Does One-Against-All or One-Against-One Improve the Performance of Multiclass Classifications?

R. Kyle Eichelberger, Victor S. Sheng

Department of Computer Science, University of Central Arkansas, Conway, Arkansas, USA
rke06001@uca.edu, ssheng@uca.edu

Abstract

One-against-all and one-against-one are two popular methodologies for reducing multiclass classification problems into a set of binary classifications. In this paper, we are interested in the performance of both one-against-all and one-against-one for classification algorithms, such as decision tree, naïve bayes, support vector machine, and logistic regression. Since both one-against-all and one-against-one work like creating a classification committee, they are expected to improve the performance of classification algorithms. However, our experimental results surprisingly show that one-against-all worsens the performance of the algorithms on most datasets. One-against-one helps, but performs worse than the same iterations of bagging these algorithms. Thus, we conclude that both one-against-all and one-against-one should not be used for the algorithms that can perform multiclass classifications directly. Bagging is better approach for improving their performance.

Introduction

Two popular methods of simplifying the decisions made within multiclass classification are one-against-all (OAA) and one-against-one (OAO). These two methods are thought to increase performance among multiclass classification algorithms by reducing a multiclass problem to binary ones, as it is simpler to make predictions for binary sets than ones with multiple classes. There has been much research done on these two approaches to determine how they function against each other (Rifkin and Klautau 2004; Tsujinishi et al. 2007) but few compared the two approaches against multiclass algorithms directly (all-at-once, abbreviated as AAO). Also, these comparisons were non-comprehensive at best, only comparing the performance of these methods on specific algorithms. Hsu and Lin (2002) made comparisons among the three approaches (OAA, OAO, and AAO) with least squares SVMs and concluded that OAO performs the best while OAA and AAO are almost identical, but they introduce fuzzy membership functions into OAA and OAO, therefore we cannot determine which is better among the simplest versions of the approaches. This also presents the question as to why OAA does not outperform AAO, as it should. It has also been said that OAO does not perform better than AAO whatsoever, which contradicts most previous work (Sulzmann et al. 2007).

This lack of comparative information presents some problems for other researchers. It cannot be determined which approach to multiclass reduction performs the best. It also cannot be said which algorithm performs the best for either approach, as various tests show various results. Also it cannot be determined if it is worth it to use a multiclass reduction approach at all. This paper proposes to clarify these problems by testing and comparing the performances of one-against-all, one-against-one and all-at-once using four well known algorithms, decision trees, naïve bayes, support vector machine, and logistic regression. The accuracies for these experiments will be determined using twenty five unique multiclass datasets with varying numbers of iterations, attributes, and classes.

Experiments

The datasets used in the experiments were retrieved from the UCI Machine Learning Repository. These were chosen based on the criteria stated before, but also because of the varying number of nominal and numeric attributes within these datasets. The experiments were done using WEKA. Therefore, J48 was used for decision trees, NaiveBayes for naïve bayes, SMO for support vector machines, and Logistic for logistic regression. None of these algorithms were optimized in any way because the purpose of these experiments is to determine the performance of these methods through direct comparison. Each method of classification was run on each dataset ten times, randomizing the dataset after each run. During these runs, the sets were split into 30%, used for the test sets, and 70%,

The work was supported by the National Science Foundation (IIS-1115417).

used for the training set.

It was determined that OAA and OAO held an unfair advantage over AAO. This advantage comes from the number of models used to make predictions for each approach. OAA uses n number of models to make its predictions, where n is the number of classes. This results in n binary problems being solved by the algorithm in use, while only one model is used by AAO. The same can be said for OAO, only it uses $\frac{n(n-1)}{2}$ models to make its predictions. This number is derived from the number of combinations necessary to make comparisons for every class value in the dataset. In order to level this advantage, comparisons were made with two additional approaches. These approaches will be called AAOvsOAA (for comparison with OAA) and AAOvsOAO (for comparison with OAO). These approaches use bagging to build multiple models. AAOvsOAA uses bagging to build n number of models while AAOvsOAO uses bagging to build $\frac{n(n-1)}{2}$ number of models. Comparisons are only done between these approaches and their respective counterparts.

The accuracies, including the standard deviation, were determined for each dataset. Those accuracies and standard deviations were averaged across the four algorithms and are shown in Table 1. Table 2 shows the comparisons amongst the three methodologies (OAA, OAO, AAO) and its variants (AAOvsOAA and AAOvsOAO) via the two-tailed t-test using a confidence level of 90%. We can see that the comparison result of OAA against AAO using J48 as the base learner is (3/16/6). That is, OAA only wins on three datasets, but loses on six datasets, and ties on the other datasets. Using this format for wins/ties/losses several things can be determined. First, OAA is outperformed in every comparison, losing 39 times to OAO, 30 times to AAO and 39 times to AAOvsOAA with only 16 wins total.

OAO performs much better that AAO, winning 22 times over AAO, but fails to outperform AAOvsOAO, only winning 6 times and losing a total of 15. These results are reflected in Table 1, with the average accuracies for OAA being the lowest method on all of the algorithms except for naïve bayes, on which it only outperforms AAO slightly. OAO performs the best on average for J48 and Logistic, but when compared to the averages for AAOvsOAO, it falls short on every algorithm but Logistic.

Thus it can be determined that OAA is not a good approach for improving the performance of multiclass classification algorithms while OAO performs much better than OAA across the board and can improve the performance of standard multiclass classification algorithms. The performance of OAO is, however, due to the high number of models it builds. If bagging is used to build the same number of classes as OAO, it is shown that the bagging technique performs better. This can be applied to both OAO and AAO.

Table 1. Average accuracies (%) of all the approaches with the four algorithms over the 25 datasets.

	OAA	OAO	AAO	AAOvs OAA	AAOvs OAO
J48	77.32 ± 1.81	79.65 ± 2.79	78.02 ± 2.68	80.69 ± 1.77	81.31 ± 1.47
NaiveBayes	73.49 ± 3.04	73.40 ± 2.81	73.12 ± 2.57	73.63 ± 2.26	73.66 ± 2.31
SMO	66.67 ± 2.71	77.10 ± 2.69	77.98 ± 2.31	77.73 ± 1.77	77.88 ± 2.06
Logistic	76.55 ± 2.60	78.90 ± 2.43	77.75 ± 2.60	78.42 ± 1.71	78.79 ± 2.34

Table 2. Summary comparisons (#wins/#ties/#loses) among the methodologies with the four algorithms over the 25 datasets.

		OAO	AAO	AAOvs OAA	AAOvs OAO
J48	OAA	1/12/12	3/16/6	1/11/13	
	OAO		10/15/0		1/16/8
NaiveBayes	OAA	1/20/4	2/20/3	1/20/04	
	OAO		1/24/0		1/24/0
SMO	OAA	1/11/13	0/11/14	0/12/13	
	OAO		2/22/1		0/21/4
Logistic	OAA	1/14/10	3/15/7	2/14/9	
	OAO		9/16/0		4/18/3
Summation	**OAA**	**4/57/39**	**8/62/30**	**4/57/39**	
	OAO		**22/77/1**		**6/79/15**

Future work will be done in order to evaluate the performance of variants of OAA and OAO, such as weighted OAA and OAO. OAA creates imbalanced training datasets since it merges the examples of the rest classes into one. Is it a potential reason why OAA performs worse than AAO? Another interesting topic is to investigate the potential problems for OAA and OAO if the original datasets are imbalanced. Besides, we will further study the possibility of extending the existing approaches of reducing multiclass classifications into multi-label classifications.

References

Hsu, C.W., Lin, C.W. 2002. A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Tran. On Neural Networks* 13(2):415-425.

Rifkin, R., Klautau, A. 2004. In Defense of One-vs-All Classification. *Journal of Machine Learning Research* 5:101-141.

Sulzmann, J., F¨urnkranz, J., and H¨ullermeier, E. 2007. On Pairwise Naive Bayes Classifiers. *In Proceedings of the18th European Conference on Machine Learning*, 371-381. Warsaw, Poland.

Tsujinishi, D., Koshiba, T., and Abe, S. 2007. Why Pairwise is Better Than One-Against-All or All-at-Once. *IEEE International Joint Conference*, 693-698.