Technology Review – Statistic Language Models

Overview of Neural Net Language Models – Past Present and Future

Yuriy Kotskyy – ykotsk2

Introduction

The goal of this paper is to provide a high-level overview of neural network language models (NNLM). Some historical background is provided to establish context for the development of NNLM followed by an overview of modern models culminating in a brief analysis of potential future developments in the field.

Historical Background

Neural networks were initially inspired by the biological neurons found in the human brain although the architecture and implementation of them has drastically diverged from their biological roots. They have many practical uses and are a high activity research topic in the field of computer science. In natural language processing (NLP), statistical language models have been standard for many decades, but they have a fundamental problem which can't be solved through smoothing. The so called "curse of dimensionality" limits modeling on a larger corpus of text and is obvious in the case when one wants to model the joint distribution in discrete space. For example, an n-gram language model with a vocabulary of size 10,000 can have up to (10000^n) -1 free parameters.

Starting in the 1980s, the development of deep and recurrent neural networks combined with exponential increases in computing power and training data have allowed NNLM to become the dominant paradigm in NLP today. The goal of NNLM is the same as that of other language models: reduce high-dimensional discrete probability distributions to low-dimensional continuous distributions that have many practical uses such as recognizing speech, handwriting, and recognizing spelling errors. While the ideas behind feedforward and recurrent Neural Networks go back as far as the 1950s, the first feed forward neural network language model (FFNNLM) was presented in 2003 by [Bengio et al.,] and the first recurrent neural network language model (RNNLM) was presented in 2010 by [Mikolov et al.,].

FFNN Language Models

FFNNLM implement modeling on continuous space by learning a distributed representation for each word and overcomes the curse of dimensionality by converting words into low-dimensional vectors. Although it is currently the leading trend in NNLM research it has a few drawbacks which have yet to be overcome. The size of the context is specified before training which is different from realist where people have access to a virtually unlimited context library to make predictions. Moreover, words in a sequence are time related and this language model does not account for timing information when making its predictions. Finally, it is still expensive and inefficient as there are many parameters for the neural network to learn although the performance is a significant improvement when compared to n-gram language models.

RNN Language Models

RNNLM take several steps towards solving the contextual problem. By introducing more structure and parameter sharing into the neural network they can capture longer contextual information and outperform FFNN in several benchmarks. Their inherent advantage is being able to receive variable length inputs which significantly reduces the number of calculations needed. Variable length inputs are also able to generate timing information and significantly reduce the number of parameters. This creates a further problem as variable input lengths can lead to slow training huge inconsistencies in runtime as parameters can disappear or go to infinity repeatedly during the training process.

LSTM-RNN Language Models

Long Short-Term Memory RNNLM were able to solve these problems using three gate structures which control the flow of information. This makes them the current state of the art language models although depending on performance needs RNN based language models are still widely used because training them is much more time consuming on a large corpus as the gate structures need to normalize and consider all the words when computing log-likelihood gradients. Current research is dedicated to developing different techniques that mimic the way humans process natural language with the goal of increasing performance of LSTM-RNN and RNN based language models.

Further Improvements

The field of linguistics and human language processing is playing an increasingly important role in improving NNLM in several ways. Character-aware models are based on the idea that in natural language some words with similar forms have the same or similar meanings. For example, “man” in “superman” and “policeman” both have the same meaning which people can understand but language models can’t. Unfortunately, character level models have worse performance than word level models, but years of research have created character level models which are more accurate in prediction with rather poor performance.

Factored models attempt to integrate word shape features into the NNLM by looking at factors such as affixes, uppercase, hyphens or parts of speech. Incorporating these factors yields mixed results and, in some cases, they can perform extremely well. However, these models require a lot more effort to run as the factors are not able to be dynamically selected and time must be spent modifying the text corpus with this information before it is fed into the model.

Caching relies on the idea to keep track of words which have appeared recently in case they may appear again. Performance impact varies greatly based on the technical implementation of the cache itself, but it has been proven to reduce computational complexity and improve model learning. Additionally, it is easy to implement in existing models, but the challenge lies in determining the appropriate size of the cache. One that is too small would negatively impact performance while one that is too large would be a waste of computational resources.

Conclusion

NNLM were a major improvement over statistical language models by solving a fundamental problem which plagued NLP for a long time. However, they are themselves not perfect and after 20 years of research still have flaws which scientists are attempting to solve by drawing inspiration from the fields of linguistics and human language processing. Today, NNLM remain a highly active research field in NLP with a bright and interesting future.

References

1. A Short Overview of Statistical Language Models
2. Neural Language Models
3. Recurrent neural network-based language model
4. A Survey on Neural Network Language Models
5. A Brief History of Neural Nets and Deep Learning