

# A Survey on Neural Language Modeling Techniques

Anuar Maratkhan

## I. INTRODUCTION

A *language model* is an algorithm for predicting next word in a text given preceeding ones. It is also mostly known as a *statistical language model* for its richness in statistical approaches in previous works. However, as soon as neural networks became popular, a *neural language modeling* approaches has been invented. The language modeling is used in speech recognition, handwriting recognition, machine translation, information retrieval and other natural language processing applications.

First, we will discuss earlier approaches to language modeling, particularly statistical ones that are *unigram*, *bigram*, *trigram*, and more generalized *N-gram* models. Further, we proceed by reviewing nowadays' state-of-the-art - neural language models. And by the end, we will summarize the topic and give future suggestions on language models.

Neural Language Models, also known as continous-space language models, make use of neural networks. These are current state-of-the-art approach in language modeling. One reason of Neural Nets' use is that they can learn from unstructured data. The studies show that it is possible to learn from unlabeled data [1]. One of the common applications of such new machine learning techniques is search engines, which is mainly information retrieval that consists of unlabeled data. That means neural nets are possible to learn from that unstructured data and perform well on that.

## II. WORD-LEVEL LANGUAGE MODELING

Recurrent Neural Networks (RNN) are used in sequential data analysis such as video processing, text processing, predicting stock prices. RNN architecture is an improvement of feed-forward networks that takes into account previous output results by storing them in memory. An invention of backpropogation caused an exciting use of RNNs [2]. But the problem in simple RNN was that the backpropogated gradients vanshish or explode [2].

The improvement in RNN architecture further was introduced with a Long-Short Term Memory (LSTM). The LSTM solved the vasnishing gradient problem by having special cells with activation gates. Moreover, Gated Recurrent Unit (GRU) improved the architecture of the LSTM by combining two gates. The study shows that RNNs perform better than complex statistical approaches, N-grams, on a real-world data [3].

## III. SUBWORD-LEVEL LANGUAGE MODELING

In progress...

## IV. REINFORCEMENT LEARNING

This section will show some Reinforcement Learning (RL) approaches to language modeling tasks and applications such as dialogue systems, machine translation, text generation.

## V. NEURAL TURING MACHINES

This section will discuss Neural Turing Machines (NTM) approach to different natural language processing tasks with controllers like feed-forward network, GRU, LSTM, and further will show applications in learning language models. Related papers are []

## VI. CONCLUSION

Over the past few years, it was clearly seen that continous-space language models outperformed statistical modeling techniques significantly. The ease of feature learning without need in extraction of those features simplified lives of scientists. The neural approach offers such opportunity to its users.

Recurrent Networks are expected to improve natural language understanding in real-world applications.

To be continued...

## REFERENCES

- [1] Q. V. Le, M. A. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Deau, and A. Y. Ng., "Building high-level features using large scale unsupervised learning," in *Proceedings of the 29 th International Conference on Machine Learning*, Edinburgh, Scotland, UK, 2012.
- [2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Macmillan Publishers Limited*, vol. 521, pp. 436–444, 2015.
- [3] A. Karpathy, J. Johnson, and L. Fei-Fei, "Visualizing and understanding reccurent networks," 2015, under review as a conference paper at ICLR 2016.