Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 4, April 2014, pg.65 – 69

REVIEW ARTICLE

Multiclass Classification: A Review

[1]Ruchida Sonar, [2]Dr. P.R. Deshmukh

[1]Student, Sipna COET, Amravati

ruchidas.sonar@gmail.com, p_rdeshmukh@yahoo.com

ABSTRACT – Support Vector Machines (SVMs) are with success applied to resolve an outsized variety of classification and regression issues. SVMs were at first developed to perform binary classification; though, applications of binary classification are terribly restricted. The way to effectively extend it for multiclass classification remains Associate in nursing current analysis issue. Most of the sensible applications involve multiclass classification, particularly in remote sensing land cowl classification. Many ways are planned wherever usually we tend to construct a multiclass classifier by combining many binary classifiers. Because it is computationally costlier to resolve multiclass issues, comparisons of those ways victimization large-scale issues haven't been seriously conducted. This paper compares the performance of six multi-class approaches to resolve classification drawback with remote sensing information in term of classification accuracy and machine price. we tend to then compare their performance with four ways supported binary classifications: "one-against-all," "one-against-one," and directed acyclic graph SVM (DAGSVM) and Error Corrected Output secret writing (ECOC) and combines their results to work out the category label of a take a look at pixel. Results from this study conclude that the "one-against-one" and DAG ways are additional appropriate for sensible use than the opposite ways. And conjointly indicate that for finding massive issues ways by considering all information quickly generally want less variety of support vectors.

I. INTRODUCTION

In machine learning, support vector machines (SVMs, additionally support vector networks) are supervised learning models with associated learning algorithms that analyze information and acknowledge patterns, used for classification and multivariate analysis. Given a collection of coaching examples, every marked as happiness to at least one of 2 classes, associate SVM coaching algorithmic program builds a model that assigns new examples into one class or the opposite, creating it a non-probabilistic binary linear classifier.

A new system supported applied math learning theory (Vapnik, 1995), referred to as the support vector machine (Boser et al., 1992) has recently been applied to the matter of remote sensing information classification [1] [2]. The most plan behind this classification technique is to separate the categories with a surface that maximize the margin between them, mistreatment boundary pixels to form the choice surface. The info points that are highest to the hyperplane are termed "support vectors". The amount of support vectors is therefore tiny as they're points near the category boundaries (Vapnik, 1995). Support vector machines (SVMs) are with success used during a range of world issues.

- SVMs are useful in text and machine-readable text categorization as their application will considerably cut back the necessity for tagged coaching instances in each the quality inductive and transductive settings.
- SVMs also are helpful in life science to classify proteins with up to ninetieth of the compounds classified properly.
- Hand-written characters are recognized mistreatment SVM.

One major advantage of support vector classifiers is that the use of quadratic programming, that provides international minima solely. The absence of native minima may be a vital distinction from the neural network classifiers. Like neural classifiers, applications of SVMs to any classification drawback need the determination of many user-defined parameters. a number of these parameters are the selection of an appropriate multiclass approach, alternative of associate applicable kernel and connected parameters, determination of an appropriate price of regularization parameter (i.e. C) and an appropriate improvement technique.

SVMs were initially developed to perform binary classification; though, applications of binary classification are terribly restricted. Most of the sensible applications involve multiclass classification, particularly in remote sensing land cowl classification. Variety of ways is projected to implement SVMs to provide multiclass classification. Most of the analysis in generating multiclass support vector classifiers is divided in 2 classes. One approach involves in constructing many binary classifiers and brushing their results whereas alternative approach considers all information in one optimization formulation. This paper compares the performance of a number of the multi category approaches in term of classification accuracy and therefore the process price for land cowl classification victimization remote sensing information.

II. SUPPORT VECTOR MACHINES

Support Vector Machine could be a sensible learning technique supported applied math Learning Theory. An easy SVM may beat a complicated neural network with elaborate options in a very handwriting recognition task. SVM have the aim of crucial the situation of call boundaries that turn out the best separation of categories (Vapnik 1995). Within the case of a two-class pattern recognition drawback during which the categories are linearly severable the SVM selects from among the infinite range of linear call boundaries the one that minimizes the generalization error. The greatest margin between the two classes, where margin is defined as the sum of the distances to the hyperplane from the closest points of the two classes (Vapnik, 1995) [1]. This drawback of maximising the margin is often solved victimization commonplace Quadratic Programming (QP) optimization techniques. The data points that are closest to the hyperplane are used to measure the margin; hence these data points are termed 'support vectors'.

SVM also can be extended to handle non-linear call surfaces. If the 2 categories don't seem to be linearly severable, the SVM tries to seek out the hyperplane that maximizes the margin whereas, at identical time, minimizing a amount proportional to the quantity of misclassification errors. The trade-off between margin and misclassification error is controlled by a user-defined constant. Boser et al. (1992) propose a way of protruding the computer file onto a high-dimensional feature house victimization kernel functions (Vapnik 1995) and formulating a linear classification drawback in this feature house.

III. SVM for Multiclass Classification

SVM were at first designed for binary (two-class) issues. Once coping with multiple categories, associate acceptable multi-class technique is required. Vapnik (1995) instructed scrutiny one category with the others taken along. This strategy generates n classifiers, wherever n is that the range of categories. The ultimate output is that the category that corresponds to the SVM with the biggest margin, as outlined on top of. For multi-class issues one has got to confirm n hyperplanes. Thus, this technique needs the answer of n QP optimization issues, every of that separates one category from the remaining categories. This strategy are often represented as 'one against the rest'.

A second approach is to mix many classifiers ('one against one'). Knerr et al. (1990) perform pair-wise comparisons between all n categories [3]. Thus, all potential 2 category classifiers are evaluated from the coaching set of n categories, every classifier being trained on solely 2 out of n categories, giving a complete of n (n-1)/2 classifiers. Applying every categoryifier to the take a look at information vectors provides one vote to the winning class. The information is assigned the label of the category with most votes. The results of a recent analysis of multi-class methods are provided by Hsu and architect (2002) [4].

Originally, SVMs were developed to perform binary classification. However, applications of binary classification are terribly restricted particularly in remote sensing land cowl classification wherever most of the classification issues involve quite 2 categories. Variety of ways to get multiclass SVMs from binary SVMs are projected by analyzers and continues to be an unbroken research topic. This section provides a quick description of some ways enforced to unravel multi-class classification drawback with SVM in present study.

a) One against One Approach

In this technique, SVM classifiers for all potential pairs of categories are created. Therefore, for M categories, there'll be binary classifiers. The output from every classifier within the variety of a category label is obtained. The category label that happens the foremost is assigned to it purpose within the information vector. Just in case of a tie, a tie-breaking strategy is also adopted. A standard tie-breaking strategy is to indiscriminately choose one in all the category labels that are tied. The quantity of classifiers created by this technique is mostly abundant larger than the previous technique. However, the quantity of coaching information vectors needed for every classifier is way smaller. The quantitative relation of coaching information vector size for one category against another is additionally. Therefore, this technique is taken into account a lot of bilateral than the One against-the-rest technique. Moreover, the memory needed to make the kernel matrix is way smaller. However, the main disadvantage of this method is the increase in the number of classifiers as the number of class increases. For instance, for seven categories of interest, twenty one classifiers are going to be created.

b) One against the remainder approach

This methodology is additionally referred to as winner-take-all classification. Suppose the dataset is to be classified into M categories. Therefore, M binary SVM categoryifiers could also be created wherever every classifier is trained to tell apart one class from the remaining M-1 categories. For instance, category one binary categoryifier is intended to discriminate between class one information vectors and therefore the information vectors of the remaining categories. Different SVM classifiers square measure created within the same manner. Throughout the testing or application part, information vectors square measure classified by finding margin from the linear separating hyperplane. The ultimate output is that the category that corresponds to the SVM with the biggest margin. However, if the outputs adore 2 or additional categories square measure terribly near one another, those points square measure labeled as unclassified, and a subjective call could got to be created by the analyst. Otherwise, a reject call (Schölkopf and Smola, 2002) might also be applied employing a threshold to choose the category label [5]. This multiclass methodology has a plus within the sense that the quantity of binary classifiers to construct equals the quantity of categories. However, there square measure some drawbacks. First, throughout the coaching part, the memory demand is extremely high and amounts to at the sq. of the full range of coaching samples. This could cause issues for giant coaching information sets and will result in computer storage issues. Second, suppose there square measure M categories and every have Associate in nursing equal range of coaching samples. Throughout the coaching part, the magnitude relations of coaching samples of 1 category to remainder of the categories are going to be 1: (M −1). This ratio, therefore, shows that coaching sample sizes are going to be unbalanced. Thanks to these limitations, the one against one approach of multiclass classification has been planned.

c) Decision Directed Acyclic Graph based Approach

Platt et al. (2000) planned a multiclass classification methodology referred to as Directed Acyclic Graph SVM (DAGSVM) supported the choice Directed Acyclic Graph (DDAG) structure that forms a tree-like structure [6]. The DDAG methodology in essence is comparable to combine wise classification such, for Associate in Nursing M category classification drawback, the quantity of binary classifiers is adequate to 1/2 M (M – 1) and every classifier is trained to classify 2 categories of interest. Every classifier is treated as a node within the graph structure Nodes in DDAG are organized in a triangle with the single root node at the top and increasing thereafter in an increment of one in each layer until the last layer that will have *M* nodes. The DDAG evaluates Associate in Nursing input vector x beginning at the foundation node and moves to successive layer supported the output values. As an example, it exits to the left edge if the output from the binary classifier is negative, and it exits to the correct edge if the output from the binary classifier is positive. The binary classifier of successive node is then evaluated. The trail followed is termed the analysis path. The DDAG methodology primarily eliminates one category out from an inventory. At first the list contains all categories. Every node evaluates the primary category against the last category within the list. For instance, the foundation node evaluates category one against category M. If the analysis leads to one category out of 2 categories, the opposite is eliminated from the list. The method then tests the primary and therefore the last category within the new list. It's terminated once just one category remains within the list. The category labels related to the input file are going to be the category label of the node within the final layer of the analysis path or the category remained within the list. Though the quantity of binary classifiers still equals the combine wise classification methodology, the inputs square measure evaluated M −1 times instead 1/2 M (M – 1) times as is that the case with combine wise classification.

d) Error-Correcting Output Code based approach

The construct of Error- Correcting Output Coding (ECOC) based mostly multi-class methodology is to use binary (two-class) classifiers to resolve the multi-class classification issues [5]. This approach works by changing M category classification drawback into an oversized range L of 2-class classification issues. ECOC assigns a singular code word to a class rather than distribution every class a label. A (L, M, d) error correcting code may be a L bit long, having C distinctive code words with a acting distance of d. The acting distance between 2 code words is that the range of bit positions during which each differs. In a very classification drawback M is that the range of categories and L may be a range determined by the tactic accustomed generate error-correcting codes. Many ways like Hadamard- Matrix codes, BCH codes (Bose and Ray-chauduri, 1960) [7]; and complete codes (Dietterich and Bakiri, 1995) square measure planned to get error correcting codes. Dietterich and Bakiri, (1995) planned to use codes with most acting distance between one another and recommended that it (d - 1)/2 errors are often corrected within the code words for a acting distance d between the codes [8].

e) Multiclass Objective Function

Instead of making several binary classifiers to see the category labels, this methodology makes an attempt to directly solve a multiclass drawback (Weston and Watkins, 1998, Lee et al., 2001; Crammer and Singer, 2001; Schölkopf and Smola, 2002) [9] [10]. This is often achieved by modifying the binary category objective perform and adding a constraint to that for each category. The changed objectives perform permits coincident computation of multiclass classification and are given by (Weston and Watkins, 1998),

$$\min w,b,\xi \quad \left[1/2\sum_{i=1}^{m} \| w \|^2 + C\sum_{i=1}^{k}\sum_{r \neq yi} \xi_i^r \right]$$

Subject to the constraints,

$$w_{y_i}.x_i + b_{y_i} \geq w_r.x_i + b_r + 2 - \xi_i^r \quad \text{for,}$$

$$\xi_i^r \geq 0 \quad \text{for i=1….,k}$$

Where $Yi \in \{1\ldots,M\}$ are the multiclass labels of the data vectors and $r \in \{1,...,M\} \setminus yi$ are multiclass labels excluding yi. Lee *et al.* (2001) and Schölkopf and Smola (2002) showed that the results from this method and the one-against-the-rest are similar [11]. However, in this method, the optimization algorithm has to consider all the support vectors at the same time. Therefore, it may be able to handle massive data sets but the memory requirement and thus, the computational time may be very high. To summarize, it may be said that the choice of a multiclass method depends on the problem in hand. A user should consider the accuracy requirement, the computational time, the resources available and the nature of the problem. For example, the multiclass objective function approach may not be suitable for a problem that contains a large number of training samples and classes due to the requirement of large memory and extremely long computational time.

IV. CONCLUSION

Present study examined six approaches for the answer of multiclass classification downside victimization remote sensing information. One against one and DAG approach give a comparable accuracy and needs virtually same machine resources. The coaching time taken by one against one and DAG techniques is a smaller amount than that with the one against the remainder strategy. This study additionally concludes that the classification accuracy is achieved with thoroughgoing ECOC approach however needs very giant coaching time. A comparison of accuracy achieved by thoroughgoing ECOC approach suggests no important improvement as compared to at least one against one approach. The most downside with the 'one against the rest' strategy is that it should turn out unclassified information, and thus lower classification accuracies. Finally, results counsel the quality of 1 against one approach for this sort of knowledge in term of classification accuracy and therefore the machine price. More study is needed to check the utility of this approach with varied variety of remote sensing information additionally as information with sizable amount categories.

Following table shows the classification of multiclass approaches. ECOC approach doesn't seem to be considerably higher as compared to at least one vs. one approach in term of classification accuracy. Approach steered by Crammer and Singer (2001) needs an outsized coaching time (approx. 347 minutes) with no considerable gain in term of classification accuracy as compared to different multiclass approaches.

Table 1. Classification OF multiclass approaches

Sr. No	Name of Algorithm	Key idea	Advantages	Disadvantages
1	One Vs All(OVA)	With n classes, M binary problems are classified, where each problem discriminates a given class from the other M-1 class	1. Simple, 2. Provide comparable performance with other complicated approach when binary classifier is tuned well.	1. Training complexity is high, as the numbers of training samples are large. 2. Memory requirement is very high during training phase.
2	One Vs One(OVO)	Binary classification requires discriminating between each pair of classes, requiring M (M-1)/2 binary classifications.	1. Kernel matrix required smaller memory. 2. Better than OVA method shorter training time.	1. Slower in training, when number of classes is big as every test sample has to be presented to large number of classifier.
3	DAGSVM	Same idea as OVO, and in recognition phase, the algorithm depends on a rooted Binary DAG to make a decision.	1. Faster Testing and achieving similar recognition rate as OVO	1. Memory requirement and accuracy similar to OVA and OVO.
4	Error Correcting Codes	Based on idea of error correcting code for neural network.	1. Improve generalization ability.	

REFERENCES

[1] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
[2] Boser, H., Guyon, I. M., & Vapnik, V. N., 1992. A training algorithm for optimal margin classifiers. In Haussler, D., *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory* 144-152. Pittsburgh, PA: ACM Press.
[3] U. Kreßel, "Pairwise classification and support vector machines," in *Advances in Kernel Methods—Support Vector Learning*, B.
[4] Schölkopf, C. J. C. Burges, and A. J. Smola, Eds. Cambridge, MA: MIT Press, 1999, pp. 255–268.
[5] Hsu, C.-W., and C.-J. Lin, C.-J., 2002, A comparison of methods for multi-class support vector machines, *IEEE Transactions on Neural Networks*, 13, 415-425.JAMES, G., 1998, *Majority vote classifiers: Theory and Applications*. Ph.D. Thesis, Department of Statistics, Stanford University, Stanford, CA.
[6] Schölkopf, B. and Smola, A. J., 2002, *Learning with Kernels – Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA: The MIT Press.
[7] J. C. Platt, N. Cristianini, and J. Shawe-Taylor, "Large margin DAG's for multiclass classification," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2000, vol. 12, pp. 547–553.
[8] Bose, R. C., and Ray-Chauduri, D. K., 1960, on a class of error-correcting binary group codes. *Information and Control*, 3, 68-79.
[9] Dietterich T. G. and Bakiri G., 1995, solving multiclass learning problems via error correcting output codes. Journal of Artificial Intelligence Research, 2:263–286.
[10] J. Weston and C. Watkins, "Multi-class support vector machines," presented at the Proc. ESANN99, M. Verleysen, Ed., Brussels, Belgium, 1999.
[11] Crammer, K. and Singer, Y., 2001, on the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, vol. 2, pp. 265-292
[12] Lee, Y., Lin, Y., and Wahba, G., 2001, *Multicategory support vector machines* Tech. Rep. 1043, Department of Statistics, University of Wisconsin, Madison, WI. Peterson, W. W., and Weldon, E. J. Jr., 1972, *Error correcting codes*. MIT Press, Cambridge, MA.

