585*	*0.805*
PWC3	*0.544*	*0.776*
PWC4	*0.520*	*0.790*

Table 4. Classification results for the most similar classes using pairwise coupling

5.4 Comparison of the one-against-one strategies

To evaluate the one-against-one strategies, we performed a set of experiments. We used both local and global image representation as mentioned in the previous paragraph. Two extreme cases have been investigated namely the classification of the most and the least similar classes. Classification results in terms of accuracy for the least similar classes using global and local representations are given in Table 5.

Methods	*Local*	*Global*
DAG	*0.767*	0.875
PWC1	*0.775*	0.892
ECOC	*0.780*	*0.900*
OAO-MaxWins	*0.780*	0.915

Table 5. Classification results for the least similar classes using one-against-one strategies.

According to the classification accuracies, the ordering of the multi-class SVM methods in ascendant performance is DAG, PWC1, ECOC, and OAO-MaxWins. The ordering is not influenced by the type of image representation. It seems that we can consider two classes of strategies with similar performances. The most performing class contains ECOC and OAO-MaxWins, while the least performing class group DAG and PWC1.

An increase of performance is observed for all methods comparing to the classification based on local texture feature with the same classes. This shows that global representation is more discriminative then local.

Authorized licensed use limited to: ULAKBIM UASL - Galatasaray Universitesi. Downloaded on February 26, 2009 at 04:44 from IEEE Xplore. Restrictions apply.

The results obtained using the most similar dataset are in accordance with the previous results. The only difference is that the accuracy values are lower.

5.5 Comparison of OAO against other strategies

We performed the same experiments for HAH and OAA. Considering the least similar classes, we first regrouped Highway and Street classes against Forest and Coast. For the most similar classes, Street and Inside of city are joined and Tall building and Mountains formed the other group. In Table 6 the performances of both strategies with the best OAO method in terms of accuracy are compared.

Methods	*Local*	*Global*
HAH	*0.770*	*0.878*
OAA	*0.785*	0.905
OAO-MaxWins	*0.780*	0.915

Table 6. Classification results for the least similar classes of OAO against other strategies.

For all the experiments, HAH is always the least performing strategy. As shown in Table 6, OAO-MaxWins rank first for global representation while OAA wins for local representation. The results are inverted when we performed the experiment with the most similar classes. In fact the performances are so close that it is very difficult to choose between the two of them.

6 Conclusion

We evaluated Divide-and-combine multi-class SVM methods. Results show that OAO-MaxWins and OAA are the most performing strategies for combining binary classifiers in natural scene classification task. This conclusion is confirmed by experiments performed on two separate groups of images using two different types of representation, local and global. Our work shows that the usage of multiple image representations and multiple test groups is not only preferable but also necessary because considering experiments separately leads to different performance orderings.

References

1. B. Boser, I.Guyon, and V. Vapnik. A training algorithm for optimal margin classi¯ers. In D. Haussler, editor, *5th Annual ACM Workshop on COLT*, pages 144{152, 1992.

2. K. Crammer and Y. Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, vol. 2:265{292, 2001.

3. Vapnik, V.: The Nature of Statistical Learning Theory. New York: Wiley, (1998)

4. J.H.Friedman. Another approach to polychotomous classification, Technical report, Standford University, Department of Statistics, 10:1895-1924,1998

5. Hastie, T., Tibshirani, R.: Classification by pairwise coupling. In: Jordan, M.I., Kearns, M.J., Solla, A.S. (eds.): *Advances in Neural Information Processing Systems 10*. MIT Press (1998)

6. Dietterich, T., Bakiri, G.: Solving multiclass problem via error-correcting outputcode. *Journal of Artificial Intelligence Research*, Vol. 2 (1995) 263–286

7. Platt, J., Cristianini, N., Shawe-Taylor, J.: Large Margin Dags for Multiclass Classification. Advances in Neural Information Processing Systems 12, (2000) 547-553

8. Lei. H, Govindaraju, V., 'Half-Against-Half Multi-class Support Vector Machines' Lecture Notes in Computer Science, Vol. 3541, pp. 156-164. (2005)

9. Hsu, C., Lin, C.: A Comparison of Methods for Multi-class Support Vector Machines. IEEE Transactions on Neural Networks, Vol. 13. (2002) 415-425

10. Ren, J., Shen, Y., Ma, S., Guo, L.: Applying Multi-Class SVMs into Scene Image Classification. Proceedings of the 17th International Conference on Innovations in Applied Artificial Intelligence, (2004) 924-934

11. He, L., Kong, F., Shen, Z.: Multiclass SVM Based Land Cover Classification With MultiSource Data. Machine Learning and Cybernetics, Proceedings of 2005 International Conference, Vol. 6. (2005) 3541 – 3545

12. Duan, K.B., Keerthi, S.S., 'Which Is the Best Multiclass SVM Method? An Empirical Study' Lecture Notes in Computer Science, Vol. 3541, pp. 278-285. (2005)

13. Demirkesen, C., Cherifi, H.: Local or Global Image Representation for Support Vector Machine Image Categorization. To appear in: IEEE The 15th International Conference on Systems, Signals and Image Processing, IWSSIP (2008)

14. Oliva, A. Torralba, A.: Building the Gist of a Scene: The Role of Global Image Features in Recognition. *Progress in Brain Research: Visual perception,* Vol. 155. (2006). 23-36

Authorized licensed use limited to: ULAKBIM UASL - Galatasaray Universitesi. Downloaded on February 26, 2009 at 04:44 from IEEE Xplore. Restrictions apply.