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. 2, Issue. 5, May 2013, pg.307 – 312

REVIEW ARTICLE

REVIEW OF ENSEMBLE CLASSIFICATION

V. Chitraa[1], Dr. Antony Selvadoss Thanamani[2]
[1]Assistant Professor, CMS College of Science and Commerce, Coimbatore, Tamilnadu, India
[2]Reader in Computer Science, NGM College (AUTONOMOUS), Pollachi, Coimbatore, Tamilnadu, India

[1] vchit2003@yahoo.co.in; [2] selvdoss@yahoo.com

Abstract*— Data mining techniques like classification is effectively for used for prediction. Due to technological up gradation, the datasets which are large are distributed over different locations and classification has become a difficult task. The single classifier models are not sufficient for these types of datasets. So the recent research concentrates on combination of various classifiers and creates models. Ensemble methods combine multiple models and are useful in both supervised and unsupervised learning. This paper discusses the framework of ensemble and two types of ensemble models. A review of various algorithms of these two models is given. Combination methods which are used for combining outputs and few applications where it can be used effectively are also discussed.

Key Terms: - Bagging; Boosting; Classification; Ensemble; Prediction

I. INTRODUCTION

Ensemble is the combination of multiple models to achieve better performance than any of the individual model [7]. The base of this concept is well known Condorcet's Jury Theorem, for democracy which refers to a Jury of voters who need to make decision regarding a binary outcome. If each voter has a probability 'P' of being correct and the probability of voters being correct is L, then P > 0.5 implies L > P and L approaches 1 for all P > 0.5 as the number of voters approaches infinity.

Ensemble methods are useful in both supervised and unsupervised learning. The purpose of supervised learning is to classify patterns into a set of categories called classes or labels in a given training set. The labels of the instance in the training set are known and the goal is to construct a model in order to label new instances. Inducer is an algorithm which constructs the model and classifier is an instance of an inducer for a specific training set.

There are two approaches in ensemble model construction. They are dependent models and independent models. In dependent model the output of a classifier is used in construction of a next classifier. In independent model, each classifier is built independently and their outputs are combined in some function.

II. FRAMEWORK OF ENSEMBLE METHODS

The building blocks for ensemble classification are as follows

-A training set which can be described in different way and it is the dataset for ensemble learning. 'D' is the set of 'n' input attributes D = {d1,d2…dn } and 'v' to represent the class variable or target attribute.

-Base inducer is an induction algorithm that obtains a training set and a classifier is formed which represents the generalized relationship between the inputs attributes and the target attribute. Let I represent an inducer and M=I(S) where 'M' is a classifier which was induced by I on a training set S.

-Diversity Generator is component which is responsible for generating diverse classifiers.

-Combiner component is responsible for combing the classification of the various classifiers.

III. Ensembling Models

There are several factors that differentiate between the various ensembles methods[6]. The main factors are:

- Inter-classifiers relationship — how does each classifier affect the other classifiers? The ensemble methods can be divided into two main types: Dependent and Independent.
- Combining method — the strategy of combining the classifiers generated by an induction algorithm. The simplest combiner determines the output solely from the outputs of the individual inducers. Several combination methods exist like uniform voting, Bayesian combination, distribution summation and likelihood combination. [1]. Moreover, theoretical analysis has been developed for estimating the classification improvement.
- Diversity generator — In order to make the ensemble efficient, there should be some sort of diversity between the classifiers. Diversity may be obtained through different presentations of the input data, as in bagging, variations in learner design, or by adding a penalty to the outputs to encourage diversity.
- Ensemble size — the number of classifiers in the ensemble.

A. Dependent Model

This model is also called as sequential model. There are different approaches for dependent models by the way they train their base models. The flow of dependent model is shown in figure 1.

1) Incremental Batch Learning: The output of first iteration is given as prior knowledge to the second iteration and the process repeats 'n' times. The learning algorithm uses the current training set together with the classification of the former classification of building the next classifier. Last iteration's classifier is taken as final classifier.

2) Model Guided Instance Selection: The classifiers that were constructed in previous iteration are used for manipulating the training set for following iterations

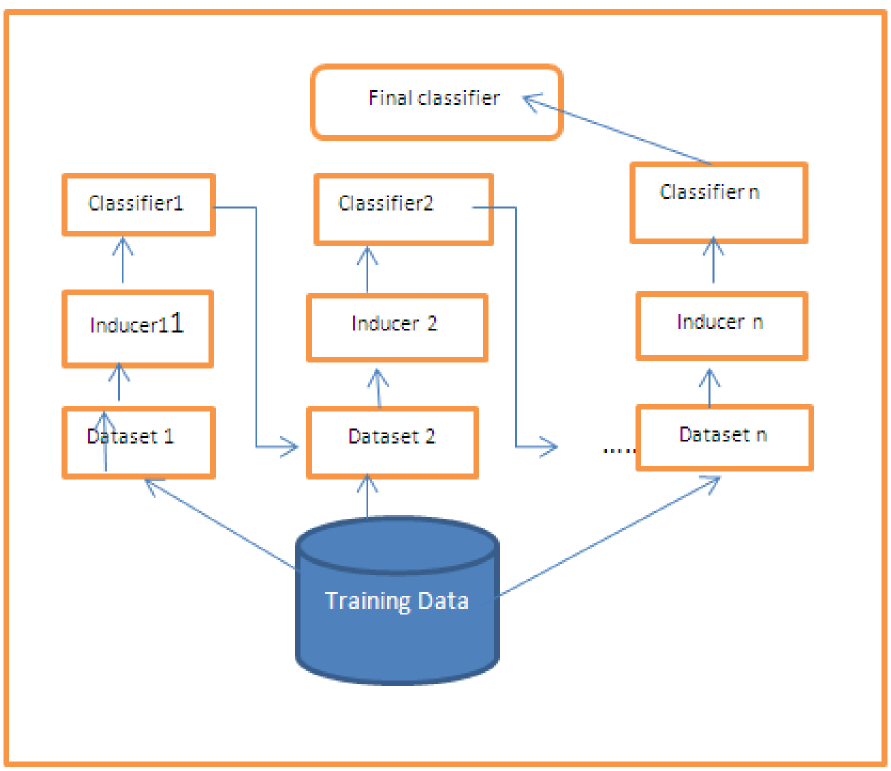

Fig 1. Dependent Model

B. Independent Models:

It is also called as concurrent models. In this methodology the original dataset is transformed into several datasets from which several classifiers are trained. All the datasets are mutually exclusive set. In this approach the models are combined by using combination methods and these methods are independent of the induction algorithms and different inducers can be used with each dataset. The advantage of independent methods over dependent is it can be parallelized which improves predictive power of classifiers. Figure 2 describes the flow of independent model.

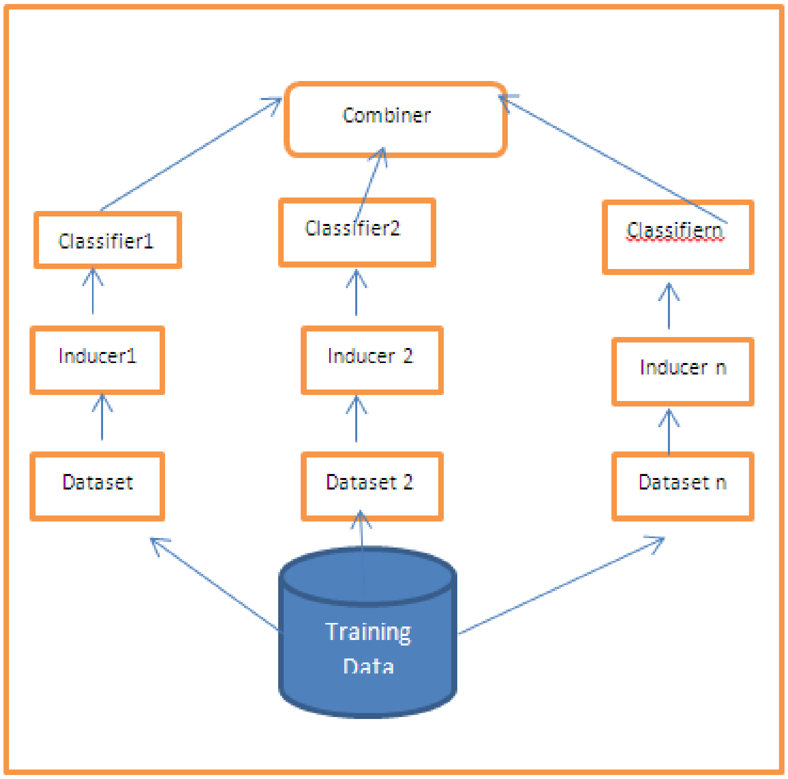

Fig 2. Independent Model

IV. DEPENDENT AND INDEPENDENT ENSEMBLING ALGORITHMS

Various dependent and independent algorithms are devised by researchers. A few important algorithms are

A. Dependent Methods

Boosting Algorithm:

This is a well-known dependent model. It is also known as Arcing (Adaptive Resampling and Combining). In these method weak learners repeatedly runs on various distributed data. The classifiers produced by weak learners are then combined into a single composite strong classifier to achieve accurate dataset labels.

1) Adaboost: Adaptive Boosting was introduced by Freud & Schapire[3] is a popular boosting algorithm. In this more focus is given on patterns which are harder to classify. A weight is assigned to every pattern in the training set. Equal weight is assigned to all patterns. As the iteration proceeds the weights of misclassified instances are increased and the correctly classified instance weights are decreased. Weak learners focus on the difficult instances of the training set by performing additional iterations and creating more classifiers. A weight is also assigned to a classifier which measures the overall accuracy of the classifier and is a function of the total weight of the correctly classified patterns. Weights are directly proportional to the accuracy of classifier. These weights are used for the classification of new patterns.

There are two versions of AdaBoost Algorithm as AdaBoost.M1 and AdaBoostt.M2 in which one is designed for binary classification and other for multi-class classification.

Many researchers enhance the algorithm for better performance and less computation cost. A revised version of AdaBoost called Real AdaBoost was given by Fried [4] to reduce computation cost and lead to better performance. In this output probabilities are combined using additive logistic regression model. For parallel classification Pivoting is given by Breiman by not using weight as a base for the classifier.

P-AdaBoost algorithm was developed by Merler et.al [5] which is a distributed version of AdaBoost. It works in two phases. In the first phase, the AdaBoost algorithm runs in its sequential standard fashion for a limited number of steps. In the second phase the classifiers are trained in parallel using weights that are estimated from the first phase.

Another algorithm revised by Zhang and Zheig[9] called local Boosting algorithm in which a local error is calculated for each training instance and it is used to update the probability and this instance is chosen for training set of the next iteration. Noisy instances are tackled in this method and give more accuracy.

Gradient Tree Boosting is a generalization of boosting to arbitrary differentiable loss function. It can be used for both classification and regression. It handles mixed type data. Prediction power is more and robustness to outliers in input space.

2) Stochastic Attribute Selection Committee builds different classifiers by stochastically modifying the set of attributes considered during induction while the distribution of the training set is kept unchanged.

3) Input Decimal Ensemble is a method of choosing different subsets of the original features based on the correlation between individual features and class labels and training classifiers on those subsets prior to combining. This method reduces the dimensionality of data to reduce the correlations among the classifiers in an ensemble which improves the classification performance of the ensemble.

B. Independent methods:

Original dataset is transformed into several data sets from which several classifiers are trained. The dataset created from the original training set may be disjoined or overlapping. For final classification combination methods are used. There are many methods available in concurrent fashion [2].

Bagging:

It produces a combined model that often performs better than the single model built from the original single data. Each classifier is trained on a sample of instances taken with a replacement from the training set. The size of all sample sizes is equal to the original training set. The training sets are different but are not independent of each other. To classify a new instance, each classifier returns the class prediction for the unknown instance.

1) Wogging: It is a variant of bagging in which each classifier is trained on the entire training set but each instance is assigned a weight.

2) Random Forest: It uses a large number of individual unpruned decision trees. Individual trees are constructed with input variables whose number is less than number of attributes. Instead of all instances, a sub-sample of the instances is used to create different trees which can be combined in ensemble.

3) Cross validated Committees: The training set is divided into k-subsets and training is done to produce k-classifiers. Each partition induced from the partition 'p' is tested on the instances in partition "p+1" in order to improve accuracy.

V. COMBINATION METHODS

There are two methods for combining the outputs. They are weighting methods and Meta learning methods.

A. Weighting Methods:

Weights are assigned as fixed or dynamic for an instance which is to be classified. When combining classifiers with this weight the strength is proportional. Various methods are available depending on weight.

1) Majority voting: Class that obtains highest number of votes is taken. This is the basic ensemble method.

2) Performance Weighting: The weight of each classifier can be set proportional to its accuracy performance on a validation set.

3) Discrimination Summation: Summing of conditional probability vector of each classifier is done in this and the class is selected by choosing highest value in the total vector.

4) Bayesian Combination: Posterior probability of the classifier is considered as weight associated with each classifier.

5) Vogging: Optimization of a linear combination of base-classifiers to aggressively reduce variance and to preserve a prescribed accuracy.

6) Density-based weighting: Various classifiers were trained using datasets obtained from different regions of the instance spaces.

7) Gating Network: The output of each expert is the conditional probability of the target attribute given the input instance. A function is assigned to each network which is considered as weight. Selection of classifier is based on the most appropriate class distribution.

B. Meta-Combination Methods:

Learning from the classifiers produced by the inducers and from the classifications of these classifiers on training data.

1) Stacking is a technique which finds the reliable and unreliable classifiers and it combines models built by different inducers [8]. The best classifier is selected by cross validation.

2) Arbiter Trees: The training set is portioned into k disjoint subsets. An arbiter is induced from a pair of classifiers which is used to provide an alternate classification when the base classifiers provide diverse classifications. This is a recursive process since a new arbiter is induced from the output of two arbiters.

3) Combiner Trees: The difference from arbiter trees lies in the placement of a combiner in each non-leaf node of a combiner tree. In this the classifications of the learned base classifiers form the basis of the meta-learner's training set.

4) Grading: The dataset is portioned into k sets, builds k-1 classifiers by dropping one set at time and then uses it to find a misclassification rate. The set with smallest misclassification is chosen and its classifier is selected.

VI. APPLICATIONS OF ENSEMBLE METHODOLOGY

Ensemble is applied in applications where standard pattern recognition algorithm is harder to apply. A few applications are

Person Recognition is the problem of identifying a person using characteristics of that person, especially security application. They are effectively applied to recognize iris, fingerprint, face and behavior. The results are more accurate than single classifier's output.

Medical Applications are more challengeable due to small training and test data set and its imbalanced nature. Too many attributes are present in data set. So ensemble is the more appropriate method. Applications like pharmaceutical molecule classification, MRI classification; ECG classification is enhanced by using this method.

Remote Sensing Applications data is so voluminous and challenging since a large number of feature set is collected across hundreds of terrain and man-made objects.

Intrusion Detection applications also called as misuse detection in which networks are always attacked by intruders and detection is an important task. It can be done in two ways. One is to find unusual patterns and another is to find correct patterns which are suitable into an identified pattern. The detection is done effectively in mobile ad-hoc networks, different types of attacks like probe, Denial of service etc.

VII. CONCLUSION

Accuracy improving in predictions is the main goal of any prediction research. Instead of one individual model, combiner models leverage the power of multiple models. Ensemble is the concept of combining various models to improve accuracy. Types of ensemble models and various algorithms are discussed in this paper. In future the models are checked with real time data set. The classifiers are to be taken in various domains.

REFERENCES

[1] Ali, Kamal M., and Michael J. Pazzani. "Error reduction through learning multiple descriptions." Machine Learning 24.3 (1996): 173-202.

[2] Bauer E, Kohavi R (1999) "An empirical comparison of voting classification algorithms: bagging, boosting, and variants", Mach Learn 5:1-38.

[3] Freund Y. and Schapire R. E., "Experiments with a new boosting algoritm in Machine Learning", Proceedings of the Thirteenth International Conference, pages 325-332, 1996.

[4] Friedman J. H.and Popescu. "Predictive learning via rule ensembles" Annals of Applied Statistics, 3(2):916-954, 2008.

[5] Merler S,Caprile B, Furianello C, "Parallelising AdaBoost by Weights Dynamics". Comput Stat Data Analysis 51. 2487-2498

[6] Rokach, Lior. "Ensemble methods for classifiers." Data Mining and Knowledge Discovery Handbook. Springer US, 2005. 957-980.

[7] Rokach, Lior. "Ensemble-based classifiers." Artificial Intelligence Review 33.1-2 (2010): 1-39.Nets, A. Sharkey (Ed.), pp. 127-162, Springer-Verlag, 1999

[8] Wolpert, D.H., "Stacked Generalization", Neural Networks, Vol. 5, pp. 241-259, Pergamon Press, 1992.

[9] Chun-Xia Zhang , Jiang-She Zhang, "A local boosting algorithm for solving classification problems", Computational Statistics & Data Analysis, v.52 n.4, p.1928-1941, January, 2008

AUTHORS PROFILE

Mrs. V. Chitraa is a doctoral student in Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu. She is working as an Associate Professor in CMS college of Science and Commerce, Coimbatore and a member of IEEE. Her research interest lies in Database Concepts, Web Mining and knowledge mining. She has published papers in reputed international journals and conferences.

Dr. Antony Selvadoss Thanamani is working as a Reader in the department of Computer Science in NGM college with a teaching experience of about 25 years. His research interests include Knowledge Management, Web Mining, Networks, Mobile Computing, and Telecommunication. He has guided more than 41 M.Phil. Scholars and presented more than 30 papers. He has attended more than 15 workshops, seminars and published about 8 books and 16 papers