Mandeep Sarangal

251000108

Data Analytics Foundations

Review for: *Wafa Anani* and *Shuo Li* (April 4, 1:50 pm)

Summary

The Problem being addressed is related to modelling a system for Network Intrusion detection. Recurrent Neural network (RNN) model was built using two architectures i.e. Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). LSTM solves for vanishing gradient problems and also maintains more constant error flow in the Backpropagation problem. On the other hand, GRU offers less computation cost and also merges cell state and hidden state. The main aim was to identify optimal parameters such as hidden layers and learning rate for achieving high accuracy.

The dataset used was KDD CUP 99. Data was divided into ratio 80:20 for training and test respectively. Feature selection employed 2 approaches: RandomForest and Principal Component Analysis (PCA). Finally, 12 out of 49 features were selected. The preprocessing of the data included scaling and normalization. Different sets of parameters were tried which included training rates, number of hidden layers, Backpropagation and training cycles.

To conclude, Feature selection with Random forest showed better results than PCA. Although GRU takes less computational time and trains the model faster, LSTM model outperforms GRU in every cycle of training and has a better accuracy.

Criticism

The presentation delivery was good and the content was relevant to the problem at hand. Both the speakers were articulate and did a good job at keeping the whole presentation coherent. Data pre-processing was done with good amount of forethought. Feature selection was also done carefully in order to minimize redundancy and get high accuracy model. Final claims about RandomForest being a better feature selection approach and LSTM being a better overall model were backed by some solid experimental data like graphs and tables.