**Deep Learning using Big Data**

Research Question:

An exploration of pruning on different types of Neural Networks (ANNs, CNNs, RNNs).

**Overview**

*Deep Learning*

Deep learning (DL) is the subfield of artificial intelligence that focuses on creating large neural network models that are capable of making accurate data-driven decisions, particularly suited to contexts where the data available is complex and large (Kelleher, DL). There have been three waves of developments in DL: cybernetics in the 1940s-1960s, connectionism in the 1980s-1990s and the current resurgence beginning in 2006 (Goodfellow, DL). DL has attracted much attention from the academic community due to its state-of-the-art performance in many research domains such as speech recognition (G. E. Dahl, D. Yu, L. Deng and A. Acero, "Context-dependent pretrained deep neural networks for large-vocabulary speech recognition", IEEE Trans. Audio Speech Lang. Process., vol. 20, no. 1, pp. 30-41, Jan. 2012), (G. Hinton et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups", IEEE Signal Process. Mag., vol. 29, no. 6, pp. 82-97, Nov. 2012.), collaborative fultering (R. Salakhutdinov, A. Mnih and G. Hinton, "Restricted Boltzmann machines for collaborative filtering", Proc. 24th Int. Conf. Mach. Learn., pp. 791-798, 2007.), and computer vision (D. Cireşan, U. Meler, L. Cambardella and J. Schmidhuber, "Deep big simple neural nets for handwritten digit recognition", Neural Comput., vol. 22, no. 12, pp. 3207-3220, 2010.),(M. Zeiler, G. Taylor and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning", Proc. IEEE Int. Conf. Comput. Vis., pp. 2018-2025, Nov. 2011.). Successful applications of DL are found in industry products that take advantage of large volumes of data. Google, Apple, and Facebook are examples of companies that gather and analyze vast amounts of data on a daily basis and implement DL algorithms to provide specific services (**EXAMPLES?**) (Big Data Deep Learning: Challenges and Perspectives).

Contrary to popular belief, and media accounts emphasizing the similarity of DL to the brain, rough guidelines are drawn from neuroscience to the fundamentals of DL. The basic idea of having many computational units that become intelligent via interaction with each other is inspired by the brain. While neuroscience is an important source of inspiration, it need not be taken as a rigid guide. We know actual neurons compute very different functions than modern rectified linear units, but greater neural realism has not yet led to an improvement in machine learning performance. Also, while neuroscience has successfully inspired several neural network architectures, we do not yet know enough about biological learning for neuroscience to offer much guidance for the learning algorithms we use to train these architectures (**SUMMARIZE**).

This paper takes a primary focus on the application of pruning methods on different types of Neural Networks. The motivation behind this research focus was to assess the tradeoff between model size reduction, and hence run time of neural networks, versus model performance. The three types of Neural Networks explored in this research paper are Artificial Neural Networks (ANNs), Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs).

The purpose of this literature review is to summarize and comment on existing literature in the areas of; applications of pruning methods related to different neural network architectures,