

Deep learning architectures in speech processing

December 9, 2016

1 Outline

1. A (very) brief overview of 'black box' and 'white/glass box' methods would be needed, starting with the old 'statistical vs. rule-based' fight. We can pool ideas about this
2. If possible we should have some kind of working demos to illustrate what we are concerned with.
3. A conclusion could contain something like tasks which are ideal for DNN etc. and tasks which are out of its scope."

```
\begin{enumerate}  
\item  
\end{enumerate}
```

What is ML?

%What is rule-based learning (i.e. generalisation,
% subsumption, automaton learning)?

% What is statistical learning?

% What is "training"?

% What is the difference between supervised and unsupervised learning?

% What is the difference between classification and sequence learning?

% What are NNs/ANNs?

What are DNNs and what is the essential literature on them?

How do they relate to ML?

```
\begin{enumerate}
```

```
\item How are DNNs developed and evaluated in practice?
```

```
\item Where can working demos of DNNs be found?
```

```
\item Who does ML and who does DNNs?
```

```
\item What are the results of ML and DNNs in comparable tasks?
```

```
\item How do ML and DNN relate to current topics such as big data,  
surveillance, chess-playing, etc.?
```

```
\item What kind of insights can ML and DNN provide for linguists?
```

```
\item How do DNNs relate to first language acquisition and foreign  
language learning?
```

```
\end{enumerate}
```

2 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, ?? outline the use of multilayered neural networks in ASR and speaker recognition, respectively.