Multiple Activity Recognition through Ensemble Learner

Ruchi Pandey
Department of CSE [M.Tech] (S.E)
RITEE ,Raipur (C.G) India.
E-mail:ruchipandey831@gmail.com

Professor Shahana Gajala Qureshi
Department of CSE
RITEE, Raipur (C.G),India
E-mail: shahana.rit@gmail.com

Abstract :The Internet of Things (IOT) is an emerging research area promising many interesting solutions to various problems encountered in various domains. Smart homes applications is one such branch that evolve from IOT with the huge challenges of data storage and handling. Multiple activity recognition is the major challenge in smart homes application that incorporates multiclass learning. The efficiency of ensemble learner in handling multiclass problem and collective dissicion delivered prompt its uses in the smart homes application In this paper we deal with multiple activity recognition problem on various ensemble learner including bagging ,boosting and random forest. The standard van Kasteren dataset contains three housedata with eight activities of thirty days. We perform our experiment on the pre-processed collected data and applied six learners i.e. three from individual learner and three from ensemble learners. On performing the extensive experiment it was found that the group of ensemble learner outcast the simple learners.

Keywords: IOT, Ensemble learner, Bagging ,Boosting, Random forest.

I. Introduction and Motivation

A smart home is an application of present computing in which the home environment is monitored by ambient intelligence. Its technologies are partially connected with one additional current trend - the Internet of things (IOT). The process of activity recognition aims to recognize the actions and goals of one or more agents from a series of observations on the agents' actions and the environmental conditions. The process of activity reorganization can be achieved using sensors. Sensor-based activity always recollect, the emerging area of sensor networks with novel data mining and machine learning techniques to model a wide range of human activities.

Recognizing human activities based on sensor readings is interesting since sensors can capture living environments of humans and human-to-environment interactions. Identify activities of multiple users is more challenging than that of a single user .The key idea behind active learning is that a machine learning algorithm can achieve greater accuracy with fewer training labels if it is allowed to choose the data.

Active learning is well-defined in many modern machine learning problems. Different from traditional machine learning approaches, an approach based on data mining has been recently proposed. In that the problem of activity recognition is formulated as a pattern-based classification problem. They gave a data mining approach based on discriminative patterns which describe significant changes between any two activity classes of data to recognize sequential, interleaved and concurrent activities in a unified solution. Attention can then be focused on developing an effective technique for the multi-class case.

A key challenge in inferring human-centric notions of context from multiple sensors is the fusion of low-level streams of raw sensor data. With the availability of sensors and the advancement of wireless sensor networks, researchers in ubiquitous computing are recently interested in deploying various sensors to collect observations, and recognizing activities based on these observations. This in turn supports many potential applications such as monitoring activities of daily life.

In this paper the new methods explained here build upon these basics to construct more powerful prediction models, and remedy some of the drawbacks of classical methods. These methods include bagging, boosting and random forest.

1.1Bagging

It turns out that bootstrapping can be used in a different context that usual for improving tree based learning methods. Recall that decision trees suffer from high variance. This means that if we split the training data into two parts at random, and fit a decision tree to both halves, the results that we may get could be quite different. What we really want is a result that has low variance if applied repeatedly to distinct data sets. Bootstrapping of course is a natural solution, since it is designed as a general-purpose procedure for reducing variance. Bagging is essentially taking repeated samples from the single training set in order to generate B different bootstrapped training data sets. We then train our method on the both training set and average all the predictions.

For decision trees, we simply construct B regression trees using B bootstrapped training sets, and average the resulting predictions. The trees are grown deep and are not pruned. Thus each individual tree has high variance, but low bias. Averaging these B trees reduces the variance. Bagging can dramatically reduce variance by combining hundreds or thousands of trees into a single procedure. For a qualitative Y, the simplest solution is for a given observation, we record the class predicted by each B tree and take a majority vote. Generally, the value of B is not critical. In practise we use a value of B sufficiently large that the error has settled down.

1.2 Boosting

Boosting is another approach for improving the prediction power from a decision tree. Boosting works similarly to bagging except that the trees are grown sequentially each tree is grown using information from previously grown trees. Boosting also does not involve bootstrapping, instead each tree is fit on a modified version of the original data. Instead of fitting a single large decision tree, which results in hard fitting the data, and potentially overfitting. The boosting approach learns slowly. Given the current model, we fit a decision tree to the residuals from the model, rather than the outcome Y. We then add this new decision tree into the fitted function in order to update the residuals. Each of these trees can be rather small, determined by a tuning parameter d. By fitting small trees to the residuals, we slowly improve the fit in areas where it does not perform well. A second tuning parameter lambda slows the processes even further by allowing more and different shaped trees to attack the residuals.

1.3 Random Forest

Random forest (RF) methodology is a machine learning technique useful for prediction problems. The RF algorithm, developed by Leo Breiman , applies bootstrap aggregation (bagging). There are many studies showing that RFs have impressive predictive performance in regression and classification problems in various fields, including financial forecasting, remote sensing, and genetic and biomedical analysis . The research work in the area of random forest aims at either improving accuracy, or reducing time required for learning and classification or both. Random forest is now known to be one of the most efficient classification methods. By minimizing the correlation among these trees, the classification accuracy of the random forest can be improved. RANDOM FOREST ALGORITHM Random forest is an ensemble classification method by voting the result of individual decision trees.

In summary, the paper makes the following contributions.

–The aim is to capture the perfect result. With the help of sensor we are capable of capturing the observations of human activity. Based on this platform we conduct and evaluation of the results through different classifiers mention above to, demonstrate the result accuracy.

– We investigate the problem of sensor-based, multi-user activity recognition in a smart home setting, and propose a method using different classifiers to get the accuracy in the result.

– We conduct experimental studies to evaluate our proposed model for multi-user Activity recognition The rest of the paper is organized as follows. Section 2 description of the literature survey. In Section 3, description of problem identification. Section 4 describes our proposed activity model, and Section 5 and 6 reports our empirical studies. Finally, Section 6 concludes the paper.

II. LITERATURE SURVEY

Table 2.1: Literature survey

S.NO	PAPER	YEAR	REMARK
1	*In Defence of One vs. All Classification*	2004	*It consider the problem of multiclass classification*
2	*Accurate Activity Recognition in Home setting*	2008	*A sensor system capable of automatically recognizing activities would allow many potential ubiquitous applications.*
3	*A Study of the Behaviour of Several Methods for Balancing Machine Learning Training Data*		*Our experiments provide evidence that class imbalance does not systematically hinder the performance of learning systems.*
4	*Improving Classification Accuracy based on Random Forest Model with Uncorrelated High Performing Trees*	2014	*Random forest can achieve high classification performance through a classification ensemble with a set of decision trees that grow using randomly selected subspaces of data.*
5	*Layered Representations for Human Activity Recognition*	2010	*It present the use of layered probabilistic representations using Hidden Markov Models for performing sensing ,learning, and inference at multiple levels of temporal granularity.*
6	*Ensemble learning for HMM*	2009	*The standard method for tanning HMM optimizes a point estimate of a model.*
7	*A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches*	2012	*Results show empirically that ensemble-based algorithms are worthwhile since they outperform the mere use of pre-processing techniques before learning the classifier, therefore justifying the increase of complexity by means of a significant enhancement of the result.*
8	*ACTIVITY RECOGNITION ON STREAMING SENSOR DATA*	2015	*The experiments conducted to evaluate these techniques on real-world smart home datasets suggests that combining mutual information based weighting of sensor events and adding past contextual information into the feature leads to best performance for streaming activity recognition.*
9	*An empirical evaluation of bagging and boosting*	2015	*In this paper both bagging and booting are classified using neural network and decision tree as the classification algorithm.*
10	*A Short Introduction to Boosting*	2000	*Boosting is a general method for improving the accuracy of any given learning algorithm*

2.1Literature summary

In Defence of One vs. All Classification: Its main thesis is that a simple "one-vs.-all" scheme is as accurate as any other approach, assuming that the underlying binary classifiers are well-tuned regularized classifiers such as support vector machines. This thesis is interesting in that it disagrees with a large body of recent published work on multiclass classification.

Accurate Activity Recognition in Home setting: In this paper, we present an easy to install sensor network and an accurate but in expensive annotation method. We achieve a timeslice accuracy of 95.6% and a class accuracy of 79.4%.

A Study of the Behaviour of Several Methods for Balancing Machine Learning Training Data: Results show that these trees are usually more complex than the ones induced from original data. Random over-sampling usually produced the smallest increase in the mean number of induced rules and Smote + ENN the smallest increase in the mean number of conditions per rule, when compared among the investigated over sampling method.

Improving Classification Accuracy based on Random Forest Model with Uncorrelated High Performing Trees: In this paper an attempt has been made to improve the performance of the model by including only uncorrelated high performing trees in a random forest.

Layered Representations for Human Activity Recognition: We review the representation, present an implementation, and report on experiments with the layered representation in an office-awareness application.

Ensemble learning for HMM: In this paper it derive and test and ensemble learning algorithm for HMM building on Neal and Hinton observation that expectation maximization algorithm can be viewed as variation free energy minimization method.

A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches In this paper, our aim is to review the state of the art on ensemble techniques in the framework of imbalanced data-sets, with focus on two-class problems. We propose taxonomy for ensemble-based methods to address the class imbalance where each proposal can be categorized depending on the inner Ensemble methodology in which it is based.

Activity Recognition on Streaming Sensor Data: In this paper we propose and evaluate a sliding window based approach to perform activity recognition in an on line or streaming fashion; recognizing activities as and when new sensor events are recorded.

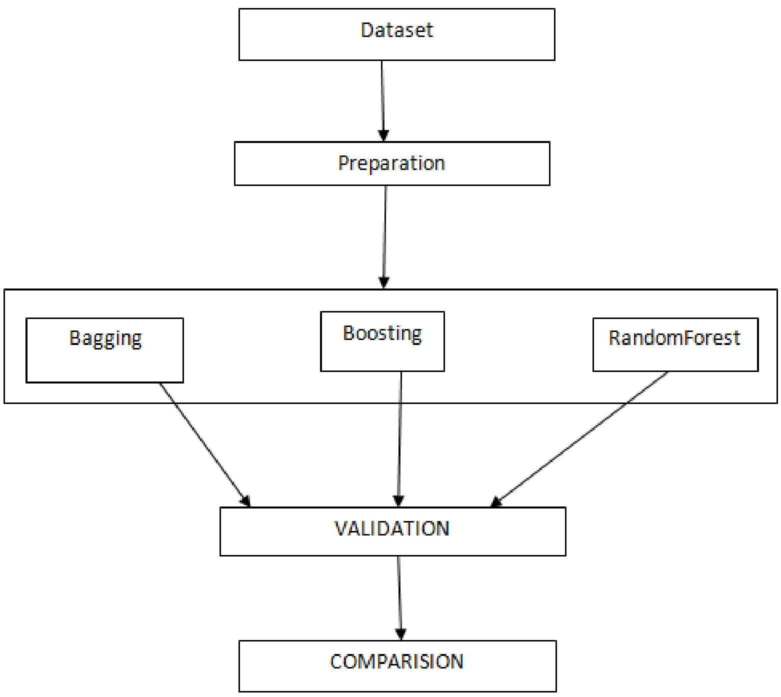
An empirical evaluation of bagging and boosting: The result clearly shows the important fact booting is a powerful technique that produces a better ensembles then bagging.

A Short Introduction to Boosting: This short overview paper introduces the boosting algorithm Ada Boost, and explains the underlying theory of boosting, including an explanation of why boosting often does not suffer from over fitting as well as boosting's relationship to support-vector machines.

III. PROBLEM IDENTIFICATION

The objective is to produce a multi-class stream ensemble method using bagging ,boosting and random forest classifiers as the base learner. Compare to the other classifiers is result is more prompt. This method is easy to implement and has a simple conceptual justification.

IV. METHODOLOGY



V. EXPLAINATION

5.1 Dataset

This paper addresses the standard kasteeren dataset and established benchmarking problems for physical activity monitoring. A new dataset of activity reorganization recorded with the sensors that include number of classes and instances. The benchmark shows the difficulty of the classification tasks and exposes new challenges for physical activity monitoring.

5.2Classifiers

1) **Bagging:** Bagging is a ``bootstrap" ensemble method that creates individuals for its ensemble by training each classifier on a random redistribution of the training set. Each classifier's training set is generated by randomly drawing, with replacement, N examples - where N is the size of the original training set; many of the original examples may be repeated in the resulting training set while others may be left out. Each individual classifier in the ensemble is generated with a different random sampling of the training set.

2) **Boosting:** Boosting is a machine learning ensemble meta-algorithm for primarily reducing bias, and also variance in supervised learning, and a family of machine learning algorithms which convert weak learners to strong ones. A weak learner is defined to be a classifier which is only slightly correlated with the true classification (it can label examples better than random guessing). In contrast, a strong learner is a classifier that is arbitrarily well-correlated with the true classification.

3) **Random Forest:** Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set

Validation:

Comparision: All the results that are achieved using the above classifiers are compared with the actual result. This comparison shows the accuracy of the result using different classifier.

VI. EXPIRIAMENTAL SETUP AND RESULT

VII. EVALUATION

1)Accuracy:
2)Precetion
3)Re-call
4) F-measure
5)G-mean

REFERENCES

[1]Technologies for an Aging Society: A Systematic Review of "Smart Home" Applications G. Demiris1, B. K. Hensel21Biomedical and Health Informatics, University of Washington, Seattle, WA, USA 2Health Management and Informatics, University of Missouri-Columbia, Columbia, MO, USA

[2] Multioccupant Activity Recognition in Pervasive Smart Home Environments ASMA BENMANSOUR, University of Tlemcen ABDELHAMID BOUCHACHIA, Bournemouth University MOHAMMED FEHAM, University of Tlemcen

[3] Activity Recognition on Streaming Sensor Data Narayanan C Krishnan and Diane J Cook

[4] A Short Introduction to Boosting Yoav Freund Robert E. Schapire AT&T Labs Research Shannon Laboratory 180 Park Avenue Florham Park, NJ 07932 USA www.research.att.com/_fyoav, schapiregfyoav

[5] An empirical evolution on Bagging and boosting.RichardMaclin,DaidOptiz

[6] A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches Mikel Galar, Alberto Fern´andez, EdurneBarrenechea, Humberto Bustince, Member, IEEE,and Francisco Herrera, Member, IEEE

[7] Ensemble learning for HMM. David J.C Mac Kay

[8] Layered Representations for Human Activity Recognition Nuria Oliver Eric Horvitz Adaptive Systems & Interaction Microsoft Research Redmond, WA98052,USAfnuria,horvitzg@microsoft.com AshutoshGarg Beckman Institute Univ. Illinois Urbana-Champaign Champaign, IL 61801

[9] Improving Classification Accuracy based on Random Forest Model with Uncorrelated High Performing Trees S. Bharathidason Department of Computer Science Loyola College, Chennai -34 C. JothiVenkataeswaran, Ph.D Department of Computer Science Presidency College, Chennai -05

[10] A Study of the Behavior of Several Methods for BalancingMachine Learning Training Data Gustavo E. A. P. A. BatistaRonaldo C. PratiMaria Carolina MonardInstituto de Ciˆencias Matem´ aticas e de Computac¸ ˜aoCaixa Postal 668, 13560970S˜ao Carlos SP,Brazil{gbatista, prati.

[11] Accurate Activity Recognition in a Home SettingTim van Kasteren, AthanasiosNoulas, Gwenn Englebienne and Ben Kr¨oseIntelligent Systems Lab AmsterdamUniversity of AmsterdamKruislaan 403, 1098 SJ, Amsterdam , The NetherlandsT.L.M.vanKasteren@uva.nl