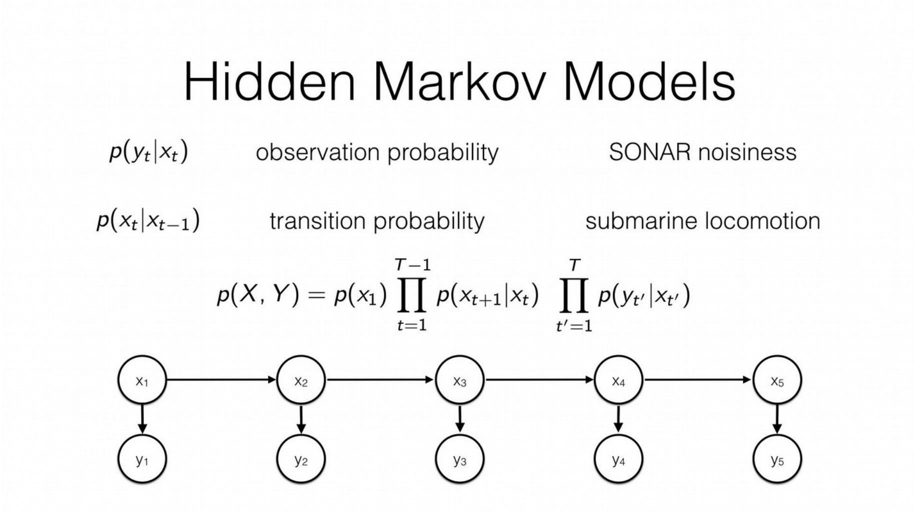
Speech Recognition

Hidden Markov Model

References: http://www.anthology.aclweb.org/J/J93/J93-1016.pdf

Hidden Markov Models (HMMs) provide a simple and effective framework for modelling time-varying spectral vector sequences. As a consequence, almost all present day large vocabulary continuous speech recognition (LVCSR) systems are based on HMMs. the approximations and simplifying assumptions involved in a direct implementation of these principles would result in a system which has poor accuracy and unacceptable sensitivity to changes in operating environment.



However, the brief paper doesn’t have a detailed description of the methods and results. Moreover, there are multiple reasons for RNN to be better than HMM as revealed by our further survey.

Recurrent Neural Networks

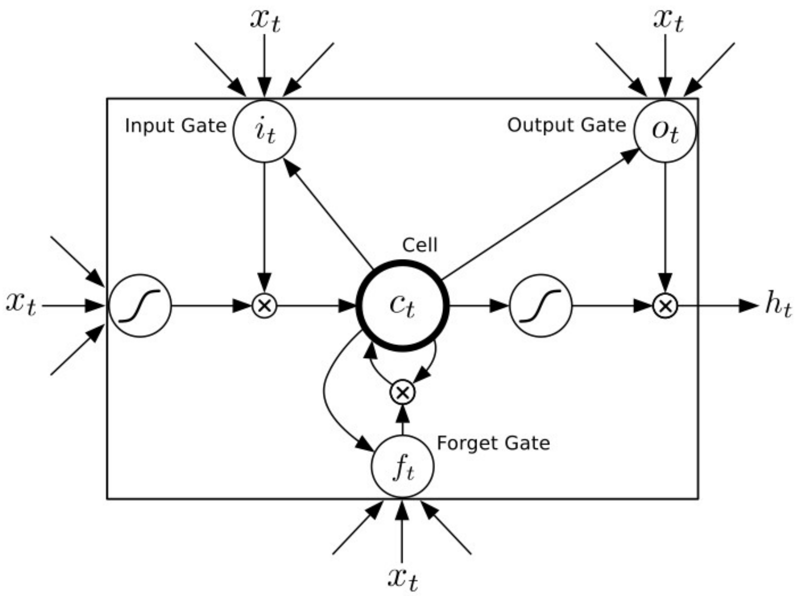
References: https://www.cs.toronto.edu/~fritz/absps/RNN13.pdf

Recurrent neural networks are powerful model for sequential data. End-to-end training methods make it possible to train RNNs for sequence labelling problems where the input-output alignment is unknown. The combination of these methods with the Long Short-term Memory RNN architecture has proved particularly fruitful, delivering state-of-the-art results in cursive handwriting recognition. However RNN performance in speech recognition has so far been disappointing, with better results returned by deep feedforward networks.

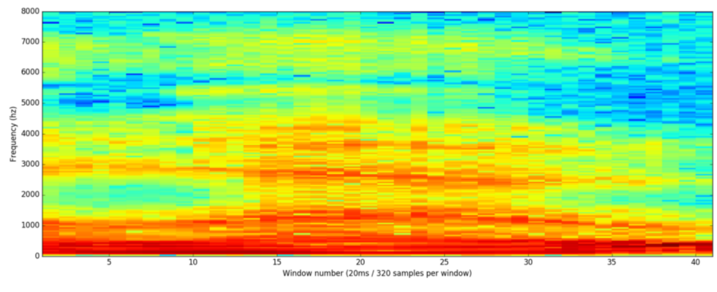
RNN:

A recurrent neural network (RNN) is a class of artificial neural network where connections between units form a directed graph along a sequence. This allows it to exhibit dynamic temporal behavior for a time sequence. Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition.

A single LSTM unit is shown below.



The pre processing of data is done in the way as described in previous report, by generating spectograms of the speech samples. This paper has reported the best accuracy ever acheived for speech recognition by RNNs.



A Spectogram

Natural Language Processing

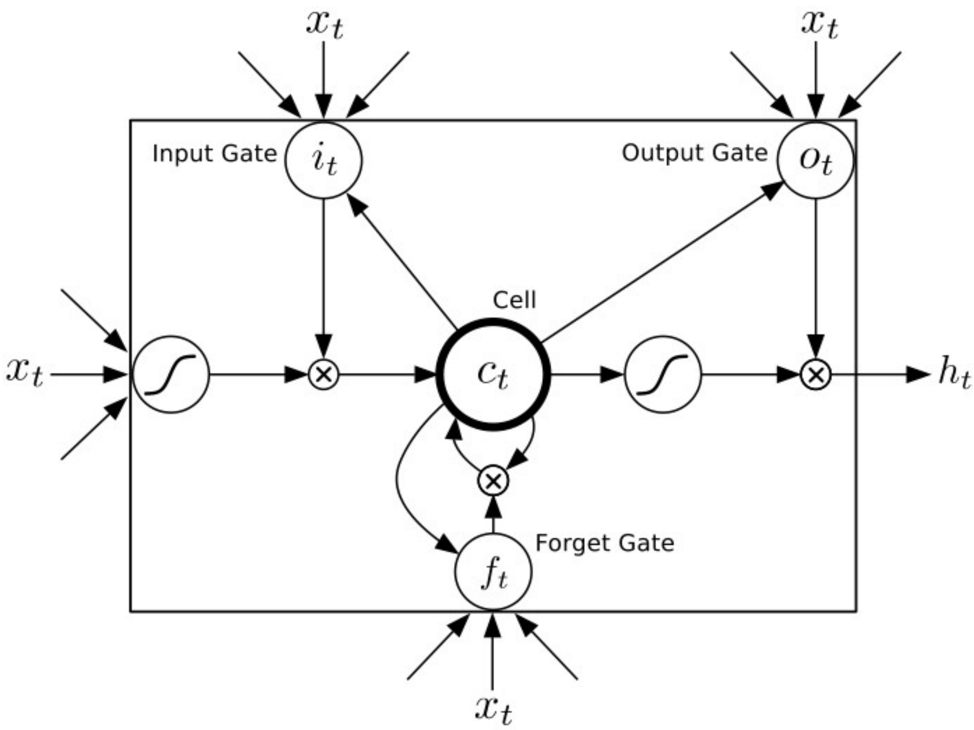
Recurrent Neural Network

References: https://nlp.stanford.edu/pubs/SocherLinNgManning_ICML2011.pdf

RNN

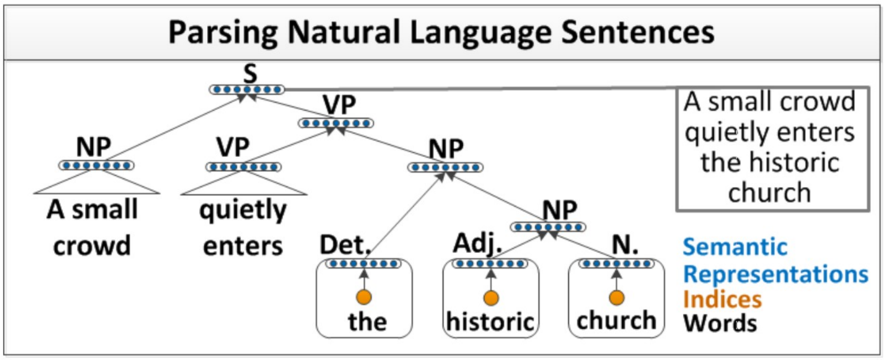
A recurrent neural network (RNN) is a class of artificial neural network where connections between units form a directed graph along a sequence. This allows it to exhibit dynamic temporal behavior for a time sequence. Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition.

A single LSTM unit is shown below.



Natural Language and RNN

Recursive structure is commonly found in the inputs of different modalities such as natural scene images or natural language sentences. Discovering this recursive structure helps us to not only identify the units that an image or sentence contains but also how they interact to form a whole.



The above figure is an Illustration of recursive neural network architecture which parses natural language sentences. Segment features and word indices (orange) are first mapped into semantic feature space (blue) and then recursively merged by the same neural network until they represent the entire sentence. Both mappings and mergings are learned.

Syntactico-Semantic Space

In order to efficiently use neural networks in NLP, neural language models map words to a vector representation. These representations are stored in a word embedding matrix. This matrix usually captures cooccurrence statistics and its values are learned.

Group members

111501029 Ashutosh Upadhye

111501023 Rahul Dhawan

111501028 Siddhardha SST

More References:

http://andrew.gibiansky.com/blog/machine-learning/speech-recognition-neural-networks/

https://www.slideshare.net/xavigiro/endtoend-speech-recognition-with-recurrent-neural-networks-d3l6-deep-learning-for-speech-and-language-upc-2017