**A Convolutional Neural Network for Modelling Sentences**

This paper develops a model for modelling sentences using a Dynamic Convolutional Neural Network (DCNN). This approach takes in input sentences of different lengths and induces a feature graph that determines both the long and short term relationships between the words. They tested the approach in 4 different experiments and saw significant improvement over existing methods in their results. This is applicable since it deals with generating synthetic data, in this case synthetic sentences, but the process remains the same.

[1]

**Improved Training of Wasserstein GANs**

This paper discusses the stability improvements of a Wasserstein Generative Adversarial Network (GAN). They also discuss the limitations that it produces with weight clipping and demonstrate some of the flaws is produces. To alleviate this problem, they propose a new method of a gradient penalty to maintain the 1-Lipschitz property of the Wasserstein GAN without the instability produced by weight clipping. This is applicable since it is a modification of a GAN for discrete data like I will be dealing with.

[2]

**Complex log file synthesis for rapid sandbox-benchmarking of security- and computer network analysis tools**

* Generation Method
  + Data Points are single log lines
  + Run Log Line Clustering to build clusters and then assign log lines to clusters
  + Select clusters from candidates
  + Model clusters as Markov chains
* Evaluation Method
  + Mean Coverage rate of clustering algorithm
  + Number of Outliers
  + Number of Clusters
  + These metrics evaluated on semi-synthetic data & reference logs

[3]

**Generative Model for Category Text Generation**

* Uses SGD as optimization method
* Mapped to real interval with SoftMax function
* Generation Method
  + Long Short-Term Memory (LSTM) w/classifier
  + RNN & RL as discriminator & classifier
* Evaluation Method
  + Negative Log-likelihood
  + Classification results on synthetically generated sentences

[4]

**Adversarial Feature Matching for Text Generation**

* Optimization technique is to minimize a NN-based MMD distance of real data distribution & synthetic data distribution
* Evaluate with BLEU score & Kernel Density Estimation

[5]

**Generating Synthetic Mobility Traffic Using RNNs**

* Data points are trajectories of users (in this case trajectories is a collection of lat/longs)
* Develop a synthetic traffic generator based on an LSTM RNN
* Coordinates mapped to a grid then feature exploration is performed on the grid
* Extracted feature vectors then used for training

[6]

**Toward Controlled Generation of Text**

* idea is to generate realistic (plausible sentences) with Machine Learning
* Model is a variational autoencoder fed into a generator/discriminator
* Generator is LSTM-RNN
* Variable Auto Encoder uses sleep-wake algorithm
* code at https://github.com/asyml/texar/tree/master/ examples/text\_style\_transfer

[7]

**Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization**

Discusses the lack of good and available datasets for training Network Intrusion Detection Systems (NIDS). Comprehensively analyzes the available datasets and then presents a framework for evaluating the quality of datasets. They then develop their own dataset and analyze it according to their framework. Main contribution to work is the analysis of available datasets and the evaluation framework for datasets.

[8]

**Generative Adversarial Nets**

Seminal paper on the idea of a Generative Adversarial Network. Presents the idea of an adversarial game played between a generator and a discriminator in which the generator tries to generate data that will trick the discriminator into thinking it is real data. Solid idea that has become the basis for lots of data generation techniques with machine learning. Basic idea is the basis of the work I will be doing.

[9]

**Improved Techniques for Training GANs**

Improves on the framework of GANs by updating its framework and adding in semi-supervised learning. This reduces some of the mode collapse issues of the original GAN. Quality of generated data is significantly improved. Applicable to my work since I will be utilizing the GAN idea.

[10]

**DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning**

[11]

**Wasserstein GAN**

[12]