ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

An Analysis on Ensemble Methods In Classification Tasks

D.Gopika[1], B.Azhagusundari[2]

Research Scholar, Dr.Mahalingam Centre for Research and Development, N.G.M College, Pollachi, India[1]

Assistant Professor, Dr.Mahalingam Centre for Research and Development, N.G.M College, Pollachi, India[2]

Abstract: Ensemble approaches in classification is a very popular research area in recent years. An ensemble consists of a set of individually trained classifiers such as neural networks or decision trees whose predictions are combined for classifying new instances. It integrates multiple classifiers to build a classification model, and also used for improving the prediction performance. "Diversity" is one of the elements required for accurate prediction when using an ensemble. It is used in wide area of research such as statistics, pattern recognition, and machine learning. This paper presents an updated survey of ensemble methods in classification tasks, and introducing a new taxonomy for characterizing them. The new taxonomy presented here, is based on five dimensions: inducer, combiner, diversity, size, and members dependency.

Key words: Ensemble-methods, Classification, Boosting, Bagging, Random Subspaces, Random Subspaces, Rotation Forest, and Extended space Forest.

I. INTRODUCTION

Classification is a data mining (machine learning) technique used to predict group membership for data instances. The four techniques in classification are Decision tree induction, Nearest neighbor classifier, Artificial neural network and Support Vector Machines. Combining classifiers is a very popular research area known under different names in the literature such as committees of learners, mixture of experts, classifier ensembles, multiple classifier systems, and consensus theory [1]. The basic idea is to use more than one classifier, hoping that the overall accuracy will be better.

Ensemble performances depend on two main properties: the individual success of the base learners of the ensemble, and the independence of the base learners' results from each other (low error, high diversity) [5]. It is possible to build diverse base learners by using same or different type base learners. When the same type base learners are used, the diversity is created by using different training data set for each base learner in the ensemble. There are several methods for creating different training data sets such as Bagging (BG), Boosting, Random Subspaces (RSs), Random Forests (RFs), and Rotation Forest. Polikar [14] and Rokach [9] have two major surveys about ensemble methods.

The existing ensemble methods create different training data sets by deleting/weighting samples or deleting or rotating features. Based on this observation, adding new features (extended spaces) to the original data set is proposed.

The main idea of an ensemble methodology is to combine a set of models, each of which solves the same original task, in order to obtain a better composite global model, with more accurate and reliable estimates or decisions than can be obtained from using a single model [13, 1, 11].

Ensemble methods are very effective, mainly due to the phenomenon that various types of classifiers have different "inductive biases".

II. TRADITIONAL ENSEMBLE ALGORITHMS

This section describes the previous ensemble algorithms.

Bagging. It was proposed by Breiman [3]. Bagging creates a new training data set for each base learner by resampling the original training data set with replacement.

Random Subspaces. Ho proposed [5] that there are two forms of the method. At the first form, each base learner is trained with a different feature subspace of the original training data set. At the second form, only decision trees can be used as the base learner.

Random Forest. It was proposed by Breiman [3]. It can be formulated as bagging plus the second form of random subspaces.

Rotation Forest. It was first proposed by Rodriguez and Alonso [6]. Each base learner is trained with a slightly rotated original training data set. The rotation matrix is calculated differently for each base learner by bootstrapping samples from the training data and from the classes.

This method works with only numeric features. If the data set has other types of features, they are transformed to numeric representation.

III. THEORETICAL FOUNDATIONS

3.1 Ensemble Classifiers

An ensemble consists of a set of individually trained classifiers (such as neural networks or decision trees) whose predictions are combined for classifying new instances.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

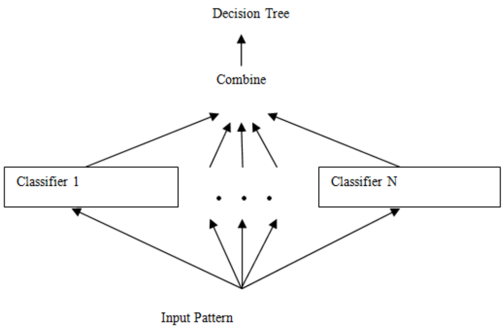

Fig 1. Ensemble of Decision Tree.

Fig 1 states the different classifiers are taken as input and it is combined to form a Decision tree.

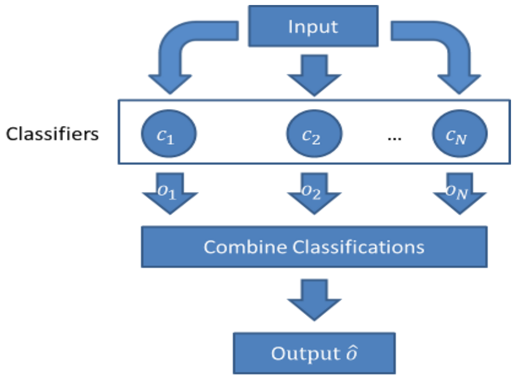

Fig 2. General concept of Ensemble Classifier.

In Fig 2 the input is taken from different classifiers such as C_1, $C2_{....}C_N$ and it is combined together such as O_1, $O2….O_N$ to combine the classifications to from the output Ô.

An ensemble Classifier is often more accurate than any of the single classifiers in the ensemble [8]. Bagging [7] and Boosting are two popular methods for producing ensembles. These methods use re-sampling techniques to obtain different training sets for each of the classifiers. Bagging stands for bootstrap aggregating which works on the concept of bootstrap samples. If original training dataset is of size N and m individual classifiers are to be generated as part of ensemble then m different training sets- each of size N, are generated from original dataset by sampling with replacement.

The multiple classifiers generated in bagging are independent to each other. In case of boosting, weights are assigned to each sample from the training dataset. If m classifiers are to be generated, they are generated sequentially such that one classifier is generated in a single iteration. For generating classifier C_i, weights of training samples are updated based on classification results of classifier C_{i-1}. The classifiers generated by boosting are dependent on each other.

The theoretical and empirical research related to ensemble has shown that an ideal ensemble consists of highly correct classifiers that disagree as much as possible. Opitz and Shavlik [12] empirically verified that such ensembles generalize well. Breiman [7] showed that bagging is effective on unstable learning algorithms. The four approaches for building ensembles of diverse classifiers are presented as:

(1) Combination level: Design different combiners.
(2) Classifier level: Use different base classifiers.
(3) Feature level: Use different feature subsets.
(4) Data level: Use different data subsets.

3.2 Extended Space Forest

In extended space forest, the training sets used in the training of the base learners of an ensemble are generated by adding new features to the original ones. For each base learner, a different extended training set is generated. Bagging, Random Subspace, Random Forest, and Rotation Forest are used. Any other ensemble algorithm and base learner algorithm can be used in the Extended Space framework. The extended training sets are generated by adding new features obtained from the original features. So each base learner is trained with a different training set. The extended space method is a general framework that can be used with any ensemble algorithm [9]. For example, if bagging is chosen as ENSemble algorithm (ENS), the training data for each base learner would be obtained from a random sample, with replacement, from data set.

3.3 Random Forest

Random Forest is a classifier consisting of a collection of tree-structured classifiers. It generates an ensemble of decision trees. To achieve diversity among base decision trees, Breiman selected the randomization approach which works well with bagging or random subspace methods [7], [5].

To generate each single tree in Random Forest Breiman followed following steps: If the number of records in the training set is N, then N records are sampled at random but with replacement, from the original data, this is bootstrap sample. This sample will be the training set for growing the tree. If there are M input variables, a number m << M is selected such that at each node, m variables are selected at random out of M and the best split on these m attributes is used to split the node. The value of m is held constant during forest growing. Each tree is grown to the largest extent possible. There is no pruning.

In this way, multiple trees are induced in the forest; the number of trees is pre-decided by the parameter N tree. The number of variables (m) selected at each node is also referred to as mtry or k in the literature. The depth of the tree can be controlled by a parameter node size (i.e. number of instances in the leaf node) which is usually set to one [7]. Once the forest is trained or built as explained above, to classify a new instance, it is run across all the trees grown in the forest.

Each tree gives classification for the new instance which is recorded as a vote. The votes from all trees are combined and the class for which maximum votes are counted (majority voting) is declared as classification of the new instance. This process is referred to as Forest RI in the literature [8].

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014
ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

IV. EXISTING SURVEYS ON ENSEMBLE OF CLASSIFIERS

Given the potential usefulness of ensemble methods, it is not surprising that a vast number of methods are now available to researchers and practitioners. The variety of ensemble techniques have arisen several taxonomies in the literature which aim to categorize ensemble methods from the algorithm designer point of view. Sharkey [15] proposed taxonomy for ensemble of neural networks. This taxonomy suggests three dimensions:

(1) The first dimension indicates if the ensemble's members are competitive or cooperative. In the competitive mode, a single member is selected to provide the classification. In cooperative mode the classifications of all members are combined.
(2) The second dimension indicates if the ensemble is created top-down or bottom-up. In top-down mode the combination mechanism is based on something other than the classifiers outputs. Bottom-up techniques take the outputs of the members into account in their combination. Bottom up methods are subdivided into fixed methods (such as voting), and dynamic methods (such as stacking).
(3) The third dimension indicates if we combine ensemble, modular, or hybrid components; it is distinguish between modular systems and pure ensemble systems. The main idea of pure ensemble systems is to combine a set of classifiers, each of which solves the same original task and to obtain a more accurate and reliable performance than using a single classifier. On the other hand, the purpose of modular systems is to break down a complex problem into several manageable problems.

The ensembles techniques are divided into two main categories are Decision optimization and Coverage optimization.

Brown et al. [5] divides up the ensemble methods according to whether they choose to implicitly obtain diversity by randomization methods or whether they explicitly gain diversity via some metric. They then grouped techniques according to three factors: how they initialize the inducers in the hypothesis space, what the space of accessible hypotheses is, and how that space is traversed by the inducer.

Although several surveys on ensemble for classification tasks are available in the literature [4] and there are several papers which suggest taxonomy for ensemble methods [5], in this paper the four main contributions are introduced:

(1) A new unified taxonomy is suggested to categorize all significant ensemble methods developed in the field. As indicated in [4]: Still there is no agreed upon structure or taxonomy of the whole field, although a structure is slowly crystallizing among the numerous attempts." On the one hand because existing taxonomies usually concentrate on some aspects (for example [5] concentrates on diversity), the new proposed taxonomy tries to organize existing taxonomies into a coherent and unified taxonomy.
(2) Due to the fact that ensemble learning is an active research field, this paper proposes an updated survey which refers to new researches from the last three years that have not been previously covered by existing surveys.
(3) This paper covers efficient and mature ensemble methods that do not belong to the mainstream, and therefore are usually not mentioned in existing surveys.
(4) It also proposes several selection criteria, presented from the practitioner's point of view, for choosing the most suitable ensemble method.

V. TAXANOMY OF ENSEMBLE CLASSIFIER

A typical ensemble method for classification tasks contains the following building blocks:

Training set
A labeled dataset used for training the ensemble. Most frequently the training set is a collection of instances (also known as samples or observations). Each instance is described by attribute-value vectors. The input space is spanned by the attributes used to describe the instances. In semi-supervised methods of ensemble generation, such as ASSEMBLE [10], unlabeled instances can be also used for the creation of the ensemble.

Inducer
The inducer is an induction algorithm that obtains a training set and forms a classifier that represents the generalized relationship between the input attributes and the target attribute.

Ensemble generator
This component is responsible for generating the diverse classifiers.

Combiner
The combiner is responsible for combining the classifications of the various classifiers.

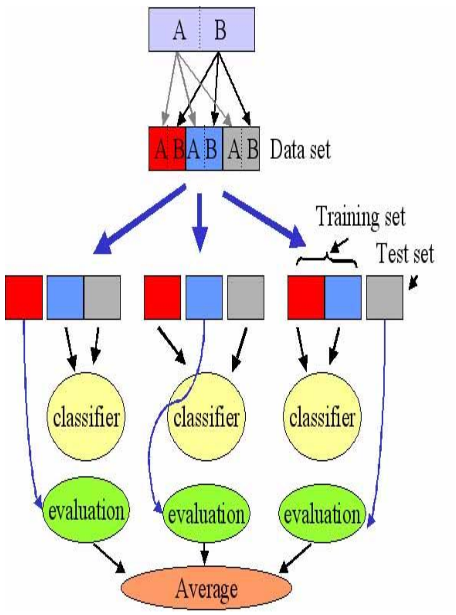

Fig 3: Taxonomy of classifiers

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

In Fig 3 the different level of taxonomy is combined and classified to evaluate an average of the classifiers used. The nature of each building block and especially the relation among them can be used to categorize ensemble methods.

It consists of the following dimensions:

(1) **Combiner usage**: This property specifies the relation between the ensemble generator and the combiner.
(2) **Classifiers dependency**: Classifiers may be dependent or in-dependent.
(3) **Diversity generator**: In order to make the ensemble more effective, there should be some sort of diversity between the classifiers. Brown et al. [2] indicate that for classification tasks the concept of "diversity" is still an ill-defined concept. Nevertheless it is believed to be closely related to the statistical concept of correlation. Diversity is obtained when the misclassification events of the base classifiers are not correlated. Several means can be used to reach this goal: different presentations of the input data, variations in learner design, or by adding a penalty to the outputs to encourage diversity.
(4) **Ensemble size**: The number of classifiers in the ensemble and how the undesirable classifiers are removed from the ensemble.
(5) **Cross-Inducer**: A Cross-inducer ensemble techniques could run on all common inducers, or simply more than one. Some ensembles have been specifically designed for a certain inducer and cannot be used for other inducers [11].

The issues of classifiers' dependency and diversity are closely linked. More specifically, it can be argued that any effective method for generating diversity results in dependent classifiers (otherwise obtaining diversity is just luck).

It is very simple to independently build ensemble of diverse classifiers. For example, it can be train in parallel a number of different classifiers by randomly drawing their training instances from the original set. Because none of these classifiers in anyway affects the training of others, they remain "independent". It can independently create the classifiers and then, as a post-processing step, select the most diverse classifiers. Instead of using the notion of classifiers' dependency, Brown et al. [2] make a distinction between explicit and implicit diversity methods. In implicit techniques no measurement is taken to ensure diversity will emerge (thus, the classifiers can be independently trained). Explicit methods, on the other hand, explicitly try to optimize some metric of diversity during building the ensemble (thus, the classifiers are usually built in some dependent manner as they need together maximize diversity).

VI. ENSEMBLE DIVERSITY

In an ensemble, the combination of the output of several classifiers is only useful if some of the inputs are different [10]. Creating an ensemble in which each classifier is as different as possible while still being consistent with the training set is theoretically known to be an important feature for obtaining improved ensemble performance. A diversified classifiers lead to uncorrelated errors, which in turn improve classification accuracy.

In the regression context, the bias-variance-covariance decomposition has been suggested to explain why and how diversity between individual models contributes toward overall ensemble accuracy. In the classification context, there is no complete and agreed upon theory [5]. More specifically, there is no simple analogue of variance-covariance decomposition for the zero- one loss function. Instead, there are several ways to define this decomposition. Each way has its own assumptions. Sharkey [15] suggested taxonomy of methods for creating diversity in ensembles of neural networks. More specifically, Sharkey's taxonomy refers to four different aspects: the initial weights, the training data used, the architecture of the networks, and the training algorithm used.

Brown et al. [5] suggest a different taxonomy which consists of the following branches: varying the starting points within the hypothesis space; varying the set of hypotheses that are accessible by the ensemble members (for instance by manipulating the training set); and varying the way each member traverses the hypothesis space.

The components of the following are not mutually exclusive, namely, there are a few algorithms which combine two of them.

(1) *Manipulating the Inducer:* It is manipulated in the way in which the base inducer is used. More specifically each ensemble member is trained with an inducer that is differently manipulated.
(2) *Manipulating the Training Sample:* The input is varied and that is used by the inducer for training. Each member is trained from a different training set.
(3) *Changing the target attribute representation:* Each classifier in the ensemble solves a different target concept.
(4) *Partitioning the hypothesis space:* Each member is trained on a different hypothesis subspace.
(5) *Hybridization:* Diversity is obtained by using various base inducers or ensemble strategies.

VII. CONCLUSION

This paper presents an updated taxonomy survey of ensemble methods for classification problems. Ensemble algorithms are mostly used in machine learning because of its well versed performances than other single algorithms.

A description of five dimensions of ensemble taxonomy and traditional algorithms are also included in this paper. The ensemble algorithms using all the features (Bagging and Rotation Forest) have more accurate base learners.

REFERENCES

[1] Bauer, E. and Kohavi, R., "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants". Machine Learning,35: 1-38, 1999.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 7, July 2014

[2] Brown G., Wyatt J., Harris R., Yao X., Diversity creation methods: a survey and categorization, Information Fusion, 6(1):5{20}.
[3] C.E. Brodley, Recursive automatic bias selection for classifier construction, Machine Learning 20 (1995) 63-94.
[4] Dietterich T., Ensemble methods in machine learning. In J. Kittler and F.Roll, editors, First International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science, pages 1-15. Springer-Verlag, 2000.
[5] G. Brown, J.L. Wyatt, and P. Tino, "Managing Diversity in Regression Ensembles," The J. Machine Learning Research, vol. 6,pp. 1621-1650, 2005.
[6] J.J. Rodriguez and C.J. Alonso, "Rotation-Based Ensembles," Proc.10th Conf. Spanish Assoc. Artificial Intelligence, pp. 498-506, 2004.
[7] L. Breiman,"Bagging Predictors," Machine Learning, vol. 24, no. 2, pp. 123-140, 1996.
[8] L.I. Kuncheva and C.J. Whitaker, "Measures of Diversity in Classifier Ensembles and Their relationship with the Ensemble Accuracy," Machine Learning, vol. 51, no. 2, pp. 181-207, 2003.
[9] L. Rokach, "Taxonomy for Characterizing Ensemble Methods in Classification Tasks: A Review and Annotated Bibliography," Computational Statistics & Data Analysis, vol. 53, no. 12, pp. 4046-4072, 2009.
[10] Opitz, D., Feature Selection for Ensembles, In: Proc. 16th National Conf. on Artificial Intelligence, AAAI,1999, pp. 379-384.
[11] Opitz, D. and Maclin, R., Popular Ensemble Methods: An Empirical Study, Journal of Artificial Research, 11: 169-198, 199
[12] Opitz D. and Shavlik J., Generating accurate and diverse members of a Neural network ensemble. In David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo, editors, Advances in Neural Information Processing Systems, volume 8.
[13] Quinlan, J. R., Bagging, Boosting, and C4.5. In Proceedings of the Thirteenth National Conference on Artificial Intelligence, pages 725-730, 1996.
[14] R. Polikar, "Ensemble Based Systems in Decision Making," IEEE Circuits and Systems Magazine, vol. 6, no. 3, pp. 21-45, Third Quarter 2006.
[15] Sharkey, A., Multi-Net systems, In Sharkey A. (Ed.) Combining Artificial Neural Networks: Ensemble and Modular Multi-Net Systems. pp. 1-30, Springer-Verlag,1999.

BIOGRAPHIES

D.Gopika is a research scholar in Nallamuthu Gounder Mahalingam College, Pollachi. She received her Master of Computer Science in 2012. She has presented papers in International/National Conferences and attended Workshop, Seminar. Her research interest focuses on Data Mining.

B.Azhagusundari received her B.Sc Mathematics and Master of Computer Applications from NGM College, Pollachi, and Coimbatore, India. She completed her Master of Philosophy in Bharathidasan University, Trichy. Presently she is working as an Assistant Professor in the P.G Department of Computer Applications in NGM College (Autonomous), Pollachi. Her area of interest includes data Mining and Networking. Now she is pursuing her Ph.D Computer Science in Mother Teresa University, Kodaikannal.