Statistical Methods in AI

Project Proposal

Title: Intrusion Detection using Machine Learning Methods

Team Id: 14

Github Link: https://github.com/vivek-2018201078/Intrusion_Detection_smai_proj

Members: Tarun Vatwani, Vivek Patare, Rajat Yadav, Ankit Mishra

I. Goal

The main goal of this project is to build a complete machine learning model for solving the problem of intrusion detection using supervised and unsupervised learning approaches. The dataset which will be using is KDD Cup 1999 dataset. Since it's a very large dataset with large no of attributes and training samples we will also implement feature reduction for our dataset to observe a tradeoff between Accuracy and Learning Time.

II. Problem Statement

The key idea is to observe a set of meaningful patterns to describe user activity on a system based on some relevant features, and build some powerful classifiers using those patterns. In this project we will focus on some of the popular classifiers such as *Support Vector Machines(SVMs)* and *Neural Networks*.

These learning methods are used because of their adaptability and generalization capability to learn new attacks once discovered by the system. But one of the major challenges is training time for these methods. The SVMs perform very well for a binary classification but for a multiclass classification usually neural networks are preferred over SVM's. Therefore we will try to use some feature detection techniques such as *Correlation Coefficient, Least Square Regression Error and Maximal Information Compression Index*, to generate a tradeoff between learning rate and training time.

But we know that in real life working on labeled data is not enough to prepare our system for some unseen threat therefore we will also explore some of the unsupervised learning techniques for our project, using the same dataset without output labels. We will be implementing *K-Means clustering* algorithm for unsupervised learning.

III. End Result:

Our end goal is to develop a model to detect network intrusions protects a computer network from unauthorized users, including perhaps insiders. The intrusion detector learning task is to build a predictive model (i.e. a classifier) capable of distinguishing between ``bad'' connections, called intrusions or attacks, and ``good'' normal connections.

IV. Project Milestone:

A. **First Week**
   Reading and understanding all three papers

B. **Second-Third Week**
   Implementation of the first version of the code as mentioned in the papers

C. **Fourth-Fifth Week**
   Trying for more optimized methods and advancement in existing implementation

D. **Sixth Week**
   Final testing and report writing