Machine Learning: Network Traffic Identification

By
Alex Baker
Erik Granger
Tarek Embree

University of New Mexico
Dr. Lydia Tapia
CS 427
Fall 2016

Introduction[1]

I. Motivation: why use machine learning to analyse network traffic?
    A. Contributions:
        1. Security
        2. Quality of Service
        3. Firewall
        4. Intrusion Detection
        5. Network Optimization
        6. Analyzing Network Traffic

[1] Written by group

Paper 1[2]

Summary for reference [1].

Contribution
Applies machine learning to automatically identify and classify network traffic based on attributes such as size, duration, packet length, time distribution, and idle times of traffic flows.

Method
Use machine learning to identify the optimal set of traffic flow attributes by minimizing processing and maximizing accuracy.

Results and Experiment
The results of the machine learning algorithm will be compared to existing traces of traffic flows to determine the accuracy of the classification and identification of the traffic flow.

[2] Written by Alex Baker

Paper 2[3]

Mobile Network Traffic Prediction Using MLP, MLPWD, and SVM
by Ali Yadavar Nikravesh Samuel, A. Ajila, and Chung-Horng Lung

The goal of this experiment is to compare the accuracy of MLP, MLPWD and SVM algorithms at predicting the future behavior of mobile traffic. The acronyms of the algorithms stand for MLP - Multi-layer Perceptron, MLPWD - Mulit-layer Perceptron with Weight Decay, and SVM - Support Vector Machine.

Contribution
The contribution of this experiment are stated as follows.

- Comparing the accuracy of MLP, MLPWD and SVM algorithms to predict future behavior of mobile network traffic.
- Analyzing the impact of the sliding windows size on the prediction accuracy of MLP, MLPWD, and SVM algorithms.
- Comparing the accuracy of MLP, MLPWD, and SVD algorithms in predicting multidimensional and unidimensional model network traffic datasets.

Method
They use a variety of algorithms to finds the most accurate regression model that predicts the future behavior of mobile traffic. They also use real-life dataset from commercial trial mobile network. They used the data to compare the accuracy of MLP, MLPWD and SVM algorithms by assuming future values of attributes or known or unknown.

Result / Experiments
The results from the experiments concluded that SVM works better than MPL and MPLWD at predicting multidimensionality of future behavior of mobile network traffic. MLPWD is better at predicting future behavior using unidimensional data. They have plans for future experiments to use parallel algorithms to facilitate real-time analytics such as for a large size of data.

[3] Written by Erik Granger

Paper 3[4]

Performance Analysis of Unsupervised Machine Learning Techniques for Network Traffic Classification by Hardeep Singh

I. Contribution
    A. Comparing K-Means and Expectation Maximisation to find which works better
II. Method
    A. Analyse network traffic packets using K-Means and Expectation Maximisation algorithms and compare accuracy
III. Experiments
    A. K-Means is more accurate than Expectation Maximisation especially with fewer clusters

[4] Written by Tarek Embree

Conclusion[5]

I. Machine learning is useful in network traffic applications for:
   A. Security
   B. Quality of Service
   C. Firewall
   D. Intrusion Detection
   E. Network Optimization
   F. Analyzing Network Traffic

[5] Written by group

References

1. S. Zander, T. Nguyen, and G. Armitage, "Automated traffic classification and application identification using machine learning," The IEEE Conference on Local Computer Networks 30th Anniversary (LCN'05)l, 2005.
2. H. Singh, "Performance Analysis of Unsupervised Machine Learning Techniques for Network Traffic Classification," *2015 Fifth International Conference on Advanced Computing & Communication Technologies*, Haryana, 2015, pp. 401-404. doi: 10.1109/ACCT.2015.5.
3. 3. A. Y. Nikravesh, S. A. Ajila, C. H. Lung and W. Ding, "Mobile Network Traffic Prediction Using MLP, MLPWD, and SVM," *2016 IEEE International Congress on Big Data (BigData Congress)*, San Francisco, CA, 2016, pp. 402-409. doi: 10.1109/BigDataCongress.2016.63