

The Performance of Single Classifier, Ensemble without Diversity and Ensemble with Diversity

Imad Alhadi Ganan
Singidunum University

Vladislav Miskovic
Singidunum University

ABSTRACT

The performance of ensemble depends on the single classifiers chosen. Diversity in ensemble could be a factor that may influence the results or the performance of ensemble. In this study we have employed bagging and boosting as ensemble classifier, DECORATE to tackle diversity in ensemble. We have chosen random forest, random tree, j48 and j48 grafts mainly as a base classifier for the ensemble methods. The empirical evidence has shown that Boosting algorithm without diversity do not improve the test performance of the single classifier.

Keywords

Bagging, boosting, ensemble, diversity, DECORATE

1. INTRODUCTION

Bagging

Bootstrap aggregating, often abbreviated as bagging [1] and known as one of the earliest ensemble algorithms [2]. It involves having each model in the ensemble vote with equal weight. In order to promote model variance, bagging trains each model in the ensemble using a randomly drawn subset of the training set [3]. As an example, the random forest algorithm combines random decision trees with bagging to achieve very high classification accuracy [4].

Kulkarni & Kelkar (2014), conducted research on Ensemble Techniques of Bagging, Boosting and Ada-Boost. The experiment is observed that the performance of ensemble classifiers is better than individual classifiers and bagging often perform better [5].

Fraz et al., (2012) conducted study on retinal vessel segmentation using ensemble classifier of bagged decision tree based on supervised classification using an ensemble classifier of bagged decision trees. The performance, effectiveness and robustness along with its simplicity and speed in training as well as in classification, make this ensemble based method a suitable tool to be integrated into a complete retinal image analysis system for clinical purposes and in particular for large population studies [6].

Ye & Suganthan (2013), conducted an empirical comparison of bagging-based ensemble classifiers. The comparison has been done empirically four bagging-based ensemble classifiers, namely the ensemble adaptive neurofuzzy inference system (ANFIS), the ensemble support vector machine (SVM), the ensemble extreme learning machine (ELM) and the random forest. The empirical results also showed that the bagging is the most favorable ensemble classifier among them [7].

Boosting

The boosting algorithm does not create each base classifier independently. Instead, the classifiers are created sequentially where the next base classifier assigns more weights to the mistakes that the previous classifier made and classification is based on weighted base classifiers [3]. Adaptive boosting (Adaboost) is one of the most popular boosting algorithms [8].

García-Pedrajas & de Haro-García (2014), conducted a study on Boosting instance selection algorithms and proposed approach of opening a new field of research in which to apply the many techniques developed for boosting classifiers, for instance selection and other data reduction techniques such as simultaneous instance and attribute selection. Using 60 datasets for balanced problems and 45 datasets for class-imbalanced problems, the experiments reported show a clear improvement in several state-of-the-art instance selection algorithms using the proposed methodology [9].

Bagging and boosting are two well-known methods of developing classifier ensembles. It is generally agreed that the ensemble methods that utilize the boosting concept can improve the quality and robustness [10].

Rashedi & Mirzaei (2013), introduced a new boosting based hierarchical ensemble method called Bob-Hic. This method is utilized to create a consensus hierarchical on a dataset, which is helpful to improve the accuracy. Bob-Hic includes several boosting iterations. In each iteration, first, a weighted random sampling is performed on the original dataset. Experiments on popular synthetic and real datasets confirm that the proposed method improves the results of simple classifiers. In addition, our experimental results confirm that this method provides better consensus quality compared to other available ensemble techniques [10].

Diversity

Diversity is one of the most challenging issue in classifiers [11]. Various studies have provided theoretical and empirical evidence that diversity is a key factor for yielding satisfactory accuracy-generalization performance with classifier ensembles. As a consequence, in the last years, several approaches for bagging and boosting reasonable levels of diversity have been investigated [12]. To address the diversity DECORATE is proven to be a better algorithm [13]. Hence, this research focus to study the previous works on DECORATE.

2. DATA ANALYSIS TOOLS

Microsoft Excel has been used to keep the records of the data and Special Package for Social Scientist (SPSS) has been employed to analysis the data for this study.

3. SINGLE CLASSIFIER

Table 1: Summary of training and test performance of Single Classifier with respect to the algorithm and data set.

Algorithm	Weblog filtered		Spam		Weblog	
	Train	Test	Train	Test	Train	Test
Random Forest	99.35	86.80	99.86	70.10	99.56	75.86
Random Tree	100	83.02	99.86	70.10	100	72.41
J48	92.26	79.25	98.14	68.04	84.11	76.25
J48 Graft	92.26	79.26	98.14	64.95	84.11	76.62

The best two methods that yield the highest test performance for weblog filtered data set are Random Forest (86.80%), followed by LMT (84.90%). For Spam, the highest test performance is obtained by LMT (74.22%), followed by both Random Forest and Random Tree (70.10%). As for weblog data set, once again LMT (79.31%) supersedes other methods (Fig. 5.1). In general, for single decision tree classifier, LMT has shown the best test performance across the data sets.

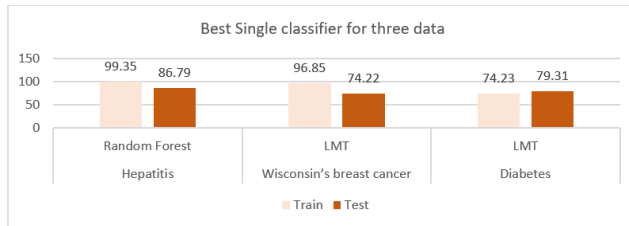


Fig 5.1: Summary of best Single Classifier with respect to the data set

4. ENSEMBLE CLASSIFIER

As mentioned earlier, TWO (2) types of ensemble methods were utilized in the experiments. The results for Boosting method are presented first, followed by the bagging methods.

Boosting without Diversity

The results shown in Table 5.2 indicate that for weblog filtered data set, LMT (84.90%) obtains the highest test accuracy, followed by Random Tree (83.02%). Like single classifier, the highest test performance is shown by LMT (95.79%), followed Random Tree (70.10%). However, Random Tree suffers from over-fitting since the gap between training and test performance is reasonably high (27.8%). Although, J48, J48 Graft and Random Forest obtain the first and second highest of test performance for Weblog data set, all of these methods are affected by over-fitting. Hence, none of the method is successful in improving the test performance of the decision tree.

Table 2: Summary of best Boosting algorithms without diversity with respect to data set

Algorithm	Weblog filtered		Spam		Weblog	
	Train	Test	Train	Test	Train	Test
AdaBoost: Random Forest	100	74.34	97.90	64.95	100	73.18
AdaBoost: Random Tree	100	83.02	97.90	70.10	100	69.35
AdaBoost: J48	100	75.47	97.20	65.97	97.90	77.78
AdaBoost: J48 Graft	100	77.36	94.06	63.92	100	73.18
AdaBoost: LMT	100	84.90	100	95.79	100	71.26

Boosting with Diversity

The test results for Boosting algorithms with diversity are shown in Table 5.3. Clearly, the results indicate J48 Graft and LMT have obtained the first and second highest test performance for weblog filtered and weblog data set. However, Random Forest and Random Tree yield test performance more than 96%.

Table 3 : Summary of best Boosting algorithms with diversity and with respect to data set

Algorithm	Weblog filtered		spam		Weblog	
	Train	Test	Train	Test	Train	Test
DECORATE AdaBoost: Random Forest	100	84.90	99.86	96.22	99.74	77.01
DECORATE AdaBoost: Random Tree	100	75.47	100	96.22	100	73.56
DECORATE AdaBoost: J48	100	83.02	100	95.80	100	76.25
DECORATE AdaBoost: J48 Graft	98.06	86.79	98.57	95.38	86.59	80.84
DECORATE AdaBoost: LMT	100	86.79	100	95.79	100	78.16

Bagging without Diversity

The experimental results exhibited in Table 5.4 indicates that Random Forest (86.79%), Random Tree (86.79%), and LMT (84.9%) obtain the first and second highest test performance for weblog filtered data set. LMT method obtains the highest test performance for Spam (95.79%), followed by J48 (73.20%). For Weblog data set, Random Tree (77.78%) and J48 (76.63%) exhibit the highest and second highest of test performance. Across the board, THREE (3) boosting methods have shown high test performance namely, Random Tree, J48 and LMT.

Table 4: Summary of best Boosting methods with respect to data set

Algorithm	Weblog filtered		Spam		Weblog	
	Train	Test	Train	Test	Train	Test
Bagging: Random Forest	98.06	86.79	95.81	67.01	97.66	76.26
Bagging: Random Tree	98.71	86.79	96.85	65.97	98.31	77.78
Bagging: J48	94.19	81.13	81.47	73.20	94.92	76.63
Bagging: J48 Graft	94.19	81.13	81.12	70.10	94.67	76.25
Bagging: LMT	100	84.9	100	95.79	100	71.26

Bagging with Diversity

The analysis of Bagging without diversity indicates that Bagging: LMT obtain the highest test performance. The test performance of Bagging with diversity is presented in Table 5.6. TWO (2) best algorithm have been selected for each dataset. For weblog filtered data set, both DEORATE Bagging: Random Tree and DECORATE Bagging: LMT obtained 86.79%. The test performance for Spam is higher than weblog filtered data set for which Random Forest and Random Tree performance are better than others (Table 5.5). Among the Bagging with diversity, both of these algorithms are the best, even when their results are compared with weblog filtered and weblog. For weblog data set, Random Tree and J48 Graft test performance is almost 80%.

Table 5: Summary of best Bagging algorithms with diversity and with respect to data set

Algorithm	Weblog filtered		Spam		Weblog	
	Train	Test	Train	Test	Train	Test
DECORATE Bagging: Random Forest	99.35	84.91	99.71	96.22	98.83	78.93
DECORATE Bagging: Random Tree	100	86.79	100	96.64	99.87	80.08
DECORATE Bagging: J48	98.06	83.02	99.14	95.80	97.27	79.69
DECORATE Bagging: J48 Graft	98.06	84.91	99.14	96.22	97.14	80.84
DECORATE Bagging: LMT	100	86.79	99.57	95.79	96.87	80.07

5. FINDINGS

The summary of the findings is presented based on the data set used in the study. From previous analysis, the test performance of the selected algorithms of the single classifier versus bagging (without and with diversity) and boosting (with and without diversity) are shown in Fig. 5.2

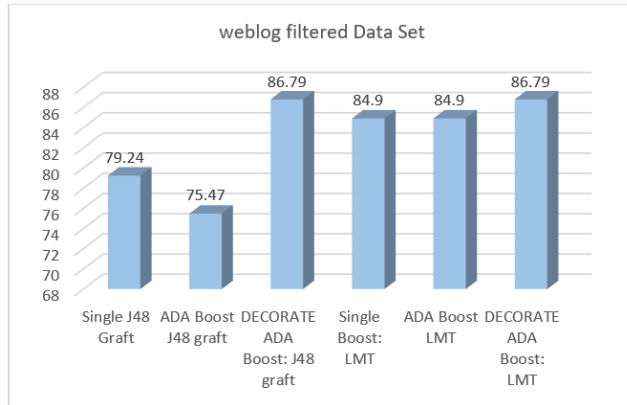


Fig 5.2. Summary of best Bagging algorithms with diversity and with respect to data set

Fig 5.2 reveals that J48 Graft and LMT without diversity are unable to improve the test performance. When diversity is considered in the decision trees, both J48 Graft and LMT have improved the test performance by 7.55% and 1.89% respectively. Based on the difference of test performance with single classifier, J48 Graft is a better ensemble classifier than LMT for weblog filtered data.

Following previous approach in performing the analysis, the test performance for the algorithms with respect to Spam is shown in Fig. 5.3.

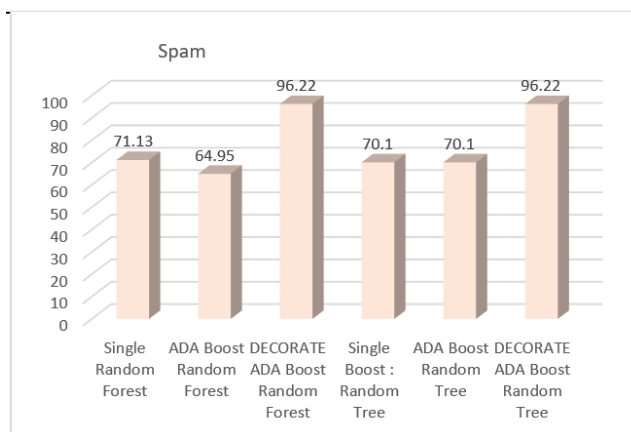


Fig 5.3. Summary of best Boosting algorithms with diversity and with respect to data set

Results shown in Fig. 5.3 indicates that Boosting algorithms without diversity (in this case, Random Forest and Random Tree) do not improve the test performance of single classifier of Random Forest and Random Tree. The bar charts also reveal that Boosting algorithms with diversity improve the test performance to 96.22%. For Random Forest, boosting algorithm with diversity is able to increase the test performance by 25.09% while Random Tree 26.12%. Based on the increment of test performance, Random Tree is a better ensemble classifier for Spam data set.

The summarized results of the test performance of the single classifiers versus boosting and bagging algorithms are shown in Fig. 5.4.

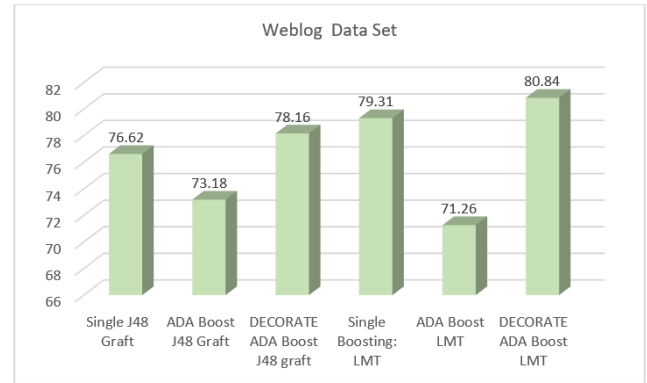


Fig 5.4. Summarized results of the test performance of the single classifiers versus boosting and bagging algorithm

Note that the performance of Boosting algorithms without diversity do not improve the performance of a single classifier. In fact, this finding is the same regardless of data set used in the experiments. The results shown in Fig. 5.4 also shows that J48 Graft and LMT have improved the test performance of the single classifier. The difference in test performance of J48 Graft versus single classifier is 1.54% while LMT versus single classifier is 1.53%. In this case, both ensemble classifiers are considered the best decision tree classifiers for weblog data set.

The results of Bagging algorithms for single classifier, Boosting and Bagging algorithms are shown in Fig. 5.5.

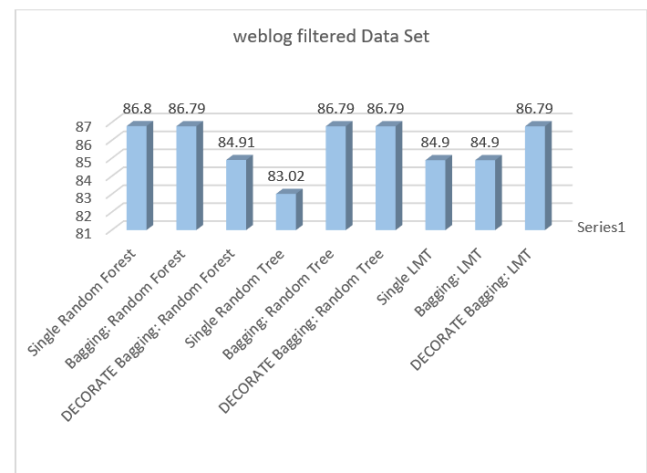


Fig 5.5 Summary of best Boosting algorithms with diversity and with respect to data set

Among Random Forest, Random Tree and LMT for weblog filtered data set, only the test performance of DECORATE Bagging: Random Forest is lower than the performance of Single Random Forest (decreases by 1.89%). Fig. 5.5 also indicates that both DECORATE Bagging: Random Tree and DECORATE Bagging: LMT improve the test performance of respective single classifiers by 3.77% and 1.89%.

The test performance results for the selected classifiers with respect to Spam data set are shown in Fig. 5.6. Although the test performance for Bagging: Random Tree is lower than its single classifier, the performance of DECORATE Bagging:

Random Tree as well as DECORATE Bagging: LMT increases by 26.54% and 21.57% respectively.

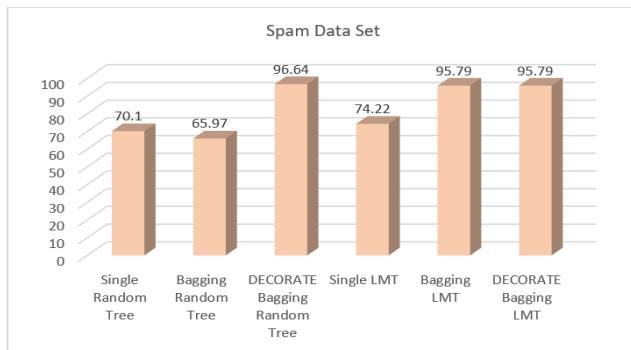


Fig 5.6. Summary of best Boosting algorithms with diversity and with respect to data set

The test performance of the single and ensemble classifiers for weblog data set are shown in Fig. 5.7. The results reveal that Bagging algorithms are able to improve the test performance of single classifiers of Random Tree, J48 and J48 Grafts. Similar observations are also identified for Bagging algorithms with diversity. In fact, Bagging algorithms with diversity exhibits better performance than Bagging algorithms without diversity.

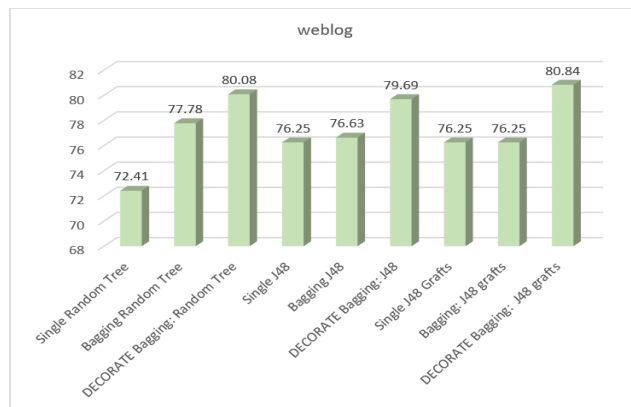


Fig 5.7. Summary of best Boosting algorithms with diversity and with respect to data set

The difference between the test performance of single classifier and its respective bagging algorithms is illustrated in the graph as shown in Fig. 5.8

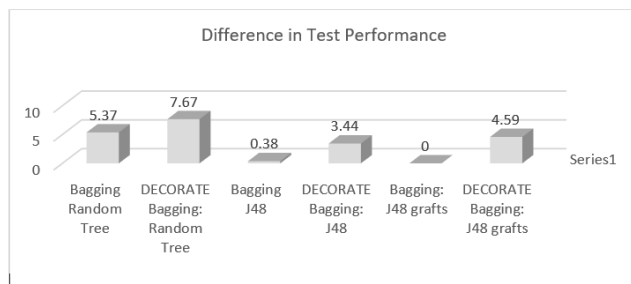


Fig 5.8. Summary of best Boosting algorithms with diversity and with respect to data set.

The results exhibited in Fig. 5.8 indicate that Bagging algorithms with diversity has inevitably improve the test performance of the single classifiers.

6. CONCLUSION

As a conclusion, the empirical evidence has shown that Boosting algorithm without diversity do not improve the test performance of the single classifier. The analysis also indicate that boosting algorithm without diversity has negative impact on the test performance. Comparing Boosting algorithms with Bagging, the analysis has proved that Bagging algorithms are able to increase the test performance of single classifiers. In addition, the empirical analysis also indicate that Boosting and Bagging algorithms when coupled with diversity show a higher increment in test performance.

7. REFERENCES

- [1] M. Kusum and M. Rupali, "A REVIEW ON VARIOUS CLASSIFICATION ALGORITHMS FOR AN INCREMENTAL SPAM FILTER," *Int. J. Appl. or Innov. Eng. Manag.*, vol. 2, no. 11, pp. 325–331, 2013.
- [2] C. Zhang and Y. Ma, *Ensemble Machine Learning: Methods and Applications*. New York: Springer New York, 2012.
- [3] D. Che, Q. Liu, K. Rasheed, and X. Tao, "Software Tools and Algorithms for Biological Systems," *Adv. Exp. Med. Biol.*, vol. 696, no. 1, pp. 191–199, 2011.
- [4] D. Darwish, "Data Mining: Concepts, Models, Methods, and Algorithms," *Int. J. Comput. Sci.*, vol. 10, no. 4, pp. 103–111, 2013.
- [5] S. Kulkarni and V. Kelkar, "Classification of Multispectral Satellite Images Using Ensemble Techniques of Bagging , Boosting and Ada- Boost," pp. 253–258, 2014.
- [6] M. M. Fraz et al., "Retinal vessel segmentation using ensemble classifier of bagged decision trees," *IET Conf. Image Process. (IPR 2012)*, pp. B9–B9, 2012.
- [7] R. Ye and P. N. Suganthan, "Empirical Comparison of Bagging-based Ensemble Classifiers," *Int. J. Inf. Technol. Converg. Serv.*, vol. 1, no. 1, pp. 917–924, 2013.
- [8] R. E. Schapire, "Explaining AdaBoost," Princeton University, 2010.
- [9] N. García-Pedrajas and A. de Haro-García, "Boosting instance selection algorithms," *Knowledge-Based Syst.*, vol. 67, pp. 342–360, Sep. 2014.
- [10] E. Rashedi and A. Mirzaei, "A hierarchical clusterer ensemble method based on boosting theory," *Knowledge-Based Syst.*, vol. 45, pp. 83–93, Jun. 2013.
- [11] B. Krawczyk and M. Woźniak, "Diversity measures for one-class classifier ensembles," *Neurocomputing*, vol. 126, pp. 36–44, Feb. 2014.
- [12] D. S. C. Nascimento, A. L. V. Coelho, and A. M. P. Canuto, "Integrating complementary techniques for promoting diversity in classifier ensembles: A systematic study," *Neurocomputing*, vol. 138, pp. 347–357, Aug. 2014.
- [13] J. Stefanowski and M. Pachocki, "Comparing Performance of Committee Based Approaches to Active Learning," *Recent Adv. Intell. Inf. Syst.*, vol. 1, no. 1998, pp. 457–470, 2013.