INTERNATIONAL JOURNAL OF SYSTEMS APPLICATIONS, ENGINEERING & DEVELOPMENT
DOI: 10.46300/91015.2022.16.10 Volume 16, 2022

Statistical Analysis of Different Artificial Intelligent Techniques applied to Intrusion Detection System

Hind Tribak, Olga Valenzuela, Fernando Rojas, Ignacio Rojas.
University of Granada, Spain

Abstract— Intrusion detection is the act of detecting unwanted traffic on a network or a device. Several types of Intrusion Detection Systems (IDS) technologies exist due to the variance of network configurations. Each type has advantages and disadvantage in detection, configuration, and cost. In general, the traditional IDS relies on the extensive knowledge of security experts, in particular, on their familiarity with the computer system to be protected. To reduce this dependence, various data-mining and machine learning techniques have been used in the literature. The experiments and evaluations of the proposed intrusion detection system are performed with the NSL-KDD intrusion detection dataset. We will apply different learning algorithms on NSL-KDD data set, to recognize between normal and attack connections and compare their performing in different scenarios-discretization, features selections and algorithm method for classification- using a powerful statistical analysis: ANOVA. In this study, both the accuracy of the configuration of different system and methodologies used, and also the computational time and complexity of the methodologies are analyzed.

Keywords—Intrusion detection, NSL-KDD intrusion detection dataset, learning algorithms, ANOVA.

I. Introduction

Security is an important issue for all the networks of companies and institutions at the present time and all the intrusions are trying in ways that successful access to the data network of these companies and Web services and despite the development of multiple ways to ensure that the infiltration of intrusion to the infrastructure of the network via the Internet, through the use of firewalls, encryption, etc. But IDS is a relatively new technology of the techniques intrusion detection methods that has emerged in recent in recent years. Intrusion detection system's main role in a network is to help computer systems to prepare and deal with the network attacks. An intrusion detection system is a component within the security model of an organization. Consists in detecting inappropriate, incorrect or anomalous activities from the outside-inside, of a computer system [1].

Intrusion detection system is classified into two categories: signature based detection systems and anomaly based detection systems. Signature based detection system (also called misuse based): This type of detection is very effective against known attacks, and it depends on the receiving of regular updates of patterns and will be unable to detect unknown previous threats or new releases.

Anomaly based detection system: This type of detection depends on the classification of the network to the normal and anomalous, as this classification is based on rules or heuristics rather than patterns or signatures and the implementation of this system we first need to know the normal behavior of the network. This type use learning to build the normal behavior profile of the system.

II. Data preprocessing techniques

The pre-processed data is an important phase in the data mining, which is explained in the next paragraph, thanks to which is meant to prepare the data for the phase of knowledge discovery, and to her be quick and easy. Some of the techniques used are:

A. Feature Selection

It is a data reduction technique. A preconstruction stage of the classifier is to determine which attributes or characteristics of the data set are relevant, to carry out a good differentiation between classes [2]. The process by which these attributes are selected is called feature selection. This process consisted of 4 key stages: Selection Procedure: This stage determines the possible subset of features to perform the representation of the problem. Function Assessment: This stage evaluates the subset of features selected in the previous section. Stopping criterion: it checks whether the selected subset satisfies the criteria for stopping the search. Validation Procedure: This stage is used to verify the quality of the subset of features that were identified. Depending on the way of evaluation, they are classified into 2 groups: Filter, which uses a heuristics to determine to what degree are related, the attribute and the class

that owns the data, and wrappers that make use of a classifier to evaluate the subset of attributes.

We will use two Filters method, one CFS [3] (based correlation filter) and the other the CNS. This selection of variables is based on the correlation of patterns with class variable. This method tries to find the optimal subset of attributes highly correlated with the class and at the same time, with a low degree of redundancy between them. The other method, CNS [4], look for combinations of attributes whose values divide the data into subsets containing a strong single class majority. Usually the search is biased towards small feature subsets with high class consistency. For wrapper method's we choose Naïve Bayes and a C4.5 classifier.

B. Discretization Methods

This is a data reduction technique by which it intends to build models of classification compact and simple. The discretization is very useful because thanks to it, algorithms that use qualitative or categorical variables can be used on numeric data sets on the other hand, algorithms that handle discrete data, can be evaluated on data sets with continuous variables.

There are various classifications of classification techniques, but that we will use, can be categorized into two groups: supervised and unsupervised, the difference is that the first use the class information to create breakpoints. We use two different methods. The discretization method of Fayyad and Irani [6], which works without a predefined number of intervals. During the training phase discretization a table for each of the measured characteristics is constructed. This table provides a set of numerical levels for each feature, later the actual values will be replaced by ranges to Fayyad which they belong.

This table is obtained by distributing the discretization of numeric attributes values obtained in this phase in a set of ranges. The algorithm uses the supervised discretization method based on MDL ("Minimum Description Length"), also known as entropy-based discretization [5], which is measured as where m is the total number of intervals, pi are the probabilities of different codes, and b is the unit of measure.

The feature is recursively divided into intervals in each phase with respect to minimizing the entropy of the intervals and the information required to specify these intervals. Separation is stopped when the entropy cannot be further reduced. On the other side, the other method use is the so called: Equal frequency intervals. This method is an unsupervised method which basically operates as follows: requires a feature's values to be sorted, therefore assuming that the discretized attribute has m distinct values, this method divides the domain of each variable into n parts, and each part has m/n continuous values of the attribute.

III. Learning algorithm based on Artificial Intelligent Techniques

Data mining has been defined as "The nontrivial extraction of implicit, previously unknown, and potentially useful information from data"[7]. It uses machine learning, statistical and visualization techniques to discovery and present knowledge in a form which is easily comprehensible to humans. The DM contains many study areas such as machine–learning, pattern recognition in data, databases, statistics, artificial intelligence, data acquisition for expert systems and data visualization. The most important goal here is to extract patterns from data and to bring useful knowledge into an understandable form to the human observer. Typically, a data mining algorithm constitutes some combination of the following three components:

1.- The model: The function of the model (e.g., classification, clustering) and its representational form (e.g. linear discriminants, neural networks). A model contains parameters that are to be determined from the data.

2.-The preference criterion: A basis for preference of one model or set of parameters over another, depending on the given data.

3.- The search algorithm: The specification of an algorithm for finding particular models and parameters, given the data, model(s), and a preference criterion. Though, there are lots of techniques available in the data mining, few methodologies such as Artificial Neural Networks, K nearest neighbor, K means approach, are popular currently depends on the nature of the data. Below we will present learning algorithms that have been used.

A. Genetic Algorithm

The Genetic Algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of Evolutionary Algorithm (EA), which generates solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection and crossover.

B. Artificial Immune Systems

The biological immune system is a robust, complex, adaptive system that defends the body from foreign pathogens. It is able to categorize all cells (or molecules) within the body as self-cells or non-self cells The immune Network theory had been proposed in the mid-seventies[8]. The hypothesis was that the immune system maintains an idiotype network of interconnected B cells for antigen recognition. These cells both stimulate and suppress each other in certain ways that lead to the stabilization of the network. Two B cells are

connected if the affinities they share exceed a certain threshold, and the strength of the connection is directly proportional to the affinity they share. The algorithm used in this paper has been CLONALG[9].

C. Artificial Neural Network

Artificial Neural Networks (ANN) is systems inspired by the research on human brain[10]. Artificial Neural Networks (ANN) networks in which each node represents a neuron and each link represents the way two neurons interact. Each neuron performs very simple tasks, while the network representing of the work of all its neurons is able to perform the more complex task. A neural network is an interconnected set of input/output units where each connection has a weight associated with it. The network learns by fine tuning the weights so as able to predict the call label of input samples during testing phase. For our experiments we have used MLP- Multilayer Perceptron-. Besides we have evaluated our data set using RBF-Radial Basis Networks-.

The idea of Radial Basis Function (RBF) Networks derives from the theory of function approximation. The Multi-Layer Perceptron (MLP) networks with a hidden layer of sigmoid units can learn to approximate functions. RBF Networks take a slightly different approach. Their main features are: a) They are two-layer feed-forward networks; b) The hidden nodes implement a set of radial basis functions (e.g. Gaussian functions); c) The output nodes implement linear summation functions as in an MLP; d) The network training is divided into two stages: first the weights from the input to hidden layer are determined, and then the weights from the hidden to output layer; e) The training/learning is very fast; f) The networks are very good at interpolation.

D. SVM

The theory of SVMs was initially developed by V. Vapnik [11] in the early 80's and focuses on what is known as Statistical Learning Theory. SVM initially appeared to separate two classes, although its application has extended to any finite number of classes. These techniques perform a linear classification, upon vectors processed on a higher dimensional space, ie in the transformed space, separating the different classes using an optimal hyperplane to maximize the margin between classes. We have evaluated SVM using different algorithms like SMO-Sequential Minimal Optimization- and using LIbSVM [12]library, both with different types of kernel function.

E. Fuzzy Logic

Fuzzy control systems, can describe the set of rules, which would use a human being, who control the process, with all the inaccuracies that have languages natural. Fuzzy logic relaxes the idea of membership of an element to a set. In traditional logic, an element belongs or not to a set, however, fuzzy logic, an element belongs to a set with a certain degree of membership. In the detection of intruders, fuzzy logic can be applied in various ways. One of them is in learning systems with fuzzy rules. This is the case of algorithm Furia[13].

F. Decision Trees

A classification tree consists of nodes, arcs and leaves. Each node represents a decision on a particular attribute value, being the terminal nodes where a decision is made about the class map. When classifying a new case will have to be compared the values of the attributes with the decisions taken at the nodes, following the branch, that match those values, in every decision. Eventually you will reach a terminal node or leaf that predicts the class for the case treated. One of the algorithms based on decision trees most used is the C4.5 [14] whose approach is TDIDT (Top Down Induction of Decision Trees), which is characterized by using a strategy of divide and conquer descending. Another algorithm, is the Random Forest introduced by Breiman in 1999, uses a set of trees (forest) classification. To classify a new data, is taken as input for each tree and produces the corresponding output classification. As a final decision for the whole of trees, takes the class with the most votes [15]. Other algorithms have been evaluated like CART[16]- Classification and Regression Trees- and NBTREE[17] which is a Naïve Bayes decision tree.

G. Rule Induction

Rule induction algorithms offer an approach to data-driven knowledge discovery from labeled data. Patter classifiers that are induced by rule learning algorithms are often simpler and easier to comprehend by humans than those induced using neural network or Support vector machines. We have used two algorithms for rule induction like RIPPER [18], or PART[19]- which build partial decision trees to induce a rule- to evaluate its performance over the data set NSL.

H. Nearest Neighbor

The basics of neighborhood classification were established by [20] in the early 50's. The nearest neighbor method and its variants are based on the intuitive idea that similar objects belong to the same class, so that the class to which it belongs an object can be inferred from the class they belong objects of the learning sample that most resembles him. The idea of similarity is reflected formally in the concept of distance. The algorithm k-NN [21], is included within the so called lazy learning techniques, and does not generate a knowledge structure, which shapes the information inherent in the training set, but the dataset itself, represents the model, i.e. is not built any model, the model is the database itself or the training set. We have evaluated KNN algorithm using 2 different variants, one using k= 1 and the other one with k =50.

I. Bayesian Networks

A Bayesian network is a directed acyclic graph and annotated, describing the joint probability distribution that governs a set of random variables. The topology or the network structure not

only provides information on probabilistic dependencies between variables, but also on the conditional independence of a variable or set of them, given one or more other variables. Provide flexible methods of reasoning based on the propagation of probabilities over the network in accordance with the laws of probability theory. The algorithm TAN (Tree Augmented Naive) [22], builds a classifier, where there is a tree structure of dependencies among the predictors. Based on a model similar to dependencies Naïve Bayes, adding conditional dependencies between nodes, the algorithm TAN forming a tree between them. The mixture of both strategies enables the relaxation of independence between the predictor variables. This model, proposed by Friedman [22], is based on the calculation of the conditional mutual information between pairs of variables. This is a probabilistic paradigm, we have selected Naïve Bayes algorithm and TAN algorithm to evaluate NSL data set.

J. Hidden Markov Models

Introduced by L. E. Baum in the 70's [23], Baum proposes this model as a statistical method of estimation of probabilistic functions of a Markov chain. A HMM can be represented as a directed graph of transitions / emissions. A discrete hidden Markov model is defined in terms of the following elements [24]: 1. Number of states model "N" 2. Number of observations different "M". 3 transition probability matrix A = {aij}. 4 observation probability matrix B = {bi (k)}. 5 Probability initial state Π = {π}. Because the number of hidden states can affect the performance of classification, we evaluate HMM with different numbers of hidden states for classification. The number of states taken is based on the number of attributes that different feature selection methods have resulted.

IV. Data Set

NSL data set is based on KDD-CUP'99 data set consists of single connection vectors where each single connection vectors consists of 41 features [25]. These features had all forms of continuous and symbolic with extensively varying ranges falling in four categories:

1. In a connection, the first category consists of the intrinsic features which comprises of the fundamental features of each individual TCP connections. Some of the features for each individual TCP connections are duration of the connection, the type of the protocol (TCP, UDP, etc.) and network service (http, telnet, etc.).

2. The content features suggested by domain knowledge are used to assess the payload of the original TCP packets.

3. Within a connection, the same host features observe the recognized connections that have the same destination host as present connection in past two seconds and the statistics related to the protocol behavior, service, etc are estimated.

4. The similar same service features scrutinize the connections that have the same service as the current connection in past two seconds.

A variety of attacks incorporated in the dataset fall into following four major categories:

A) Denial Of Service Attacks (DOS): A denial of service attack is an attack where the attacker constructs some computing or memory resource fully occupied or unavailable to manage legitimate requirements, or reject legitimate users right to use a machine.

B) User to Root Attacks (U2R): exploits are a category of exploits where the attacker initiate by accessing a normal user account on the system (possibly achieved by tracking down the passwords, a dictionary attack, or social engineering) and take advantage of some susceptibility to achieve root access to the system.

C) Remote to User Attacks (R2L): takes place when an attacker who has the capability to send packets to a machine over a network but does not have an account on that machine, makes use of some vulnerability to achieve local access as a user of that machine.

D) Probes: Probing is a category of attacks where an attacker examines a network to collect information or discover well-known vulnerabilities. These networks investigations are reasonably valuable for an attacker who is staging an attack in future. An attacker who has a record, of which machines and services are accessible on a given network, can make use of this information to look for fragile points.

V. Experimental Study

As shown in Fig.1 we can see the distribution and frequency of attacks in the data set being observed that there are certain types of attacks that take precedence over other.

For this case, the attacks are mapped to the category to which they relate and the system is trained with records like "Normal", "Two", "Probe", "R2L" or "U2R" (Fig.2). This classification is the most studied in the literature.

DOI: 10.46300/91015.2022.16.10 Volume 16, 2022

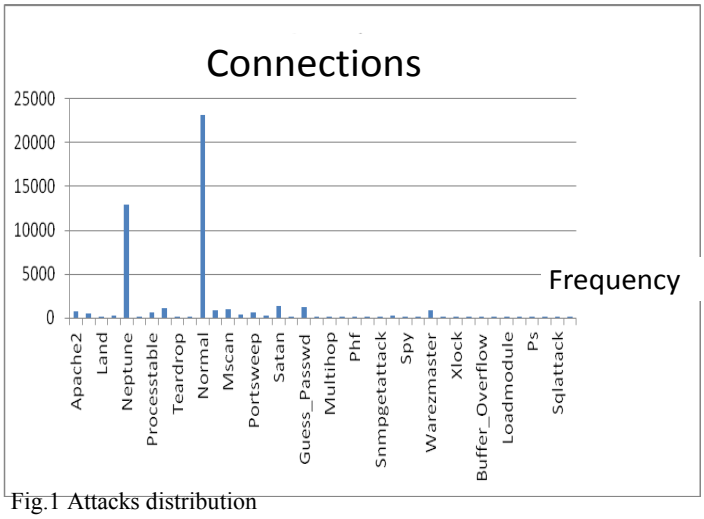


Fig.1 Attacks distribution

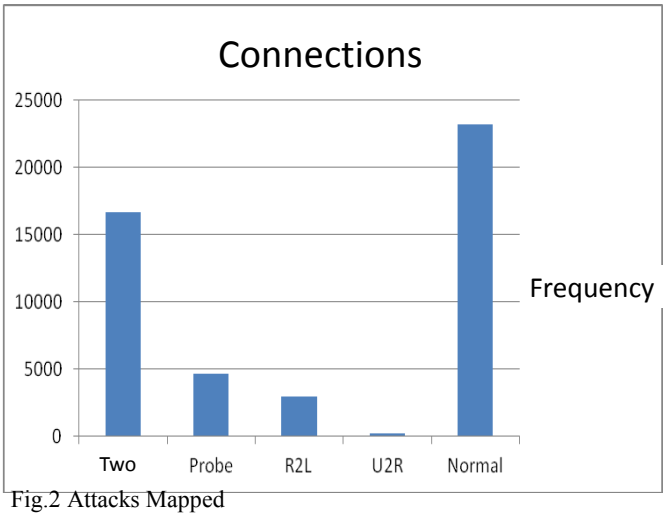
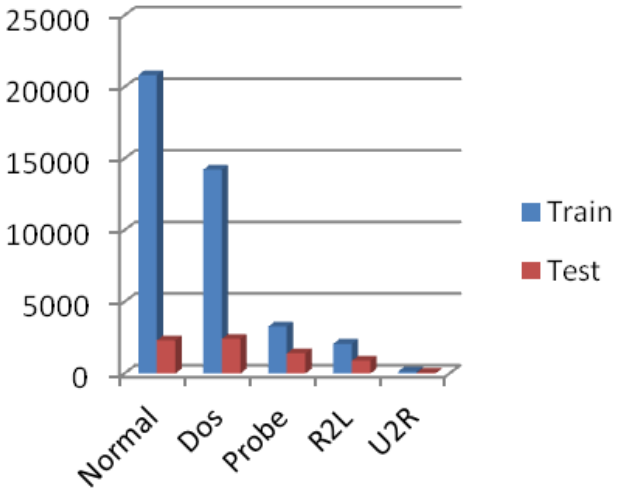
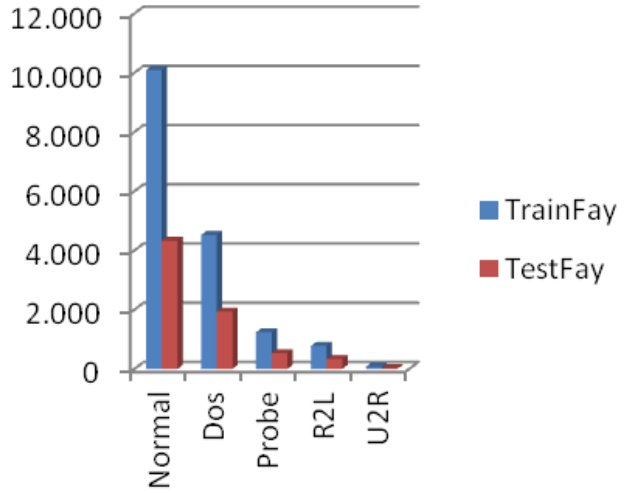


Fig.2 Attacks Mapped

As shown in the figures below we built data set without discretization, data set with Fayyad & Irani discretization and data set with Equal Frequency Intervals discretization. It is important to note that discretization significantly reduces the size of the dataset.





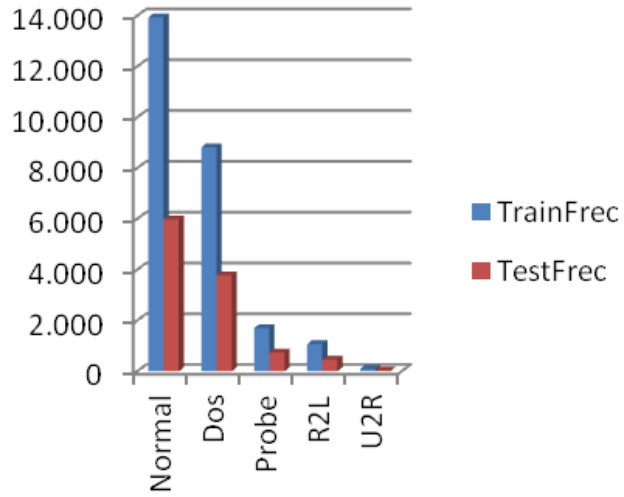


Fig.3 a) Without Discretization; b) Fayyad & Irani Discretization; c) Equal Intervals Frequency Discretization

We train and test a first kind of file without selection features and discretization, a second pair, using Fayyad Discretization and a third pair using Equal Intervals Frequency. On the other hand for each training and testing data set mentioned we will apply selection features.

The total files numbers is 30 (15 files for training and 15 for testing). For each one we will build a model using 20 different learning techniques. Finally we build a table which contains the information about the global accuracy of a classifier and the time consuming to do build it.

The evaluation of a pattern recognition experiment is based on the measure of success (percentage of well classified samples or instances) of a data set, also called the test set. Therefore in this study were obtained construction time of the models and the confusion matrices. It should be noted that 315 executions has been carried out.

VI. Statistical Analysis

We will study how the selection of different values or levels for:

a) Filter Type;
b) Discretization Type;
c) Algorithm Type

Influence on two output variables:

a) Error classification system global (denoted as AcGlobal)
b) Time required to build the model (denoted as TimeTr).

Filter, discretization and algorithm are called a factor. A factor is an independent treatment variable whose settings (values) are controlled and varied in the experiment. The different values settings are called levels.

For Filter are: 1) All: all the features (no filter is applied); 2)FC4.5: filter C4.5 (25 features are used); 3)FCFS: filter CFS is used (10 features); 4)FCNS: filter CNS (15 features); 5) FNB: filter Naive Bayes is used (11 features).

For Discretization are: 1) 0: means that no discretization is performed; 2) Fay: discretization of Fayyad and Irani; 3) Fre: Equal frequency intervals is used.

Finally, for the type of intelligent classifier or algorithm, the levels are the following 20 methods: PML-multilayer Pervceptron-, RBFNet (Radial basis Network), Bayesian algorithm (Naïve Bayes and TAN), FURIA (fuzzy), Rule induction (RIPPER and PART), K-nearest neighbor (KNN with 1 and 50 neighbors), Markov-11, Markov-15, Markov-25, (Hidden Markov Model with different parameters), Decision Trees (Random Forest, SimpleCART, NBTree and C4.5), Genetic Algorithm, Clonalg(Immune) , and SVM (SMO using polynomial (SMOPoly) and Gaussian (SMORBF) kernel, and Libsvm with sigmoid kernel (C-SVCSigmoide) and gaussian (C-SVCRBF)).

To know if a factor has an influence or not on the output variable we analyze ANOVA table. This procedure execute the ANalysis Of VAriance of the different factors for ACGlobal and TimeTr.

Main effects	*Sum of squares*	*D.F.*	*Mean square*	*F-ratio*	*P-value*
A:Filtro	827,943	4	206,986	2,46	0,0458
B:Discr	5438,28	2	2719,14	32,29	0,0000
C:ALG	41924,9	21	1996,42	23,71	0,0000
Residue	24166,6	287	84,2041		
TOTAL	71471,6	314			

Table 1. ANOVA table analysis for Acglobal

Main effects	*Sum of squares*	*D.F.*	*Mean square*	*F-ratio*	*P-value*
A:Filtro	2,23405E8	4	5,58512E7	1,33	0,2586
B:Discr	1,92204E8	2	9,61022E7	2,29	0,1031
C:ALG	5,59739E9	21	2,66543E8	6,35	0,0000
Residue	1,20455E10	287	4,19703E7		
TOTAL	1,80374E10	314			

Table 2. ANOVA table analysis for

For both cases, we can see that Filter, Discretization and Algorithm have the value P (in red) more less than 0.05, that mean this factors have an statistically significant effect on AcGlobal and TimeTr (output variables) with a 95% confidence level.

The following figure (Fig.4) shows how the variables Filter and Discretization influence on AcGlobal.

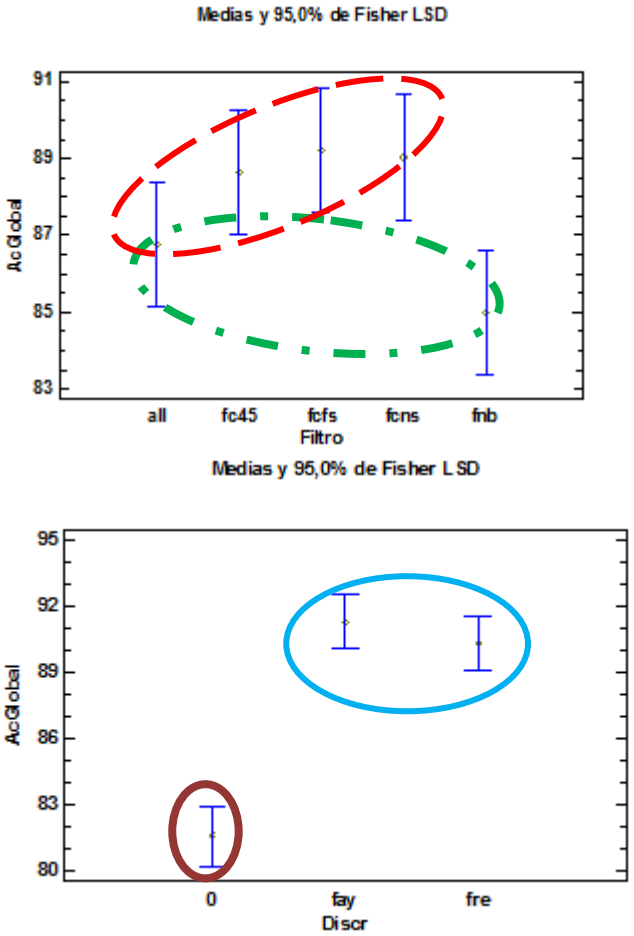


Fig.4 a) Evolution of the AcGlobal for different level of factor Filter; b) Evolution of the AcGlobal for different level of factor Discretization.

The discontinuous lines represent groups of levels that are homogeneous (i.e. this levels from a statistical point of view have the same repercussion on the output variable).

We can see that there are two homogeneous groups for the filter factor with intersection, one with FNB filter and ALL, and the other with the rest of filters and ALL. From the statistical point of view, the worst performance is for the first group which include FNB (wrapper with Naïve bayes) and ALL (without selection features). The other filters CFS

(correlation filter), CNS(consistency filters), FC4.5 (wrapper with C4.5) and All (without filter) for the output variable, overall accuracy, these four types of filters are equivalent or similar, and have the same behavior.

For the discretization methods, the worst behavior is for 0 – without discretization- and Fayayd & Irani and equal frequency interval have the best. The different between the use of discretization or not, is significant.

Finally, for the algorithm used for the classification, there are 10 different groups:

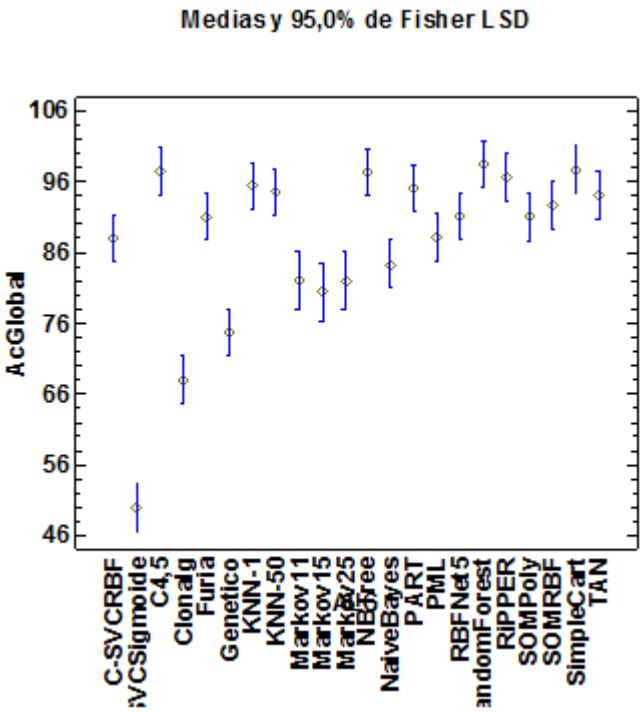


Fig.5 Evolution of the AcGlobal for different level of factor intelligent classifier.

As can be seen in Fig.5, there are ten different groups in which you can group the algorithms used, mentioning that there are no statistically significant differences between those levels sharing the same column of X's. As some methods belong to more than one group, therefore exist intersection between them.

In the first group, comments that the method: C-SVCSigmoide, SVM with sigmoid kernel. In the second group: Artificial Immune Algorithm Clonalg-. In the third group the different variants of Markov models: MARKOV15, MaARKOV25 and MARKOV11, along with the genetic algorithm.

Successively we can mention the other groups until the last that would have the algorithms: SOMRBF, TAN, KNN-50, PART, KNN-1, RIPPER, NBTREE, C4.5, CART, RANDOM FOREST, this group is the best results obtained. As important conclusion, the best algorithm is RANDOM FOREST.

VII. Conclusion

We have a set of 15 different files with a combination of selection features methods, discretization and without them. On the other hand we have carried out 315 different executions with different algorithms and in some cases its variants.

And our goal has been study how this factors influence on the accuracy of classifier and the time to build it.

We have many techniques to help us to build a good model to classify different kind of attacks into 5 different categories.

Thanks to ANOVA we have demonstrated that The discretization method (supervised, unsupervised and without discretization), the selection of attributes method (filter, wrapper, and without using feature selection) and the selected algorithm, influence the behavior of the classifier, both, from the point of view of the error and the time required to obtain the build the model.

For example, for the error of the classifier, ANOVA revealed that, the unsupervised discretization type -FRE,-equal frequency intervals-, and the supervised Fayyad method improves the accuracy of the system. Regarding the features selection, wrapper method based on Naive Bayes got worse performance and the CFS- filter method- has improved it. It was found that the filter methods have had better results. In this first case we can say that the selection features provides significant improvement in the classification task. Finally regarding the algorithm employed, have been shown that the simple methods like, Decission trees and nearest neighbor have shown better results than other more complex methods. Random Forest algorithm has had better performance.

The same methodology was used for the computational time, observing that in this case the most significant conclusion was extracted from the method used to build the classifier. In this case, MLP- Multilayer perceptron- is the most time-consuming.

The most important conclusion here is that we have used different strategy to build robust and reliable systems to detect intrusion, attacks or threats, and the result have shown that simple and straightforward techniques have obtained the best results, like Decision trees (Random Forest, C4.5) followed by induction of rules and KNN methods.

References

[1] Jian Pei, Shambhu J. Upadhyaya. Faisal Farooq, Venugopal Govindaraju "Data mining for intrusion detection: techniques applications and systems " In Proceedings of the 20th International Conference on Data Engineering, pp: 877 - 87, 2004

[2] Dash, M., Liu, H., 1997. Feature selection for classification. Int. J. Intell. Data Anal., 1:131-156.

[3] Hall, M. A. & Smith, L. A. Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper in `Proceedings of the Twelfth International Florida Artificial Intelligence Research Society Conference, Orlando, USA' pp. 235-239. 1999

[4] H. Liu and R. Setiono. A probabilistic approach to feature selection: A filter solution. In Proceedings of the 13th International Conference on Machine Learning, pp 319–327. Morgan Kaufmann, 1996.

[5] U.M. Fayyad and K.B. Irani. Multi-interval discretization of continuous valued attributes for classification learning. In Proceedings of the 13th International Joint Conference on Artificial Intelligence, pp 1022-1027, Morgan Kaufmann, 1993.

[6] Ellis J. Clarke, Bruce A. Barton,"Entropy and MDL Discretization of Continuous Variables for Bayesian Belief Networks" International Journal Of Intelligent Systems, VOL. 15, 61]92 _2000.
[7] W. Frawley and G. Piatetsky-Shapiro and C. Matheus, Knowledge Discovery in Databases: An Overview. AI Magazine, Fall 1992, pgs 213-228.
[8] Jerne, N. K. (1974). Towards a Network Theory of the Immune System, Ann. Immunol. (Inst. Pasteur) 125C, 373 – 389
[9] L. Ñ unes de Castro and J. Timmis. An artificial immune network for multimodal function optimization. In Proccedings of the 2002 Congress on Evolutionary Computation (CEC'2002), volume 1, pp 669–674, Honolulu, Hawaii, May 2002.
[10] Hammerstrom, D., 1993. Neural networks at work, IEEE Spectrum, June, 26–32
[11] Vapnik, V. N. The Nature of Statistical Learning Theory. Springer. 1995
[12] Software available at http://www.csie.ntu.edu.tw/_cjlin/libsvm.
[13] Jens Christian Hühn and Eyke Hüllermeier. Furia: an algorithm for unordered fuzzy rule induction. Data Mining and Knowledge Discovery, 19(3):293–319, 2009.
[14] Quinlan J. C4.5: Programs for machine learning, Morgan Kaufmann Pub., 1993 (ISBN: 1558602380).
[15] Leo Breiman. Random forests. Machine Learning Journal, 45:5–32, 2001
[16] L. Breiman, J. Friedman, R. Olshen, and C. Stone. Classification and regresion trees. Wadsworth Int. Group, Belmont, CA, 1984.
[17] R. Kohavi. Scaling up the accuracy of naive-Bayes classifiers: a decision-tree hybrid. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, pp 202-207, 1996.
[18] William W. Cohen. Fast effective rule induction. In In Proceedings of the Twelfth International Conference on Machine Learning, pp 115–123. Morgan Kaufmann, 1995.
[19] Frank, E. Y Witten, I.H. (1998): "Generating Accurate Rule Sets Without Global Optimization", en J. SHAVLIK (ed.): Proceedings of the Fifteenth International Conference on Machine Learning, Madison, Wisconsin. Morgan Kaufmann, San Francisco, pp. 144-151
[20] E. Fix and J. Hodges. Discriminatory analysis, nonparametric discrimination consistency properties. Technical Report 4, US Air Force, School of Aviation Medicine, Randolph Field, TX, 1951
[21] T. M. Cover and P. E. Hart. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, IT-13(1):21–27, 1967
[22] Friedman, J. S., C. A. Tepley, P. A. Castleberg, and H. Roe, Middle-atmospheric Doppler lidar using an iodine-vapor edge filter, Optics Letters, 22, 1,648-1,650, 1997.
[23] Baum, L. E. An inequality and associated maximizaton technique in statistical estimation for probabilistic functions of markov processes. Inequalities, 3:pp 1-8, 1972.
[24] Rabiner, Lawrence R. A tutorial on hidden markov models and selected applications in speech recognition. pp 267-296, 1990.
[25] NSL-KDD data base. http://www.iscx.ca/NSL-KDD/