Aspect-based Financial Sentiment Analysis with Deep Neural Networks

Shijia E
Tongji University
Shanghai, China
436_eshijia@tongji.edu.cn

Mohan Zhang
Tongji University
Shanghai, China
zhangmohan@tongji.edu.cn

Li Yang
Tongji University
Shanghai, China
1452238@tongji.edu.cn

Yang Xiang
Tongji University
Shanghai, China
shxiangyang@tongji.edu.cn

ABSTRACT

Aspect-based financial sentiment analysis, which aims to classify the text instance into a pre-defined aspect class and predict the sentiment score for the mentioned target. In this paper, we propose a neural network model, Attention-based LSTM model with the Aspect information (ALA), to solve the financial opinion mining problem introduced by the WWW 2018 shared task. The proposed neural network model can adapt to the financial dataset so that the neural network can effectively understand the semantic information of the short text. We evaluate our model with the 10-fold cross-validation, and we compare our model with a variety of related deep neural network models.

CCS CONCEPTS

• **Computing methodologies** → **Natural language processing**; **Neural networks**; *Learning latent representations*;

KEYWORDS

sentiment analysis; long short-term memory network; representation learning

ACM Reference Format:

Shijia E, Li Yang, Mohan Zhang, and Yang Xiang. 2018. Aspect-based Financial Sentiment Analysis with Deep Neural Networks. In *The 2018 Web Conference Companion*, April 23–27, 2018, Lyon, France. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3184558.3191825>

1 INTRODUCTION

Aspect-based financial sentiment analysis is a special task of short text classification problem with financial data. The core of the task is to model the sentence effectively and obtain the opinion of the text instance. Sentence modeling aims at representing sentences as meaningful features for tasks such as sentiment classification. Traditional sentence modeling uses the bag of words model which often suffers from the curse of dimensionality.

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23–27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3191825>

Recently, deep learning approaches have achieved a lot of success in many kinds of research including the sentence representation [9], text parsing [11], machine translation [2], and sentiment classification [5] [7]. Among neural network models, long short-term memory neural network (LSTM) is an effective neural network model to represent the sentence, and it can handle sequences of any length and capture long-term dependencies.

In this paper, we use a new deep neural architecture called Attention-based LSTM model with the Aspect information (ALA) to model sentences. Our proposed solution is an end-to-end, unified architecture, and it does not need much feature engineering and human intervention.

To apply neural architectures for modeling sentence, the common method of current research is using word2vec method [8] to convert the sentence into a word embedding sequence. In order to improve the generalization ability of this specific task, all the stock targets are processed into the same target work. Therefore, the proposed model can also handle the classified targets that have not been seen before.

The rest of this paper is structured as follows. Section 2 contains related work. In Section 3, we describe our proposed solution. Experimental results and discussions are presented in Section 4, and finally, we give some concluding remarks in Section 5.

2 RELATED WORK

For the general short-text classification task, the CNN model is still widely used. The capability of capturing local correlations along with extracting higher-level correlations through pooling empowers CNN to model sentences naturally from consecutive context windows. Collobert et al. [1] apply convolutional filters to successive windows for a given sequence to extract global features by the max-pooling operation. Kim et al. [5] propose a CNN architecture with multiple filters and varying window sizes to capture word relations of varying sizes. Kalchbrenner et al. [4] propose a dynamic k-max pooling mechanism that is capable of explicitly capturing short and long-range relations. As for aspect level sentiment classification, Tang et al. [12] introduce a deep memory network which is able to capture the importance of each context word. However, it requires the aspects to appear in the training data many times, otherwise, the model cannot have a strong generalization ability. For the financial dataset provided by the task, many stock names

appear only a few times, so we cannot apply the deep memory model directly.

3 MODEL DESCRIPTION

In this section, we describe the proposed ALA model to solve the aspect-based financial sentiment analysis problem and how we preprocess the training data.

3.1 Data preprocessing

For our proposed neural network architectures, word embedding is one of the necessary inputs. We use the provided word-level train set corpus to generate word vectors with word2vec [10]. For the word which does not appear in the provided train set, we simply use zero vector to represent them. The pre-trained word vectors will be the initial weights of the embedding layer, and the word embedding can be updated during the training process. Another important thing is that, for all the different aspects, we change them into the same symbol “T” (it means *Target*). Therefore, our model can learn how the context words affect the sentiment of the aspect in the text instance.

3.2 Long Short-Term Memory Network Layer

The Long Short-Term Memory (LSTM) architecture contains special units called memory blocks in the recurrent hidden layer [3]. At each time step, the output is controlled by a set of gates, i.e., the previous hidden state h_{t-1} , the forget gate f_t , the input gate i_t , and the output gate o_t . With the input at the current time step x_t , those gates collectively decide how to update the current memory cell c_t and the current hidden state h_t . We use d to denote the memory dimension in the LSTM, and all vectors in this architecture share the same dimension. The LSTM transition functions are defined as follows:

$$\begin{aligned} i_t &= \sigma(W_i * [h_{t-1}, x_t] + b_i) \\ f_t &= \sigma(W_f * [h_{t-1}, x_t] + b_f) \\ q_t &= \tanh(W_q * [h_{t-1}, x_t] + b_q) \\ o_t &= \sigma(W_o * [h_{t-1}, x_t] + b_o) \\ c_t &= f_t \odot c_{t-1} + i_t \odot q_t \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \quad (1)$$

where σ is the logistic sigmoid function that has an output in $[0, 1]$, $\tanh$ denotes the hyperbolic tangent function that has an output in $[-1, 1]$, and $\odot$ denotes the element-wise multiplication. We can view f_t as the function to control what extent the information from the old memory cell is going to be thrown away, i_t to control how much new information is going to be stored in the current memory cell, and o_t to control what to output based on the memory cell c_t .

LSTM is explicitly designed for time-series data with learning long-term dependencies, and therefore, we choose LSTM to learn such dependencies in the sequence.

3.3 Attention-based LSTM with the aspect information

We believe that the sentiment score of a text instance is not only determined by the content but is also highly dependent on the

given aspect. Therefore, it is worthwhile to explore the correlation between an aspect and the content of a text instance. To address this issue, we introduce an attention mechanism that captures the importance of context words in response to a given aspect.

As shown in Figure 1, the implementation of the attention-based LSTM is that keeping the intermediate outputs of the LSTM from each step of the input sequence and training the model to learn to pay selective attention to these input steps and relate them to the final distributed representation. On the basis of the typical LSTM layer, the vector representation of the aspect (Emb_{aspect}), which is randomly initialized by sampling from a uniform distribution and optimized during the training phase, will be used to determine the attention weights α . Therefore, r is the weighted representation of the input sequence. It is defined as follows:

$$r = H\alpha^T \quad (2)$$

where H is the hidden vector. The final feature representation of the input sequence is computed as follows:

$$h_{final} = \tanh(W_r R + W_h H) \quad (3)$$

where W_r and W_h are the parameters to be learned during the training phase. By leveraging attention mechanism, the ALA model incorporates the aspect information effectively and can learn the potential connection between context words and given aspect, which enables the model to focus different parts of a text instance when different aspects are concerned.

In our experiments, the different aspects presented in the training set can be used to assist the prediction of sentiment score. As for the FiQA task, the sentiment score prediction and aspect classification are two independent task. Therefore, when we submit our test result, we assume that all the samples in the test set have the same aspect. Even with that kind of pre-processing, the performance with the aspect information is still better than models without using any aspect information.

3.4 Dense Layer

Suppose the output vector of ALA is $[x_{final\ 0}, x_{final\ 1}, \dots, x_{final\ n}]$, the outputs of neurons in the dense layer are defined as follows:

$$O_i = f(\sum w_{final\ i} * x_{final\ i} + b) \quad (4)$$

where f is the $\tanh$ non-linear activation function. The final output of the dense layer is $[O_0, O_1, \dots, O_m]$ that will be the vector to be classified (for aspect classification) or predicated (for sentiment score). At last, we get the final label (or score) of the sentence using *softmax* with the standard categorical cross-entropy loss function (or using *logistic regression* with the MSE loss function).

4 EXPERIMENTS

4.1 Dataset

The dataset provided by this task contains 504 news headlines and 675 microblog messages, i.e. 1179 training samples. In this dataset, the sentiment scores of all text instances are ranged from -1 to 1, and there are 27 classification labels. The limited size of the training data also poses a challenge to this task.

Figure 1: The overall framework of ALA model.

Table 1: The experimental results of the 10-fold cross-validation for the FiQA task-1.

Model	Aspect Classification	Sentiment Score Prediction	
	Accuracy	MSE	R^2
CNN	0.3512 (+/- 0.0032)	0.1164 (+/- 0.0108)	0.2580 (+/- 0.0805)
LSTM	0.3547 (+/- 0.0028)	0.1160 (+/- 0.0146)	0.2689 (+/- 0.0578)
DeepMem	0.3621 (+/- 0.0018)	0.1059 (+/- 0.0167)	0.3289 (+/- 0.0901)
ALA	0.378 (+/- 0.0016)	0.1045 (+/- 0.0079)	0.3363 (+/- 0.0471)

Table 2: The final results with the ALA model.

Dataset	Aspect Classification	Sentiment Score Prediction	
	Accuracy	MSE	R^2
News headline	0.2688	0.1345	0.4579
Microblog message	0.8485	0.1041	0.0924

4.2 Experimental Setup

4.2.1 Parameter settings. We evaluate our model with 10-fold cross-validation. During the training process, we utilized the pre-trained 300-dimensional word embedding as the initial embedding matrix. The batch size was 64. The number of hidden units for each LSTM layer is 300. We trained our model with Adam [6] as the final optimization strategy. To avoid model overfitting, we add L2 regularization to the parameters of our model.

4.2.2 Baseline. In our experiments, we also compare our model with several baseline models, such as the *Deep Memory* model proposed by Tang et al [12], vanilla CNN, and LSTM models.

4.3 Results and Discussions

The performance of the aspect classification and sentiment score prediction in our experiments is shown in Table 1. The final test results with the ALA model are shown in Table 2.

As the results shown in Table 1, we can find that: (1) For aspect classification, all the deep neural network models have not achieved high accuracy. (2) The ALA model can capture the key part of the sentence in response to the target aspect, and it can be well adapted to the unseen stock names. Therefore, among the several neural network architectures, ALA model can always get the best performance.

5 CONCLUSION

In this paper, we use the ALA neural network model for the aspect-based financial sentiment analysis. The ALA model can capture the most important part of a sentence for a target aspect. Experiments show that the ALA model obtains superior performance over several baseline models. For future work, a possible direction would be incorporating a few rule templates in the financial domain with deep neural networks to improve the performance of current methods.

ACKNOWLEDGMENTS

The work is supported by the National Basic Research Program of China under Grant No.: 2014CB340404, National Natural Science Foundation of China under Grant No.: 71571136 and Science and Technology Commission of Shanghai Municipality under Grant No.: 16JC1403000, 14511108002.

REFERENCES

- [1] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research* 12, Aug (2011), 2493–2537.
- [2] Jacob Devlin, Rabih Zbib, Zhongqiang Huang, Thomas Lamar, Richard M Schwartz, and John Makhoul. 2014. Fast and Robust Neural Network Joint Models for Statistical Machine Translation.. In *ACL (1)*. 1370–1380.
- [3] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [4] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188* (2014).
- [5] Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882* (2014).
- [6] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [7] Ji Young Lee and Franck Dernoncourt. 2016. Sequential short-text classification with recurrent and convolutional neural networks. *arXiv preprint arXiv:1603.03827* (2016).
- [8] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
- [9] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
- [10] Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, 45–50. <http://is.muni.cz/publication/884893/en>.
- [11] Richard Socher, John Bauer, Christopher D Manning, and Andrew Y Ng. 2013. Parsing with Compositional Vector Grammars.. In *ACL (1)*. 455–465.
- [12] Duyu Tang, Bing Qin, and Ting Liu. 2016. Aspect Level Sentiment Classification with Deep Memory Network. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 214–224.