A Transformer and Recurrent Neural Network based Sentiment Analysis Method

Abstract

The dominant sequence transduction models are based on complex recurrent neural networks or Transformer models that include an encoder and a decoder. In recent years, language models are rising with attention mechanism proposed in Transformer before. The best performed language model also connect encoder and decoder with attention layers. The source of this research is based on BERT (Bidirectional Encoder Representation from Transformers), SHA-RNN (Single Headed-Attention with RNN), and attention mechanism from Transformer as the main axis. In previous work, Transformer works well whenever long range sequence is available, but they cannot be used to capture short dependency from a sequence. In this research, we present a general language model to improve the defect of capturing short-range sequence in Transformer. Our model replaces the attention layer in decoder from Transformer with SHA-RNN because doing so introduced many short-term dependencies between the source and the target sentence which make the optimization problem easier. The result of this research is developed for improving specific tasks well by applying to sentiment analysis, social media analysis, and disease spreading prediction.

1. Introduction

Natural language processing is the main objective of combining deep neural networks and linguistics, focusing on the communication between natural language and computers. Natural language processing includes natural language understanding (NLU), and natural language generation (NLG). NLU and NLG are introduced to understand input made in the form of sentences in text or speech format, and transform structured data into natural language respectively. It's important to realize that language is far more than human language. Language have many forms of encoding, each word is a signifier that maps to a signified.

Recurrent neural networks (RNN), Long short-term memory (LSTM), and Transformer have been resolutely established as state-of-the-art approaches in language modeling. Numerous efforts have since continued to push the limits of language model quality estimation. In RNN based models, words in the sequence are read in order and assigned with different wights. As the distances between words and depths of networks become further and deeper, the weights input earlier would be diluted. With the attention mechanism in Transformer, researchers

created the techniques to pay attention to specific words. For RNNs, instead of only encoding the whole sentence in a hidden state, each word has a corresponding hidden state hat is passed all the way to the decoding stage.

Compare with RNN and Transformer, Transformer introduced attention mechanism to improve time series problem which is a major defect in RNN. For an input token, its input representation is construct by summing the corresponding token, segment, and position embeddings. An input representation would pass through multi-head attention, feed-forward neural networks, and layer normalization. An output representation from encoder (also known as an input of decoder) would then pass through masked multi-head attention and feed-forward neural networks which are connected with residual connection.

What if the feature extraction technique of Transformer never existed, what happened would it be to the development in NLP ? Perhaps RNN still take the lead, and most optimized techniques would be RNN related. Thus, SHA-RNN was popular when it was introduced.

In this research, we suggest to rebuild the encoder-decoder architecture to improve the defect of capturing short range sequence in Transformer.