Recurrent Neural Network Language Model

CS410 Tech Review, Fall 2021, University of Illinois at Urbana Champaign

Ryan Baker (NetID: rtb2)

Recurrent neural networks are a group of machine learning networks that form a directed graph. The inputs are processed through that graph until the final layer. Recurrent neural networks are one of the most common types of deep neural networks. Their defining feature is the ability to process information temporally and preserve some context information via an internal state mechanism that is often referred to as memory because of the network's ability to "remember" what came before the current input. This feature makes recurrent neural networks ideal for processing natural languages, with applications such as handwriting recognition. As has been taught in this course, there are various forms of language models that are used in text and information retrieval and text mining. These language models can be used to perform tasks such as identifying the most relevant document in a search query. There have been significant efforts during the last decade to identify how recurrent neural networks can be used as a language model in place of other methods, in particular with real world applications such as speech recognition and transcription. The remainder of this review will focus on describing several different research efforts in this area and summarizing the major findings of these efforts with respect to using recurrent neural networks as a language model.

One of the first important research works published on this subject was in 2010 from Mikolov, et al [1]. This research presents a language model built from several individual recurrent neural networks language models, comparing it on several datasets with the state of the art at that time, which was a back-off probabilistic n-gram language model. With n-gram based language models, there have been significant efforts to preserve history via various forms of caching. One of the strengths of the recurrent neural network language model is that this ability is native to the approach. The key finding of this research is that recurrent neural network language models have an inherent ability to perform better at real-life applications such as speech recognition. Compared to the state of the art n-gram based method of the day, the recurrent neural network approach recorded up to an 18% lower word error rate, a very significant improvement over the n-gram approach. These results and others have been repeated in various studies to find that recurrent neural networks are indeed a very viable language model for information retrieval and text mining tasks.

Subsequent to this research, other efforts have been made to more fully identify the various strengths and weaknesses of the recurrent neural network language model. One of the most immediate and apparent drawbacks of the recurrent neural network language model is that deep learning is not as approachable as other more "intuitive" methods. It is generally well accepted within the community that deep learning approaches to problems require higher levels of expertise to deliver, and they are more difficult to interpret or explain. They can also take significant time to train in the first place. Fortunately, in the same initial research, Mikolov et al.

were able to show that using recurrent neural networks for a language model requires significantly less data in order to achieve the same, or better, performance. Even with 5 times more data, an n-gram model was not able to achieve the same performance as the recurrent neural network language model [1].

In subsequent research, experts were also able to determine that recurrent neural networks are robust to being able to add additional features to the inputs and consistently improve the performance of the model. This is an important breakthrough in working with recurrent neural networks. In traditional n-gram based approaches, it has been shown in various circumstances that adding additional information such as part-of-speech (POS) tagging, stemming or syntactic information to the inputs is able to increase model performance. Researchers were able to build on these efforts and apply such information to a recurrent neural network. After doing so, specifically with POS tagging and stemming, the language model saw a modest improvement in word error rate over the state of the art recurrent neural network language model, which in turn performs significantly better than the state of the art n-gram based model [2]. Another group was able to perform similar research, with the addition of socio-circumstantial features, and was able to once again achieve better performance than the current best recurrent neural network language model [4].

Another example of applying adjustments from traditional language models to recurrent neural network models is enhancing performance by feeding not just words, but phrases or entities into a recurrent neural network [3]. Standard recurrent neural networks process one word at a time, similar to how the traditional n-gram based approaches review one word at a time. Those traditional approaches were shown to improve when evaluating phrases or identified personal entities as one item instead of individual words. These improvements were replicated when allowing recurrent neural networks to process not just single words, but phrases and entities. All of these results are important because they show that even if direct changes to recurrent neural network models are difficult, as advancements are made to traditional n-gram based language models, these can subsequently be applied to the current best recurrent neural network based language models to achieve even better performance.

Finally, another important element of recurrent neural networks is that they can be effectively combined with other approaches that will address limitations of recurrent neural networks. The one advantage of n-gram based approaches of neural networks is their ability to model explicit position-dependent word interactions. These are common in conversational speech. Fortunately, these are well captured with common feedforward neural networks with a fixed window size. Oualil, et al were able to combine the architectures of a feedforward neural network with a recurrent neural network to create a unified language model that is able to preserve contextual information from the recurrent neural network as well as the position-dependent word information from the feedforward network with improved overall performance [5]. Once again, this shows how recurrent neural network language models are able to significantly improve upon existing traditional language model approaches without significant downsides.

Deep learning networks such as recurrent neural networks can certainly still be complicated for many different real-world tasks. They are often criticized as being “black-boxes” that cannot be explained or understood. And yet, they continue to be more commonplace and the benefits have been shown again and again, including in the realm of language modeling applied to things like information retrieval and mining. Recurrent neural network language models have significantly better performance and are able to continue evolving by incorporating known improvements from traditional n-gram based approaches as well as novel improvements specific to recurrent neural networks. And there still remains much more that can be understood and learned about their applications for yet further improvement.

References

[1] Mikolov, T., Karafiát, M., Burget, L., Černocký, J., Khudanpur, S. (2010) Recurrent neural network based language model. Proc. Interspeech 2010, 1045-1048, doi: 10.21437/Interspeech.2010-343
[2] Wu, Y., Yamamoto, H., Lu, X., Matsuda, S., Hori, C., Kashioka, H. (2012) Factored recurrent neural network language model in TED lecture transcription. Proc. International Workshop on Spoken Language Translation (IWSLT 2012), 222-228
[3] Levit, M., Parthasarathy, S., Chang, S. (2016) Word-Phrase-Entity Recurrent Neural Networks for Language Modeling. Proc. Interspeech 2016, 3514-3518, doi: 10.21437/Interspeech.2016-44
[4] Shi, Y., Wiggers, P., Jonker, C.M. (2012) Towards recurrent neural networks language models with linguistic and contextual features. Proc. Interspeech 2012, 1664-1667, doi: 10.21437/Interspeech.2012-456
[5] Oualil, Y., Greenberg, C., Singh, M., Klakow, D. (2016) Sequential Recurrent Neural Networks for Language Modeling. Proc. Interspeech 2016, 3509-3513, doi: 10.21437/Interspeech.2016-422