

2007

Contents

1	Introduction	3
2	The architecture of Deep Neural Netowrks	3

1

Introduction

Deep learning as a broad framework of methods has taken the speech and natural language processing systems by a storm. The deluge of work that has happened in the last 25 years or so has shown remarkable promise and has also underscores the need to present here the progress that has been made since the late 80's. One of the primary goals of our paper, therefore, is to highlight some of the most significant approaches and the findings therein. While our approach mostly will be chronological, we will also focus on how deep neural networks (DNN) have fundamentally revolutionized our approach to both Automatic Speech Recognition (ASR) and Text-to-Speech Synthesis (TTS). In addition, we want to focus on a very specific aspect within the use of DNNs in speech processing, namely the integration of linguistic knowledge in achieving some of the remarkable successes in the core tasks of speech processing.

In a series of seminal papers, Bengio *et al.* (1989a,b) outline the use of multilayered neural networks in ASR and speaker recognition, respectively.

2

The architecture of Deep Neural Netowrks

Typically, DNNs refer to feedforward multi-layered artificial neural networks (ANN) with more than one layer of hidden units with a logistic function to traverse between the hidden layers and the output. Here are relying heavily on Hinton *et al.* (2012) to outline the general architecture of DNNs. We will illustrate the functioning of the algorithms and the processes with an acoustic modelling task as discussed in Hinton *et al.* (2012).

$$y_j = \text{logistic}(x_j) = \frac{1}{1 + e^{-x_j}}, x_j = b_j + \sum_i y_i w_{ij}, (1)$$

$$p_j = \frac{\exp(x_j)}{\sum_k \exp(x_k)} \quad (2)$$

$$C = -\sum_j d_j \log p_j \quad (3)$$

$$\Delta w_{ij}(t) = \alpha \Delta w_{ij}(t-1) - \epsilon \frac{\delta C}{\delta w_{ij}(t)} \quad (4)$$

References

- Bengio, Y., Cardin, R., De Mori, R., and Merlo, E. (1989a) Programmable execution of multi-layered networks for automatic speech recognition. *Commun. ACM*, **32** (2), 195–199, doi:10.1145/63342.63345. URL <http://doi.acm.org/10.1145/63342.63345>.
- Bengio, Y., De Mori, R., and Cardin, R. (1989b) Speaker independent speech recognition with neural networks and speech knowledge, in *Proceedings of the 2nd International Conference on Neural Information Processing Systems*, MIT Press, Cambridge, MA, USA, NIPS’89, pp. 218–225. URL <http://dl.acm.org/citation.cfm?id=2969830.2969857>.
- Hinton, G., Deng, L., Yu, D., Dahl, G., rahman Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., and Kingsbury, B. (2012) Deep neural networks for acoustic modeling in speech recognition. *Signal Processing Magazine*.