Intrusion Detection based on Incremental Combining Classifiers

Dipali Bhosale
Department of Computer Engineering,
Dr. D. Y. Patil School of Engg. & Tech., Pune
Savitribai Phule Pune University, Pune

Roshani Ade
Department of Computer Engineering
Dr. D. Y. Patil School of Engg. & Tech., Pune
Savitribai Phule Pune University, Pune

ABSTRACT
Intrusion detection (ID) is the task of analysis the event occurring on a network system in order to detect abnormal activity. Intrusion Detection System has increased due to its more constructive working than traditional security mechanisms. As the network data is dynamic in nature, it leads to the problem of incremental learning of dynamic data. Now, combining classifiers is a new method for the improving classifiers robustness and accuracy. Most of ensemble methods operates in batch mode. For this purpose, proposed system incremental combining classifiers that combines three classifiers that operates incrementally on dynamic data, Naïve Bayes, K-star, Non Nested Generalised Exemplars classifiers based on voting approach. In incremental learning process, numbers of hypotheses are generated during classification; an ensemble decision method is required to aggregate all the votes from multiple hypotheses for the final decision process which produces better accuracy in most of the cases in experiments.

General Terms
Security, Algorithms

Keywords
Intrusion detection, incremental, combining classifier, naïve bayes, k-star

1. INTRODUCTION
An intrusion detection method is a component of the information security and its main aim is to detect normal activities of the system and behaviour that can be classified as abnormal activity [1]. ID method required because of the large number of abnormal activity reported every year. Training data does not predict better performance in classification task. Number of classifiers with same training performances may or may not have different performance in classification. In such cases, aggregate the outputs of different classifiers may reduce the risk of poorly performing classifier. The averaging may beat the performance of the better classifier in the classifier combining method, but it certainly omitted the risk of making a accurate weak selection. The strategy in ensemble decision systems is to create multiple classifiers, and combine these outputs for improving the performance of individual classifier.

This needed because; single classifier can make errors on different instances. The purpose of this is, if each and every classifier makes different errors, then combination of all classifiers can decrease the total amount of error. Incremental learning task has attracted from both academia and industry. Ensemble learning has become active research, within the computational intelligence community. Ensemble decisions learning also have been widely used in many real applications, including decision making, supporting system, financial engineering, Web mining, remote sensing [2, 3]. There is no matter what type of mechanisms are applied to generate the number of classifiers, a classifier combination is needed to aggregate all the votes for the final decision making task. Ensemble decision learning has benefit of better accuracy and robustness compared to the individual hypothesis [4-6]. In the ensemble decision learning [7-9] scenario, different models are generated, and these decisions are aggregated with a classifier combination method to predict the testing instances. Because multiple hypotheses for multiple classifiers may provide different views of the targeted function, the combined decision of each and every classifier will provide robust and accurate prediction compared to the individual.

Some developments in ensemble decision learning, like Elite, an ensemble decision learning algorithm based on a global optimization Trusttech, is proposed to analyse good quality ensembles decision [10]. In this technique, Trusttech is applied to detect two important issues in neural network techniques, i.e. network architecture criteria selection and optimal weight of training data invtances. A Bayesian algorithm artificial immune system (BAIS) is used to learn the ensemble decision of neural networks to detect classification problems in paper [11, 12]. This shows, BAIS is applied to generate a high quality networks and then combine classifiers on the basis of some rule.

The rest of this paper is organized as, section 2 gives detail information of classifier combination approach. Section 3 includes proposed work with the system architecture. Section 4 deals with the result comparison based on various parameters. Section 5 includes the conclusion of overall work with future scope.

2. CLASSIFIER COMBINATION
In paper [13], there are 3 methods in which combining classifiers can achieve good performance and these are statistical, representational and computational. The analysis starts with the observing any learning algorithm which try to find high accuracy on the training data instances. Constructing ensemble decision learning is out of all these classifiers can allow the algorithm to omit the risk. The effective combining rules [14] which are encapsulate below contains AA rule, Majv rule, Max rule, Min rule and BC rule.

GA rule

GA rule uses P (Ym | xt) to reduce the average Kullback–Leibler (KL) separation between probabilities

$$D_{ab} = \frac{1}{L}\sum_{n=1}^{L} D_n \quad (1)$$

where

$$D_n = \sum_{m=1}^{C} P(Y_n|x_t) \ln \frac{P(Y_n|x_t)}{P_j(Y_n|x_t)} \quad (2)$$

Lagrange multiplier ∑Cm=1 P(Ym|xt)= 1, optimization of (1) w. r. t. P(Ym | xt) :

$$P(Y_m|x_t) = \frac{1}{A} \prod_{n=1}^{L} (P_n(Y_m|x_t))^{\frac{1}{L}} \quad (3)$$

where, A is a class-independent number.

Based on (3), GA rule divine the testing instant xt to the class identity label that maximizes the product of Pn (Ym | xt).

GA Rule:

$$x_t \rightarrow Y_m \text{ satisfy } \max_{Y_m} \prod_{n=1}^{L} P_n(Y_m|x_t) \quad (4)$$

AA rule

The probability distance of an alternative KL separation a follows:

$$D_n = \sum_{m=1}^{C} P_n(Y_m|x_t) \ln \frac{P_n(Y_n|x_t)}{P(Y_n|x_t)} \quad (5)$$

$$P(Y_m|x_t) = \frac{1}{L} \sum_{n=1}^{L} P_n(Y_m|x_t) \quad (6)$$

Therefore, the AA rule explained as finding the maximum value of the arithmetic average of Pn (Ym | xt).

$$x_t \rightarrow Y_m \text{ satisfy } \max_{Y_n} \frac{1}{L} \sum_{n=1}^{L} P_n(Y_m|x_t) \quad (7)$$

MV rule

In the condition of probability of Pn (Ym | xt), the AA rule may have weak combination performance. In such a case, MV rule will predict final class label along with max median value.

MV Rule:

$$x_t \rightarrow Y_m \text{ satisfy } \max_{Y_n} \{\text{median}(P_n(Y_m|x_t))\} \quad (8)$$

MajV rule

Each and every single classifier can directly predicts the class label of the testing instances.

$$x_t \rightarrow Y_m \text{ satisfy } \max_{Y_n} \sum_{n=1}^{L} \Delta_n(Y_m|x_t) \quad (9)$$

Where, $\Delta_n(Y_m|x_t) = \begin{cases} 1, & \text{if } x_n(x_t) = Y_n \\ 0, & \text{otherwise} \end{cases}$

Max rule

Max rule uses data provided by the max value of Pn (Ym | xt) over class labels.

$$x_t \rightarrow Y_m \text{ sati sfy } \max_{Y_n} \{\max_{n} (P_n(Y_m|x_t))\} \quad (10)$$

Min rule

Like Max rule, Min rule uses vote from final predicted class label which have max of the min values of Pn (Ym |xt) across all potential class labels.

$$x_t \rightarrow Y_m \text{ satisfy } \max_{Y_n} \{\min_{n} (P_n(Y_m|x_t))\} \quad (11)$$

BC rule

The BC rule uses categorized order of class labels which are provided by separate Pn (Ym |xt). Based on the classifier output, each classifier ranks class labels.

$$x_t \rightarrow Y_m \text{ satisfy } \max_{Y_n} \sum_{n=1}^{L} \Omega_n(Y_m|x_t) \quad (12)$$

where Ωn (Ym |xt) = C −p if classifier h n categorized xt in pth position of class label Yn, and C is the number of classes.

There are 2 censorious issues in such weighted combining methods; 1. Classifier variety 2. Combining weights. For instance, if all classifiers in an ensemble decision learning system generates same vote then classifier cannot give any benefit by integrating all decisions from each and every individual classifier.

For this reason it is most important to understand how to create multiple classifier ensembles decision and measure such variety of ensemble decision learning method.

3. PROPOSED WORK

This section describes detail flow of system with algorithm for proposed work. Here, on ID data user has to apply incremental learning that includes K-star, NNge, NB classifiers, then apply the voting rule (Average of Probability) to combine the votes from different classifier is shown in fig. 1 and it represents proposed structure for intrusion detection system.

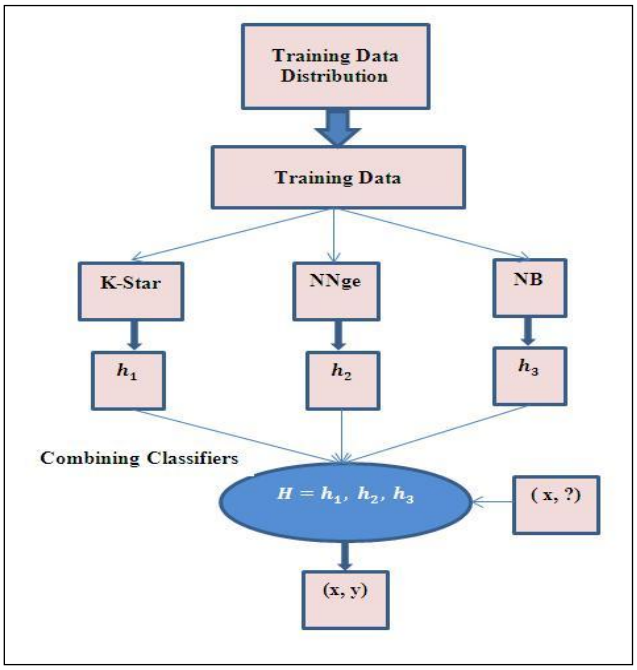

Fig 1: Proposed approach for combining classifier

Incremental learning algorithms allow dynamic approach for these types of datasets which learns data incrementally and produces multiple hypotheses. The approach of combining classifiers is proposed new way for improvement the robustness and classification accuracy. In most of the cases, ensemble methods operates in batch mode, to avoid this a new method has been produce to maintain a dataset and to produce ensemble for classifier.

In this paper, incremental learning combines three classifiers: Naïve Bayes, K-star, NNge based on voting approach. Experiments performed based on accuracy for different 11 dataset and two well-known classifiers: SMO and IBK. In most of cases, proposed system gives better results compare to other classifier.

It is computationally more costly than other techniques, because, it appears fast using standard computational power which is available now days, to accommodate just incremental applications in the real world.

4. COMPARISION AND RESULTS

In this section, for comparison purpose dataset comes from UCI Repository: contact-lenses, credit g, diabetes, glass, ionosphere, iris, labor, segment-challenge, vote, soyabean, NSL-KDD [15]. Table 1 gives short description of these datasets based on number of instances, attributes and classes in dataset. From these datasets some datasets are binary it means they have only two classes and some has multiple classes.

In order to compute the final accuracy of classifier, the whole set of data has been divided into 10 equal sized subset of data. For performing the operation on these data subset cross validation runs 10 times for each and every classifier then finally mean value has been calculated. During first experiment, each incremental learning classifier (k-star, NNge, NB) is compared with proposed system.

Table 1. Dataset Description

Dataset	Instances	Attributes	Classes
contact-lenses	24	5	3
credit g	1000	21	2
diabetes	768	9	2
glass	214	10	7
ionosphere	351	35	2
iris	150	5	3
labor	57	17	2
segment-challenge	1500	20	7
vote	435	17	2
soyabean	683	36	19
NSL-KDD	500	42	2

Table 2. Accuracy for different datasets on different classifiers (in %)

Dataset	SMO	IBK	Proposed
contact-lenses	70.83	79.16	70.83
credit g	75.10	72.00	74.50
diabetes	74.30	70.18	**74.73**
glass	56.07	70.56	**76.63**
ionosphere	88.60	86.32	**92.59**
iris	96.00	95.33	**96.00**
labor	89.47	82.45	**91.28**
segment-challenge	91.93	96.20	**96.63**
vote	96.09	92.41	93.79
soyabean	91.85	91.21	**92.38**
NSL-KDD	96.20	98.20	**97.20**

Table 3. Comparison of proposed system with well-known classifiers

Dataset	NB	KStar	NNge	Proposed
contact-lenses	70.83	70.83	70.83	**70.83**
credit g	75.40	69.40	70.50	74.50
diabetes	76.30	69.10	73.95	74.73
glass	48.59	75.23	70.09	**76.63**
ionosphere	82.62	84.61	90.02	**92.59**
iris	96.00	94.66	96.00	**96.00**
labor	89.47	89.47	77.19	**91.28**
Segment-challenge	81.06	96.60	95.80	**96.63**
vote	90.11	93.33	96.09	93.79
soyabean	92.07	87.99	91.80	**92.38**
NSL-KDD	92.40	96.20	96.20	**97.20**

In this case, proposed system gives better results than other classifiers in case of six datasets and these are contact-lenses, glass, ionosphere, iris, labor, segment-challenge, soyabean, NSL-KDD and equal result in case of two datasets describe in Table 2 and figure 2. For credit g, diabetes and vote dataset this proposed system does not gives better results and results are 74.50, 74.73 and 3.7 respectively.

Based on comparisons with other well-known classifiers (SMO, IBK), proposed system gives better results in case of 7 datasets and gives equal result for iris dataset that describes in Table 3 and figure 3.

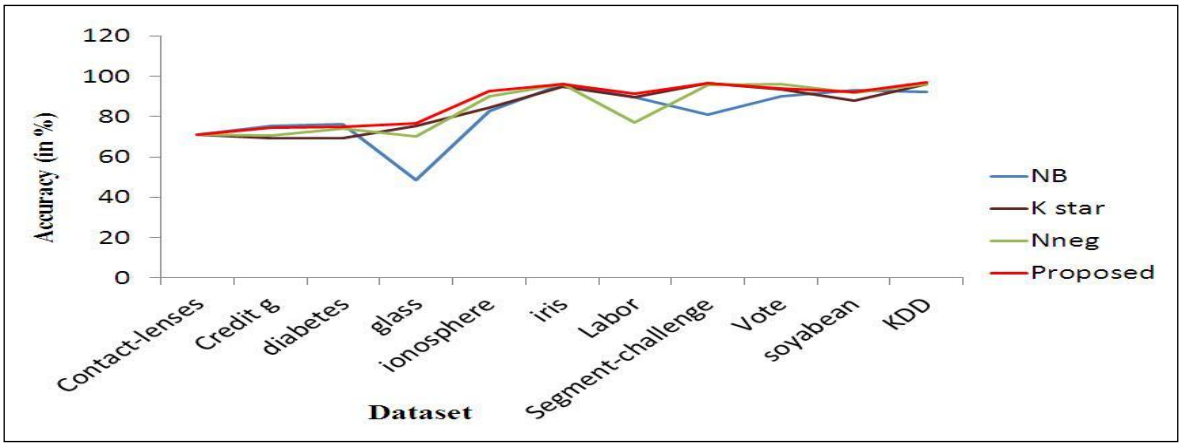

Fig 1: Accuracy for different datasets on different classifiers (in %)

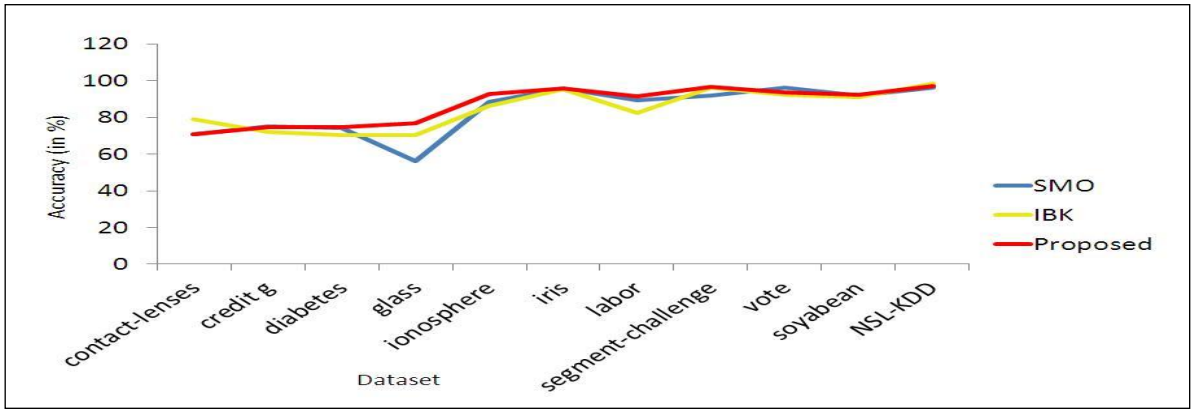

Fig 3: Accuracy for different datasets on different classifiers (in %)

5. CONCLUSION

The In areas such as, data mining, e-commerce, networks data generated dynamically over dynamic environment for classify these data new incremental learning algorithm has been proposed. These algorithms allow dynamic approach for these types of datasets which learns data incrementally and produces multiple hypotheses. The approach of combining classifiers is proposed new way for improvement the robustness and classification accuracy. In most of the cases, ensemble methods operates in batch mode, to avoid this a new method has been produce to maintain a dataset and to invoke it offline to produce ensemble for classifier. In this paper, incremental learning combines three classifiers: Naïve Bayes, K-star, NNge based on voting approach. Experiments performed based on accuracy for different 11 dataset and two well-known classifiers: SMO and IBK. In most of cases, proposed system gives better results compare to other classifier.

In future work, better voting rule can be applied to integrate the output of single classifier to provide final ensemble decision making task in classification, which can affects on final decision..

6. REFERENCES

[1] Zhang, Y., Lee, W. 2000. "Intrusion Detection in wireless ad-hoc networks", In Proceedings of the 6th annual international conference on Mobile Computing and networking, ACM , 275-283.

[2] Calinon, S., Billard, A. 2007. "Incremental Learning of gestures by imitation in a humanoid robot", In Proceedings of the ACM/IEEE international conference on human-robot interception, 255-263.

[3] R. Elwell and R. Polikar. 2011. "Incremental learning of concept drift in nonstationary environments", In IEEE Trans. Neural Netw., vol. 22, no. 10, 1517-1531.

[4] Wang, H., Fan, W., Yu, P. S., Han, J. 2003. "Mining concept-drifting data streams using ensemble classifiers", In Proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining. ACM. 226-235.

[5] Kittler, M. Hatel, R. P. W. Duin, and J. Matas. 1998. "On combining classifiers", in IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 3, 226-239.

[6] Hulten, G., Spencer, L., Domingos, P. 2001. "Mining time-changing data streams", In Proceedings of the 7th ACM SIGKDD international conference on knowledge

discovery and data mining. ACM. (August, 2001), 97-106.

[7] Haibo He, Sheng Chen, Kang Li And Xin Xu. 2011. "Incremental Learning From Strem Data", In IEEE Transaction On Neural Networks, vol. 22 No. 12., (December, 2011).

[8] Syed, N. A., Sung, K. K. 1999. "Handling Concept drift Incremental Learning with support vector machine", In Proceedings of the 5th ACM SIGKDD international conference on knowledge discovery and data mining. ACM. (August, 1999), 317-321.

[9] Street, W. N., Kim, Y. 2001. "A streaming ensemble algorithm (SEA) for large-scale classification", In Proceedings of the 7th ACM SIGKDD international conference on knowledge discovery and data mining. ACM. (August, 2001), 377-382

[10] Ade, R., & Deshmukh, P. "Efficient Knowledge Transformation for Incremental Learning and Detection of New Concept Class in Students Classification System", In Information Systems Design and Intelligent Applications. Springer India. 757-766.

[11] Ade, Roshani, and P. R. Deshmukh. "An incremental ensemble of classifiers as a technique for prediction of student's career choice", ICNSC.

[12] M. D. Muhlbaier, A. Topalis, and R. Polikar.. "Learn ++.NC: Combining ensemble of classifiers with dynamically weighted consult-and-vote for efficient incremental learning of new classes", In IEEE Trans. Neural Netw., vol. 20, no. 1, , (Jan., 2009). 152-168.

[13] Janssens D, Brijs T, Vanhoof K,Wets G. "Evaluating the performance of cost-based discretization versus entropy- and error-based discretization", Comput Oper Res 33(11), 3107–3123

[14] N. Littlestone and M.Warmuth. "Weighted majority algorithm", Information and Computation, vol. 108, 212-261.

[15] NSL-KDD dataset for network-based intrusion detection systems available on http://iscx.info/NSL-KDD/