International Journal of Research in Advent Technology (IJRAT)
(E-ISSN: 2321-9637)
Special Issue National Conference "CONVERGENCE 2017", 09th April 2017

Security of Big Data Using Multitier Classifier: A Review

Sujit R. Borey[1], Dr.P.M Jawandhiya [2].
M.E.CSE PLIT [1] ,PRINCIPAL PLIT [2]
Email:sujeetbore3@gmail .com[1] , principal@plit.ac.in[2]

Abstract- This topic introduces Big Data achieves more attention from researchers in recent years because it has become in numerous application. The svm (support vector machine) and J48 classifiers with base classifiers for improving performance of classification. Svm is higher accuracy and it can produce powerful result in range from excellent. The planed LIME classifier is large because it is tailored for handling Big Data. These ensemble classifiers are combined at each tier. Next tier will collect output from previous tier, combine and analyses to and send result in next tier. Multitier are use to many tiers, work are divided into each of these tiers so that speed and accuracy increases. So easy to run. It included many ensemble classifiers to several levels. This classifier is also use to security of Big Data. It is generated automatically as a result of several iteration in applying ensemble Meta classifiers. Ensemble Meta classifier into several tiers simultaneously and combine them into one generated iterative system so that many ensemble Meta classifiers function as integral parts of other ensemble Meta classifiers at Higher Tiers.

Index Terms- : j48 and svm, LIME classifiers, Multitier.

1. INTRODUCTION

This five-tier Large Iterative Multitier Ensemble (LIME) classifiers Designed for application concerning information of security of Big Data. The aim of this is to develop LIME classifiers as a general technique that may be useful for the analysis of Big Data in various application domains. Technology to extract the knowledge from the pre-existing databases. It is used to explore and analyses the same data. The data which is to be mined varies from a small data-set to a large data-set i.e. Big Data. Big data is so large that it does not fit in the main memory of a single machine, and it need to process big data by efficient algorithms [1]. Modern computing has entered the era of Big Data. The investigation of this new construction is important, because the role of algorithms for analysis of Big Data has been growing. It helps in improving security of big data.

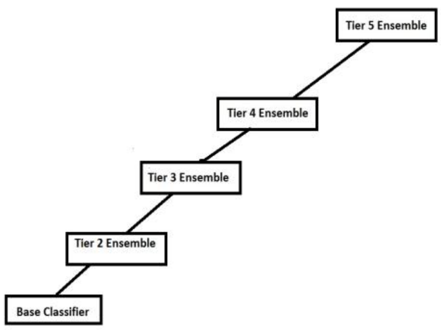

Fig.1.Initialization of five-tier classifier.

This demonstrates that our new technique of combining diverse ensemble Meta classifiers into one unified five-tier ensemble incorporating diverse ensemble Meta classifiers as elements of different ensemble Meta classifiers can be applied to enhance classifications.

2. RELATED WORK

Big data is small loss in accuracy was observed. Researchers in proposed the four-tier Large Iterative Multitier Ensemble (LIME) classifier which is used for security of the big data.

- This classifier puts the idea of combining multiple classifiers at several levels [1].
- The first classification method integrating static and dynamic features into a single test was presented.
- The approach proposed there improved on previous results using individual features collected separately. The time required for the test was reduced by half.

The paper investigates an iterative hierarchical key exchange scheme for secure scheduling of big data. The lead to a reduction in false negative rates by factor[2]. The speed of running the tests improved by factor of approximately. The privacy preservation over big data considered.

1. ***Motivating and Research***

There are multipal base classifiers available in WEKA. BayesNet, FURIA, SVM, J48, DTNB, Random Forest, Multiboost.

A. ***Random Forest***:

The Random Forest builds a forest of random trees by generating many decision tree predictors with randomly selected variable subsets and utilizing a different subset of training and validation data for each of these trees, as partitioning[7].

To control the variation in creating the set of random trees, Random Forest uses the process of random selection of features proposed. After creating many trees, the resulting class prediction is based on votes from the strees [7]. The variables are ranked and variables with lower rank are eliminated based on the basis of empirical performance heuristics.

B. ***MultiBoost:*** The MultiBoost use to extend the different approach of the Adaboost with wagging technique [6]. The Wagging is variant for bagging with the weights for training instance permeation during boosting are used in select the bootstrap samples. The diverse collect of UCI data sets to demo on Multiboost achieves high accuracy significantly more often the wagging [6].

3. PROPOSED WORK AND OBJECTIVES

This classifier will mix numerous ensemble Meta classifiers into one reiterative gradable system and may be terribly massive. Each ensemble Meta classifier combines a group of base classifiers into a typical arrangement. Economical multi-tier classifiers and additional general multi-classifier systems are explored, for instance, within the previous publications [4]. Our construction of this classifier was impressed by previous analysis within the literature, however is completely different. It is simple to set up and generate this classifier. All fourth tier ensemble Meta classifiers are generated by the fifth tier ensemble Meta classifier given just one instance of a third tier ensemble as an input parameter for the generation stage. The forth tier ensemble meta classifier generates all third tier ensemble meta classifiers and executes them in exactly identical manner because it usually handles base classifiers[5]. Similarly, every third tier ensemble Meta classifier applies its method to generate and mix its second tier ensemble Meta classifiers. Finally, the second tier ensemble Meta classifiers generate, execute and combine their base classifiers in their standard fashion.

To start the method a designer should initialize a five-tier classifier by specifying which ensemble Meta classifier can operate at the fifth tier. Then the designer provides a parameter to the fifth tier ensemble Meta classifier indicating, which fourth tier ensemble Meta classifier is to be used as a part of the quality generation method of the fifth tier ensemble Meta classifier. After that, the designer specifies the second tier ensemble Meta classifier method to be used by the third tier ensemble meta classifier, and the base classifier handled by the second tier ensemble meta classifier.

4. DESIRED IMPLICATIONS

In this paper we used various ensemble Meta classifiers and base classifiers implemented within the Waikato environment for knowledge Analysis (WEKA). After choosing proper choices, the whole system is generated automatically by the SimpleCLI, using the embedded iterative and recursive capability of Java programming. The base classifiers analyze the features of the main instances and pass on their output to the second tier ensemble Meta classifiers. The second tier ensemble Meta classifiers collect all outputs of the base classifiers, combine them, and send their own output to their parent third tier ensemble Meta classifiers. As same the third tier ensemble Meta classifiers gathered the outputs of the second tier ensemble Meta classifiers analyze and combine them, and send their own output to the fourth tier ensemble Meta classifier. The Fourth tier ensemble classifier collects the output from third tier Meta classifier and send their own output to the fifth tier ensemble Meta classifier.

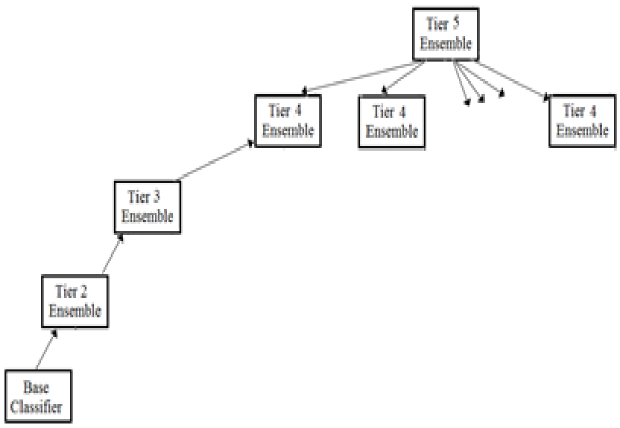

Fig. 2 A generating Five-tier classifier.

The fifth tier ensemble Meta classifier analyses the results of the third tier ensemble Meta classifiers and produces the final decision of the classifier.

Our work shows that large five-tier LIME classifiers are quite easy to use and can be applied to improve classifications, if diverse ensemble Meta classifiers are combined at different tiers. It is an interesting question for future research to investigate

LIME classifiers for other large datasets. Random Forest outperformed other base classifiers for the malware data set, and Decorate improved its outcomes better than other ensemble Meta classifiers did. The best outcome of AUC 0.998 was obtained by the fourth-tier LIME classifier where Multi Boost was used at the fourth tier, Decorate was used at the third tier and Bagging was applied at the second tier[3]. The performance of ensemble Meta classifiers considered in this paper depends on several numerical input parameters. In all experiments we used them with the same default values of these parameters in order to have a uniform equivalent comparison of outcomes across all of these ensemble Meta classifiers.

❖ ***SVM:*** Support Vector Machines (SVM) recently became one of the most popular classification ways. They have been employed in a good variety of applications. Support Vector Machines may be thought of as a technique for constructing a special kind of rule, known as a linear classifier, in a way that produces classifiers with theoretical guarantees of excellent predictive performance (the quality of classification on unseen data)[2]. The theoretical foundation of this technique is given by statistical learning theory.

❖ ***J48*:** Creates a C4.5 decision tree by adding attributes to the tree as explained. At every step the feature with the highest information gain is added. This means that every next attribute is chosen so that it is best in discriminating the instances in the training set[1].

5. CONCLUSION

New evaluating performances of such large five-tier classifiers. In SVM and J48 is performed best in this setting, in five-tier LIME classifiers will used to more improvement of the classification outcomes. This five-tier classifiers supported svm and J48 achieved higher performance compared the other classifier it gives best result

REFERENCES

1. Jemal H. Abawajy, Andrei Kelarev, Morshed Chowdhury," Large Iterative Multitier Ensemble Classifiers for Security of Big Data" , in IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING, 30 October 2014.
2. Laura Auria, Rouslan A. Moro, "Support Vector Machines (SVM) as a Technique for Solvency Analysis", in Berlin, August 2008.
3. C. Liu et al."An iterative hierarchical key exchange scheme for secure scheduling of big data applications in cloud computing," in Proc. 12th IEEE Int. Conf. Trust Security Privacy Comput. Commun. Melbourne, Australia, Jul. 2013, pp. 9-16.
4. R. Islam, J. Abawajy, and M. Warren, "Multi-tier phishing email classification with an impact of classifier rescheduling," in Proc. 10th ISPAN, 2009, pp. 789-793.
5. R. Islam and J. Abawajy, ``A multi-tier phishing detection and filtering approach," J. Netw. Comput. Appl., vol. 36, no. 1, pp. 324_335, 2013.
6. G. I. Webb, ``MultiBoosting: A technique for combining boosting and wagging," Mach. Learn., vol. 40, no. 2, pp. 159-196, 2000
7. L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5-32,2001.