

Aspect-level text sentiment analysis method combining Bi-GRU and AIBERT

Zhigang Xu, Bo Jin, Honglei Zhu*

School of Computer and Communication, Lanzhou University of Technology, Lanzhou, Gansu, China

ABSTRACT

As more and more industries and enterprises begin to develop the online/offline business model, the demand for efficient sentiment analysis of online service user comments is increasing day by day. In view of the problem that the aspect emotion analysis model is difficult to combine the semantic information of the text effectively, and the accuracy of aspect emotion classification of the texts need to be improved, this paper uses GRU model to improve the ATAE-LSTM model and improves AIBERT model through the attention mechanism based on aspect level. The ATAE-AIBERT-BiGRU model is proposed. In this paper, comparative experiments are carried out on the SemEval 2014 data set, and the results show that the accuracy of the proposed model is significantly improved compared with the comparison model in the aspect level emotion analysis task, and it performs well in reducing the computational load of the model.

Keywords: Sentiment analysis, aspect-based-attention, Bi-GRU, ATAE-LSTM, AIBERT

1. INTRODUCTION

Sentiment Analysis is an important branch of Natural language Processing (NLP), alternatively known as Opinion Mining or Subjectivity Analysis. The main purpose of Sentiment Analysis is to obtain the sentiment bias of text by utilizing Tokenization, Induction, Semantic Extraction, and Reasoning etc.¹. With the development of mobile Internet technology, the demand of the analysis of the sentiment tendency of sort text is growing rapidly, and the automatic sentiment analysis of short texts using NLP has become a major research trend. The core of text sentiment analysis is to classify a given text according to the sentiment tendency contained in the text. Mainstream sentiment analysis includes the following three methods which will be listed below.

Sentiment dictionary based: Sentiment dictionary-based sentiment analysis appeared relatively early. WordNet² offers a more comprehensive emotional dictionary. The accuracy of sentiment analysis is highly relied on the perfection of the sentiment dictionary, thus current research work mainly focused on the expansion and improvement of the sentiment dictionary. The WordNet sentiment dictionary is expanded in quantity and enriched in emotion by adding polarities other than positive, negative, and neutral.

Machine learning based analysis mainly extracts the statistical rules of corpus data through statistical learning methods, and later accomplish the sentiment analysis of the text. It is relatively easy to implement and timely efficient due to less required training time which lead to many applications. Popular machine learning methods include K-nearest neighbour (KNN)³, Support Vector Machine (SVM)⁴⁻⁵, Naïve Bayesian⁶, Conditional Random Field (CRF)⁷ etc. Traditional machine learning methods have the capacity of capturing corpus features in an automatic behave. However, it cannot capture the features in depth meaning less complex semantic information can be extracted based on the simple modelling structures. Modelling result highly depends on the selection of features and the quality of the samples, making the model less suitable for more general scenarios.

Deep Learning based text sentiment analysis mainly uses deep neural networks to perform semantic analysis on text information, and then analyses the sentiment tendency of the corpus through semantics. Socher et al.⁸ improved the Recurrent Neural Network (RNN) and applied it to the sentiment analysis of the text. Zhu et al.⁹ used the Word2Vec technology proposed by Mikolov¹⁰ to convert the input text into word vectors and train them in the way of Sequence to Sequence (Seq2Seq)¹¹. Zhou et al.¹² spliced CNN and LSTM into one channel in a chained manner, which improved the performance of text classification. Wang et al.¹³ proposed the ATAE-

*z_honglei@126.com

LSTM model specifically for aspect-level sentiment classification tasks, and innovatively added an aspect-level attention mechanism, which greatly improved the prediction accuracy of aspect-level sentiment classification tasks. Google introduced the Attention mechanism, proposed the Transformer model¹⁴, and later proposed the BERT¹⁵ pre-training (Pre-Train) Model based on the Transformer, to solve the problem that the models mentioned above cannot distinguish the weights of key features. Zhang et al. Introduced a new language model ERNIE¹⁶, which is used to enhance the expressive ability of the language model through knowledge shielding, based on BERT. Google later proposed a simplified version of the BERT model, AI-BERT¹⁷, which greatly reduces the parameters and training time of the BERT model. AI-BERT model does not affect the classification effect. Better training effects can be achieved via the technologies such as Factorized Embedding Parameterization, Cross-layer Parameter Sharing, with fewer parameters and training time. Wu et al.¹⁸ tried to improve task performance by adding contextual information to the self-attention model. Mao et al.¹⁹ enhanced the task performance by improving the multi-task learning framework.

This article proposes the ATAE-AIBERT-BiGRU applied to aspect-level sentiment analysis. The ATAE-AIBERT-BiGRU network mainly improves the efficiency of ATAE-LSTM network by enhancing the training efficiency and making full use of the semantic information of the context in the text. The experimental results show that the ATAE-AIBERT-BiGRU proposed in this paper can reduce the amount of model calculations and improve the accuracy of model classification by reducing the number of parameters, reducing the depth of the model, and adding targeted aspect-level information.

2. RELATED TECHNOLOGY

2.1 Attention mechanism

The attention mechanism was first introduced in image classification tasks by Mnih et al.²⁰, and later introduced into natural language processing by Bahdanau et al.²¹ through the field of machine translation. At present, this technology has been widely used in NLP tasks. The main purpose of the attention mechanism is to 1) allow the model to extract more important information from the current learning task, 2) add weights onto the model parameters to categorize the level of importance on the tasks which have a greater impact on the task results. Parameters are used to improve the training effect of the model. In NLP tasks, the main function of the attention mechanism is to calculate the relevance weight (Value, V) of an input word (Query, Q) and some other keywords (Key, K). The specific calculation method is as follows:

$$\text{Attention} = \text{Softmax}(QK^T)V \quad (1)$$

2.2 ATAE-LSTM model

The ATAE-LSTM model²¹ is a fine-grained text sentiment polarity analysis network mode. LSTM has only one sentiment type output result for the text sentiment classification since the LSTM model cannot extract the aspect-level feature information of the text. If the aspect-level sentiment information appears in the sentence and the sentiment classification is different, the LSTM network will provide poor results, or even wrong classification. Therefore, aspect-level vectors and attention mechanisms are added onto the ATAE-LSTM model to improve the performance of the network in aspect-level sentiment classification tasks.

2.3 AIBERT model

Transformer structure. The transformer structure abandons CNN and RNN by involving Self-Attention, Feed Forward Neural Network, and Residual Structure.

Position embedding. Position embedding is used to solve scenario where data needs to be added into the model for serialization model (such as RNN, etc.) resulting in the in-parallelizable modelling. Such a calculation method cannot only add the position information of the words to the text, but also ensure the relative position of different words in the text does not change, the difference between the position codes will also remain unchanged, which is conducive to the model learning Relative position information between words. The residual structure allows the gradient as the depth of the model increases.

Optimization of BERT by AIBERT model. In order to solve the problem of difficulty in BERT training, Google proposed the AIBERT model, which is essentially an optimization of the BERT model. The main optimization methods contain the following two methods.

Factorized embedding parameterization of the presentation layer. In BERT, the size of the parameter is determined by the size of the dictionary and the dimension of the hidden layer. The specific formula is: $O(V \times H)$. However, the input of BERT and AIBERT network is a single word and does not need the context information originally contained in the hidden layer, so the dimension should be set much larger than the dimension of the word embedding layer. This specific method is used to change the original independent One-Hot Encoding is mapped to a vector space of distributed representation where it relates to the hidden layer so that the parameter scale is reduced to: $O(V \times E + E \times H)$.

Cross-layer parameter sharing method shares parameters for both the fully connected layer and the attention layer. Transformers of the same magnitude sacrifice very little experimental accuracy. To ensure the improvement of training speed.

3. ASPECT-LEVEL SENTIMENT CLASSIFICATION BASED ON ATAE-ALBERT-BIGRU

ATAE-LSTM does not use the attention layer to learn the relative importance of context words. In the process of model training, additional calculations of the relationship between aspect levels and words are added. Because the aspect-level representation vector and the hidden layer are also in a splicing relationship, this leads to an increase in the dimensionality and an increase in the amount of calculation. The AIBERT network focuses on the overall context of sentences and cannot be optimized for aspect-level emotional features. Therefore, this article proposes a new network model: ATAE-AIBERT-BiGRU.

ATAE-AIBERT-BiGRU network includes five parts: input layer, AIBERT layer, presentation layer, ATAE-BiGRU layer and output layer. The network structure is shown in Figure 1.

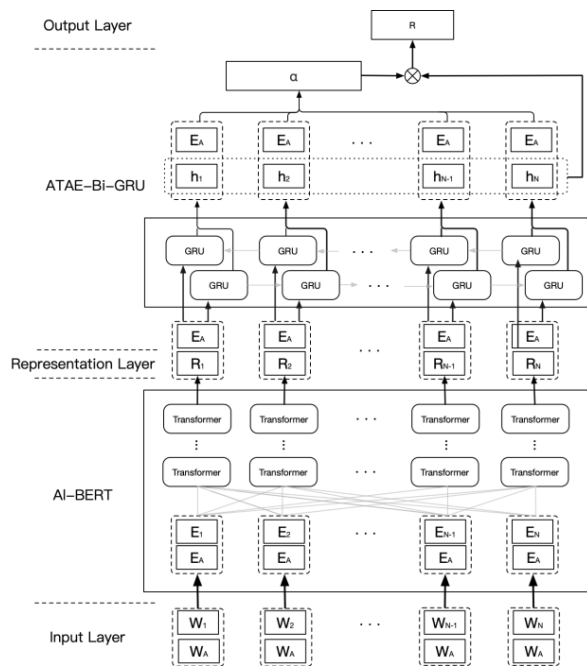


Figure 1. Structure of ATAE-ALBERT-BiGRU, this model includes 5 parts, input layer, ALBERT layer, presentation layer, ATAE-BiGRU layer and output layer.

3.1 Input section

The attention mechanism of the AIBERT model does not contain aspect-level information, and the influence of aspect-level information on emotional polarity cannot be further considered. The ATAE-AIBERT-BiGRU model optimizes AIBERT in the input part and adds aspect-based vector information to the attention mechanism of AIBERT. The purpose is to improve its concentration on the aspect-level information of the corpus so that it does not need to spend extra power for computation to analyse the aspect-level emotional information in the corpus.

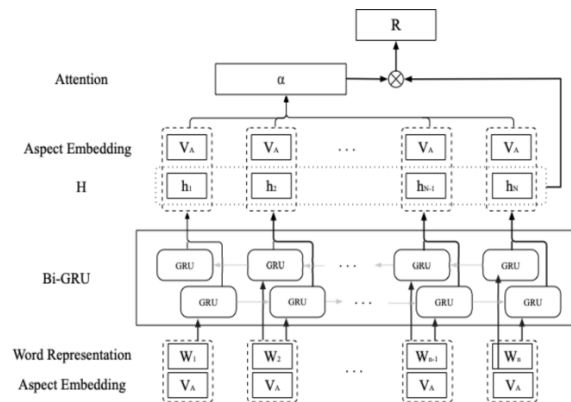


Figure 2. Structure of ATAE-BiGRU.

3.2 Output section

Compared with the LSTM model, the GRU model has a better effect when dealing with relatively long text information. On one hand, the model can retain a longer memory of semantics. On the other hand, it offers higher computational efficiency than LSTM. It is necessary to analyse the context of the text in order to better understand the emotional tendency of different sections in the text. In summary, this article improves the ATAE-LSTM model and constructs the ATAE-BiGRU model. The specific structure of this part is shown in Figure 2.

In this paper, two aspects are optimized in the ATAE-BiGRU model. On one hand, the original LSTM structure of the model is modified to the GRU structure. On the other hand, the unidirectional GRU structure is modified to the bi-directional Bi-GRU structure to calculate the aspect-level information from both the beginning and the end of the corpus at the same time.

3.3 Network processing flow

The specific processing flow of using ATAE-AIBERT-BiGRU network for aspect-level text sentiment classification is as follows:

Step 1: Normalize the text in the corpus into sentences of length N , truncate the excess part, and convert each word into a sentence vector: $W = [w_1, w_2, \dots, w_N]$, and at time convert the Aspect words W_a into a word vector E_a in parallel.

Step 2: Concatenate each word vector w_i and aspect vector E_a in the sentence vector V_a into the AIBERT model for training, and then splice the output result $R = (R_1, R_2, \dots, R_N)$ with the aspect vector E_a as the presentation layer.

Step 3: Input the presentation layer into the ATAE-Bi-GRU model and obtain the hidden layer $H = (h_1, h_2, \dots, h_N)$ after the training of the Bi-GRU model.

Step 4: The concatenation of hidden layer H and aspect vector E_a are put into a full connected layer to achieve a middle layer α .

Step 5: The result of Step 4, fully connected with the hidden layer H , is passed through a *Softmax* layer, to achieve the sentiment classification.

3.4 The loss function of the network

The loss function of the ATAE-AIBERT-BiGRU network adopts the cross-entropy loss function:

$$Loss = -\frac{1}{N} \sum_i p(x_i) \log q(x_i) \quad (2)$$

The calculation process is as follows:

Step 1: Split all samples into n mini batches.

Step 2: Forward spread (pick one of the mini batches):

Step 3: Calculate the loss by the Loss Function.

Step 4: Backward spread.

Step 5: Update parameters of weights.

Step 6: Repeat Step 2 to Step 5, till finish all mini batches.

To prevent the model from over-fitting, a random deactivation rate Dropout is introduced in the process of linking each layer to the lower layer, and the links of some nodes are randomly removed. The method described above improve the generalization of the network while model does not need to excessively rely on training samples.

4. EXPERIMENT

4.1 Experimental data set and evaluation indicators

The experimental data set uses the SemEval-2014Task4 dataset²². Each comment in the data set corresponds to an Aspect list and corresponding Aspect sentiment tendencies. This article uses some of the Restaurant data for experiments. The specific data distribution is shown in Table 1:

Table 1. Distribution of SemEval-2014 task 4.

	Restaurant	
	Train	Test
Pos.	2164	728
Neg.	807	196
Neu.	637	196
Total	3608	1120

Training set is divided into 5 parts and cross-validation is utilized for training. Test set is used for verification. Experimental results are achieved after the execution of steps mentioned above. The result of the experiment is evaluated by the accuracy rate and the F_1 Score.

4.2 Experimental hyperparameter settings

For the ATAE-AIBERT-BiGRU model mentioned in this article, the main hyperparameters used are divided into two parts, including AIBERT layer hyperparameters and Bi-GRU layer hyperparameters. Among them, the hyperparameters of the AIBERT model use the parameter settings in the AIBERT-Large pre-training model, as shown in Table 2.

Table 2. Hyperparameters of AI-BERT.

Parameter	Value
Dimension of word vector	128
Dimension of hidden layer	1024
Number of hidden layer	12
Size of vocabulary table	30000
Activation function	ReLU
Number of heads of attention	12

The hyperparameters of Bi-GRU layer are shown in Table 3:

Table 3. Hyperparameters of Bi-GRU.

Parameter	Value
Dimension of hidden layer	1024
Number of hidden layer	1
Number of mini-batch	32
Dropout rate	0.2
Learning rate	1e-4
Activation function	ReLU

4.3 Comparative experiment

To better evaluate the result of the ATAE-AIBERT-BiGRU on aspect-level sentiment analysis, SVM, ATAE-LSTM, ATT-CNN and BERT are used for comparative experiments. Details are as follows:

- (1) SVM: Use SVM as a classifier for feature extraction, and inject the sentiment dictionary, so that SVM can analyse aspect-level sentiment.
- (2) ATAE-LSTM: Achieve aspect-level sentiment analysis by adding attention mechanism and Aspect information to the LSTM network.
- (3) ATT-CNN: A specific aspect-level attention mechanism is added to the traditional convolutional neural network to obtain the attention information of the emotions in the sentence, to realize the aspect-level sentiment analysis²³.
- (4) BERT: Directly use the BERT pre-training model to input sentiment classification labels and sentence vectors into the network for training on aspect-level sentiment analysis tasks.

4.4 Experimental results

For the above four models and the ATAE-AIBERT-BiGRU proposed in this article, a comparative experiment was carried out on the data set SemEval2014. The experimental results are shown in Table 4.

Table 1. Experimental results.

	Acc.	F ₁ Score
SVM	80.00	61.32
ATAE-LSTM	78.50	77.81
ATT-CNN	68.19	-
BERT	81.54	76.91
ATAE-AIBERT-BiGRU	81.59	77.93

The ATAE-AIBERT-BiGRU model achieved satisfactory results on the SemEval2014-Res data level. Compared to the other four models, it improved in Acc. and F_1 Score indicator.

The loss function and accuracy of the model can converge faster due to the reduction of the two parameters during training. Figures 3 and 4 respectively show the tendency of the loss function and the accuracy rate of the ATAE-AIBERT-BiGRU model. It clearly states the model can converge to a more ideal result after 500 epochs, and the result is completely stable after 1000 epochs, which shows that the ATEAE-AIBERT-BiGRU model can quickly converge to a more ideal state.

To verify the effectiveness of the model after structural adjustment and optimization, the experiment analysed the accuracy change trend of the prediction results of the four models, as shown in Figure 5.

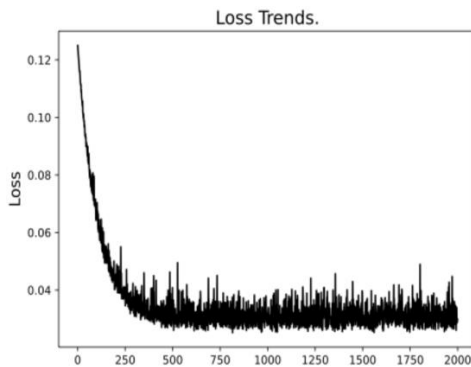


Figure 3. The change trend of the loss function of the TAE-AIBERT-BiGRU model.

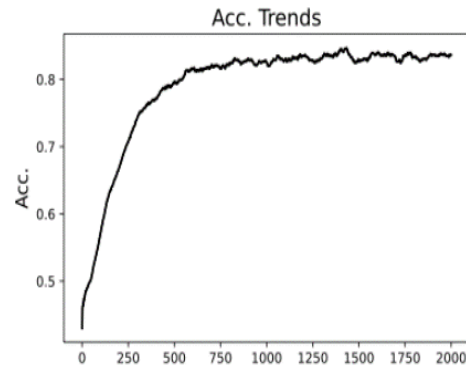


Figure 4. The accuracy rate change trend of the ATAE-AIBERT-BiGRU model.

Figure 5 indicates that although the accuracy of the BERT model after the training process is close to that of the model in this paper, its convergence speed is not as good as the model in this paper. The accuracy of the ATAE-LSTM model and the ATT-CNN model is obviously not as good as the ATAT-LSTM model and the model in this paper. Compared with the other three models, the accuracy of the model in this paper has basically reached a stable peak after 500 epochs. Therefore, the model in this paper can use less calculations and achieve a better result. Comparing to the ATAE-LSTM model, the model in this paper allows using the context information of the corpus at the same time, it can understand the emotional information in the corpus more accurately, which further improves the classification accuracy of the model.

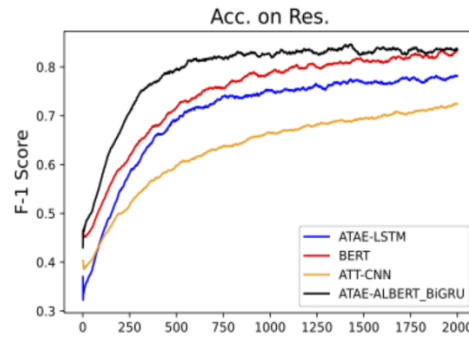


Figure 5. Accuracy changes trend of each model on the Res. data set.

4. CONCLUSION

This paper proposes a network model for aspect-level sentiment analysis: ATAE-AIBERT-BiGRU. This model improves the ATAE-LSTM model and optimizes its core structure to the BiGRU structure. In addition, the network performance is improved by combining the ability of the AIBERT model to understand sentence semantics and the ability of ATAE-BiGRU model to process aspect-level emotions. Experimental analysis shows the model in this paper has higher computational efficiency and classification accuracy in text-level sentiment classification tasks comparing to the other models discussed in the previous context. The next step is to apply the model to a wider range of data sets, including some other English data sets and Chinese data sets, and to verify its versatility.

ACKNOWLEDGMENTS

This paper is supported by the National Key R&D Program of China (No. 2018YFB1702900).

REFERENCES

- [1] Poria, S., Cambria, E., Bajpai, R. and Hussain, A., "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion*, 37, 98-125 (2017).
- [2] Miller, G. A., [WordNet: An Electronic Lexical Database], MIT press, Cambridge.
- [3] Murty, M. R., Murthy, J., Reddy, P. P. and Satapathy, S. C., "A survey of cross-domain text categorization techniques," *1st Inter. Conf. on Recent Advances in Information Technology (RAIT)*, 499-504 (2012).
- [4] Pang, B., Lee, L. and Vaithyanathan, S., "Thumbs up? Sentiment classification using machine learning techniques," *arXiv preprint cs/0205070*, (2002).
- [5] Peng, M., Wang, Q., Huang, J. M., Zhou, L. and Hu, X. H., "Stock analysing based on sentiment analyse," *Journal of Wuhan University*, 124-30 (2015). (in Chinese)
- [6] Jain, A. P. and Dandannavar, P., "Application of machine learning techniques to sentiment analysis," *2016 2nd Inter. Conf. on Applied and Theoretical Computing and Communication Technology (iCATccT)*, 628-32 (2016).

- [7] Wang, R. Y., Ju, J. P., Li, S. S. and Zhou, G. D., "Opinion targets based on CRFs," *Journal of Chinese Information*, 26, 56-62 (2012). (in Chinese)
- [8] Socher, R., Huval, B., Manning, C. D. and Ng, A. Y., "Semantic compositionality through recursive matrix-vector spaces," *Proc. of the 2012 Joint Conf. on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1201-11 (2012).
- [9] Zhu, X., Sobihani, P. and Guo, H., "Long short-term memory over recursive structures," *Inter. Conf. on Machine Learning*, 1604-12 (2015).
- [10] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J., "Distributed representations of words and phrases and their compositionality," *Advances in Neural Information Processing Systems*, 3111-9 (2013).
- [11] Sutskever, I., Vinyals, O. and Le, Q. V., "Sequence to sequence learning with neural networks," *Advances in Neural Information Processing Systems*, 3104-12 (2014).
- [12] Zhou, C., Sun, C., Liu, Z. and Lau, F., "Category enhanced word embedding," *arXiv preprint arXiv:151108629*, (2015).
- [13] Wang, Y., Huang, M., Zhu, X. and Zhao, L., "Attention-based LSTM for aspect-level sentiment classification," *Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing*, 606-15 (2016).
- [14] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I., "Attention is all you need," *Advances in Neural Information Processing Systems*, 5998-6008 (2017).
- [15] Devlin, J., Chang, M. W., Lee, K., and Toutanova, K., "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:181004805*, (2018).
- [16] Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M. and Liu, Q., "ERNIE: Enhanced language representation with informative entities," *arXiv preprint arXiv:190507129*, (2019).
- [17] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P. and Soricut, R., "Albert: A lite bert for self-supervised learning of language representations," *arXiv preprint arXiv:190911942*, (2019).
- [18] Wu, Z. and Ong, D. C., "Context-guided bert for targeted aspect-based sentiment analysis," *Association for the Advancement of Artificial Intelligence*, 1-9 (2020).
- [19] Mao, R. and Li, X., "Bridging towers of multi-task learning with a gating mechanism for aspect-based sentiment analysis and sequential metaphor identification," *Proc. of the AAAI Conf. on Artificial Intelligence*, 13534-42 (2021).
- [20] Mnih, V., Heess, N., and Graves, A., "Recurrent models of visual attention," *Advances in Neural Information Processing Systems*, 2204-12 (2014).
- [21] Bahdanau, D., Cho, K. and Bengio, Y., "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:14090473*, (2014).
- [22] Kiritchenko, S., Zhu, X., Cherry, C. and Mohammad, S., "NRC-Canada-2014: Detecting aspects and sentiment in customer reviews," *Proc. of the 8th Inter. Work. on Semantic Evaluation (SemEval 2014)*, 437-42 (2014).
- [23] Yin, W., Schütze, H., Xiang, B. and Zhou, B., "ABCNN: Attention-based convolutional neural network for modeling sentence pairs," *Transactions of the Association for Computational Linguistics*, 4, 259-72 (2016).