

# 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 *statistical language model* for its richness in statistical approaches in previous works. However, as soon as neural networks became popular, *neural language modeling* approaches were invented. Language modeling is used in speech recognition, handwriting recognition, machine translation, information retrieval and other natural language processing applications.

This part of introduction will conclude earlier approaches to language modeling, particularly statistical ones that are *uni-gram*, *bigram*, *trigram*, and more generalized *N-gram* models.

Further, we will proceed by reviewing today's 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 continuous-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 study [1] show that it is possible to learn from unlabeled data and perform well on that. One of the common applications of such new machine learning techniques is search engines, which is mainly information retrieval that consists of unlabeled data.

In second section we review word-level language modeling techniques, starting from [2] proceeding by [3], [4], [5], and [6]. Next section presents different approach on neural language modeling, division of words into subwords. Fourth section includes reinforcement learning techniques on language modeling tasks. Lastly, neural Turing machines approach is reviewed.

## II. WORD-LEVEL LANGUAGE MODELING

### A. Recurrent Neural Network

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 backpropagation caused an exciting use of RNNs [7]. The problem of simple RNN is that the backpropagated gradients vanish or explode [8].

The improvement in RNN architecture further was introduced with invention of Long-Short Term Memory (LSTM). The LSTM solved the vanishing 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 [8].

## III. SUBWORD-LEVEL LANGUAGE MODELING

This section will show how subword language modeling exceeds the performance of word-level and character-level language modeling by presenting previous approaches from [9], etc...

## 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 [10], etc...

## VI. CONCLUSION

Over the past few years, it was clearly seen that continuous-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. Dean, and A. Y. Ng., "Building high-level features using large scale unsupervised learning," in *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, Scotland, UK, 2012.
- [2] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, and S. Khudanpur, "Recurrent neural network based language model," in *Interspeech*, vol. 2, 2010.
- [3] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," 2014, arXiv preprint arXiv:1409.2329.
- [4] Y. Gal and Z. Ghahramani, "A theoretically grounded application of dropout in recurrent neural networks," in *Advances in neural information processing systems*, 2016, p. 10191027.
- [5] H. Inan, K. Khosravi, and R. Socher, "Tying word vectors and word classifiers: A loss framework for language modeling," 2016, arXiv preprint arXiv:1611.01462.
- [6] Z. Yang, Z. Dai, R. Salakhutdinov, and W. W. Cohen, "Breaking the softmax bottleneck: A high-rank rnn language model," 2017, under review as a conference paper at ICLR 2018.
- [7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Macmillan Publishers Limited*, vol. 521, pp. 436–444, 2015.
- [8] A. Karpathy, J. Johnson, and L. Fei-Fei, "Visualizing and understanding recurrent networks," 2015, under review as a conference paper at ICLR 2016.

- [9] T. Mikolov, I. Sutskever, A. Deoras, H.-S. Le, S. Kombrink, and J. Cernocký, "Subword language modeling with neural networks," 2011.
- [10] T. Deleu and J. Dureau, "Learning operations on a stack with neural turing machines," in *1st Workshop on Neural Abstract Machines Program Induction (NAMPI)*, Barcelona, Spain, 2016.