Abstract

Fake news detection is the task of classifying news according to its veracity. In a simple setting, it is a binary classification task, whereas in a more difficult setting it is a fine-grained classification task. Fake news detection is one of the hottest research topics in artificial intelligent recently. Due to the availability of the Internet and the readiness of sharing information through social media, it is easy to make fake news and propagate it worldwide. When being spread widely, the fake news may cause substantial adverse impact to many aspects of life. Consequently, an amount of research has been conducting recently to spot fake news. Despite receiving significant attention from the crowd, fake news detection did not gain much improvement for a while due to the insufficiency of fake news data. In fact, it was not until 2017 with the introduction of Liar dataset, has fake news detection shown some noticeable performance.

Most of the current methods applied to tackle fake news detection are based on deep learning on the ground that deep learning models have been achieving state-of-the-art results in a wide range of artificial intelligence such as natural language processing, computer vision, speech processing. Concretely, Convolutional Neural Network(CNN) and Long Short-Term memory network(LSTM) were used in fake new detection and yielded remarkable results. Having said that, fake news detection is still in its infancy with the accuracy of fine-grained being under 50%. In an attempt to improve the performance of fake news detection, we explore a range of techniques focusing on attention-based neural networks. In this thesis, our contributions are: (1) the implementation of single attention with an intermediate representation, (2) the modeling of two-way relationship between main text and connected information, (3) the proposal of memory network in word level and character level.

First of all, we present the simplest approach called single attention neural network for fake news detection. In this model, the statement is merely considered a sequence of words, and different types of side information are also considered a kind of sequential data. The side information is then represented as a unique vector. For clarity, we simply sum over all types of side information. This vector then act as an attention factor over the main text. Our experiment demonstrates that the single attention neural network surpasses the more complicated hybrid CNN which employs two CNNs and one LSTM to drive intermediate representations.

Our second approach stems from the question that whether the main text has reverse effect on the given dataset. To clarify this doubt, we try to model the two-way interaction between the text and its connected information. First of all, side information is processed as described in the single attention neural network. In the reverse fashion, the text is also twisted into a unique vector to act as an attention component. Similar to the case of single attention, we take summation of all vectors for words in the statement. Our experiment show that the mutual relationship between text and its connected information is relevant to detecting fake news. In fact, the performance increases more nearly 10%.

Still, there is a gap to arrive at the same level of the state-of-the-art.

Our next implementation is the word level memory network. In this module, we try to construct a memory comprising a number of cells. The structure of the network includes two types of memories which are input and output memory. These memories allow for learning different representations from the given input. In the attempt to mimic the operation of a memory, we also allow update of memory cells. Moreover, we do not treat all types of side information as a whole. Rather, they are examined separately. Our experimental results show that our memory network helps improve the performance of the state-of-the-art. The results also reveal that different types of side information do not contribute equally to the task, and the mixture of all side type information may not be useful.

Our final effort in detecting fake news is the proposal of character level memory network. In this model, we examine another way to construct memory cell. In fact, we build each memory cell from all character of words, instead of words themselves. In doing so, we take advantage of a weighting encoding to fill value for a memory cell from characters of a word. Unfortunately, our experiments exhibit that the encoding scheme from character level does not further enhance the result.

Overall, we introduce a number of approaches to fake new detection task. Those can provide a view attention-based neural network approach for further advances. The most crucial contribution is that we push the accuracy to a higher level, which surpasses the state-of-the-art.

Keywords: fake news detection, attention mechanisms, attention-based neural network, natural language processing.