

Attention-based ResNet for Chinese Text Sentiment Classification

Hu Han*, Xuxu Bai and Jin Liu

School of Electronic & Information Engineering, Lanzhou Jiaotong University, Lanzhou, Gansu, 730070, China

*Corresponding author

Abstract—Identifying sentiment polarity of a document is a building block of sentiment classification and natural language processing tasks, it aims to automate the prediction of sentiment orientation in a document. In general, recently fast-growing Deep Neural Networks(DNN) method has been extensively used as a sentiment learning approach. But the dominant approach for sentiment classification tasks are recurrent neural networks, in particular LSTM, and convolutional neural networks. However, these architectures are rather shallow in comparison to the Residual Neural Networks(ResNet) which have pushed in computer vision. We present a model using ResNet for high-level document representation, and attention mechanism to capture the crucial components for document. The experimental results show that using up to 2 ResNet block and attention mechanism achieve state-of-the-art performance on three public sentiment classification datasets.

Keywords—sentiment classification; attention mechanism; ResNet

I. INTRODUCTION

Opinions are key influencers to almost all of our activities, since whenever we need to make a decision, e.g., when we purchase products, we want to know others' opinions. From the industry point of view, grasping know the feelings of customers about products may support strategic market decisions. Moreover, individuals may want to know others' opinions about political candidates before making a voting decision in a political election. Nowadays, many people present their opinions on the Internet, which benefits the researchers devising automatic procedures to identify these textual reviews/opinions to be positive or negative [1,2]. Mining opinions from these texts is known as sentiment classification, which is a special task of text classification whose objective is to automatically classify a document according to the sentiment polarities of opinions it contains, e.g., positive or negative, like or dislike.

Most existing methods for document sentiment classification consists of two parts: (a) designing text features such as bag-of-words [2] and (b) training some machine learning models using text features and annotated sentiment polarities. Since the performance of the classifiers relies heavily on the effectiveness of feature engineering, these methods usually focus on extracting features from texts and sentiment lexicons [3].

Motivated by the success of deep learning's application in computer vision, speech recognition [4] and natural language

processing [5], neural networks have been employed in sentiment classification and exhibited powerful ability of learning text features without feature engineering. In the canonical neural sentiment classification models [5,6], a sentence or a document is simply vectorized and then taken as input of the models which generate the semantic representations of the text. By embedding additional information e.g. user's preference and product's characteristics into input, several models [6,7] have proved the improvement of the accuracy of sentiment classification.

Convolutional neural networks, in Residual Neural Networks(ResNets), are very successful in computer vision. This fundamental idea of ResNets is to answering a question notorious problem of vanishing/exploding gradients. The best networks are using more than 150 layers as in [8]. Nowadays, [9] using ResNet to generate the semantic representations of document, this is the first time very deep Convolutional nets have been applied to text processing. Attention has become an effective mechanism to obtain superior results, as demonstrated in image recognition [10], machine translation [11] and text summarization [12]. Even more, attention can improve the ability to question answer.

In this work we propose a new neural sentiment classifier which combines Residual Neural Networks and Attention mechanism. We compare our model with neural sentiment classifiers PCNN and PCNN+Dropout [13]. Experimental results show that, our method outperforms most of the baseline methods and the state-of-the art approaches. The main contribution of this work are as follows:

- We present a new neural network model by combining ResNet and Attention for document-level sentiment classification.
- We validate the influences of the number of ResNet blocks for sentiment classification.
- We report empirical results on three datasets and show that our proposed model outperforms state-of -the-art models for sentiment classification.

II. RELATED WORK

With the huge success of deep learning in computer vision, speech recognition and natural language processing, in NLP field, neural network due to its ability of text representation has introduced into sentiment classification. Existing neural network methods can be divided into two groups: word

embedding and semantic composition. For learning word embeddings, [14] use local and global contexts, [7,15-18] further incorporate sentiment of texts. [5] adopt convolution neural network (CNN) to learn sentence representations and achieve outstanding performance in sentiment classification.

[13] adopt piecewise CNN(PCNN) to combine the structural features, in which the feature vector of a sentence is divided into three segments and does the max-pooling for each of them achieve higher than all of the baseline performance of Chinese hotel datasets and Stanford Sentiment TreeBank.

Recurrent neural network (RNN) also benefits sentiment classification because it is capable of capturing the sequential information. There are also some hierarchical models proposed to deal with document-level sentiment classification [17,19] which generate different levels (e.g., phrase, sentence or document) of semantic representations within a document. Moreover, attention mechanism is also introduced into sentiment classification, which aims to select important words from a sentence or sentences from a document [20,21] proposed user product attention based(UPA-NSC) method improve performance of sentiment classification.

[22] use convolutional layer capturing the character local connectivity of scene text recognition tasks and use RNN to learn sequential information. [9] propose using very deep convolutional network(VDCNN) for text classification, which using 2 ResNet block and k-max pooling to detect the k most important features in a sentence. Inspired by this, we feed word embedding vector to ResNet block which learning word context feature and using attention mechanism to compute words contribute from input sentence.

III. METHODS

In this section, we will introduce our Attention-based Residual Network(ARN) model in detail. First, we give the formalizations of document-level sentiment classification. Afterwards, we discuss how to obtain document semantic representation via the Attention-based ResNet, which incorporates the forward and backward information. The enhanced sentence representation is used as input for document representation. An overall illustration of Attention-based ResNet model is shown in Fig.I, which consist of four modules: embedding module, ResNet Block, attention module and dropout module.

A. Formalization

We represent the review as a document d with n words $\{w_1, w_2, \dots, w_n\}$. Document-level sentiment classification aims to inferring the numeric rating(1-5 or 1-10 stars) or sentiment polarity(positive, negative) of these reviews according to their text information.

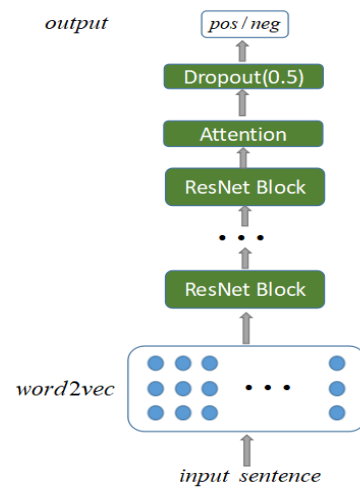


FIGURE I. THE ARCHITECTURE OF OUR ATTENTION-BASED RESNET SENTIMENT CLASSIFICATION MODEL

B. Embedding Module

We use **word2vec** embed each word into a low dimension semantic space. Let W be an word embedding generated by an unsupervised method such Glove or SkipGram, where d is the dimension of word vectors and $|V|$ is the vocabulary size. Word embedding may be fine tuned in the training of our framework because of corpus limited for training word embedding.

C. ResNet Block

The role of the ResNet Block is to read the embedding vectors for input sentence and construct the high-level representation. Formally, in this paper we consider a building block defined as:

$$y = F(x, w_i) \cdot x \quad (1)$$

Here x , and y are the input and output vectors of the ResNet block. An detail illustration of ResNet Block is given Fig.II. ResNet block is a sequence of three convolutional layers, each one followed by a BatchNormalization layer and an ReLU activation. The kernel size and filters of all the convolution is 5 and 200, respectively.

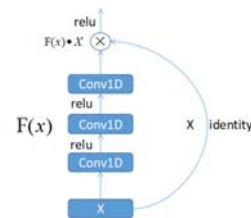


FIGURE II. THE ARCHITECTURE OF OUR RESNET BLOCK DETAIL.

D. Attention

While the ResNet Block has richer capacity, however, not all words contribute equally to the sentiment classification. Attention mechanism has huge success to capture the crucial components over high-level semantic.

Formally, the enhanced sentence representation is a weighted sum of hidden states as:

$$s_i = \sum_{j=1}^{l_i} \alpha_j^i h_j^i \quad (2)$$

where α_j^i measures the importance of the j-th word or sentence for current sentence. Thus, the attention weight α_j^i for each hidden state can be defined as :

$$\alpha_j^i = \frac{\exp(e(h_j^i))}{\sum_{k=1}^{l_i} \exp(e(h_k^i))} \quad (3)$$

where e is a score function which scores the importance of words for composing document representation. The score function e is defined as :

$$e(h_j^i) = v^t \tanh(W_H h_j^i + b) \quad (4)$$

where W_H, v^t are weight matrices for hidden state.

E. Sentiment Classification

We apply our model to document level sentiment classification under supervised learning framework. Document representation d is extracted from the words in the documents, it is a high level representation of the document. The sentiment classifier is built from documents with gold standard sentiment labels.

We use sigmoid to build the classifier because its outputs can be interpreted as conditional probabilities. Softmax is calculated as given in Equation 5, where C is the category number(e.g. positive or negative):

$$p_c = \frac{\exp(d_c)}{\sum_{k=1}^C \exp(d_k)} \quad (5)$$

Where p_c is the predicted probability of sentiment class c . In our model, cross-entropy error between gold sentiment distribution and our model's sentiment distribution is defined as loss function for optimization when training:

$$L = -\sum_{d \in D} \sum_{c=1}^C p_d^c(d) \cdot \log(p_c(d)) \quad (6)$$

IV. EXPERIMENTS

We evaluate the effectiveness of our model on Chinese hotel sentiment classification datasets, which are built by Songbo Tan, and Stanford Sentiment TreeBank(SST). SST use real number for [0,1] to represent document sentiment score. We split real number to binary and five classification problem which binary classification problem(SST-2) means [0, 0.5] is negative and [0.5, 1] means positive, five classification problem(SST-5) is [0, 0.2], [0.2, 0.4], [0.4, 0.6], [0.6, 0.8], [0.8, 1]. We split each corpus into training, validation and testing sets in the proportion of 8:1:1. We use metrics Accuracy which measures the overall sentiment classification performance.

$$accuracy = \frac{T}{N} \quad (7)$$

where T is the numbers of predicted sentiment rating that are identical with gold sentiment ratings, N is the numbers of documents. We use pre-trained the 100-dimensional word embeddings on each corpus separately with SkipGram [14]. We set the widths of convolutional filters as 5, number of filters set to 200. We tune the hyper parameters on the validation sets and use ADAM to update parameters when training. We select the best configuration based on performance on the validation set, and evaluate the configuration on the test set.

A. Baseline Methods

We compare our neural network model with several baseline methods for Chinese hotel and Stanford Sentiment TreeBank(SST) document sentiment Classification:

a) *PCNN [13]: implement the piecewise convolution neural networks(PCNN) for document sentiment classification.*

b) *PCNN+Dropout [13]: implement the PCNN and use the Dropout to enhance the performance.*

For all baseline methods above, we report the results in [13] since we use the same data sets.

B. Experimental Results

TABLE I. EXPERIMENTAL RESULTS

Model	Data Sets		
	Hotel	SST-2	SST-5
CNN	0.885	0.819	0.455
PCNN	0.897	0.833	0.456
PCNN+Dropout	0.910	0.854	0.459
our model	0.930	0.872	0.470

Experimental results are given in Table 1. We obtain state-of-the-art results for all three data sets. Overall, compared to previous PCNN+Dropout model, our best architecture with 2 ResNet block and attention mechanism has growth almost 1.1% and 1.6% in the Stanford Sentiment TreeBank dataset. To be specific, Chinese Hotel dataset has a test accuracy of 0.93 compared to 0.91. This represents a gain of 2% absolute

accuracy. The significant improvements which we obtain on Chinese Hotel dataset compared to PCNN+Dropout models do not include any data oversampling technique or re-weight technique, even though this is imbalance data sets, which means the number of negative is the double of positive.

C. Model Analysis: Number of ResNet Block

To investigate the performance of our model with various number of ResNet block. Table 2 shows the performance of various number of ResNet block and Figure III shows the convergence process of different number of ResNet block. From the table, we can observe that various number of ResNet block has different ability to capture document semantic. In this work, we have explored five different depths for our networks: 1, 2, 3, 4 and 5, which we define as being the number of ResNet block.

As we increase the number from 1 to 2 of block, the predicting accuracy increase can be improved from 91.8% to 93.0%. As we continue to increase number of block, from 2 blocks to 5 blocks, the convergence time and algorithms performance both decreased.

TABLE II. MODEL ANALYSIS: NUMBER OF RESNET BLOCK

Number of ResNet block	Hotel Data Set
1	0.918
2	0.930
3	0.918
4	0.911
5	0.912

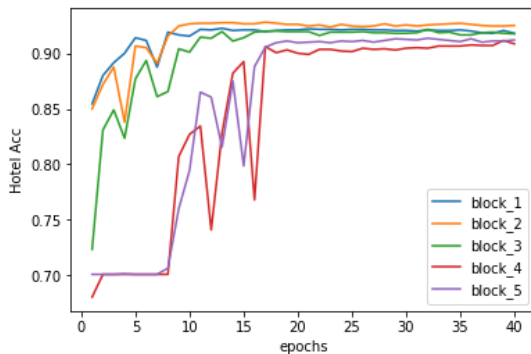


FIGURE III. MODEL ANALYSIS WITH DIFFERENT NUMBER OF RESNET BLOCK.

V. CONCLUSION

In this paper, we propose an Attention-based ResNet network which incorporates ResNet block and attention with document level. Our model can take account of document information with ResNet block. In experiments, we evaluate our model on three sentiment classification task. The experimental results show that our model achieves significant and consistent improvements compared to other state-of-the-

art methods. Our model are much deeper than previously published neural network in sentiment classification field. This result make us believe future research should invest into making text processing models deeper to capture semantic information.

We will explore more in future as follows:

- (1) In this paper, we only consideration CNN to capture semantic representation, we will explore why our model is so powerful and why very deep neural network can capture so much high-level semantic.
- (2) We will explore the effectiveness of our model on aspect level sentiment classification.

VI. ACKNOWLEDGMENT

This work was supported by the National Social Science Foundation of China(No.17BXW071) , National Natural Science Foundation of China(No.71461017) and the Technology Program of Lanzhou Science and Technology Bureau(No.2014-1-168).

REFERENCES

- [1] Bo Pang and Lillian Lee, Opinion Mining and Sentiment Analysis, 2008.
- [2] Bo Pang, Lillian Lee and Shivakumar Vaithyanathan, Thumbs up? Sentiment Classification using Machine Learning Techniques, EMNLP, 2002, pp.79–86.
- [3] I. S. Jacobs and C. P. Bean, Lexicon-Based Methods for Lexicon-Based Methods for Sentiment Analysis in Magnetism, ACL, 2011, pp. 271–350.
- [4] Amodei, Dario and Anubhai, Rishita and Battenberg, Eric and Case, Carl and Casper, Jared and Catanzaro, Bryan and Chen, Jingdong and Chrzanowski, Mike and Coates, Adam and Diamos, Greg, Deep Speech 2: End-to-End Speech Recognition in English and Mandarin. 2015.
- [5] Yoon Kim, Convolutional neural networks for sentence classification. EMNLP. 2014.
- [6] Richard Socher, Jeffrey Pennington, Eric H Huang, Andrew Y Ng, and Christopher D Manning. Semi-supervised recursive autoencoders for predicting sentiment distributions. EMNLP, 2011, pp. 151-161.
- [7] Duyu Tang, Bing Qin and Ting Liu. Learning Semantic Representations of Users and Products for Document Level Sentiment Classification. ACL, 2015.
- [8] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, Deep residual learning for image recognition. In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.
- [9] Alexis Conneau, Holger Schwenk, Yann Le Cun, Loic Barrault, Very Deep Convolutional Networks for Text Classification, arXiv preprint arXiv:1606.01781v2.
- [10] Volodymyr Mnih, Nicolas Heess and Alex Graves, Recurrent models of visual attention. Neural Information Processing Systems, 2014, pp. 2204-2212.
- [11] Dzmitry Bahdanau, Kyunghyun Cho and Yoshua Bengio, Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
- [12] Jiwei Tan, Xiaojun Wan and Jianguo Xiao, Abstractive Document Summarization with a Graph-Based Attentional Neural model, ACL, 2017, pp. 1171-1181.
- [13] Du Chang-shun, Huang Lei, Sentiment analysis with piece convolution neural network, 2017.
- [14] Tomas Mikolov, Kai Chen, Greg Corrado and Jeffrey Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781.

- [15] Andrew L Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts., Learning word vectors for sentiment analysis, ACL, 2015, pp. 142-150
- [16] Igor Labutov and Hod Lipson, Re-embedding words. ACL, 2013.
- [17] Duyu Tang, Bing Qin and Ting Liu, Document modeling with gated recurrent neural network for sentiment classification, EMNLP, 2015, pp. 1422-1432.
- [18] Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin, Learning sentiment-specific word embedding for twitter sentiment classification, ACL, 2014, pp. 1555-1565.
- [19] Parminder Bhatia, Yangfeng Ji, and Jacob Eisenstein, Better document-level sentiment analysis from rst discourse parsing, EMNLP, 2015.
- [20] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola and Eduard Hovy. Hierarchical attention networks for document classification, NAACL, 2016.
- [21] Huimin Chen, Maosong Sun, Cunchao Tu, Yankai Lin and Zhiyuan Liu., Neural Sentiment Classification with User and Product Attention, EMNLP, 2016.
- [22] Baoguang Shi, Xiang Bai and Cong Yao, An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.