

Improving ADA-boost as a Popular Ensemble in Classification Problems

M. Sravan Kumar Reddy, K.E. Naresh Kumar, Dharmendra Singh Rajput

Abstract: In Data Mining, several classification algorithms are used to perform classification based on single learner but classification accuracy is not in an effective manner. To increase the accuracy of the classification then multiple learners are combined to get better results. The multiple learners are trained and combined into an ensemble. The ensemble can increase generalization ability and robustness [3]. Based on the advantage of ensemble, the ensemble classification is a major concern in research directions of machine learning. Another importance of ensemble is that it is much stronger than single base learner to produce accurate hypothesis. The ensembles are divided into homogeneous or heterogeneous, dependent or independent ensembles. The dependent ensemble methods like boosting and AdaBoost algorithms are promisingly provide an accurate hypothesis. Finally, AdaBoost can be a better classifier ensemble to generate accurate results.

Index Terms: AdaBoost, ensemble classification, classifier ensemble, accuracy, bagging

I. INTRODUCTION

An Ensemble consists of multiple learners to solve the similar problem. These multiple learners are base learners in which some of them are weak learners but the ensemble can improve learners from weak to strong and strong learners are more efficient than random guessing. The construction of an ensemble can be performed in two steps. First, a group of classifiers are to be trained in sequence or parallel manner. These base learners are training data by base learning algorithms like decision trees, back propagation and other machine learning algorithms. Second, base learners hypothesis are combined to generate an accurate hypothesis. The combination of hypothesis can be performed by popular methods like majority voting, cross fold validation and averaging.

There is no historical proof for determining starting point of ensembles in machine learning. Because, the basis for ensemble is Occam's razor which derives simple hypothesis to complex one's with scientific proof of empirical observations. The Greek philosopher Epicurus introduced principle of multiple explanations [1] consists of consistent hypothesis with empirical derivations. The ensemble learning can be contributed into various disciplines based on its implementation. If an ensemble is defined by combining Classifiers used in pattern recognition. If an ensemble is specified by boosting weak learners used in machine learning paradigm. Otherwise, an ensemble is derived by mixture of experts can be used in neural networks community.

Revised Manuscript Received on July 22, 2019.

M. Sravan Kumar Reddy, Assistant Professor, CSE Dept, RGM CET, Nandyal, Kurnool, India

K.E. Naresh Kumar, Assistant Professor, CSE Dept, RGM CET, Nandyal, Kurnool, India

Dr. Dharmendra Singh Rajput, Associate Professor, SITE, VIT University, Vellore, India

In 1990's Ensemble methods become popular learning paradigm. The key implementation of ensemble is to combine individual opinions of various classifiers to make a final decision by combining them. The idea of ensemble is implementing in the period of 1970's within supervised learning algorithms. To understand its importance the study of Condorcet Jury Theorem derives a solution to the question of "Can a collection of weak classifiers create a single strong one?" with accurate prediction. Later, English Philosopher Sir Francis Galton Conceived Standard Deviation and Correlation. James Micheal Surrowiecki, written a book entitled "The wisdom of crowds: Why the many are smarter than few and How Collective Wisdom shapes Business, Economics, Societies and Nation" states that a strong learner as crowd become wise decision with criteria of Diversities of Opinion, Opinion Independence, Opinion Decentralization and Opinion Aggregation.

In the past, 1977 Tukey proposed linear regression models to perform combining of two models. After two years, 1979 Dasarathy and Sheela proposed a partition dataset combines two or classifiers. The major improvement takes place in 1990's Hansen and Salamon introduced an ensemble consists of configure networks to increase accuracy of stronger one. Finally, in the same time of 1990 Scaphire laid the foundation for AdaBoost and its practical implementation derived in 1995 by another scientist Yoav Freund.

II. TYPES OF ENSEMBLES

An Ensemble framework consists of major building blocks to perform the classification are:

- Training Set
- Base Classifier/Learner
- Diversity Generator
- Combiner

The Ensemble Training set consists of labeled dataset described in various formats and its instances are defined in attributes based vectors. The base classifier or learner makes use of training set to generalize the relationship between test data and training data. Diverse Generator generates the diverse classifiers. Finally, Combiner combines the classification of various classifiers.

There are several alternative types of ensembles are defined based on type of base learner used for combination [2] and type of framework used for combination [3]. According to type of base learner used for ensemble method is as similar type then it leads to Homogeneous Ensembles. Otherwise, if an ensemble uses different types of learning algorithms for base learners determines Heterogeneous Ensembles.

Based on type of framework used for ensemble can divides into dependent and independent frameworks. The independent framework in which set of classifiers



can be executed in parallel or independent but the final decision can be derived by combining them.

The dependent frameworks in which set of classifiers are executed in sequential manner by maintaining the prior knowledge of previous classifier the errors can be corrected in the next classifier. It performs number of iterations for learning a classifier in next iterations. In [5] the dependent learning framework may again divided into two approaches.

- i. Incremental Batch Learning
- ii. Model Guide Instance Selection

In incremental Learning approach the classification performed in each iteration used as prior knowledge used for constructing next classifier. The classifier produced at final iteration can be used as final classifier to perform the classification.

In Model Guide Selection, Classifiers constructed used in previous iterations used for changing the training set and it allows to learn for misclassified instances also.

The example of incremental batch learning algorithm is Logistic Regression, Decision Trees and Support Vector Machines [7]. To model guided instance selection the best example is Boosting. It can also referred as Arcing (Adaptive Resampling and Combining).

The boosting can be used to create strong learners from set of weak learners. The general idea of boosting is to perform the correction of mistakes from one classifier on subset of Dataset, based on which the next classifier can be constructed.

The word boosting used from the answer given to questions posed by Kearns and Valiant [6][8]. In their paper, the model exhibits discrete hierarchy of desired accuracy can be achieved by proposed efficient algorithm converts poor hypothesis into strong hypothesis. Kearns and Valiant proposed three algorithms for converting weak learners to strong learners using filters. Boosting can be performed repeatedly to generation of weak learners are finally combined into a single strong learner. Later, in 1990 Schapire proposed a Probable Approximately Correct (PAC) model of learning to identify a strong learner polynomial algorithm with low error and high confidence and also to achieve high accuracy [9].

To extending the boosting algorithm, the new algorithm, proposed to make use of prior knowledge of accuracies of weak learners for generating weighted majority strong learner called AdaBoost. The Adaboost (Adaptive Boosting) refers to popular ensemble algorithms improves boosting algorithm performance using an iterative process. It was proposed by Freund and Scaphire in 1996. It allows weights to an individual classifier and performing additional iterations to create a strong classifier. The basic Adaboost algorithm was introduced to perform binary classification problems. Adaboost estimates large margin classifier like SVM to perform linear classification.

Later, Freund and Schapire introduced two versions of Adaboost algorithms are Adaboost.M1 and Adaboost.M2 for handling multi class classification problems. In [10] a simple arcing algorithm, Arc-x4 demonstrated the working functionality of Adaboost performed by Adaptive Resample but not use of weigh functions. Due to these major improvements of Adaboost, several variants of Adaboost are introduced. Ensembles can also be divided into generative and Non Generative Ensembles [26].

Generative Ensembles generates set of base learners to improve accuracy and diversity of base learners. Non

generative Ensembles are not generating any new base learners but combine with output of base learners by majority voting. Examples of Non-generative ensembles are majority voting, Naïve Bayes Rule, Meta Learning. Examples of generative ensembles are Bagging, Boosting, Arcing etc.

III. LITERATURE WORK ON ADABOOST

Adaboost can introduces the process of adjusts adaptively [11] to the errors of weak hypotheses returned by weak learner combines by summing their Probabilistic Predictions to strong learner. Adaboost algorithm is the first practical boosting algorithm introduced by Yoav Freund and Robert E. Scaphire. It produces a single composite classifier by repeated executions of weak learning algorithms on various distributions of training dataset. Boosting compared to Bagging by performing sampling of training dataset with replacement in bagging but in boosting modifies the distribution of dataset on each round by voting multiple classifiers. Based on [12] Boosting performs better than bagging when weak learning algorithm generates simple classifiers.

The various improvements of Adaboost algorithm implementation are basic Adaboost algorithm will increase the accuracy of learning model. Adaboost assigns weight for each training set with distribution. It generates 'm' number of classifiers at 'm' number of rounds to perform classification. At each iteration, the error can be calculated to reassign increased weights in next iteration for incorrect classifier tuples. Otherwise, weights are decreased for correctly classified tuples. The normalized values can be used to perform the final classification [13]. When compared boosting and bagging, boosting focuses on misclassified tuples to risk over fitting but bagging does not. Sometimes, Boosting performs less accuracy than bagging rather than single model. But, Boosting achieves greater accuracy than bagging in various cases.

Adaboost achieves by minimizing exponential loss function. It applies Resampling by Boosting with Restart terminates boosting procedure before the completion of 'n' number of rounds. Adaboost can also fits an additive model using surrogate loss function leads to Logit Boost performs log loss. Later, the variant of Adaboost which finds weak learner with smallest error in weight distribution by reaching optimality forms LP Boost algorithm. It can regularizes entropy and exhibits better performance than Adaboost. The original Adaboost designed sensitive to noise. Mada Boost [14] is another variant of Adaboost used for handling noise by depressing large instance weights. Later, another variant of Adaboost is Filter Boost [15] adopts log loss function in place of exponential loss function.

P- Adaboost [16] algorithm is a distributed version of Adaboost by changing its implementation in sequential for limited steps and trained in parallel. Adaboost-r is another variant of Adaboost implements that r is a parameter used for selection of weak classifiers used for classification. It does not generate a strong classifier.

IV. RECENT ADVANCES OF ADABOOST

Adaboost can be consistent by stopping early with limited number of iterations [17]. Several variants of Adaboost algorithms are Real Adaboost, used for



implementing classification based on degree or confidence as real probability [18]. Adaboost.NC performs the implementation of Adaboost with random sampling of training dataset with improved predictive accuracy on minority dataset [19].

The Performance of Classifier can be measured or compared with other algorithms can be performed with following criteria:

1. Precision
2. Sensitivity
3. Specificity
4. Accuracy
5. Root Mean Square Error

Accuracy can be measured by correct and incorrect classified tuples.

V. CONCLUSION

An ensemble is better than single classifier. The computational cost of an ensemble is not larger than a single classifier. Most popular ensembles of bagging, boosting and others in which AdaBoost can be considered as powerful ensemble. Adaboost can improve the accuracy of classification problems. Several variants of AdaBoost concentrate on various aspects of classification to improve the accuracy.

REFERENCES

1. Epicurus' Scientific Method. Cornell University Press, 1984. Asmis, Shadi Bartsch, and Martha Nussbaum, Seneca's Complete Works, 2010- 17, University of Chicago Press.
2. Ensemble Methods: Foundations and Algorithms, by Zhi-Hua Zhou (New York: Chapman & Hall/CRC, 2012, 222 pp.) ISBN: 978-1-4398-3003-1.
3. Lior Rokach, Ensemble based Classifiers, Artificial Intelligence Rev (2010) 33:1-39.
4. R. Polikar, Ensemble based systems in decision making, IEEE Circuits and Systems Magazine, Volume 6 Issue 3 • Third Quarter 2006.
5. Provost and Koplluri, 1997.
6. Keans, Valiant, "Thoughts on Hypothesis Boosting", 1988
7. Jesse Read , Albert Bifet , Bernhard Pfahringer , Geoff Holmes, Batch-Incremental vs. Instance-Incremental Learning in Dynamic and Evolving Data, UC3M,UoW.
8. L.Valiant, "A theory of the learnable", Comm. ACM, 27(100), 1984.
9. Robert E. Scaphire, "The Strength of weak learnability", 1990.
10. L. Brieman, Bagging Predictors, Machine Learning, 24(2):123-140, 199.
11. Yoav Freund , Robert E.Scaphire, "A Decision theoretic generalization of online learning and an application to boosting", 1995.
12. Yoav Freund , Robert E.Scaphire, Experiments with a New Boosting Algorithm , 1996.
13. Jiawei Han, Micheline Kamber, Jian Pei, "Data Mining Concepts and Techniques", 2011.
14. C. Domingo, O. Watanabe, Madaboost: A modification of Adaboost, 2000.
15. Bradley, Scaphire, 2008
16. Merler, 2007
17. Peter L. Bartlett, Mikhali Traskin, "Adaboost is Consistent", Journal of Machine Learning 8 (2007) 2347-2368.
18. Artur Ferreria, Mario Figueiredo, Boosting Algorithms: A Review of Methods, Theory and Applications.
19. Rushi, Snehlatha, Latesh Malik, Class Imbalance Problem in Data Mining: Review, IJCSN, Vol 2, Feb 2013.
20. J. Xu, Q. Wu, J. Zhang, Z. Tang, "Exploiting Universum data in AdaBoost using gradient descent," Image and Vision Computing, vol. 32, pp. 550-557, 2014.

21. B. Sun, S. Chen, J. Wang, H. Chen, "A robust multi-class AdaBoost algorithm for mislabelled noisy data," Knowledge-Based Systems, vol. 102, pp. 87-102, 2016.
22. M. Baig, M.M. Awais, E. M. El-Alfy, "AdaBoost-based artificial neural network learning," Neurocomputing, vol. 16, pp. 22 – 41, 2017.
23. L. D. Miller and L. K. Soh, "Cluster-Based Boosting," IEEE Transactions on Knowledge and Data Engineering, vol. 27, pp. 1491-1504, 2015.
24. Jasmina D. Novakovic, Alempije Veljovic, Adaboost as Classifier ensemble in Classification Problems, Infoteh-Jahorina, Vol 13, March 2014.
25. B. Lipinska, A. Wosiak, Classification of Medical Data Derived from Diagnostic devices ensembles of Classifiers, 2015.
26. Matteo Re, Giorgio Valentini, Ensemble Methods: a Review, 2001.

AUTHORS PROFILE



Mr. M. Sravan Kumar Reddy, Received B.Tech Degree in Computer Science and Information Technology from Jawaharlal Nehru Technological University, Anantapur, India, in 2005 and Master of Technology in Computer Science from Jawaharlal Nehru Technological University, Hyderabad, India, in 2010. Having an experience of 11 years in teaching and 2 years in Research. Life member of ISTE. Published 07 International Papers in various conferences and Journals.

Presently working as Associate professor in CSE Dept at R.G.M.C.E.T, Nandyal, India.



Mr. K. E. Naresh Kumar, Received B.Tech Degree in Computer Science and Engineering from Jawaharlal Nehru Technological University, Anantapur, India, in 2006 and Master of Technology in Computer Science from Visweswaraya Technological University, Belgaum, India, in 2010. Having an experience of 11 years in teaching and 2 years in Research. Life member of IE (I) and CSI. Published 07 National

08 International Papers in various conferences and Journals. Presently working as Associate professor in CSE Dept at R.G.M.C.E.T, Nandyal, India.



Dr. Dharmendra Singh Rajput working as Associate Professor in the Department of Software and Systems Engineering, School of Information Technology and Engineering, VIT University. His research area are Data Mining and Big Data Predictive Analytics.