

# Recurrent Neural Network Based Language Model – Approach

Mohan Khanal, [mkhanel2@illinois.edu](mailto:mkhanel2@illinois.edu)

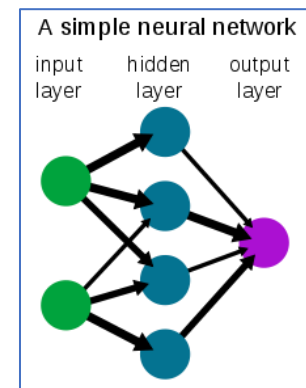
## Background

Statistical Language Modeling (LM) is building a statistical model which can estimate the distribution of natural language as accurate as possible. LM is the center piece for most of the Natural Language Processing(NLP) which includes automatic speech recognition, sentence completion, etc. One of the simplest statistical language model that assigns probabilities to sentences and sequences of words is the n-gram model. In an n-gram model, the  $P(w_1, \dots, w_n)$  of observing the sentence  $w_1, \dots, w_n$  is given as:

$$P(w_1, \dots, w_m) = \prod_{i=1}^m P(w_i | w_1, \dots, w_{i-1}) \approx \prod_{i=1}^m P(w_i | w_{i-(n-1)}, \dots, w_{i-1})$$

Till recent past, n-gram language model with smoothing techniques has been the most popular due to its simplicity and performance. But for the large vocabulary size, the parameters for n-gram model becomes large and it may become poor generalization to low frequency and unseen words. Due to that limitation many new techniques has been proposed. Among those technique neural network based langue model are one of the popular ones., like feed-forward neural network, recurrent neural network (RNN).

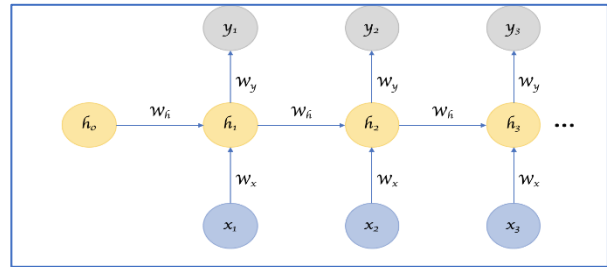
Neural Network (NN) / Artificial Neural Network(ANN) is information processing paradigms which is derived from the concept of how the human brain works. Like human brain, ANN has neurons connected with other neurons in the network. ANN has three different layers : input layer, hidden layer and output layer. In this NN, the input layer takes in input, and the output layer generates output. The hidden layers have no connection with the outer world that's why it called hidden layers. For one of the simplest network, a feed-forward neural network(FFN), every neuron in one layer is connected with each node in the next layer and there are no back-loops in the feed-forward network.



In more practical terms NN are non-linear statistical data modeling or decision making tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data. Neural network approaches are achieving better results than classical methods on language models. The key reason behind it is improved performance and ability to generalize. With all the benefits , one of the shortcoming of using simple neural network / feed-forward neural network for language processing is that it cannot capture sequential information in the input data. That is where the Recurrent Neural Network (RNN) comes into play.

## Recurrent Neural Network Language Model (RNNLM)

Recurrent Neural Network (RNN) is a variation of the neural-network model which contains the feedback loop for itself. Which means that each neurons/cell are not only dependent on the input at given time but are also dependent on the previous state of the network. This property of RNN allows us to preserve the sequential nature of the data, where-as in basic neural network model the assumption was that each input are independent of each-other. This is the reason RNN is most promising for language model compared to other neural network. As, RNN captures the sequential information present in the input data i.e. dependency between the words in the text. For example, when we are trying to predict the next word in a sentence, we need to know the previously used words first.



This how RNN Language Model can be trained:

- Get big corpus of text data with sequence of words which can be used to train the RNN
- Feed the training data to the model and compute the output distribution.
- Calculate loss in each step , which is given by the difference in predicted distribution and the actual next word
- Average the overall loss in the entire training set
- Use gradient decent to update the weights to minimize loss

Python and other popular programming languages already has a built-in library/package for carrying out the RNN model. We can simply use those packages as per our need.

With many advantages of RNN for Language model , it does have some disadvantages. The RNN with large time steps suffer from the vanishing and exploding gradient problem during training. There are methods like LSTM network which solves this problem, but those are out of scope for this article.

## Conclusion:

Recurrent Neural Network language Model (RNNLM) are one of most promising language model when it comes to text mining / text processing because of its ability to preserve the sequential nature of the data. Although it has some shortcoming but different research has shown that it has produced better results than the traditional statistical language model approaches.

### **References:**

- Recurrent Neural Networks Tutorial, Part 1 – Introduction to RNNs - <http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/>
- Recurrent neural network based language model - [https://www.isca-speech.org/archive/archive\\_papers/interspeech\\_2010/i10\\_1045.pdf](https://www.isca-speech.org/archive/archive_papers/interspeech_2010/i10_1045.pdf)
- Natural Language Processing with Deep Learning CS224 - Lecture 6: Language Models and Recurrent Neural Network : <https://web.stanford.edu/class/cs224n/slides/cs224n-2019-lecture06-rnnlm.pdf>
- Neural Network (Wikipedia) - [https://en.wikipedia.org/wiki/Neural\\_network#:~:text=A%20neural%20network%20is%20a,of%20artificial%20neurons%20or%20nodes.](https://en.wikipedia.org/wiki/Neural_network#:~:text=A%20neural%20network%20is%20a,of%20artificial%20neurons%20or%20nodes.)
- Main Types of Neural Networks and its Applications — Tutorial : <https://medium.com/towards-artificial-intelligence/main-types-of-neural-networks-and-its-applications-tutorial-734480d7ec8e>