ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

ENSEMBLE LEARNING USING INTERNET OF THINGS PLATFORM

ABINAYA.P [1], ARUNA.M [2] , KAUSALYA.S.P [3] , KARTHIKEYAN.T [4]

[1,2,3]UG Scholar, Department of Computer Science and Engineering, Knowledge Institute of Technology,Salem.
[4]Assistant professor, Department of Computer Science and Engineering, Knowledge Institute of Technology, Salem.

ABSTRACT

Ensemble learning is a process in which the multiple numbers of nodes are created strategically and compiled together to solve an intelligence problem which is selected particularly. The main idea behind this is to use multiple learning algorithms for obtaining a better predictive performance than any other that could be obtained by other constituent learning algorithms. The original ensemble method is Bayesian averaging, but the recent algorithm includes error correcting output coding, bagging and boosting. One of the main disadvantages in ensemble learning is that the loss of interaction may occur between nodes during interaction. It is possible that some of the individual networks designed do not contribute to the entire ensemble leading to a failure in the model. The most straight forward way of manipulating the training set is called bagging. Different training data subsets are randomly drawn with replacement of training data set. Boosting also creates an ensemble of classifiers by reassembling the data which will then be combined by majority voting. The best known ensemble based algorithms is Adaptive boosting which extends boosting to multi class and regression problems. This paper is about howthe concept of ensemble learning can be used in Internet of Things platform. IOT is a large network of physical objects and devices which can be embedded with software's, sensors and network connectivity that enables these objects in collecting and exchanging of data between them. The internet of things allows objects to be sensed using sensors and to control them across existing network infrastructure, thus creating environment for more amount of direct integration of the physical world into the computer systems, which results in efficiency improvement, accuracy and economic benefit.

Keywords: Ensemble learning, Adaptive Boosting, Internet of Things, Learning Algorithm

1. INTRODUCTION

The real world problems are large and complex for a single system to solve them alone. The idea behind ensemble learning is to generate multiple learners and to combine their individual predictions. Ensemble methods are learning models that achieves performance by combining the opinion of multiple learners. The principle behind Ensemble learning is that the decision of a committee, with appropriate combination of individual predictions, will have a better overall accuracy, on an average than the individual predictions [1].

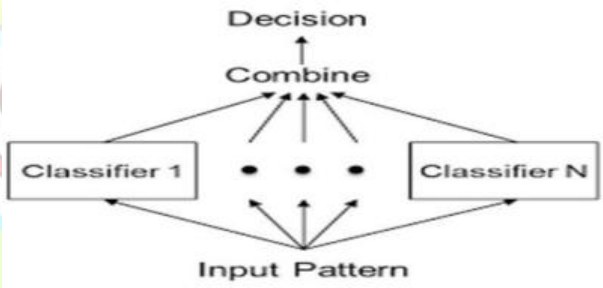


Fig. 1.1: Classification

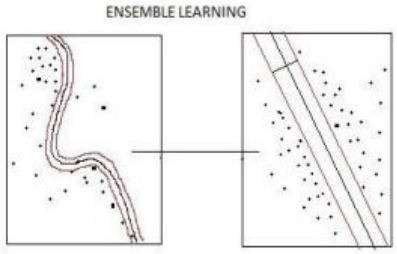


Fig. 1.2: Ensemble Learning

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

Many theoretical studies have been demonstrated that ensemble models offer more accuracy when compared with individual models.

In Learning , several alternative definitions for a concept using different training data or different learning algorithms. Witten and Frank stated the four methods of combining multiple models such as bagging, boosting, stacking and error correcting output codes. Alpaydin stated seven methods of combining multiple nodes such as bagging, boosting, stack generalisation, mixture of experts, error correcting output codes, cascading and voting [2].

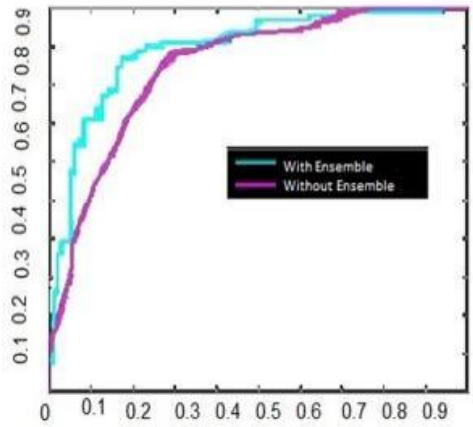


figure 1.3 Graph for Ensemble learning

Finally four ensembler methods of grouping multiple nodes include bagging, boosting, stack generalisation and the random subspace method. Ensemble methods are capable of boosting weak learners which are at least slightly better than the random performance of good learners, which will be useful in making exact predictions. The boosting concept explains the basic deal, which starts with fitting one learner and correcting the learner's mistake in subsequent learners. Boosting involves building incrementally an ensemble by training individual new model instance to emphasize the training instance which the previous model has classified. In some cases, boosting yields better accuracy than boosting, but it also tends to make more likely than the training data to over fit. The common technique of implementation of boosting includes adaboost, even though some new algorithms provide better results when compared. Error bounds of the final combined learner can be discussed based on the errors of the learners who were weak.

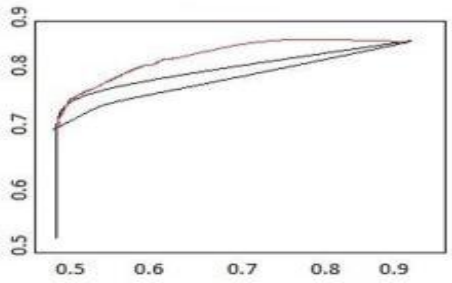


figure 1.4 ROC Curves

The bagging concept is a parallel ensemble method and lends itself in the possibility of parallel computing. Bagging uses the concept of bootstrap sampling which can be said as composing a new data set of equal size by sampling with the replacement from another base dataset. Stacking which can be commonly called as stacked generalisation involves a learning algorithm to combine the results of other learning algorithms. Bagging is based on the idea that the combination of independent based learners will lead in a substantial decrease of errors and hence we need to obtain base learners as independent as possible. The other algorithms are trained to use the available data, then a combiner algorithm is used and trained to make a final prediction using all the predictions of other algorithms as an additional inputs [3].

2. CASE STUDY

Learning can be defined as many different form of activities ranging from concept learning to reinforcement learning. The best understandable form of statistical learning is known as supervised learning. In this method each data point consists of a vector of features denoted by x and a class label denoted by y and also an assumption is made that an underlying function denoted by f such that y=f(x) for each data point(x, y). The main goal of learning algorithm is to find a better approximation h to f which can be used in assigning values for new x values. The function h can be said as a classifier, as it assigns class labels y to input data points x. such supervised learning algorithm can be used to solve many problems such as medical diagnosis, handwriting recognition and part of speech tagging in language processing. Ordinary machine learning algorithms works by searching through possible functions, which is called as

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

hypothesis, to find the one function h, which, is the best approximation for the unknown function f. To determine the best hypothesis h, a learning algorithm can be used to measure how well the function h matches f on the input points and the algorithm can also access how consistent h is with any available prior knowledge, about the problem.

A decision tree learning algorithm will search for space in trees by considering the trees first which test only for one feature and making them an immediate classification. Then they consider in expanding the tree by replacing one of the leaf node by a test case of second feature. Several heuristics methods are applied

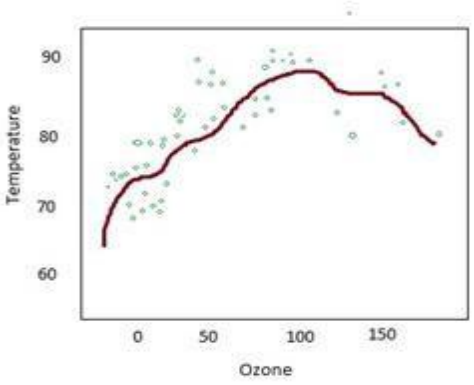
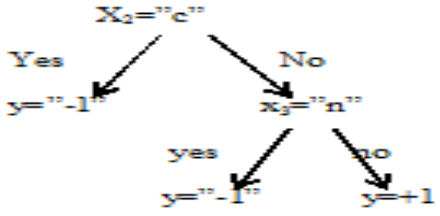


figure 1.5
Temperature

in choosing the test case for including in each iteration and when to stop designing of tree.

let us consider an decision tree for pronouncing the alphabet "K". At first the feature x2 is considered and tested whether it matches letter "c". If it doesn't match the feature x3 is taken and tested whether it matches with letter "n". The letter "k" will be pronounced only if x2 does not matches "c" and x3 does not matches "n" [4]. In addition to decision trees, there is more number



of other representations which includes perception, back propagation, etc. In all these cases, the algorithms find a best hypothesis h and give the output as the solution for the learning problem [5]. Ensemble learning algorithms have approach in different way. Rather than finding one best hypothesis for explaining the data, they construct a set of hypothesis which is called as ensemble or committee and based on that hypothesis vote in certain fashion to predict the label of new data points. To be precise, an ensemble method constructs a set of hypothesis {h1, h2,...,hk}, and a set of weights{w1,w2,…wk} and construct a voted classifier H(x)=w1h1(x)+w2h2(x)+…..wkhk(x).

3. REVIEW OF ENSEMBLE ALGORITHMS

Ensemble learning algorithm is based on running "base learning algorithm" multiple times and forming a vote among them for identifying the result hypothesis. There are two approaches of designing ensemble learning algorithms.

The first approach is to design each and individual hypothesis independently in such a way that the resulting set of hypothesis is accurate and diverse - that is each hypothesis has a considerable amount of less error rate for making new predictions and yet the hypothesis disagree with one another in most of the predictions. Such hypothesis of ensemble can be developed as they will be more accurate than any of the other component classifiers [6].

The second approach is to design ensembles for constructing the hypothesis in a coupled fashion in such a way that the weighted vote of the hypothesis gives a better fit for the data.

4. DRAWBACKS IN ENSEMBLE LEARNING

The major disadvantage in ensemble learning is that the loss of interaction may occur between nodes during interaction and learning. There is a possibility that some of the independently designed individual networks will not make enough contribution to the whole ensemble network [7].

5. ENSEMBLE LEARNING USING IOT

Let us consider there is a set of hypothesis h1, h2, .. hn. Let a new hypothesis hk enters. The hypothesis is predicted and verified by using the bagging and boosting concept of ensemble learning. Boosting involves building an ensemble by training each new model to instance to emphasize the training instances that previous model classified. When the entire hypothesis is connected over internet, the hypothesis are

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

predicted and analysed and we can arrive at a conclusion. For example in a retail shop, we cannot predict what the customer is wishing to buy. If all the purchases are connected over internet and predicative analysis is done using bagging and boosting can be done based on the user's wish list and the things he is viewing frequently and based on the analysis we can arrive at a conclusion of customers buying items [8].

CONCLUSION AND FUTURE ENHANCEMENT

To conclude, the concept of ensemble learning has some disadvantages like failure of independent node and no predicate analysis can be done about future, ensemble learning using Internet of Things (IoT) platform will be a better choice. The entire hypothesis is analysed and predicted using bagging and boosting of ensemble learning.

In future we will be working for the development of ensemble learning through different scheduling algorithms and trying to find the most efficient algorithm for ensemble learning using Internet of Things platform [9].

REFERENCES

[1] Thomas G. Dicterich, Oregon state university, to appear in the handbook of brain theory and neural networkshttp://www-vis.lbl.gov/~romano/mlgroup/papers/hbtnn-ensemble-learning.pdf

[2] Thomas G. Dicterich, Corvallis, USA, Ensemble methods in machine learning http://web.engr.oregonstate.edu/~tgd/publications/mcs-ensembles.pdf

[3] Martin Sewell, university college London, Ensemble learning http://machine-learning.martinsewell.com/ensembles/ensemble-learning.pdf

[4] Zhi - Hua zhou, A book review on ensemble methods: Foundations and algorithms reviewed by Dirk Van Den Poel http://www.comp.hkbu.edu.hk/~iib/2012/Dec/review/iib_vol13no1_review2.pdf

[5] Todd Holloway, Ensemble methods http://www.cbcb.umd.edu/~hcorrada/PracticalML/pdf/lectures/EnsembleMethods.pdf

[6] Converging Technologies for smart Environments and integrated Ecosystemshttp://www.internet-of-things-research.eu/pdf/Converging_Technologies_for_Smart_Environments_and_Integrated_Ecosystems_IERC_Book_Open_Access_2013.pdf

[7] Y.Liu, X.Yoa, the University of Birmingham, UK. Ensemble learning via negative correlation http://www.cs.bham.ac.uk/~pxt/NC/ncl.pdf

[8]http://www.gsma.com/connectedliving/wpcontent/uploads/2014/08/cl_iot_wp_07_14.pdf

[9] jayavadhana Gubbi, Rajkumar Buyya, Slaven Marusic, Marimuthu palanisamy on Internet of things(IOT) : A vision, architectural elements and future directions. http://www.buyya.com/papers/Internet-of-Things-Vision-Future2013.pdf