International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 6, Issue 3, May- June 2017 ISSN 2278-6856

Least Square Support Vector Machine based IDS, using feature selection algorithm

Rekha Preethi M.C[1], Mr.Chetan R[2]

[1]M-tech, 4th sem, ,SJBIT, Bangalore,

[2]Assistant Professor, Dept. of ISE,SJBIT, Bangalore.

Abstract: ***When the different users on the Internet access similar content which may be redundant or irrelevant data features which causes problems in network traffic classification. This retards the network traffic classification process and prevents to make accurate and optimal decisions when are dealing with big data. In this paper A hybrid feature selection algorithm is used for optimal feature classification and these mutual information based algorithms can handle both linearly and nonlinearly dependent data features. The results will be evaluated during network intrusion detection. The Least Square Support Vector Machine based IDS (LSSVM-IDS) which is an Intrusion Detection system and is developed using features of feature selection algorithm and its performance is evaluated using the data sets provided by KDD Cup 99 data sets. This improves accuracy and computational cost will be lowered as compared to other methods.***

Introduction:

Data mining is primarily used today by companies with a strong consumer focus.It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics.It enables them to determine the impact on sales, customer satisfaction, and corporate profits. Finally, it enables them to "drill down" into summary information to view detail transactional data.A feature selection algorithm can be seen as the combination of a search technique for proposing new feature subsets of huge sets, along with an evaluation measure which provides the different feature subsets. The simplest algorithm is to test each possible subset of features finding the one which reduces the error rate and increases the optimal solution rate.

Wrapper methods use a predictive model to score feature subsets and wrapper method searches for an optimal feature subset which is tested on a hold-out set. The error rate of the model gives the score for that subset.

Filter methods use a proxy measure instead of finding the error rate to score a featured subset. Common measures include the mutual information, the pointwise mutual information, Pearson product-moment correlation coefficient, inter/intra class distance or the scores of significance tests for each class/feature combinations. Filter methods have also been used as a pre-processing step for wrapper methods, allowing a wrapper to be used on larger problems.Embedded methods are a catch-all group of techniques which perform feature selection as part of the model construction process. One other popular approach is the Recursive Feature Elimination algorithm, commonly used with Support Vector Machines to repeatedly construct a model and remove features with low weights

Related work:

Gisung Kim et.al, [1] presents a new hybrid intrusion detection method that hierarchically combines a misuse detection and anomaly detection in a decomposed structure. The decision tree was used to create the misuse detection model that is used to disintegrate the normal training data into smaller subsets. Then, the one-class support vector machine (1-classSVM) was used to create an anomaly detection model in each decomposed region. Throughout the integration, the anomaly detection model can indirectly use the known attack information to enhance its ability when building profiles of normal behaviour. This is the first attempt to use the misuse detection model to enhance the ability of anomaly detection model. Decision tree does not form a cluster, which can degrade the profiling ability thus reducing the accuracy of the system.

Shi-Jinn Horng et.al, [2] proposed an intrusion detection system, which combines a clustering algorithm, a simple feature selection algorithm, and the Support Vector Machine (SVM). In this study, in addition to a simple feature selection method, it proposed an SVM-based network intrusion detection system with BIRCH hierarchical clustering for data International Journal of Advanced Computational Engineering and Networking, ISSN: 2320-2106, Volume-3, Issue-12, Dec.-2015 A Review On Intrusion Detection System Using Classification Technique 63 pre-processing. The BIRCH hierarchical clustering provides a highly qualified and reduced datasets, in place of original large dataset, for SVM training. In addition to reduction of the training time, the resultant classifiers showed better performance than the SVM classifiers using the originally redundant dataset. However, in terms of accuracy, the proposed system could obtain the best performance at 95.72%. This approach provides better performance in terms of accuracy in comparison to the other NIDS (Network based IDS). It only detects Dos and Probe attacks not U2L and R2L attacks.

Mrutyunjaya Panda et.al, [3] proposed hybrid intelligent decision technologies using data filtering by adding guided learning methods along with a classifier to make more

classified decisions in order to detect network attacks. It is seen from the results obtained that the Naive Bayes model is quite appealing because of its integrity, elegance, robustness and effectiveness. On the other hand, decision trees have proven their efficiency in both generalization and detection of new attacks. The results show that there is no single best algorithm to outperform others in all situations. In certain cases there might be dependence on the characteristics of the data. To choose a suitable algorithm, a domain expert or expert system may employ the results of the classification in order to make better decisions.

Juan Wang et.al, [4] presented an intrusion detection system based on decision tree technology. In the process of constructing intrusion rules, information gain ratio is used in place of information gain. The experiment results show that the C4.5 decision tree is feasible and effective, and has a high accuracy rate. His experimental study shows that the C4.5 decision tree is an effective technique for the implementation of decision tree and it gives almost 90% of classifier accuracy. But in this approach the error rate remains the same.

Hong KuanSok et.al,[5] presents a paper on using the ADTree algorithm for feature reduction. ADTree also gives good classification performance. In addition, its comprehensible decision rules endows the user to discover the features that heads towards better classification. This knowledge base facilitates to design a smaller dimension of support vectors for suitable classifier. The experiment supports the idea of using this algorithm as both knowledge discovery tool and classification. The classification task has been simplified and the speed increased drastically due to the reduced operations required to implement the classification.

Proposed work:

We have proposed a hybrid feature selection algorithm (HFSA).

HFSA consists of two phases.

1) The upper phase conducts a preliminary search to eliminate irrelevant and redundancy features from the original data.

2) The lower phase (wrapper method) helps to decrease the searching range from the entire original feature space to the pre-selected features. The key contributions of this paper are listed as follows.

This work proposes a new filter-based feature selection method, in which theoretical analysis of mutual information is introduced to evaluate the dependence between features and outputclasses.The most relevant features are retained and used to construct classifiers for respective classes. As an enhancement of Mutual Information Feature Selection (FMIFS) and Modified Mutual Information based Feature Selection (MMIFS).

System Architecture:

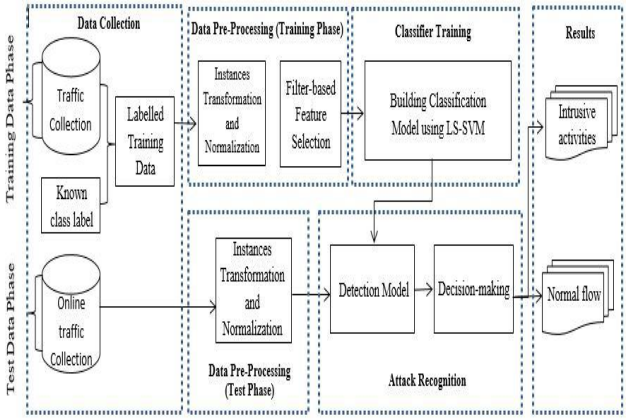

Fig 1 : system architecture

Data Collection

Data collection is the first and a critical step to intrusion detection. Using **KDD Cup 1999** 10% subset is used.

Data Preprocessing

Data Transferring - The trained classifier requires each record in the input data to be represented as a vector of real number. Thus, every symbolic feature in a dataset is first converted into a numerical value.

Data Normalization - An essential step of data preprocessing after transferring all symbolic attributes into numerical values is normalization. Data normalization is a process of scaling the value of each attribute into a well-proportioned range, so that the bias in favor of features with greater values is eliminated from the dataset. Every feature within each record is normalized by the respective maximum value and falls into the same range of [0-1].

Feature Selection - Even though every connection in a dataset is represented by various features, not all of these features are needed to build an IDS. Therefore, it is important to identify the most informative features of traffic data to achieve higher performance.

The proposed feature selection algorithm scan only rank features in terms of their relevance but they cannot reveal the best number of features that are needed to train a classifier.Determine the optimal number of required features. To do so, the technique first utilizes the proposed feature selection algorithm to rank all features based on their importance to the classification processes. Then, incrementally the technique adds features to the classifier one by one.

Modules:

Data Preprocessing

The data obtained during the phase of data collection are first processed to generate the basic features such as the ones in KDD Cup 99 dataset. The trained classifier requires each record in the input data to be represented as a vector of real number. Thus, every symbolic feature in a dataset is first converted into a numerical value. For example, the KDD CUP 99 dataset contains numerical as well as symbolic features. These symbolic features include the type of protocol (i.e., TCP, UDP and ICMP), service type (e.g., HTTP, FTP, Telnet and so on) and TCP status flag (e.g., SF, REJ and so on). The method simply replaces the values of the categorical attributes with numeric values.

Filter based feature selection

If one considers correlations between network traffic records to be linear associations, then a linear measure of dependence such as linear correlation coefficient can be used to measure the dependence between two random variables. However, considering the real world communication, the correlation between variables can be nonlinear as well. Apparently, a linear measure cannot reveal the relation between two nonlinearly dependent variables. Thus, we need a measure capable of analyzing the relation between two variables no matter whether they are linearly or nonlinearly dependent. For these reasons, this work intends to explore a means of selecting optimal features from a feature space regardless of the type of correlation between them.

Attack classification & Recognition

In general, it is simpler to build a classifier to distinguish between two classes than considering multi classes in a problem. This is because the decision boundaries in the first case can be simpler. The first part of the experiments in this paper uses two classes, where records matching to the normal class are reported as normal data, otherwise are considered as attacks. However, to deal with a problem having more than two classes, there are two popular techniques: \One-Vs- One" (OVO) and \One-Vs-All" (OVA).

Algorithm 1 Flexible mutual information based featureselectionMMIFS:

Input: Feature set $F = \{f_i, i = 1, \ldots, n\}$
Output: S - the selected feature subset
begin
Step 1. Initialization: set $S = \phi$
Step 2. Calculate $I(C; f_i)$ for each feature, $i = 1, \ldots, n$
Step 3. $n_f = n$; Select the feature f_i such that:

$$\arg\max_{f_i}(I(C; f_i)), i = 1, \ldots, n_f,$$

Then, set $F \leftarrow F \backslash \{ f_i \}; S \leftarrow S \cup \{ f_i \}; n_f = n_f - 1.$
Step 4. while $F \neq \phi$ do
Calculate G_{MI} in (4) to find f_i where $i \in \{1, 2, \ldots, n_f\}$;
$n_f = n_f - 1$;
$F \leftarrow F \backslash \{f_i\}$;
if $(G_{MI} > 0)$ then
$S \leftarrow S \cup \{ f_i \}$.
end
end
Step 5. Sort S according to the value of G_{MI} of each selected feature.
return S

Algorithm 2 Flexible Linear Correlation Coefficient based Feature Selection

Input: Feature set $F = \{f_i, i = 1, \ldots, n\}$
Output: S - the selected feature subset
begin
Step 1. Initialization: $S = \phi$
Step 2. Calculate $corr(C; f_i)$ for each feature, $i = 1, \ldots, n$
Step 3. $n_f = n$; Select the feature f_i such that:

$$\arg\max_{f_i}(corr(C; f_i)), i = 1, \ldots, n_f,$$

Then, set $F \leftarrow F \backslash \{ f_i \}; S \leftarrow S \cup \{ f_i \}; n_f = n_f - 1.$
Step 4. while $F \neq \phi$ do
Calculate G_{corr} in (7) to find f_i where $i \in \{1, 2, \ldots, n_f\}$;
$n_f = n_f - 1$;
$F \leftarrow F \backslash \{f_i\}$;
if $(G_{corr} > 0)$ then
$S \leftarrow S \cup \{ f_i \}$.
end
end
Step 5. Sort S according to the value of G_{corr} of each selected feature.
return S

Algorithm 3: Attack classification based on LS-SVM

Input: LS-SVM Normal Classifier, selected features (normal class), an observed data item x
Output: L_x - the classification label of x
begin
$L_x \leftarrow$ classification of x with LS-SVM of DoS class
if L_x == "DoS" then
Return L_x
else
$L_x \leftarrow$ classification of x with LS-SVM of Probe class
if L_x == "Probe" then
Return L_x
else
$L_x \leftarrow$ classification of x with LS-SVM of R2L class
if L_x == "R2L" then
Return L_x
else
L_x == "U2R";
Return L_x
end
end
end
end

Snapshot:

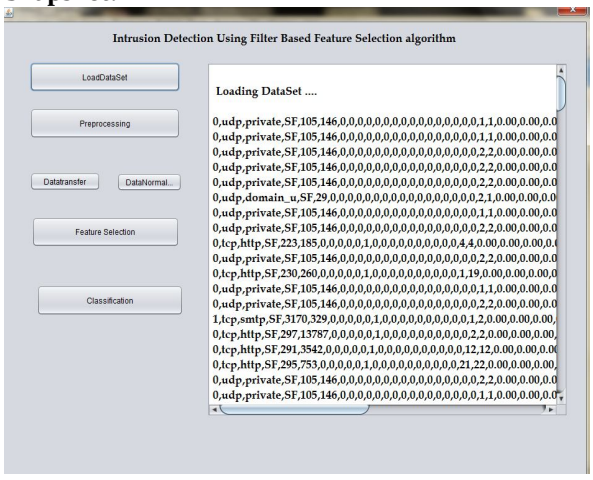

Fig2: KDDcup-99 data set loaded

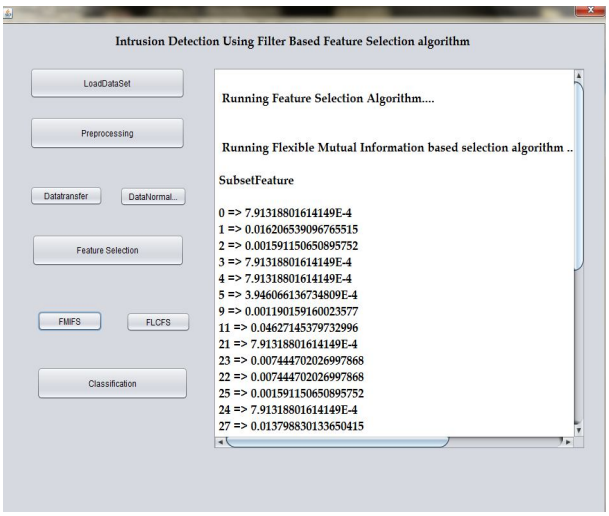

Fig 3 : feature selection using mutual information

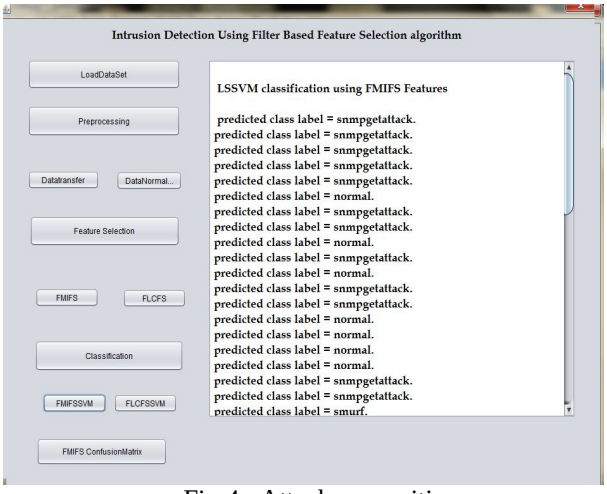

Fig 4 : Attack recognition

Conclusion:

Recent studies have shown that two main components are essential to build an IDS. They are a robust classification method and an efficient feature selection algorithm. In this paper, a supervised filter-based feature selection algorithm has been proposed, namely Flexible Mutual InformationFeature Selection (FMIFS). FMIFS is an improvement over MIFS and MMIFS. FMIFS suggests a modification to Battiti's algorithm to reduce the redundancy among features. FMIFS eliminates the redundancy parameter _ required in MIFSand MMIFS.

References

[1] S. Pontarelli, G. Bianchi, S. Teofili, Traffic-aware design of a highspeedfpga network intrusion detection system, Computers, IEEETransactions on 62 (11) (2013) 2322–2334.

[2] B. Pfahringer,Winning the kdd99 classification cup: Bagged boosting,SIGKDD Explorations 1 (2) (2000) 65–66.

[3] I. Levin, Kdd-99 classifier learning contest: Llsoft'sresultsoverview, SIGKDD explorations 1 (2) (2000) 67–75.

[4] D. S. Kim, J. S. Park, Network-based intrusion detection withsupport vector machines, in: Information Networking, Vol. 2662,Springer, 2003, pp. 747–756.

[5] A. Chandrasekhar, K. Raghuveer, An effective technique for intrusiondetection using neuro-fuzzy and radial svmclassifier,in: Computer Networks & Communications (NetCom), Vol. 131,Springer, 2013, pp. 499–507.

[6] S. Mukkamala, A. H. Sung, A. Abraham, Intrusion detection usingan ensemble of intelligent paradigms, Journal of network andcomputer applications 28 (2) (2005) 167–182.

[7] A. N. Toosi, M. Kahani, A new approach to intrusion detectionbased on an evolutionary soft computing model using neurofuzzyclassifiers, Computer communications 30 (10) (2007) 2201–2212.

[8] Z. Tan, A. Jamdagni, X. He, P. Nanda, L. R. Ping Ren, J. Hu,Detection of denial-of-service attacks based on computer visiontechniques, IEEE Transactions on Computers 64 (9) (2015) 2519–2533.

[9] A. M. Ambusaidi, X. He, P. Nanda, Unsupervised feature selectionmethod for intrusion detection system, in: InternationalConference on Trust, Security and Privacy in Computing andCommunications, IEEE, 2015.

[10] A. M. Ambusaidi, X. He, Z. Tan, P. Nanda, L. F. Lu, T. U. Nagar,A novel feature selection approach for intrusion detection dataclassification, in: International Conference on Trust, Security andPrivacy in Computing and Communications, IEEE, 2014, pp. 82–89.

[11] R. Battiti, Using mutual information for selecting features in supervisedneural net learning, IEEE Transactions on Neural Networks5 (4) (1994) 537–550.

[12] F. Amiri, M. RezaeiYousefi, C. Lucas, A. Shakery, N. Yazdani,Mutual information-based feature selection for intrusion detectionsystems, Journal of Network and Computer Applications 34 (4)(2011) 1184–1199.

[13] A. Abraham, R. Jain, J. Thomas, S. Y. Han, D-scids: Distributedsoft computing intrusion detection system, Journal of Networkand Computer Applications 30 (1) (2007) 81–98.

[14] S. Mukkamala, A. H. Sung, Significant feature selection usingcomputational intelligent techniques for intrusion detection, in:Advanced Methods for Knowledge Discovery from ComplexData, Springer, 2005, pp. 285–306.

[15] S. Chebrolu, A. Abraham, J. P. Thomas, Feature deduction andensemble design of intrusion detection systems, Computers &Security 24 (4) (2005) 295–307.

[16] Y. Chen, A. Abraham, B. Yang, Feature selection and classificationflexible neural tree, Neurocomputing 70 (1) (2006) 305–313.

[17] S.-J. Horng, M.-Y. Su, Y.-H. Chen, T.-W. Kao, R.-J. Chen, J.-L.Lai, C. D. Perkasa, A novel intrusion detection system basedon hierarchical clustering and support vector machines, Expertsystems with Applications 38 (1) (2011) 306–313.

[18] G. Kim, S. Lee, S. Kim, A novel hybrid intrusion detection methodintegrating anomaly detection with

misuse detection, Expert Systemswith Applications 41 (4) (2014) 1690–1700.

[19] P. Gogoi, M. H. Bhuyan, D. Bhattacharyya, J. K. Kalita, Packetand flow based network intrusion dataset, in: ContemporaryComputing, Vol. 306, Springer, 2012, pp. 322–334.

[20] R. Chitrakar, C. Huang, Selection of candidate support vectors inincrementalsvm for network intrusion detection, Computers &Security 45 (2014) 231–241.

[21] H. F. Eid, M. A. Salama, A. E. Hassanien, T.-h. Kim, Bi-layerbehavioral-based feature selection approach for network intrusionclassification, in: Security Technology, Vol. 259, Springer, 2011, pp.195–203.

[22] E. de la Hoz, A. Ortiz, J. Ortega, E. de la Hoz, Network anomalyclassification by support vector classifiers ensemble and non-linearprojection techniques, in: Hybrid Artificial Intelligent Systems,Vol. 8073, Springer, 2013, pp. 103–111.1226–1238.