ABSTRACT

Feature selection is a preprocessing step to machine learning, leads to increase the classification accuracy and reduce its complexity. Feature selection methods are classified into two main categories: filter and wrapper. Filter methods evaluate features without involving any learning algorithm, while wrapper methods depend on a learning algorithm for feature evaluation. Variety hybrid Filter and wrapper methods have been proposed in the literature. However, hybrid filter and wrapper approaches suffer from the problem of determining the cut-off point of the ranked features. This leads to decrease the classification accuracy by eliminating important features. In this paper the authors proposed a Hybrid Bi-Layer behavioral-based feature selection approach, which combines filter and wrapper feature selection methods. The proposed approach solves the cut-off point problem for the ranked features. It consists of two layers, at the first layer Information gain is used to rank the features and select a new set of features depending on a global maxima classification accuracy. Then, at the second layer a new subset of features is selected from within the first layer redacted data set by searching for a group of local maximum classification accuracy. To evaluate the proposed approach it is applied on NSL-KDD dataset, where the number of features is reduced from 41 to 34 features at the first layer. Then reduced from 34 to 20 features at the second layer, which leads to improve the classification accuracy to 99.2%.

Keywords: Feature Selection, Information Gain, Network Intrusion Classification, Redundancy

INTRODUCTION

Intrusion detection system (IDS) is a major research problem in network security. The IDS goal is to dynamically identify unusual access or attacks to secure the networks (Tsai, 2009; Debar, 1999). Hence, Network-based IDS (NIDS) is a valuable tool for the defense in depth of computer networks. NIDS looks for known or potential malicious activities in the network traffic and raises an alarm whenever a suspicious activity is detected.

One of the important research challenges for constructing high performance IDS is dealing with data containing large number of features. Irrelevant and redundant features of
the dataset complex the IDS and reduce the detection accuracy as well. Therefore, dataset dimensional reduction is an active research area in the field of machine learning and pattern recognition (Amiri, 2011; Li, 2012; Kuchimanchi, 2004). The dimensionality reduction of the dataset can be achieved by feature extraction or feature selection. Feature selection aims to choose an optimal subset of features that are necessary to increase the predictive accuracy and reduce the complexity of the learned results (Dash, 2002; Koller, 1996). Different feature selection methods are proposed to enhance the performance of IDS (Tsang, 2007). Information gain is one of the most common feature selection methods (Ben-Bassat, 1982).

In this paper, we propose a bi-layer behavioral-based feature selection approach which hybrid filter and wrapper method. The proposed approach aims to improve the network intrusion classification accuracy, it consists of two layers. The first layer used information gain method to rank the features and select a new set of features depending on a global maxima classification accuracy. Followed by a second layer which selects a new set of features from within the first layer redacted data by searching for a group of local maximum classification accuracy in order to increase the number of reduced features. To evaluate the performance of the proposed bi-layer behavioral-based feature selection approach several experiments are conducted on NSL-KDD datasets using. The results obtained show the effectiveness of the proposed approach over single level feature selection method. The rest of this paper is organized as follows: the third section gives an overview of data reduction and information gain. The fourth section describes the proposed Bi-layer behavioral-based feature selection approach. The experimental results and conclusions are presented in the fifth and sixth sections respectively.

**Related Work**

In 1987, Denning (1987) developed the first intrusion detection model. Denning proposed the idea that intrusions could be detected by creating a normal profile, where any sufficient deviate from the norm is considered as anomalous. Since Denning model many research efforts have been focused on how to construct an effective and accurate intrusion detection models. However, Intrusion detection classification accuracy depends on the number of features that adequately characterize the data (Dash, 1997).

In the last decade, applying data reduction techniques become a real prerequisite for classification model building (Dash, 2000; Gheyas, 2010; Park, 2005; Pudil, 1994). Dimensionality reduction techniques include either the altering of input data set by a set of lower number of features, like those based on projection (e.g. principal component analysis), Or selecting a subset of features without changing the original representation of the data.

Feature selection (FS) allows the determination of the features related to the classification problem, and though reduces the size of the input data set without affecting the internal structure of the data. Also declaring these features helps in interpreting the rules generated from machine learning techniques. Several feature selection techniques have been proposed in the literature (Guyon, 2003; Xu, 2006). FS techniques can be divided into four categories: distance, information dependency, consistency and classifier error rate (Dash, 2003). The first three types of evaluation measures are known as “filter” methods. The filter methods are independent from any learning algorithm. While, the “wrapper” methods depend on the machine learning algorithm. Where, the learning algorithm is applied to the features and the classifier error rate is used to select the best features (Kohavi, 1997).

Several authors proposed hybrid filter and wrapper approaches, t-statistics and a Genetic Algorithm (Tan, 2006), correlation based feature selection algorithm and a Genetic Algorithm (Shazzad, 2005) and mutual information and a Genetic Algorithm (Huang, 2006). However, hybrid filter and wrapper approaches suffer from the risk of eliminating good features. Since, it is not clear how to determine the cut-off point
Related Content

Classifying Behaviours in Videos with Recurrent Neural Networks
www.igi-global.com/article/classifying-behaviours-in-videos-with-recurrent-neural-networks/195006?camid=4v1a

3D Image Acquisition and Analysis of Range Face Images for Registration and Recognition
www.igi-global.com/chapter/3d-image-acquisition-and-analysis-of-range-face-images-for-registration-and-recognition/141635?camid=4v1a
Accelerating Deep Action Recognition Networks for Real-Time Applications

Entropy based Range Optimized Brightness Preserved Histogram-Equalization for Image Contrast Enhancement