

Principal applications:

- Network Behaviour
- Network Failure Detection
- Routing Protocols
- Quality of Service and Status Report
- Security Breach Detection and Firewalls
- Network Design



Figure 1. Non-Machine Learning Traffic Measurement

Introduction

Networks have increased in size in the past few years and with this increase, networks traffic and data are getting complex and diverse. It is important that networks are managed appropriately to respond to these new demands and decisions made are based accurate data assisted by current data science models and technology. Take for instant, with COVID 19 people are staying at home, working from home, learning or engaging in more online activities such as playing games or watching online streaming. However, with such remarkable event many questions are raised regarding the traffic volume increase and weather it has any impact on network performance, speed and quality. Machine learning and data science can play a critical role in answering and predicting traffic behaviour and to assist in making reliable decisions.

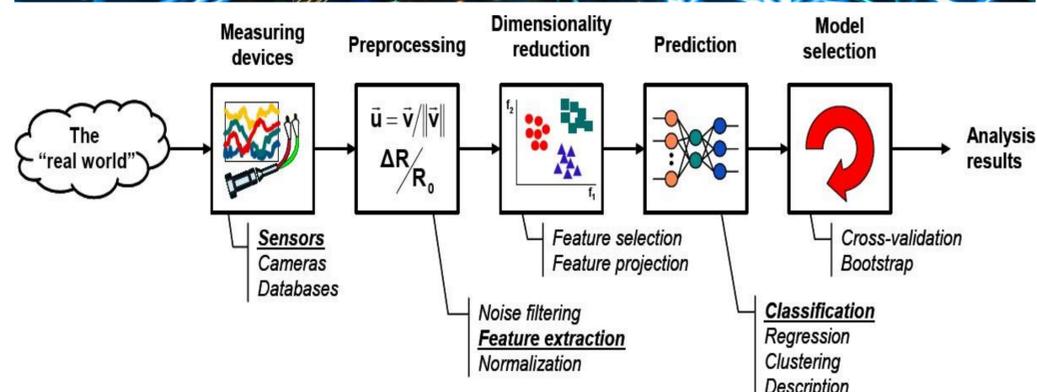


Figure 2: Building Machine Learning Model for Traffic Detection

Key research questions and focus

Are the current Machine learning models able predict and classify network traffic accurately compared to our Mixture Gaussian proposed model?

Background of study

Given that networks need to respond not only by the expanding in size but also by being able to cope with both the traffic diversity coupled with the big data demand for diversity, availability, veracity and for higher speeds as data is getting bigger. Currently, learning machine models have been deployed and evaluated but their accuracy in predicting and classifying network traffic still remain an issue specially packet losses during congestion.

Technologies

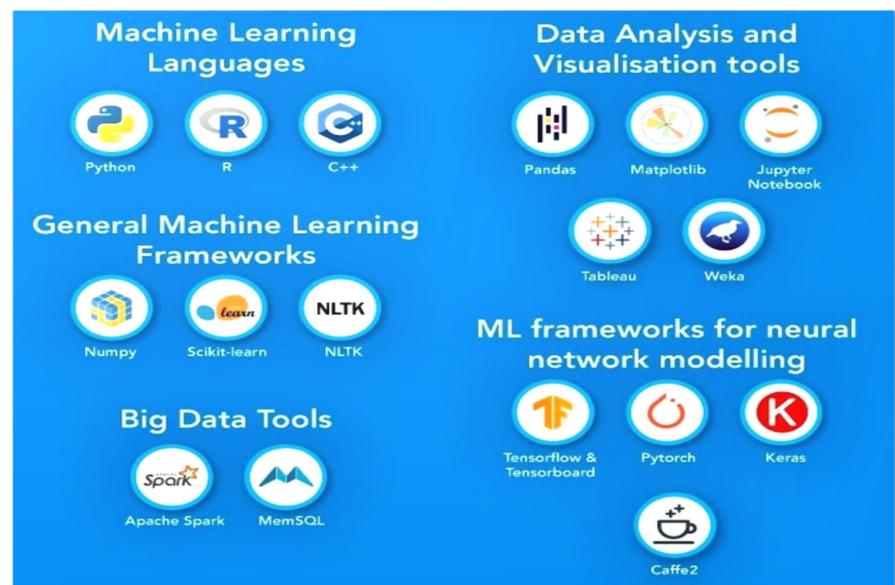


Figure 3: Tools and Technologies for data Science.

Discussion

This research proposes full implementation of Gaussian mixture model to cluster traffic and accurately predict traffic behaviour. Kornysky and et.al (2019) used Gaussian mixture model in wireless networks, achieving impressive results however it has not been used for internet traffic.

Other machine learning algorithms such as Naive Bayes, k-nearest neighbors, neural networks, logistic regression, clustering algorithms, and Support Vector Machines (SVM) have been applied for traffic detection and prediction but with varying degrees of unsatisfactory reliable and accurate results.

- The following range of data science cutting edge technologies will be used but not limited are: Python, Pandas, Numpy, Python Socket, Machine Learning Algorithms, Apache Spark, Nginx, Hadoop, Deep Learning (Karas, Tensorflow, Pytorch) Naive Bayes, k-nearest neighbors, Neural Networks, Logistic Regression, Clustering Algorithms, and Support Vector Machines (SVM).
- The data set will be a mixture of data gathered from second hand sources
- The research is at early stages. A considerable amount time will be spend on data cleansing, normalizing and feature selections.
- The data analysis and results will be used to answering the research questions
- An evaluation and recommendation will be provided.

Conclusion

Applying Gaussian mixture model to the network traffic is expected to result in higher rates of accuracy in both clustering and predicting network behaviour in comparison to current models.

References

1. Dashevskiy, M., & Luo, Z. (2014) Network Traffic Classification and Demand Prediction. 10.1016/B978-0-12-398537-8.00012-2.
2. VIAVI SOLUTION INC. (2020) State of the Network | 2019 NetOps and SecOps Converge. [Online] Available At: <https://www.stateofthenetwork.com/studies/2019.php>. [Accessed April 15 2020].
3. Yu, S., Liu, M., Dou, W., Liu, W. & Zhou, S. (2017) Networking for Big Data: A Survey. IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 19, NO. 1, FIRST QUARTER 2017531.