Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence

An Investigation of Sensitivity on Bagging Predictors: An Empirical Approach

Guohua Liang
The Centre for Quantum Computation & Intelligent Systems, FEIT,
University of Technology, Sydney, NSW 2007 Australia

Abstract

As growing numbers of real world applications involve imbalanced class distribution or unequal costs for mis-classification errors in different classes, learning from imbalanced class distribution is considered to be one of the most challenging issues in data mining research. This study empirically investigates the sensitivity of bagging predictors with respect to 12 algorithms and 9 levels of class distribution on 14 imbalanced data-sets by using statistical and graphical methods to address the important issue of understanding the effect of varying levels of class distribution on bagging predictors. The experimental results demonstrate that bagging NB and MLP are insensitive to various levels of imbalanced class distribution.

Introduction

Imbalanced class distribution refers to the training samples that are non-uniformly distributed with unequal cost among classes. Typically, in a binary classification, the minority class and majority class are regarded as a positive class and a negative class, respectively. A growing number of researchers focus on solving imbalanced class distribution problems in real world applications in a variety of domains, such as, credit card fraud detection, medical diagnosis, and biological data analysis.

(Weiss and Provost 2001) evaluated the effect of class distribution on classifier learning by assessing the relationship between training class distribution and performance of C4.5 learner to draw their conclusions as to which distribution is best for training based on two evaluation measures: error rate and Area Under the ROC curve (AUC). They however did not evaluate which learner is sensitive when the levels of class distribution vary. Moreover, imbalanced class distribution or the unequal cost of mis-classification errors often causes learning algorithms to perform poorly on the minority class; the mis-classification error rate cannot distinguish the accuracy of the minority class (He and Garcia 2009; Weiss and Provost 2001). Two evaluation measures, Receiver Operating Characteristic (ROC) Curve and Geometric mean (G-mean) of the accuracy rates for both positive and negative classes are therefore adopted for this study.

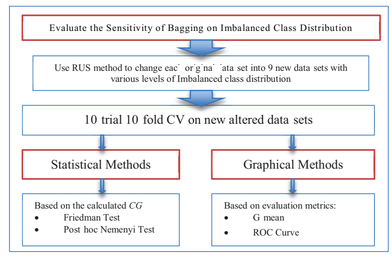

Figure 1: Designed framework

Bagging (Breiman 1996) uses bootstrap samples to build a set of classifiers to form a prediction model; the final decision is aggregated by a majority vote of the predictions of the individual classifiers in the ensemble. It has been applied to a variety of real world applications. Our previous studies investigated the performance of bagging predictors in natural class distribution (Liang, Zhu, and Zhang 2011); but we did not investigate the sensitivity of bagging, so it is unclear which bagging predictors are sensitive to various levels of class distribution.

Our main contribution is to conduct a intensive evaluation of the sensitivity of bagging predictors to understand the effect of varying levels of class distribution. The experimental results provide a useful guide for data mining practitioners to understand the sensitivity of the bagging predictors and to solve imbalanced class distribution problems for their applications.

Designed Framework

Figure 1 represents designed framework to investigate the sensitivity of bagging predictors as follows: (1) a random under-sampling (RUS) method is used to change original data-set into 9 new data-sets with different imbalanced class distribution, (2) a 10-trial 10-fold cross-validation (CV) is performed on each altered data-set, (3) statistical methods are applied to draw validated conclusions, and (4) two evaluation metrics are adopted to further visualize the sensitivity of bagging predictors.

Statistical Method: the Friedman test with the Post-hoc Nemenyi test (Demšar 2006) are used to compare multiple bagging predictors: Step1. calculate the changed G-mean (CG) between two adjacent levels of class distribution; Step 2. using CG to rank bagging predictors on each data-set (lowest value ranked as 1); Step 3. the Friedman test is used

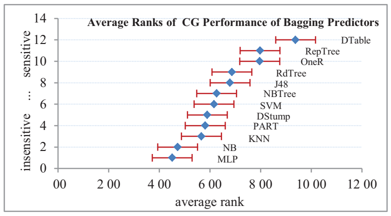

Figure 2: Comparison of all bagging predictors against each other with the Nemenyi test, where the x-axis indicates the average rank of the bagging predictors, the y-axis indicates the ascending order of the average rank of CG performance, and the horizontal bars indicate the CD.

to obtain the average rank of CG among 12 bagging predictors over all data-sets; Step 4. Post-hoc Nemenyi test is used to calculate "critical difference" (CD).

Evaluation Metrics: Two evaluation metrics are used to visualize the performance of selected bagging predictors to further examine the statistical results:

(1) **A ROC curve** is used to plot the False Positive Rate (FPR), and True Positive Rate (TPR) on the x-axis and y-axis, respectively. The point (0,1) stands for "perfect point". In the ROC space, one point is better than another if it is close to the "perfect point" (Provost and Fawcett 1997). In this study, a 10-trial 10-fold cross-validation is performed on each altered data-sets to obtain nine pairs of (FPR, TPR) to form a ROC curve for each original data-set, so a ROC curve is used to represent the performance of each bagging predictor at 9 different levels of class distribution.

(2) $G-mean$ monitor the accuracy rates of both TPR and True Negative Rate (TNR) for the minority and majority classes, respectively (Ng and Dash 2006).

$$G-mean = \sqrt{TPR * TNR} \quad (1)$$

Experimental Results

Figure 2 presents comparison of all bagging predictors against each other with the Nemenyi test, where the x-axis indicates the average rank of CG performance of the bagging predictors; the y-axis indicates the ascending order of the average rank of CG performance, which represents bagging predictors from insensitive to sensitive; and the horizontal bars indicate the CD. Groups of bagging predictors that are not significantly different (at $p = 0.05$), when the horizontal bars are overlapped. The results indicate that the group of bagging predictors, Multi-layer Proceptron (MLP) and Naïve Bayes (NB) is the most insensitive predictors, that means the performance of those bagging predictors change gradually between adjacent levels of class distribution, so they are insensitive to varying levels of class distribution; while the group of bagging predictors, Decision Table (DTable), RepTree and OneR is the most sensitive predictors, that means the performance of those bagging predictors change sharply between adjacent levels of class distribution, so they are sensitive to varying levels of class distribution. The ranking order of CG performance of those sensitive bagging predictors is therefore greater than those insensitive bagging predictors, when the levels of class distribution change. There are statistically significant differences between the two groups.

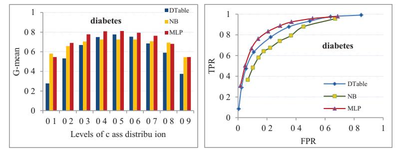

Figure 3: Comparison of ROC curve and $G-mean$ among three selected bagging predictors on diabetes data-set.

Figure 3 presents graphical comparisons of G-mean and ROC curve of three selected bagging predictors at 9 levels of class distribution on diabetes data-set in two subfigures. When the levels of class distribution are changed, the G-mean of bagging predictors, NB and MLP change gradually, while bagging predictors, DTable changes sharply. The ROC curves indicate that Bagging predictors, MLP and NB have more points close to the "perfect point" and better performance than bagging DTable at same level of imbalanced class distribution, eg., at 10%, 20%, 80%, and 90% levels imbalanced class distribution. The graphical observations confirm that bagging predictors, MLP and NB are insensitive to various levels of class distribution and perform relatively well with extremely imbalanced class distribution. The graphical results therefore are consistent with the statistical results.

Conclusion

This paper empirically investigated the sensitivity of bagging predictors with respect to various levels of imbalanced class distribution. Both graphical observations and statistical results demonstrated that the group of bagging predictors, MLP and NB is insensitive to different levels of imbalanced class distribution.

References

Breiman, L. 1996. Bagging predictors. *Machine Learning* 24(2):123–140.

Demšar, J. 2006. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research* 7:1–30.

He, H., and Garcia, E. 2009. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering* 21(9):1263–1284.

Liang, G.; Zhu, X.; and Zhang, C. 2011. An empirical study of bagging predictors for different learning algorithms. In *Proceeding of the 25th AAAI Conference on Artificial Intelligence, AAAI 2011*, 1802–1803.

Ng, W., and Dash, M. 2006. An evaluation of progressive sampling for imbalanced data sets. In *Proceedings of the 6th IEEE International Conference on Data Mining Workshops, ICDM Workshops 2006*, 657–661. IEEE.

Provost, F., and Fawcett, T. 1997. Analysis and visualization of classifier performance with nonuniform class and cost distributions. In *Proceedings of AAAI-97 Workshop on AI Approaches to Fraud Detection & Risk Management*, 57–63.

Weiss, G., and Provost, F. 2001. The effect of class distribution on classifier learning: an empirical study. Technical report, Rutgers Univ.