STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume **LXI**, Number 2, 2016

SUPPORT VECTOR MACHINE AND BOOSTING BASED MULTICLASS CLASSIFICATION FOR TRAFFIC SCENE OBSTACLES

ROXANA MOCAN AND LAURA DIOŞAN

Abstract. Multiclass classification is an extensively researched topic due to its importance in making the binary classification problems a complex and well tuned system and minimising the running time for multiple classification problems. In the traffic scenes one can encounter several types of obstacles like cars, pedestrians, animals, low elevated objects, road signs that must be detected and categorised for safety reasons regarding the driver and traffic. The purpose of this paper is two-folds: to accurately classify four obstacle types (pedestrians, cars, animals and other types of objects) and to compare some multiclass classification methods based on Support Vector Machine and Boosting algorithms. The experiments showed that the method Fuzzy Clustering with improvements using Particle Swarm Optimisation achieves great results compared to the traditional hierarchical multiclass classification and the proposed hybrid approach that combines Boosting and Support Vector Machine increases the classification accuracy even further.

1. Introduction

The main purpose of the paper is to classify multiple types of objects in traffic scenes. The focus is to accurately classify animals, pedestrians, cars and road signs given an area of interest. Objects belonging to the same class may vary from each other in view or shape, which increases the difficulty of classification. An accurate traffic scene classification cannot be done without a discriminative feature extractor and a strong multiclass system. For feature extraction phase a method that extracts from images the Histogram of Oriented Gradients (HOG) based on Aspect Ratio of Region of Interest (ROI) is presented. This method is characterised by a good discriminative power and improves not only the classification results, but the prediction time too. For

Received by the editors: September 30, 2016.

2010 *Mathematics Subject Classification.* 68T05, 91E45.

1998 *CR Categories and Descriptors.* I.2.6 [**Learning**]: – *Induction*, I.2.6 [**Learning**]: – *Concept learning*.

Key words and phrases. Multiclass classification, HOG, Decision Trees, Boosting.

the learning phase, an hierarchical approach is proposed: the learning data are distributed on layers by using a Fuzzy Clustering algorithm, whose parameters are optimised by a Particle Swarm Optimisation (PSO) method. For classification a hybrid approach that combines Support Vector Machine (SVM) as base classifier and boosting for increasing the classification rate is proposed. The comparison with the traditional methods, both for feature extraction and multiclass classification, show that the proposed approach can improve the prediction performance.

The contribution of this paper is two-fold:

- We propose a hybrid classifier that combines the SVM with boosting in order to get better results in terms of accuracy. Tests have shown that using only SVM will speed up the process, but for better results is needed a more complex classification system. The classifier is based on a Fuzzy C-Means (FCM) algorithm whose parameters are optimized by PSO.
- We present the improvements obtained by FCM Clustering and PSO over the Traditional Hierarchical Classification (THC) in order to classify multiple types of objects from traffic scenes.

The outline of the paper is as follows. After briefly reviewing related work to multiclass classification algorithms in Section 2, we present the background of our system in Section 3.The proposed approach is described in Section 4. Numerical experiments are presented in Section 5, while the conclusions are highlighted in Section 6.

2. Related work

Different methodologies have been developed until now for dealing with the multiclass classification problem.

2.1. **Extended Binary Classifiers.** A first category is represented by several binary classifiers that have been naturally extended to the multiclass case directly:

- Artificial Neuronal Networks [1] - by having more binary output neurons [2].
- SVM [3] - additional parameters and constraints are added to the optimisation problem involved in the classical SVM in order to handle the separation of the different classes.

2.2. **Binarisation techniques.** This section discusses strategies for reducing the problem of multiclass classification to multiple binary classification problems (e.g. SVM where the number of classes are divided repeatedly by two until is reached a binary classification [4]. Previous works have been done

showing the advantage of binarisation techniques [5], [6] and [7]. After decomposition, an individual classifier is designed to classify only some classes in the problem [8]. In order to label all the examples, a combination method must be used. Usually, it is easier to build a classifier to distinguish only between fewer classes than to consider all the classes of the problem. When the classifier has to distinguish between two classes only, the ensemble will be composed by some base binary learners. Different decomposition strategies can be found in the literature:

- one-against-all (OAA) - the strategy involve training a single classifier per class, with the samples of that class as positive samples and all other samples as negatives [9]. In this case, K base binary classifiers, where K is the number of classes, are required in order to discriminate one of the classes from all other classes.
- one-against-one (OAO) - one trains $K(K-1)/2$ binary classifiers for a K-way multiclass problem; each classifier receives the samples of a pair of classes from the original training set, and must learn to distinguish these two classes. At prediction time, a voting scheme is applied [7].

Different methods to combine the outputs of the base classifiers from these strategies have been developed, called in what follows aggregation methods. In the case of OAO strategy, several aggregation methods are used in order to obtain the predicted class from a score matrix computed by the binary classifiers from the ensemble. Some of them are:

- Preference relations solved by Non-Dominance Criterion (ND) [10],
- Voting (binary voting or Max-Wins rule [11]),
- Binary tree of classifiers (BTC) [12],
- Weighted voting strategy (WV) [13]),
- Learning valued preference for classification (LVPC) [14], [13],
- Nesting one-against-one (NEST) [15].

2.3. **Special formulations.** The last category includes methods that require special formulations in order to be able to solve the multiclass classification problem (classifiers that have a built-in multiclass support). Approaches that try to pose a hierarchy on the output space (the available labels) and then to perform a series of tests to detect the class label of new patterns are included in this category. Some multiclass classification problems allows arranging the classes in a tree, obtaining a hierarchical division of the output space. In this case, a special class of classifiers could be used:

- Binary Hierarchical Classifier [16] introduces a hierarchical technique to recursively decompose a K-class problem in two meta-class problems. A generalised modular learning framework is used to partition a set of classes into two disjoint groups called meta-classes. The coupled problems of finding a good partition and of searching for a linear feature extractor that best discriminates the resulting two meta-classes are solved simultaneously at each stage of the recursive algorithm. This results in a binary tree whose leaf nodes represent the original K classes.

- Divide–By–2 tree [17] is used for extending SVM to multiclass problems. Divide–By–2 offers an alternative to the standard OAO and OAA algorithms. Beginning with the whole data set, Divide–By–2 hierarchically divides the data into two subsets until every subset consists of only one class. This algorithm divides the data such that instances belonging to the same class are always grouped together in the same subset. Thus, it requires only $N - 1$ classifiers, where N represent the number of training examples. In Section 3.2 we describe in detail how these $N - 1$ classifiers are built during training. And, in Section 5 we illustrate how Divide–By–2 method classifies new data in the testing phase.

3. Theoretical Background

The approach developed in this paper is a hybrid multiclass classifier. It combines SVM with boosting in order to get better results in terms of accuracy. Tests showed that using only SVM will speed up the process, but for better results is needed a more complex classification system. Furthermore, the hybrid approach is based on a fuzzy clustering of data, when an optimised FCM algorithm is used (the optimisation of the clustering algorithm is performed by a PSO method).

Furthermore, this paper aims, also, to present the improvements obtained by FCM clustering and PSO over the THC in order to classify multiple types of objects from traffic scenes. The focus is to classify animals, pedestrians, cars and others, given a region of interest. All classification systems, which have as input images, have to perform two steps:

- feature extraction from each image and
- model learning.

Each step will be detailed in what follows.

3.1. **Feature extraction.** Depending on image width (W) and height (H) the ROI can have different aspect ratio. For example the ROI of a pedestrian image has W much smaller than the height H, an animal image can have $H = W$ or $H < W$ depending on animal's position, car images have always $W > H$ and others can have different sizes depending on the object. That is the reason Aspect Ratio Histogram of Oriented Gradients [18] has been included in this paper, for getting the same size for all HOG vectors, but not depending on the input image size.

3.2. **Multiclass classification.** Several multiclass classifiers are investigated in this paper. Some of them already exists in literature, while others are proposed here in comparison to the traditional ones.

3.2.1. *Traditional Hierarchical Multiclass Classification.* In this paper, the THC is used: the K classes are repeatedly divided in two groups for each layer, heuristically chosen, as represented in Fig. 1.

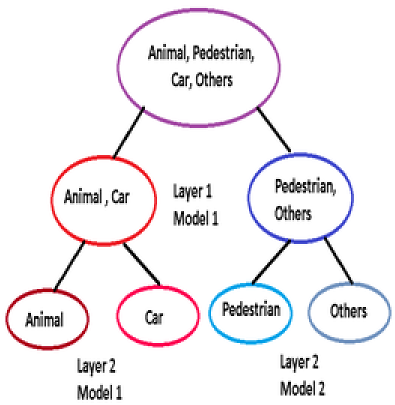


FIGURE 1. Hierarchy of classes

3.2.2. *Hybrid Multiclass Classification using FCM based on Particle Swarm Optimization (FCM&PSO).* The proposed approach is based on a hierarchical classification system. In this hierarchy, the input data are separated by using an optimised FCM clustering approach (the parameters of the FCM are optimised by a PSO algorithm). The idea of FCM-based multiclass classification have been proposed in [19], but this time, at the classification level, a hybrid method is actually used: an SVM is combined by Boosting.

3.2.3. *Fuzzy C-means.* Fuzzy clustering algorithms treat clusters as soft groups to which every data object has a membership degree. These algorithms are slower than crisp approaches, but give better results in cases where data is incomplete or uncertain and has a wider applicability.

The FCM algorithm, proposed by Bezdek [20], is the one of the most popular fuzzy clustering algorithms in the literature and it creates the non-unique partitioning of the data in a collection of clusters. FCM is using a fuzzification parameter M in the range $[1, N]$ (N is the number of data samples), which defines the degree of fuzziness in the cluster. This parameter, known as weighting exponent, indicates the width of the group as follows:

- $M = 1$ means that there is a crisp clustering of points
- $M > 1$ is the degree of fuzziness among points in the decision space[1].

When tests on Learn in Layers approach [18] have been made, results showed that most of the classification errors were made in the first layer and have spread into the next layers. This kind of misclassification can be recovered by using clustering. This approach did not removed all the false positives, but it improved the results. Therefore all the data was separated in 2 clusters and trained the samples belonging to each cluster separately, a cluster containing all 4 initial classes, as in Figure 1. In the first cluster we had the majority of the samples (the samples that are more similar, in terms of distance correlation) and in the second cluster we had some samples that are different than the others (outliers) and must be treated differently for getting a better prediction.

The need of fuzzy clustering came from the idea of removing from cluster 1 only the samples that are too dissimilar (the membership degree threshold was more relaxed than the one from simple clustering).

For prediction phase we only computed the correlation distance from both clusters (by using Eq. 1) and decided which path will follow for further prediction (the models trained for cluster 1 or 2).

$$d(H_1, H_2) = \frac{\sum\limits_{i=1}^{N} (H_1(i) - H_1^{'})(H_2(i) - H_2^{'})}{\sqrt{\sum\limits_{i=1}^{N} (H_1(i) - H_1^{'})^2 (H_2(i) - H_2^{'})^2}} \tag{1}$$

where:

- H_1 = sample (HOG vector),
- H_2 = center (also a HOG vector),
- N = number of histogram bins,
- $H_k^{'} = \frac{1}{N} \sum_{j=1}^{N} H_k(j)$, where $H_k(j)$ is a HOG vector and is a sample or center depending on k ($k = 1$ or $k = 2$) and $H^{'}$ is the mean of all samples.

The performance of the presented algorithm is strongly influenced by its parameters. Therefore, a PSO algorithm is involved in the proposed approach in order to identify the best values of these parameters.

---

[1] Common range for M is $[1.25, 2]$.

3.2.4. *PSO.* Inspired by social behaviour of birds and fish, PSO algorithm combines self-experience with social experience [21]. PSO is a global swarm algorithm which uses multiple individual particles in parallel to explore the search space for the optimal solution. This algorithm uses the overall best solution and individual particle's best solution, a particle's inertia to determinate how to move each particle through the search space. A swarm is a set of particles that maintains its global best (*gBest*). A particle is a potential solution with position and velocity. Each particle maintains individual best position (*pBest*). Each particle tries to modify its position using the following information: current position, current velocities (momentum term), distance between the current position and *pBest* (cognitive term), distance between the current position and *gBest* (social term). Basic steps of the PSO algorithm [21] are:

(1) Initialization of the swarm from the solution space;
(2) Evaluation of the fitness of each particle;
(3) Update individual and global bests;
(4) Update velocity and position of each particle;
(5) Go to step 2 and repeat until termination condition.

4. Proposed Approach

4.1. **Hybrid classifier: SVM + Boosting.** The algorithm takes each sample and extracts the information using HOG based on Abspect Ratio of the images and computes by FCM the clustering for all the samples and the centroids are initialised with random samples from dataset. The FCM from this project has 2 centroids that are changing after each iteration in order to find the best division of samples in clusters. The output of the FCM algorithm is a membership matrix. Matrix size is $N \times C$ (where N is number of samples and C is number of centers) and it contains for each sample the percentage of belonging towards one centroid or another.

First the *pBest* and *gBest* of all particles are initialised with the same matrix, but after some iterations *pBest* of each particle changes individually if some membership matrix that have a better objective function (smaller Euclidean distance between samples from the same cluster) can be found. The global best *gBest* changes only if a particle finds the smallest distance between the centroids and belonging samples.

When the Euclidean distance from center to all samples from class has reached a minimum, the membership matrix that had made this minimum to be reached, is kept and the best of all particles is *gBest*. This is computed like multi-threading, but with communication between particles.

In the project 10 particles are utilised, and maximum 100 iterations are performed. PSO fuzzy clustering is stopped when the maximum number of iterations is reached or when is no difference between $gBest$ and previous $gBest$ for 5 consecutive iterations.

Despite its simplicity and ease of implementation, FCM has some shortcomings, such as sensitivity to the initialisation of the prototypes and the possibility of being trapped into the local optima. Data clustering algorithms based on swarm perform global search and thus can be used to improve clustering for a multidimensional feature space. A simpler algorithm for clustering is not possible using this approach. The k-means algorithm, for example, has not a membership matrix as output, therefore PSO receives no input.

Boosting Ensemble was successfully applied to extreme learning and has increased generalisation performance and stability of the system. Boosting refers to a general and, probably, effective method of producing a very accurate classifier by combining rough and moderately inaccurate rules of thumb. It is based on the observation that finding many rough rules of thumb can be a lot easier than finding a single, highly accurate classifier. The boosting algorithm repeatedly calls this weak learner, each time feeding it a different distribution over the training data.

For Boosting are used 200 weak classifiers and a weak classifier has maximum depth equal to two levels. The approach developed in this paper combines the FCM technique with boosting in order to get better results in terms of accuracy. The numerical tests showed that using only SVM will speed up the process, but for better results is needed a more complex classification system. In the proposed approach, in the first layer, FCM based on PSO algorithm combined by SVM as base classification method is used because the numerical experiments showed that the most misclassification is performed in the first layer and the error will propagate along the branches of the tree. For such a complex algorithm that has multiple iterations, the run time speed of the SVM method is needed. For the second layer, where we only have on each branch two classes, we can use Boosting even if is slower than SVM, but it has better results. The main reason of using this type of computation is that for a fast, but weak classifier we must use a complex system and for models that are smaller in terms of samples we can use a slow but strong classifier.

This hybrid classification consists on a trade-off of time consuming and gained accuracy.

5. Numerical experiments

The purpose of this paper is to validate the new approach and to compare it to THC and FCM&PSO improvements using the hybridisation schemes for multiclass classification described in previous sections.

Images are gathered from more than one image processing data sets: INRIA Pedestrian Dataset [22], Caltech Pedestrian and car data set [23] and the majority are from Stereo Camera placed on car and cropped after. The entire data set used for these experiments is a combination of internet images with real life images for a more robust classification.

As programming language C++ was used; the container vector is part of std library and image reading, processing and learning uses OpenCv library.

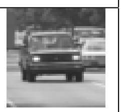
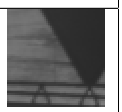
Classes	Data set for all 4 classes		
	Train images	Test images	Sample
Animal	3500	500	
Pedestrian	3500	500	
Car	3500	500	
Others	3500	500	

Table 1. Number of ROI images in the dataset

The performance measures that could be taken into account in the case of a classification problem are: the true positive rate (TPR), the false positive rate (FPR) and Global Accuracy (computed as an average over all classes) [18].

In Table 2 we compared the traditional approach with FCM based classification and the results show that the second algorithm achieved better results for global accuracy than the first one, but it has worse results for TPR on Animal class. In Table 3 we compared the FCM with PSO approach against the hybrid approach and the hybrid can show overall better results. Even if the global accuracy from FCM based classification is better than FCM with

PSO based classification with 1%, by using PSO the results are improved for Animal class with 14% and the 4 classes have about the same average accuracy.

FCM optimised by PSO performs better than traditional with 4.5% for global accuracy.

The proposed hybrid approach FCM & PSO has improved the global accuracy with 2%.

Method	THC			FCM & PSO		
	TPR	FPR	Global Accuracy	TPR	FPR	Global Accuracy
Animal	0.66	0.11	0.7906	0.70	0.09	0.8356
Pedestrian	0.78	0.06	$\pm$ 0.00039	0.87	0.05	$\pm$ 0.00036
Car	0.62	0.09		0.70	0.06	
Others	0.66	0.14		0.73	0.11	

Table 2. Experimental results for THC and classification using FCM

Method	FCM&PSO			Proposed approach		
	TPR	FPR	Global Accuracy	TPR	FPR	Global Accuracy
Animal	0.70	0.09	0.8356	0.71	0.10	0.8553
Pedestrian	0.87	0.05	$\pm$ 0.00036	0.89	0.08	$\pm$ 0.000345
Car	0.70	0.06		0.74	0.01	
Others	0.73	0.11		0.78	0.16	

Table 3. Experimental results for classification using FCM & PSO and classification using the proposed approach

6. Conclusions and further work

The aim of this paper is to improve the classification accuracy of the four types of objects that we can find most often in traffic scenes (animal, pedestrian, car and other objects). Different methods used for multiclass classification of objects in traffic scene have been presented. The classes contain ROI images with animals, pedestrians, cars and others.

First, we experiment with a multiclass THC and than we added FCM for excluding outliers. FCM method worked better than the traditional one, but for animal class the results have been weaker. Therefore, we added PSO optimisation in order to provide better balance of fuzzy clustering and to avoid falling into local minima quickly and thereby obtaining better solutions.

In the end we tried a hybrid classifier that outperforms the previously investigated methods by using different classifiers for each layer. Further work would consist in adding new feature extraction algorithms to improve the classification results.

References

[1] C. M. Bishop, *Neural Networks for Pattern Recognition.* Clarendon Press, Oxford, 1995.

[2] T. G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," *Journal of Artificial Intelligence Research*, vol. 2, pp. 263–286, 1995.

[3] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, p. 273, 1995.

[4] P. Clark and R. Boswell, "Rule induction with CN2: Some recent improvements," in *Proc. Fifth European Working Session on Learning.* Berlin: Springer, 1991, pp. 151–163.

[5] J. Frnkranz, "Round robin classification," *Journal of Machine Learning Research*, vol. 2, pp. 721–747, 2002.

[6] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE-EC*, vol. 13, pp. 415–425, Mar. 2002.

[7] R. Rifkin and A. Klautau, "In defense of one-vs-all classification," *Journal of Machine Learning Research*, vol. 5, pp. 101–141, 2004.

[8] M. Wozniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, vol. 16, pp. 3–17, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.inffus.2013.04.006

[9] T. Hastie and R. Tibshirani, "Classification by pairwise coupling," in *Advances in Neural Information Processing Systems*, M. I. Jordan, M. J. Kearns, and S. A. Solla, Eds., vol. 10. The MIT Press, 1998.

[10] A. Fernández, M. Calderón, E. B. Tartas, H. B. Sola, and F. Herrera, "Solving multi-class problems with linguistic fuzzy rule based classification systems based on pairwise learning and preference relations," *Fuzzy Sets and Systems*, vol. 161, no. 23, pp. 3064–3080, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.fss.2010.05.016

[11] J. H. Friedman, "Another approach to polychotomous classification," Department of Statistics, Stanford University, Stanford, CA, Tech. Rep., 1996. [Online]. Available: http://www-stat.stanford.edu/ jhf/ftp/poly.ps.Z

[12] B. Fei and J. Liu, "Binary tree of SVM: a new fast multiclass training and classification algorithm," *IEEE Trans. Neural Networks*, vol. 17, no. 3, pp. 696–704, 2006. [Online]. Available: http://dx.doi.org/10.1109/TNN.2006.872343

[13] E. Hullermeier and S. Vanderlooy, "Combining predictions in pairwise classification: An optimal adaptive voting strategy and its relation to weighted voting," *Pattern Recognition*, vol. 43, no. 1, pp. 128–142,, Jan. 2010. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2009.06.013

[14] J. C. Huhn and E. Hüllermeier, "FR3: A fuzzy rule learner for inducing reliable classifiers," *IEEE Trans. Fuzzy Systems*, vol. 17, no. 1, pp. 138–149, 2009. [Online]. Available: http://dx.doi.org/10.1109/TFUZZ.2008.2005490

[15] B. L. 0002, Z. Hao, and E. C. C. Tsang, "Nesting one-against-one algorithm based on SVMs for pattern classification," *IEEE Trans. Neural Networks*, vol. 19, no. 12, pp. 2044–2052, 2008. [Online]. Available: http://dx.doi.org/10.1109/TNN.2008.2003298

[16] S. Kumar, J. Ghosh, and M. M. Crawford, "Hierarchical fusion of multiple classifiers for hyperspectral data analysis," *Pattern Anal. Appl*, vol. 5, no. 2, pp. 210–220, 2002. [Online]. Available: http://dx.doi.org/10.1007/s100440200019

[17] V. Vural and J. G. Dy, "A hierarchical method for multi-class support vector machines," in *ICML*, ser. ACM International Conference Proceeding Series, C. E. Brodley, Ed., vol. 69. ACM, 2004. [Online]. Available: http://doi.acm.org/10.1145/1015330.1015427

[18] R. Mocan and L. Diosan, "Obstacle recognition in traffic by adapting the hog descriptor and learning in layers," *Studia Univ. Babes Balyai*, vol. 55, pp. 47–54, 2015.

[19] R. Mocan and L. Dioşan, "Multiclass classification based on clustering approaches for obstacle recognition in traffic scenes," in *ICCP 2016*, 2016, pp. 1–8.

[20] R. E. J. C. Bezdek and W. Full, "FCM: The fuzzy *c*–means clustering algorithm," *Computers and Geosciences*, vol. 10, no. 2–3, pp. 191–203, 1984.

[21] A. Szabo, L. N. de Castro, and M. R. Delgado, "The proposal of a fuzzy clustering algorithm based on particle swarm," in *NaBIC*. IEEE, 2011, pp. 459–465. [Online]. Available: http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6083215

[22] N. Dalal, "INRIA person dataset," 2005. [Online]. Available: http://pascal.inrialpes.fr/data/human/

[23] "Caltech dataset," 2005. [Online]. Available: http://vision.caltech.edu/archive.html

Department of Computer Science, Faculty of Mathematics and Computer Science, Babeş-Bolyai University, 1 M. Kogălniceanu St., Cluj-Napoca, Romania

E-mail address: `roxanamocan@ymail.com`

E-mail address: `lauras@cs.ubbcluj.ro`