https://upright.pub/index.php/ijrstp/

Original Contribution

Word Embedding with ConvNet-Bi Directional LSTM Techniques: A Review of Related Literature

Mahesh Babu Pasupuleti[1], Ruhul Amin[2]

Keywords: Deep Learning, Machine Learning, BiLSTM, Attention Mechanism

International Journal of Reciprocal Symmetry and Theoretical Physics

Vol. 5, Issue 1, 2018 [Pages 9-13]

The exponential growth of social evaluations of services has led many researchers to focus on emotion analysis. Having so much data helps analyze end-user behavior to improve QoS. Text categorization is a prominent language processing research area that organizes disorganized text into useable categories. The LSTM and CNN models are widely employed in text-based categorization applications and produce reliable results. By contrast, LSTM-based models gain long-term connections between text sequences and so are better suitable for text classification. In this case, the hybrid approach may memorize classification, slowing down the training process. This work proposes an optimal attention focused BiLSTM and ConvNet model. The suggested model is trained on two independent datasets to validate its performance. The proposed attention-based model outperformed previous deep learning algorithms. Compared to existing machine learning methods, the proposed approach is more accurate.

INTRODUCTION

Customer input is vital to enterprises because it allows them to enhance their services and facilities (Pasupuleti, 2016a). Users of online social media websites have been increasing in recent years. These websites have a lot of client interactions and are a great source of information. Researchers are also becoming more interested in social media sites like Facebook and Twitter (Adusumalli, 2016a). Organizations use public opinion analysis because it describes human behavior and how individuals are impacted by others' views (Adusumalli, 2016b).

Contextual text is used to determine the polarised meaning of sentences. Because of its superior modelling and classification skills, ML is a key factor in NLP difficulties. Techniques for sentiment analysis include SVM, Naive Bayes, and hybrid models. Recent sentiment categorization research utilizing neural networks has demonstrated encouraging results (Pasupuleti, 2016b). Neural networks are widely used in NLP for paraphrase identification, machine translation, and question answering. The increasing application of deep learning in recent years has improved image processing, natural language processing, and speech recognition. Due to its autonomous learning properties, deep learning is widely used in natural language processing models, including sentiment analysis (Adusumalli, 2017a). The most commonly used deep learning algorithms for sentiment analysis are CNN and RNN (RNN). Regrettably, RNN suffers from gradient vanishing and explosion difficulties. Due of these difficulties, RNN is difficult to train for long distance correlations. BiLSTM is an RNN model that has showed promise in text-based sentiment analysis. This model features two-way LSTM channels to enable the network grasp its contexts. BiLSTM features forward and backward layers that allow

[1]Data Analyst, Department of IT, Iminds Technology Systems INC, 1145 Bower Hill Rd, Pittsburgh, PA 15243, USA
[2]Senior Data Entry Control Operator (IT), ED-Maintenance Office, Bangladesh Bank (Head Office), Dhaka, BANGLADESH

the network to access previous and subsequent sequences.

The attention mechanism improves deep learning model performance, sentence summarization, and reading comprehension (Adusumalli, 2017b). Most deep learning text analysis involves word embedding to create feature vectors from the dataset. BiLSTM only gathers contextual information from features and cannot prioritize data. A convolutional layer captures vector features and decreases their dimension, unlike BiLSTM. The purpose of this research is to create a novel text categorization model that combines the best of CNN and BiLSTM using attention mechanism. The attention-based Conv-BiLSTM mechanism tries to overcome the constraints of BiLSTM by including a convolutional layer into the CNN model. The proposed model trains the input data using a Keras skip-gram model, then feeds it to a Convolution Layer that extracts low-level semantic information. To decide which features are highly linked to semantics and should be used for final classification, the Conv layer produces feature vectors that the BiLSTM layer uses to combine. Word embedding is performed using the Word2Vec and Keras embedding algorithms. The Self Attention Based Conv-BiLSTM model outperformed other deep learning models and regular machine learning methodologies in experiments.

LITERATURE REVIEW

In recent years, public opinion analysis on network platforms has emerged as a significant field of investigation, with sentiment analysis being among the most popular study topics. Many hand-coded characteristics, such as bag-of-words (BOW) and sentiment lexicons, are used in most conventional techniques (Jiang et al., 2011), which are then classified using traditional classification techniques (such as supported vector machine) on the basis of a large number of classification techniques (Perez-Rosas et al., 2012). The quality of the features utilized in these methodologies, on the other hand, has a significant impact on the findings of the study. Furthermore, these approaches are time-consuming and, in most cases, result in text representations that are high-dimensional and sparse in their content. Many neural network-based algorithms have shown to be beneficial in a range of sentimental analysis applications in recent years, with many more to come. It was proposed by Zhang et al. (2016) to partition the set of the exam into two different sections, with each segment having its own set of context-based characteristics. The target is then connected with each of these qualities individually. Increased interaction between the target and its surrounding environment had been extracted by Tang et al. (2016) using a gate-based multi-way control network, which had been constructed.

As a result of the enormous success of the attention-based method in natural language processing, its application in Sentiment Analysis tasks has gained favor in recent years. TANG et al. (2016) proposed and demonstrated the TD-LSTM model, which is a target-dependent long short-term memory network that is also a target-dependent LSTM network. By using the target as the input unit of the last LSTM, TD-LSTM is able to fully use the semantic knowledge of the target, allowing it to make more accurate assessments of the topic's sentiment polarity. The sentence splitting technique used in the LSTM sentence coding technique is included. The target acts as a dividing line for both forward and backward modeling operations. The attention mechanism in the given model allows it to pay greater attention to the relevant data, which raises the correctness of the model in the process. The attention mechanism is employed in order to examine the relationship between a pair of words in their respective contexts. A recursive self-adaptive neural network was built by Dong and colleagues (2016), which employs the structure of sentiment as well as information provided by context to learn the sentiment connections between words. RNN, on the other hand, did not do well when modeling lengthy sentences. LSTM models (Rosenthal et al., 2017) were proposed as a solution to this problem because they were capable of retaining sequence information while also producing satisfactory outcomes for various different sequence modeling problems.

At the textual level, the attention mechanism is a valuable instrument for recording the link between context and aspect. To gather information based on the interaction between an aspect and its surrounding context, Rosenthal et al. (2017) created a multi-grain attention network (MGAN), which employs fine-grained as well as coarse-grained attention processes to collect information. According to Adusumalli (2017a), the notion of feed-forward networks as well as multi-head attention (MHA) was employed to effectively extract the hidden representation of a context and aspect embedding in another research study. To extract the higher-level feature representation from an upgraded layer of word embedding, Adusumalli (2016a) used a CNN to extract the higher-level feature representation. In order to focus on significant term properties of aspects, an attention layer is used after a BiLSTM has captured local and global semantic information about the aspect. Furthermore, Pasupuleti (2015c) used an attention mechanism on top of Bi-GRU to do opinion mining on pre-trained weight for online medicine reviews utilizing the attention mechanism.

Lin et al. (2017) offer a methodology for extracting phrase embeddings that incorporates self-attention as part of the extraction process. In this unique approach based on sentence embedding, each of the embeddings conforms to a 2-Dimensional matrix, with each row of the word matrix focused on a different component of the sentence. Both the sentiment classification and the word embedding tasks are achieved with high accuracy and precision. As opposed to earlier methodologies, Chen et al. (2016) suggested a Long Short Term Memory-based model that takes into account preferences for both global user and product characteristics. An LSTM model with a hierarchical structure is used to generate representations of sentences and documents. A technique of attention is then applied to the user and product information, which successfully addresses information at the word and semantic levels as well. Adusumalli and Pasupuleti (2017) employed a lexicon to enhance the traditional LSTM-based technique for sentiment analysis, which was previously published. Overall, the lexicons increase the capability for representing words using word embeddings while also introducing an attention-based approach that makes use of the global knowledge of the entire text rather than a single emphasis to represent words. Dou et al. (2017) suggested a deep memory network for sentiment categorization, which they described as follows: At the same time, the suggested approach may be used to record information on both the user and the product. Inference-based components are distributed throughout a large long-term memory, which serves as a knowledge base in addition to its other functions. With regard to its architectural structure, the model has been separated into two parts. An LSTM has been used to represent each document, and then the ratings for each document are predicted using a deep memory network with several levels (hops), with each layer representing an attention model based on the document's content.

With the use of Keras word embedding, Pasupuleti (2015a) developed a Dual LSTM model for categorizing the sentiments of travelers, which was implemented. According to the authors, the suggested model performed word embedding using two separate word embedding strategies, namely word2Vec and Keras embedding. Word2Vec showed better outcomes than Keras embedding when compared to both methods. A hierarchy-based attention (HA) technique was developed by Yang et al. (2016) for capturing the hierarchical structure of documents at the sentence and text levels, in which information of different value was given particular care while producing document representations. In part because most past approaches focused just on local text information and overlooked worldwide user preferences and product quality, Chen et al. (2016) developed a sentiment classification attention method for global users that is dependent on product information. The combination of CNN with three separate attentions, according to Adusumalli (2017b), was recommended for text sentiment analysis. These attentions were LSTM attention, attention vector attention, and attentive pooling. The significance of each word was captured by Pasupuleti (2015b), who employed a self-attention with sparse technique for determining text emotion polarity and therefore polarity.

The following are the most significant contributions made by the research:

(1) Two distinct word embedding approaches are used. Word2Vec and Keras Skip Gram were used to show tweets as vectors of words, and the results were quite impressive. The two word embedding techniques are unsupervised word vectors that have been pre-trained and can capture the semantics of words. They have been learned from a large collection of words. The purpose of employing these two word vector models is to evaluate the effectiveness of the proposed model's efficiency.

(2) It is proposed to use a ConvNet combined with BiLSTM and Self Attention method for sentiment categorization of reviews. A combination of word embedding and the Self Attention-based BiLSTM module is used to gather local features, while the selected features are classified in the classification result, which is based on the classification result.

(3) The experimental results are compared with those of other popular deep learning-based techniques and classic machine learning approaches in order to confirm the viability of the recommended optimized strategy.

RECOMMENDED METHOD

The proposed attention based method has following layers:

- Word Vectorization Layer
- ConvNet Layer
- BiLSTM Layer
- Attention Layer

Word Vectorization Layer

Because the raw dataset contains outliers and missing values, pre-processing the data prior to presenting it to the model is required. Each word in the pre-processed dataset, which is comprised of a distinct and meaningful sequence of words, is recognized as a separate entity in the system. The Word-Vectorization layer is responsible for learning the embedding representation of each input token that has been pre-processed. Symbolic representation reflects the hidden relationships between words that are likely to appear in the same context as they are represented in the text. The input dataset was trained using the Keras Skip-gram model, which can be found here.

ConvNet Layer

The ConvNet layer's function is to select a subset of features from a given set of input data. ConvNet layers are used to select low-level semantic information from the source text and to reduce the number of dimensions in the text by using convolutional networks. It is proposed in this work to use a number of 1-Dimensional convnet kernels in order to conduct convolution for the input vectors.

BiLSTM Layer

The BiLSTM layer accepts as input the attributes collected by the ConvNet layer and outputs features by extracting the final hidden state from the final hidden state. The BiLSTM module has access to both the prior and subsequent context information, and the data acquired by the BiLSTM module can be thought of as two independent text-based representations of the same information. This sequential form is generated by inputting the ConvNet feature sets into the BiLSTM model, which outputs a sequential representation. BiLSTM obtains word annotations by aggregating information for words in both directions (ahead and backward), and as a result, the annotations contain contextual information as well as semantic information.

Attention Layer

Following BiLSTM, this output representation is passed to the attention layer, which identifies which features are highly related and should be used for final categorization. With a softmax function, which is focused on the characteristics of selected words, the attention mechanism, which is a completely connected layer with a softmax function, reduces the influence of less important, less important words on the sentiment of the text by decreasing the influence of less important, less important words.

CONCLUSION

It was decided to conduct the experiment on two different datasets. Both the Word2Vec and Keras word embedding approaches were used to train and evaluate the model on the two datasets, respectively. The properties of CNN and BiLSTM were merged with a self-attention mechanism in the proposed model. BiLSTM takes text context information from ConvNet and extracts text attributes from that information. The attention mechanism improved classification accuracy by retrieving the context of the sentence with greater precision from the phrase. Hyper-parameter tweaking was employed in order to improve the model. As a result, the accuracy and efficiency of classification were improved as a result of the proposed model. The suggested model beat other standard models on the data from US Airlines, according to the results.

REFERENCES

Adusumalli, H. P. (2016a). Digitization in Production: A Timely Opportunity. *Engineering International*, *4*(2), 73-78. https://doi.org/10.18034/ei.v4i2.595

Adusumalli, H. P. (2016b). How Big Data is Driving Digital Transformation?. *ABC Journal of Advanced Research*, *5*(2), 131-138. https://doi.org/10.18034/abcjar.v5i2.616

Adusumalli, H. P. (2017a). Mobile Application Development through Design-based Investigation. *International Journal of Reciprocal Symmetry and Physical Sciences*, *4*, 14–19. Retrieved from https://upright.pub/index.php/ijrsps/article/view/58

Adusumalli, H. P. (2017b). Software Application Development to Backing the Legitimacy of Digital Annals: Use of the Diplomatic Archives. *ABC Journal of Advanced Research*, *6*(2), 121-126. https://doi.org/10.18034/abcjar.v6i2.618

Adusumalli, H. P., & Pasupuleti, M. B. (2017). Applications and Practices of Big Data for Development. *Asian Business Review*, *7*(3), 111-116. https://doi.org/10.18034/abr.v7i3.597

Chen H., Sun M., Tu C., Lin Y., and Liu Z., "Neural sentiment classification with user and product attention," in Proc. Conf. Empirical Methods Natural Lang. Process., 2016, pp. 1650– 1659.

Chen, H.; Sun, M.; Tu, C.; Lin, Y.; Liu, Z. Neural sentiment classification with user and product attention. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 21 September 2016; pp. 1650–1659.

Dou, Z.Y. Capturing user and product Information for document level sentiment analysis with deep memory network. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 3 March 2017; pp. 521–526.

Jiang L., Yu M., Zhou M., Liu X., and Zhao T.: Target-dependent Twitter sentiment classification. in Proc. ACL, 2011, pp. 151–160.

Lin, Z.; Feng, M.; Santos CN, D.; Yu, M.; Xiang, B.; Zhou, B.; Bengio, Y. A structured self-attentive sentence embedding. arXiv preprint 2017, arXiv: 1703.03130.

Pasupuleti, M. B. (2015a). Data Science: The Sexiest Job in this Century. *International Journal of Reciprocal Symmetry and Physical Sciences*, *2*, 8–11. Retrieved from https://upright.pub/index.php/ijrsps/article/view/56

Pasupuleti, M. B. (2015b). Problems from the Past, Problems from the Future, and Data Science Solutions. *ABC Journal of Advanced Research*, *4*(2), 153-160. https://doi.org/10.18034/abcjar.v4i2.614

Pasupuleti, M. B. (2015c). Stimulating Statistics in the Epoch of Data-Driven Innovations and Data Science. *Asian Journal of Applied Science and Engineering*, *4*, 251–254. Retrieved from https://upright.pub/index.php/ajase/article/view/55

Pasupuleti, M. B. (2016a). Data Scientist Careers: Applied Orientation for the Beginners. *Global Disclosure of Economics and Business*, *5*(2), 125-132. https://doi.org/10.18034/gdeb.v5i2.617

Pasupuleti, M. B. (2016b). The Use of Big Data Analytics in Medical Applications. *Malaysian Journal of Medical and Biological Research*, *3*(2), 111-116. https://doi.org/10.18034/mjmbr.v3i2.615

Perez-Rosas V., Banea C., and Mihalcea R.: Learning sentiment lexicons in Spanish. LREC, vol. 12, p. 73, May 2012

Ren Y., Zhang Y., Zhang M., and Ji D.. Context-sensitive Twitter sentiment classification using neural network. in Proc. AAAI, Phoenix, AZ, USA, Feb. 2016, pp. 215_221.

Rosenthal S., Farra N., and Nakov P.. SemEval-2017 task 4: Sentiment analysis in Twitter. Proc. SemEval, Vancouver, BC, Canada, Aug. 2017,pp. 502_518.

Tang D., Qin B., and Liu T.: Aspect level sentiment classification with deep memory network. Proc. Conf. Empirical Methods Natural Lang. Process. 2016, pp. 214_224.

Tang D., Qin B., Feng X., and Liu T.: Effective LSTMs for target-dependent sentiment classification. in Proc. 26th Int. Conf. Comput.Linguistics, Tech. Papers (COLING), 2016, pp. 3298_3307.

Yang Z., Yang D., Dyer C., He X., Smola A., and Hovy E.. Hierarchical attention networks for document classification," in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum.Lang. Technol., 2016, pp. 1480–1489.

Zhang M., Zhang Y., and Vo D. T.: Gated neural networks for targeted sentiment analysis. Proc. AAAI Conf. Artif. Intell., 2016, pp. 3087–3093

--0--