**Sparse Neural Nets: Improving Feedforward Neural Networks Efficiency with Structured Sparse Matrices**

**Abstract**

We show empirical evidence that sparse connectivity in between layers in a Feed Forward Neural Network does not impact accuracy significantly, as compared to a fully connected layer. Using the canonical MNIST data set, we compute accuracy measures for many Feed Forward Neural Nets with different connection schemes and topologies, showing there is no significant drop off as low as 10% of connections in the fully connected setting. With sparse connection paradigms, neural networks can be stored with less memory and predictions can be made with faster or with less computation resources, suggesting suitable applications in IoT.

To speed up computation, we proved empirically that using a sparse regular matrix is equivalent to using a sparse structured matrix. (Kz needs to come up with a mathematical proof ? )

**Structure of the article**

**I Introduction: How feedforward neural networks algorithms works**

**-Current state of Neural Networks, how they’re obscure, and push to Deep Learning vs. shallow**

Neural Networks are fast becoming the learning method of choice, and have become the state-of-the-art models for many areas \* (find out which areas). Despite this, very little is known about the theoretical underpinnings of the non-parametric learning method; understandably so.

**-Feed forward paridigm briefly, and generalized complexity of FF FC with some picture**

**II Explain our method : train with only a fix number of edges chosen by random**

**II Experimental Setup**

**-What we mean by Sparsity (picture)**

**-How we impose sparsity (structured vs. random) and big O notation of methods**

**-Methods: keraspatal, computer architecture, training paridigm**

**III Explain our method : train with only a fix number of edges chosen deterministically**

**III Results**

**some graphs and some results in terms of accuracy etc.**

**IV More sparse structured matrix that would get good results**

**IV Analysis (maybe theoretical thing from Doc K)**

**Fully Connected Networks do not add as much accuracy, over parameterized, with quick wins coming with very little connections, and additional ones, adding very little.**

**V Conclusion**

This paper will revolutionize the tech industry using feedforward neural networks, Google will save millions of dollars by only training its neural networks with sparse matrix, and we will not get any of this because we did it for science !!

I hope we will at least get an A+ ! hahahaha!

**Bring home major point : most of NN accuracy, comes with few connections**

**Link to IoT**

**Future Work for other Grad Students:**

- Implementing in different Neural Nets (Convolutional neural nets, are in a sense sparse)

- Training analysis (does it take longer, shorter, more or less local minima?)

- Datasets where it succeeds, where it fails

- More information on topologies vs. sparsity

- Computational resources were a big limitation

- Maybe sparsifying intentionally during training (Drop out and Stay out)