Volume 4, Issue 5, May 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Nearest Neighbor Ensembles Combines with Weighted Instance and Feature Sub Set Selection: A Survey

Pragya Pandey*, Megha Singh
Comp.Sc & Engg, R.G.P.V. Bhopal
India

Abstract*— Ensemble learning deals with methods which employ multiple learners to solve a problem The generalization ability of an ensemble is usually significantly better than that of a single learner, so ensemble methods are very attractive, at the same time feature selection process of ensemble technique has important role of classifier. This paper, presents the analysis on classification technique of k-nearest neighbor method while applying feature reduction in subsets and further more constructing ensembles of nearest neighbor classifiers on the basis of the instance selection. Instance selection is to obtain the subset of the instances available for training capable of achieving, at least, the same performance as the whole training set will provide.

Keywords*— Ensemble learning, k-nearest neighbor, feature sub set selection and instance selection.

I. INTRODUCTION

Knowledge discovery [1] is the process of finding the knowledge from large database. Various classifiers are used to classify the unseen patterns according to the known class labels. The k-nearest neighbors' algorithm [2] is amongst the simplest of all machine learning algorithms. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors. k is a positive integer, typically small. If k = 1, then the object is simply assigned to the class of its nearest neighbor. In binary (two class) classification problems, it is helpful to choose k to be an odd number as this avoids tied votes.

Ensemble methods make predictions by combining the predictions from a set of individual classifiers. Good ensemble methods require accurate and diverse individual classifiers. The two most popular approaches to generating diverse classification models by applying single learning method are bagging [3] and boosting [4]. At the same time another random subspace ensemble method (RSM) also known as random forests [5], is used for significant improvement of the generalization error. For bagging related approaches the training sets are randomly chosen k times with replacement (bootstrap techniques) from the original data set. As a result, k training sets are obtained, each of them having the same size as the original set of data. Each classifier is then trained with one of the generated training sets.

Boosting methods construct an ensemble of classifiers iteratively focusing each new member on the most difficult instances by means of a biased distribution of the training instances. Boosting methods "boost" the accuracy of a weak classifier by repeatedly resampling the most difficult instances. Boosting methods construct an additive model.

Instance selection [6] consists of choosing a subset of the total available data to achieve the original purpose of the data mining application as if the whole data are used. Different variants of instance selection exist. We can distinguish two main models [7]: instance selection as a method for prototype selection for algorithms based on prototypes (such as: nearest neighbors (NN) [8]) and instance selection for obtaining the training set for a learning algorithm that uses this training set (such as decision trees or neural networks).

We are explaining feature selection or attribute selection with the instance selection. It is a broad subject that spans to research disciplines such as statistics, machine learning, fractals, rough sets theory and mathematical programming [9, 10]. The advantages of feature selection are that it reduces the dimensionality of the feature space and removes the redundant, irrelevant or noisy data.In the context of ensembles, the aim of the feature selection methods is to reduce the number of input features presented to the base classifiers and, at the same time, dealing with the dimensionality and diversity problems. Recently, several authors have investigated the use of feature selection methods in ensembles. For instance, the random subspace method [11] relies on a pseudorandom procedure to select a small number of dimensions from a given feature space. In each step, such a selection is made and a subspace is fixed. In this paper we have analyzed the methodologies used both for the speeding up the classification process as well as the techniques to improve the accuracy of the classifiers.

II. FEATURE REDUCTION ON NEIGHBORHOOD MODEL

K-NN learning and neighborhood classifier [27] are lazy learning algorithms, they keep all training samples and initiate classification process on receiving of the test sample. K-NN classification largely used to recognize pattern and it has been always worthy if we use relatively lower dimensional space to avoid curse-of dimension, reduce cost in measuring and processing information and enhance the interpretability of learned models. However, in the real situations as

development of information techniques, more and more samples and features are acquired and stored. Classification algorithms will be confused with a lot of features. Therefore, feature subset selection is implicitly or explicitly conducted for some learning systems.

For constructing a neighborhood classifier with improved performance there are two steps required-First we search an optimal feature subspace, which has a similar discriminating power as the original data, but the number of features is greatly reduced. Then we associate a neighborhood with each test sample in the selected subspace and assign the class label with majority samples in the neighborhood to the test.

Rough set based feature selection process includes the select a minimal attribute subset, which has the same characterizing power as the whole attribute set, and without any redundant attribute. In other words, we have to select the attributes those are capable enough to hold the dependency and removal of any further attribute will affect the dependency. Dependency function introduces to evaluate the features for selection. Neighborhood classifier [19] can be understood as a classification system which uses the samples in the neighborhood to estimate the local class probability density of the test samples. In this paper, the samples in the neighborhood have the same influence on the estimated probability density; however, if we consider the neighborhood as a window function, we can use other window functions to predict the class probability, where different weights will be assigned to the samples in the neighborhood. On the other hand, we also can use the concept of fuzzy neighborhood to generalize the proposed algorithm. Some techniques to improve the speed of lazy learner algorithms- K-Nearest Neighbor [17, 18] introduced.

III. ENSEMBLE OF CLASSIFIERS

To boost up the performance of classifier several classifiers are combined, each classifier is known as member classifier and makes predictions independently of each other. After that, these predictions are combined together by voting in order to yield the final prediction. This overall system is called ensemble of classifiers. Ensemble members can belong to different types of algorithms, but we have chosen the Nearest Neighbor learning method with the neighborhood reduction thus we are more particular for kNN classification algorithm. As a combination technique, the conventional majority vote was selected in order to demonstrate that ensembles shows good performance even when employing simple non-trainable combiners.

The main goal for any ensemble is to perform better than its most accurate member. It is well known that an ensemble is able to outperform its best performing member if ensemble members make mistakes on different cases so that their predictions are uncorrelated and diverse as much as possible. On the other hand, an ensemble must include a sufficient number of accurate classifiers since if there are only few good votes, they can be easily drowned out among many bad votes, and as a result, an ensemble can predict wrongly most of the time.

IV. BAGGING ENSEMBLE WITH K-NN

Ensemble method Bagging was devised to make work with k-NN. According to Breiman [24] bagging (bootstrap aggregating in which random sampling with replacement was preferred) is a method for generating multiple versions of a classifier and using them to obtain an aggregated classifier. The multiple versions are generated by making bootstrap samples of the original training set and considering these samples as new training sets. Each bootstrap sample has the same number of patterns as the original training set. The aggregation is combining predictions (class labels) of the individual versions by majority voting, where all votes are equally weighted.

However, bagging will not help if the accuracy of an indivudual classifier is close to the limits attainable on a given data set. In the work reported by [20] bagging was applied to kNN classifiers through the application of introducing randomness to the distance metrics. A new distance metric based on 2 parameters, referred to as Minkovdm, was developed by combining the Value Difference Metric (VDM) [21] and the Minkowsky distance. The original data set was initially bootstrap sampled. Then, for each sample, random values were assigned to the 2 parameters. In this approach, each newly generated kNN differed not only with the training set being used, however, also with the distance metric. The combination of both bagging and Minkovdm was reported as offering improved overall performance in comparison with the single kNN.

For bagging related approaches the training sets are randomly chosen k times with replacement (bootstrap techniques) from the original data set. As a result, k training sets are obtained, each of them having the same size as the original set of data. Each classifier is then trained with one of the generated training sets.

V. CONSTRUCTING ENSEMBLE OF NEAREST NEIGHBORS BY WEIGHTED INSTANCE SELECTION AND FEATURE SUBSET SELECTION

To the best of our knowledge no previous work has tried to adopt the instance selection for the purpose of constructing ensembles. Some work of Freund and Schapire [12] shares some of the ideas underlying the approach. They developed a boosting version of nearest neighbor classifier, but with the goal of increasing the classification speed and have not done anything for improving its accuracy. In their method, each weak classifier (lazy learner) is defined by a subset of the training set and a 1-NNrule. On each boosting round, a set is created for the classifier adding instances in a stepwise manner. Initially, At each round of boosting, ten random candidates are selected according to the current distribution of instances given by the boosting algorithm. The candidate that causes the largest decrease in the pseudoloss is added to the set. This process is repeated until reaches a prespecified size. The k-NN rule is a well-known and widely used method for classification. The method consists of storing a set of prototypes that must represent the knowledge of the problem.

This set of prototypes can be the whole training set, a subset obtained using instance selection [13], or different combinations of training instances [14].

The Nearest neighbor method is used mainly because of its simplicity and its ability to achieve error results comparable with much more complex methods. For instance, in computer vision, it has been applied to a wide range of problems successfully, such as the method can be used with any of the existing boosting methods. On many occasions, the reported results are able to improve the errors obtained with more sophisticated methods [15].

Previous combination methods, such as voting or bagging, are not usually useful when applied to nearest neighbor, as this method is fairly stable with respect to modification of the training set. Furthermore, error correcting output codes are also unsuccessful with Nearest Neighbor. The work has been shown that error correcting output codes are not useful with classifiers that use local information due to high error correlation On the other hand, Nearest Neighbor is very sensitive to input perturbations(changes to the input data), such as subspace selection or nonlinear projections. The test pattern can be classified in three different classes depending on the subspace considered. In this way, subspace methods have shown good results when applied to Nearest Neighbor [16] this property will provide the basic for our idea of implementing Nearest Neighbor ensemble with neighborhood reduction. We have analyzed the basic idea of using the distribution given by the boosting algorithm to optimize a weighted training error using an instance selection algorithm. The overview of this procedure is illustrated here which is used to boost NN classifier.

Algorithm 1: Ensemble with weighted instance selection

Input: A training set $T = \{(x_1,y_1),\ldots\ldots\ldots, (x_n,y_n\}$, where $Y = \{y_1,\ldots,y_n\}$ are sample labels,
A learning /classification algorithm L,
number of iterations M

Output: Final classifier $F(x)= \arg Max_{y\epsilon Y}$

1. Use the training set for learning algorithm and generate M learner in each iteration
2. Obtain weights for each instance using boosting methods
3. Find the subset of T and apply the L on this subset.

In the field of statistical discrimination [26] nearest neighbor methods are a well known, quite simple but successful nonparametric classification tool. In higher dimensions, however, predictive power normally deteriorates. In general, if some covariates are assumed to be noise variables, variable selection is a promising approach. The paper's main focus is on the development and evaluation of a nearest neighbor ensemble with implicit variable selection. In contrast to other nearest neighbor approaches we are not primarily interested in classification, but in estimating the (posterior) class probabilities. In higher dimensions simple nearest neighborhood estimates tend to be unstable when noise variables are present. Therefore we aim at selecting relevant variables or interactions between them, and combine the nearest neighborhood predictions based on single or small groups of predictors.

Domeniconi and Yan [28, 30] proposed ensembles of nearest neighbor classifiers based on random subsets of predictors, while performing adaptive sampling. In this mechanism variable selection is done by probability distribution but not completely at random. This distribution is derived from some kind of relevance measure. Thus, the approach is a two-step procedure, because in a first step all features' relevance needs to be determined and variable selection in the feature space, using nearest neighbor methods.

We have an alternative approach to producing diverse kNN classifiers involving random sampling over the feature space was proposed in [16]. In this work each base classifier was generated with a randomly selected subset of features. The final ensemble with a combining voting technique was able to improve performance in comparison with the kNN base classifiers. A similar method based on this concept was presented by [8]. In this work each kNN model was built by applying a different subset of features. Each feature had a weight assigned to reflect its relevance to the problem being considered. A weighted Euclidean distance metric was applied while selecting neighbors. To make the final decision a voting scheme was applied among the instances selected by all the kNN in the ensemble. Following evaluations it was found that the proposed approach based on sampling over the feature space provided a significant improvement in comparison with a single kNN classifier.

A specific method of ensemble: Boosting [21, 23, 29] has been proposed by Garcia-Pedrajas and Ortiz-Boyer [22, 29], in which input space was modified by a means of input selection or non-linear projection. In the first instance different subspaces, that minimized the weighted error, were selected for each kNN classifier. In the second instance a nonlinear projection was constructed using the weight vectors. In a previous work [25] we developed a new approach based on obtaining, before training a classifier, the subspace where the classification error weighted using the distribution obtained by boosting is minimized.

VI. CONCLUSIONS

Data mining is a broad area that integrates techniques from several fields including machine learning, statistics, pattern recognition, artificial intelligence, and database systems. To perform the data analysis task various data mining/classification algorithms are used. In this paper we have analyzed and briefed the several proposed methods for ensemble of k-NN classifier with bagging, boosting and random sub space method. These algorithms further combined with instance distribution during the learning process, input space modifications. And feature subset selection. In a final study we have shown that the proposed methods are able to improve k-NN performance by means of reducing bias term of the error.

REFERENCES

[1] Fayyad, Piatetsky-Shapiro, Smyth, and Uthurusamy, "Advances in Knowledge Discovery and Data Mining", AAAI/MIT Press 1996

[2] Dasarathy BV (ed) ,"Nearest neighbor (NN) norms: NN pattern classification techniques", IEEE Computer Society Press, 1991

[3] L. Breiman. "Bagging Predictors". Machine Learning , vol. 24 (2), pp.123-140, 1996.

[4] Y. Freund, R. E. Schapire. "A Short Introduction to Boosting". Japanese Society for Artificial Intelligence, vol. 14 (5), pp. 771-780, 1999.

[5] L. Breiman, "Random Forests," Machine Learning, 2001, 45:5–32.

[6] H. Liu and H. Motoda, "On issues of instance selection," Data Mining Knowl. Disc., vol. 6, pp. 115–130, 2002.

[7] J. R. Cano, F. Herrera, and M. Lozano, "Using evolutionary algorithms as instance selection for data reduction in KDD: An experimental study," IEEE Trans. Evol. Comput, vol. 7, no. 6, pp. 561–575, Dec.2003.

[8] C. Domeniconi, G. D., and J. Peng, "Large margin nearest neighbor classifiers," IEEE Trans. Neural Netw., vol. 16, no. 4, pp. 899–909, Jul.2005.

[9] M. Chen, J. Han, and P. Yu.,"Data mining: an overview from database perspective", IEEE Trans. Knowledge and Data Engineering, 8(6):866–883, 1996.

[10] Hall, Mark A. Correlation-based feature selection for machine learning. Diss. The University of Waikato, 1999.

[11] T K Ho., "The random subspace method for constructing decision forests", IEEE Trans. Pattern Anal. Mach. Intell. 20(8), 832–844 (1998)

[12] Y. Freund and R. Schapire, "Experiments with a new boosting algorithm", in Proc. 13th Int. Conf. Mach. Learn., Bari, Italy, 1996, pp.148–156.

[13] H. Brighton and C. Mellish, "Advances in instance selection for instance- based learning algorithms," Data Mining Knowl. Disc., vol. 6,pp. 153–172, 2002.

[14] Aha, David William. "A study of instance-based algorithms for supervised learning tasks: Mathematical, empirical, and psychological evaluations." Ph.D. dissertation, Dept. Inf. Comput. Sci., Univ. California at Irvine, Irvine, CA, 1990.

[15] T. K. Ho and H. S. Baird, "Large-scale simulation studies in image pattern recognition," IEEE Transaction Pattern Analysis Machine Intelligence, vol. 19, no. 10, pp. 1067–1079, Oct. 1997.

[16] S. D. Bay, "Nearest neighbor classification from multiple feature subsets," Intelligent Data Anal., vol. 3, no. 3, pp. 191–209, 1999.

[17] Tan, S. B., "Neighbor-weighted K-nearest neighbor for unbalanced text corpus" Expert Systems with Applications, 28, 667–671, 2005.

[18] Bickel, S., & Scheffer, T. , "Multi view clustering", Forth IEEE conference on Data Mining Brighton, UK,2004

[19] Fix, E. and J. L. Hodges, "Discriminatory analysis - nonparametric discrimination: consistency properties", US air force school of aviation medicine, Randolph Field Texas, 1951.

[20] Z. Zhou, Y. Yu., "Adapt Bagging to Nearest Neighbor Classifiers", Computer Science and Technology, vol. 20 (1), pp. 48-54, 2005.

[21] C. Stanfill, D. Waltz, "Toward memory based reasoning", Communications of ACM, vol. 29 (12), pp. 1213-1228, 1986.

[22] N. Garcia-Pedrajas. "Constructing Ensembles of Classifiers by Means of Weighted Instance Selection". IEEE Transactions on Neural Networks, vol. 20 (2), pp. 258-277, 2009.

[23] S. D. Bay, "Nearest Neighbor classification from multiple feature subsets", Intelligence Data Analysis, vol. 3, pp. 191-209, 1999.

[24] Breiman, L., "Bagging predictors", Machine Learning, 24(2), 123–140,(1996a).

[25] García, Pedrajas, N., & Ortiz-Boyer, D., "Boosting random subspace method." ,Neural Network, 21, 1344–1362, 2008.C.

[26] J. Merz, "Using correspondence analysis to combine classifiers," Machine Learning, vol. 36, no. 1, pp. 33–58, Jul. 1999.

[27] Qinghua Hu. , Daren Yu, Zongxia Xie, "Neighborhood classifiers" , Expert Systems with Applications xxx (2007)

[28] Gertheiss, Jan, and Gerhard Tutz. "Feature selection and weighting by nearest neighbor ensembles." Chemometrics and Intelligent Laboratory Systems 99.1 (2009): 30-38.

[29] Nicolás García-Pedrajas *, Domingo Ortiz-Boyer, "Boosting k-nearest neighbor classifier by means of input space projection", Expert Systems with Applications 36 (2009) 10570–10582

[30] Domeniconi, C. and B. Yan, "Nearest neighbor ensemble", In Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK,2004.

