


RESEARCH

Open Access



Attention-based Sentiment Reasoner for aspect-based sentiment analysis

Ning Liu^{1,2}, Bo Shen^{1,2*} , Zhenjiang Zhang^{2,3}, Zhiyuan Zhang^{1,2} and Kun Mi⁴

*Correspondence:

bshen@bjtu.edu.cn

¹ School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China

Full list of author information is available at the end of the article

Abstract

Aspect-based sentiment analysis (ABSA) is a powerful way of predicting the sentiment polarity of text in natural language processing. However, understanding human emotions and reasoning from text like a human continues to be a challenge. In this paper, we propose a model, named Attention-based Sentiment Reasoner (AS-Reasoner), to alleviate the problem of how to capture precise sentiment expressions in ABSA for reasoning. AS-Reasoner assigns importance degrees to different words in a sentence to capture key sentiment expressions towards a specific aspect, and transfers them into a sentiment sentence representation for reasoning in the next layer. To obtain appropriate importance degree values for different words in a sentence, two attention mechanisms we designed: intra attention and global attention. Specifically, intra attention captures the sentiment similarity between any two words in a sentence to compute weights and global attention computes weights by a global perspective. Experiments on all four English and four Chinese datasets show that the proposed model achieves state-of-the-art accuracy and macro-F1 results for aspect term level sentiment analysis and obtains the best accuracy for aspect category level sentiment analysis. The experimental results also indicate that AS-Reasoner is language-independent.

Keywords: Aspect-based sentiment analysis, Natural language processing, Attention mechanism, Text mining, Deep learning

Introduction

Over the last decade, aspect-based sentiment analysis (ABSA) has been a rapidly growing field of natural language processing [1]. ABSA is a fine-grained sentiment analysis task which aims at detecting the polarity of an entity or an entity's attribute [2]. In ABSA, the aspect can be divided into two levels: aspect term (also called aspect target) and aspect category. An example sentence is provided in Fig. 1. For the sentence "Best Pastrami I ever had and great portion without being ridiculous," the aspect terms are "Pastrami" and "portions," respectively. Aspect categories are "FOOD#QUALITY" and "FOOD#STYLE_OPTIONS." Obviously, a satisfactory method for an ABSA task should be applicable at both aspect levels.

In the early days, traditional machine learning methods dominated as the method for an ABSA task, such as Support Vector Machine (SVM) [3]. However, these methods need feature engineering which is time consuming and laborious. Recently, deep learning methods have become increasingly popular for the sentiment analysis task, such as

long short term memory [4], gated recurrent unit [5], recursive neural network [6], convolutional neural network [7] and memory-augmented network [8]. However, the challenge of how to make the machine reason like humans in the aspect-based sentiment analysis task remains a challenge.

One possible way to give the machine a reasoning ability for aspect-based sentiment polarity is to gradually assign an increasingly precise weight to different words according to a specific aspect. First, the importance of different words with respect to a specific aspect can be captured from a global perspective. For example, in Fig. 1, the polarity is positive with respect to the aspect term “Pastrami” and the aspect category “FOOD#QUALITY,” which are decided upon when glancing at the word “Best,” and ignoring words such as “I” and “is.” An effective deep learning approach should distinguish important degrees levels for different words according to a specific aspect.

Second, the importance of different words in the context with respect to a specific aspect can be captured by modeling the relationship between any two words in a review. Modeling the relationship between any two words in the context can learn sentiment words that are not present in a sentence. Therefore, an effective deep learning approach should have the ability to capture sentiment similarity between different words in the context, which results in the machine understanding the meaning of different sentiment expressions.

This paper proposes a multi-layered neural network architecture, named Attention-based Sentiment Reasoner (AS-Reasoner) to deal with the problem of how to make the machine reason like humans in sentiment analysis. AS-Reasoner assigns importance degrees to different words in a sentence to capture key sentiment expressions toward for specific aspects, much like how reasoning is done by humans, then utilizes these key sentiment expressions in reasoning in the next layer. To assign different degrees to different words in a sentence, two kinds of attention mechanism are proposed: intra attention and global attention. Specifically, intra attention captures sentiment similarity between any two words in a sentence to assign weights. Global attention assigns weights by a global perspective. Experiments on all datasets demonstrate that our proposed approach obtains state-of-the-art results in terms of both aspect target and aspect category level tasks. The experimental results also indicate that our proposed approach is language-independent.

Sentence 1: Best Pastrami I ever had and great portion without being ridiculous.

Aspect term: Pastrami; **Aspect category:** FOOD#QUALITY; **Polarity:** positive
Aspect term: portions; **Aspect category:** FOOD#STYLE_OPTIONS; **Polarity:** positive

Sentence 2: Worth the trip from Manhattan.

Aspect term: NULL; **Aspect category:** RESTAURANT#GENERAL; **Polarity:** positive

Fig. 1 Two example sentences of aspect term and aspect category in restaurant domain. In general, aspect category is predefined in the domain

Related work

In the field of ABSA, neural networks, especially RNN, are most commonly employed to classify the sentiment polarity of an aspect. Tang et al. [9] proposed TD-LSTM and TC-LSTM to incorporate aspects into the model. TD-LSTM divides a sentence into a left part and right part around the aspect target and each part is fed into two LSTM models with separated forward and backward sequential direction. The last hidden vectors of left LSTM and right LSTM are concatenated and fed into softmax to classify the sentiment polarity label. However, TD-LSTM does not capture the interactions between aspect target and the context. To alleviate the problem, TC-LSTM considers the semantic relation between the aspect and its context by concatenating aspect target embeddings and context word embeddings as the inputs, and sends them into two separated LSTMs in the forward and backward direction, which is similar with the strategy used in TD-LSTM. Wang et al. [10] designed AT-LSTM and ATAE-LSTM which employs the attention mechanism to explore the correlation of aspect and context. Attention mechanism can detect the import part of sentence towards a given aspect. AT-LSTM concatenates each hidden representation of LSTM and aspect target embedding as a matrix to learn attention weights by hyperbolic tangent and softmax functions. After producing attention weights, AT-LSTM can get a weighted hidden representation. The last hidden vector of LSTM and the weighted hidden representation add up to make the final sentence representation by the nonlinear function. Then, the final sentence representation is fed into softmax to decide the sentiment polarity. In order to better utilize aspect target information, ATAE-LSTM concatenates each context embedding with aspect target embedding as the input of LSTM. Using this way, the hidden vectors of LSTM can contain the information from the aspect target, which can make the model obtain more precise attention weights.

Employing only one attention mechanism may not have the ability to capture the importance of varying context words when sentiment information is distributed over a long distance in a sentence. To alleviate this problem, multiple-attention mechanisms have been adopted, such as Recurrent Attention on Memory (RAM) [11] and Interactive Attention Network (IAN) [12]. RAM utilizes multiple attentions to capture complicated features from its memory. Attention weights of memory slices in current episode can be obtained from the memory slice, the state of previous episode and the aspect target vector. The content vector can be obtained using these attention weights. GRUs are employed to combine these attention results. The last hidden vector of GRU is fed into the softmax function to predict the sentiment polarity. IAN employs multiple attentions to separately learn the representations of both contexts and targets via interactive learning. IAN utilizes average pooling to get the initial representations of context and target, and then they are injected into each other to produce attention weights. Ruder et al. [13] found that the performance of such models can be improved by modeling the inner knowledge of the review structure. Therefore, a hierarchical model, named Hierarchical bidirectional LSTM (H-LSTM) was proposed to capture the sentiment correlation of different sentences in a review. H-LSTM employs bidirectional LSTMs as a review-level LSTM and sentence-level LSTMs separately. H-LSTM stacks review-level LSTM on the top of these sentence-level LSTMs to form the hierarchical architecture.

Recently, memory networks have achieved better results in question answering [14, 15]. Tang et al. [8] introduced a multiple-hop memory network in which the context words are considered as the fact description and the aspect target is regarded as the question. Attention mechanism is employed to select related information towards the aspect target. This model can obtain better accuracy as the number of hops increases, but still may be inadequate for generating powerful attention value and word-aspect representations. Tay et al. [16, 17] proposed circular convolution and circular correlation to implement word-aspect associative fusion to mitigate the problem of weak attention weights. In circular correlation, they use the Fast Fourier Transform (FFT), inverse Fast Fourier Transform and the complex conjugate of FFT to model hidden vectors of LSTM and aspect embedding. In circular convolution, the FFT and inverse FFT are employed to model hidden vectors of LSTM and aspect embedding, in which the conjugate operator is absent.

In some cases, the aspect term is formed by multiple words. The methods mentioned above take the average sum of aspect word embeddings as the word embedding of an aspect, which may result in a new word with an irrelevant word meaning and lose the sequential information of the aspect term. Some approaches have been proposed to solve the problem, such as Target-specific Transformation Networks (TNet) [18] and Aspect Target Sequence Model (ATSM) [19]. The former employs a bidirectional LSTM to model aspect term sequence and generate target-specific representations thru the interaction of each context representation and aspect term representations. The latter treats an aspect term at three granularities: radical, character and word. ATSM firstly learns the adaptive word embeddings of aspect terms thru a LSTM. Then, aspect target sequence is fed into a LSTM to model the sequence information at three granularities. Finally, ATSM designs fusion mechanisms to combine these representations from three granularities. Yang et al. [20] proposed a coattention mechanism by learning both aspect attention weights and context attention weights alternatively, which firstly focus on important targets and then learn more effective context representation. They applied the coattention mechanism into LSTM and MemNet, and achieve better results than vanilla LSTM and MemNet.

In recent year, Bidirectional Encoder Representations from Transformers (BERT) [21] has achieved great success across a variety of NLP tasks. For taking advantage of BERT fully, Sun et al. [22] constructed additional auxiliary sentences towards the specific aspect by four ways and transformed ABSA into a sentence-pair classification task. Xu et al. [23] proposed a novel post-training approach on BERT in ABSA. Model can fine-tune the parameters of BERT by the post-training way on new dataset.

In ABSA, aspect terms need to be identified when the aspect is not given. Tang et al. [24] proposed a topic model, named joint aspect based sentiment topic (JABST), that jointly extracts aspects and opinions, then predicts the sentiment labels towards the specific aspect by modeling aspects, opinions and sentiment polarities simultaneously. In order to improve accuracy, they integrated maximum entropy into JABST, named MaxEnt-JABST, by means of the semi-supervised learning. Yu et al. [25] proposed a multi-task learning framework to implicitly capture the relations between the extracting aspect terms and opinion terms by employing a BiLSTM as the encoder. Then, they proposed a global inference method to identify aspects and opinions by

explicitly modeling several syntactic constraints between aspect term extraction and opinion term extraction. Knowledge is important in natural language processing. Integrating common knowledge or other knowledge such as domain knowledge or sentiment knowledge into the ABSA task has aroused great attention. Wu et al. [26] proposed a unified model which incorporates structure and sentiment knowledge on ABSA. Structure knowledge is extracted thru clause recognition and fused in the model via the generation of multiple context representations. Sentiment knowledge is exploited through pretraining the model with the sentiment labels of documents and freezing parameters of some layers. Ma et al. [27] proposed a knowledge-rich solution to ABSA which respectively leveraged commonsense knowledge to model the aspect and its context by employing the LSTM. In order to explicitly integrate the explicit knowledge with implicit knowledge, they extended LSTM, termed Sentic LSTM. The cell of Sentic LSTM designed a separate output gate that interpolated the token-level memory and the concept-level input. Kumar et al. [28] integrated the ontologies into convolutional neural network (CNN) for predicting sentiment polarities. The goal of ontology is to facilitate knowledge towards a specific domain in such a format that can be easily understood by machines. The semantic features extracted from the ontology model are fed into the CNN for classifying.

Intuitively, the words close to an aspect term may have more influence in predicting the sentiment polarity of the specific aspect target in a sentence. Inspired by this idea, Shuang et al. [29] incorporated the location information of context words by assigning different weights in Attention-Enabled and Location-Aware Double (AELA-DLSTM), which employed Double LSTMs to capture semantic information in both forward and backward directions. A novel attention mechanism is proposed in AELA-DLSTM to make better use of the correlations between aspect words and their context words. Zeng et al. [30] put forward a new attentive LSTM, termed PosATT-LSTM, to incorporate the position-aware vectors with Gaussian kernel in the PosATT-LSTM. These position-aware influence vectors are appended into the hidden representation of the contexts on the top of LSTM layer.

Zainuddin et al. [31] proposed a new hybrid aspect-based sentiment classification for Twitter by embedding a feature selection method. These feature selection methods are rule-based methods such as the principal component analysis (PCA), latent semantic analysis (LSA), and random projection (RP) features selection methods. When multiple aspect terms in a sentence need to predict their corresponding sentiment polarities simultaneously, these methods mentioned above are powerless in this scenario. Ma et al. [32] adopted the Frobenius norm of a matrix to regularize the attention weights of all aspects in a sentence, then they injected this pattern into both position attention and content attention mechanisms to predict sentiment labels of all aspect terms in the same sentence.

In some practical applications, there is a limited amount of labeled data in this scenario. The methods which are based on supervised learning are not suitable. To solve the problem, Fu et al. [33] proposed a novel Semi-supervised Aspect Level Sentiment Classification Model based on Variational Autoencoder (AL-SSVAE) for semi-supervised learning. The given aspect is sent into an encoder and then the output is sent into a decoder based on a variational autoencoder (VAE). Besides, AL-SSVAE

employed the ATAE-LSTM as the aspect level sentiment classifier to model the sentiment polarity information about a specific aspect. García-Pablos et al. [34] proposed an unsupervised system based on topic model, termed W2VLDA, which incorporated continuous word embeddings and a Maximum Entropy to perform aspect category classification. W2VLDA required only one seed word per domain aspect plus one positive and one negative word independent of the domain to predict sentiment polarities of aspect terms over any unlabeled corpus. These semi-supervised and unsupervised methods can effectively avoid costly and time-consuming works such as manually labelling data.

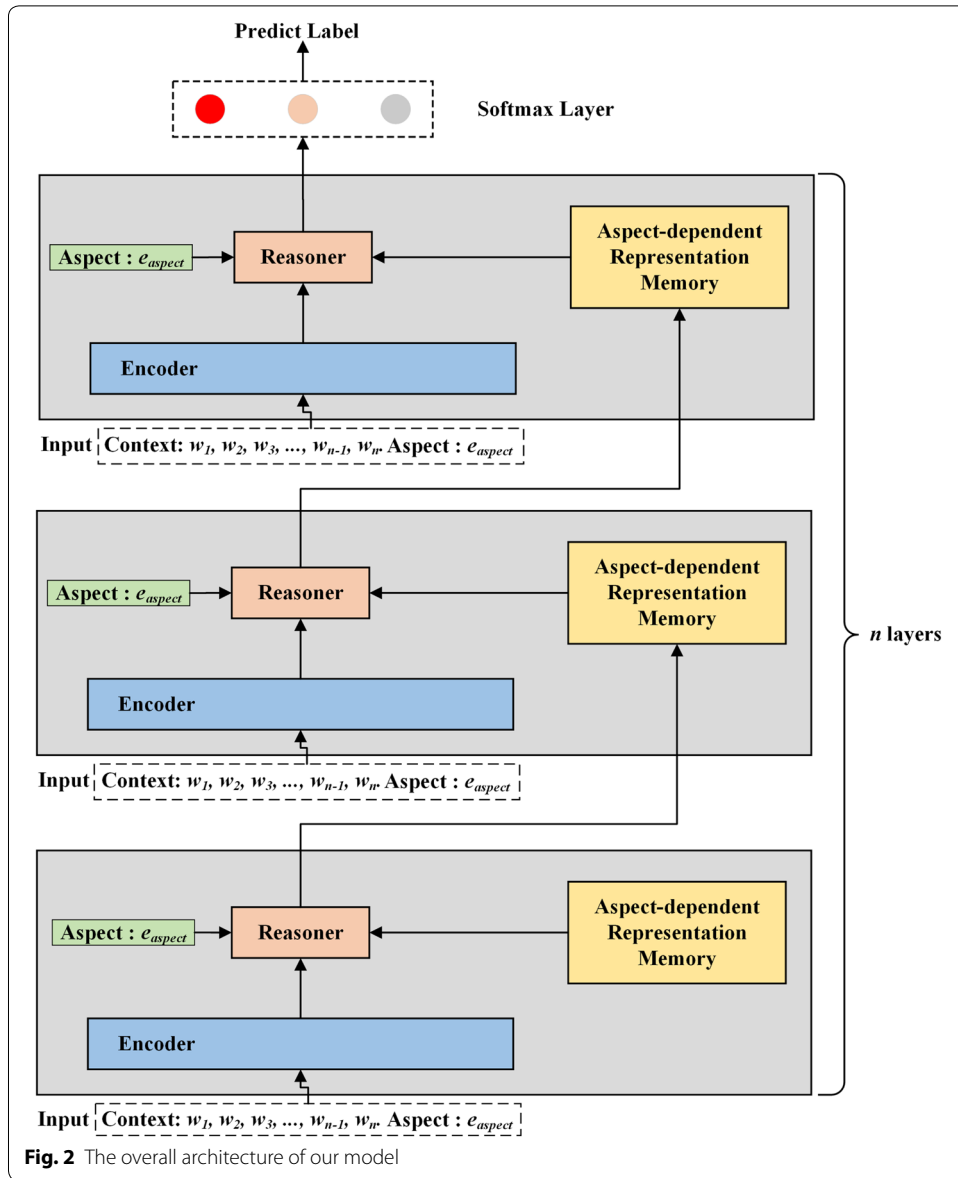
Although the models mentioned above have achieved much success on the ABSA task, none of them can reason sentiment information and learn semantically-similar words from a sentence. To the best of our knowledge, we are the first to address the ABSA task by reasoning with intra and global attentions.

Attention-based Sentiment Reasoner

The proposed AS-Reasoner model has a multi-layered architecture, as shown in Fig. 2. Each layer of AS-Reasoner can be seen as an encoder-reasoner structure that can complete the ABSA task independently. At first, the input sequence is encoded into a sequence of vector representations by the encoder. Then, making use of these generated representations, the aspect information and the current aspect-dependent representation state, the reasoner of each layer calculates a sentence vector representation to independently predict the sentiment orientation of the aspect and updates the aspect-dependent representation memory. Finally, we use the vector representation of the last layer's sentence as the final sentence representation to perform sentiment classification.

The most important part in AS-Reasoner is the reasoner module. AS-Reasoner can gradually extract informative sentiment expressions and semantically similar words in a sentence towards to the specific aspect by means of the reasoner module. The reasoner module is made up of intra attention and global attention. Intra attention is more concerned about the association relationship between any two words in a sentence. Meanwhile, global attention focuses on finding important sentiment expressions corresponding to a specific aspect from the perspective of the whole sentence. Through these two attention perspectives, the reasoner is able to gain the more precise sentence vector representation.

Meanwhile, like a cache, the aspect-dependent representation memory is only updated after output of the reasoner from the last layer, which is a vector representation. Aspect-dependent representation memory is just like an extra information source consisting of an independent sentence representation of each layer, and the memory state of each layer will not change in the multi-layered architecture. Therefore, there is no need to keep the aspect-dependent representation memory continuous. Although content in aspect-dependent representation memory is a vector representation, usage of the memory component is a more symbolic way in a neural network framework. Besides, the aspect-dependent representation memory makes it possible that AS-Reasoner can train in the larger number of layers.



Model inputs

We are dealing with the sentiment classification of an aspect target t_n and an aspect category $u_n \# z_n$, when the aspect target level task is performed, the aspect representation can be computed as the following:

$$e_{aspect} = \sum_{n=1}^m t_n, \quad t_n \in \mathbb{R}^d \tag{1}$$

where d is the dimension of the word embedding. When the aspect category level task is performed, the aspect representation can be computed as the following:

$$e_{aspect} = \sum_{n=1}^{l1} u_n + \sum_{n=1}^{l2} z_n, \quad u_n \in \mathbb{R}^d, \quad z_n \in \mathbb{R}^d \tag{2}$$

where m is the length of the aspect target, $l1$ is the length of the entity and $l2$ is the length of the attribute, d is the dimension of the word embedding. We concatenate the aspect representation with each word embedding within a sentence to get input \hat{x} .

Encoder

The encoder is an RNN model, which takes the sequence $\{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_l\}$ as the input and is unidirectional. Here we use LSTM [35] as our encoder, which alleviates the problem of a vanishing gradient occurring in a vanilla RNN. Moreover, it can efficiently capture long-term dependencies and model the word order information contained in a sentence. For the word \hat{x}_n in the input sequence, its hidden state h_n can be obtained from the following equations:

$$h_n = LSTM(\hat{x}_n, h_{n-1}, c_n), \quad h_n \in \mathbb{R}^{d \times 1} \quad (3)$$

where $LSTM$ denotes processing of the computing in the LSTM, c_n is the memory state of the LSTM.

Aspect-dependent representation memory

In the process of human reading, we read text from end-to-end and cover to cover. The sentence semantics in our brain are continually updated by the repeated reading. When a sentence is long, we may leave out some important clues and may not remember the precise expressions. We may also repeatedly read a particular sentence to deepen our understanding.

Like the process of human reading, AS-Reasoner relies on sentence representations to predict the sentiment polarity of the aspect, one important work is how to more effectively and reasonably store these vector representations. Our proposed solution is to design extra memory for storage and updating. The memory thus provides decisive information for the reasoner to produce an inference result. Aspect-dependent representation memory consists of the reasoner's output from each layer, i.e.,

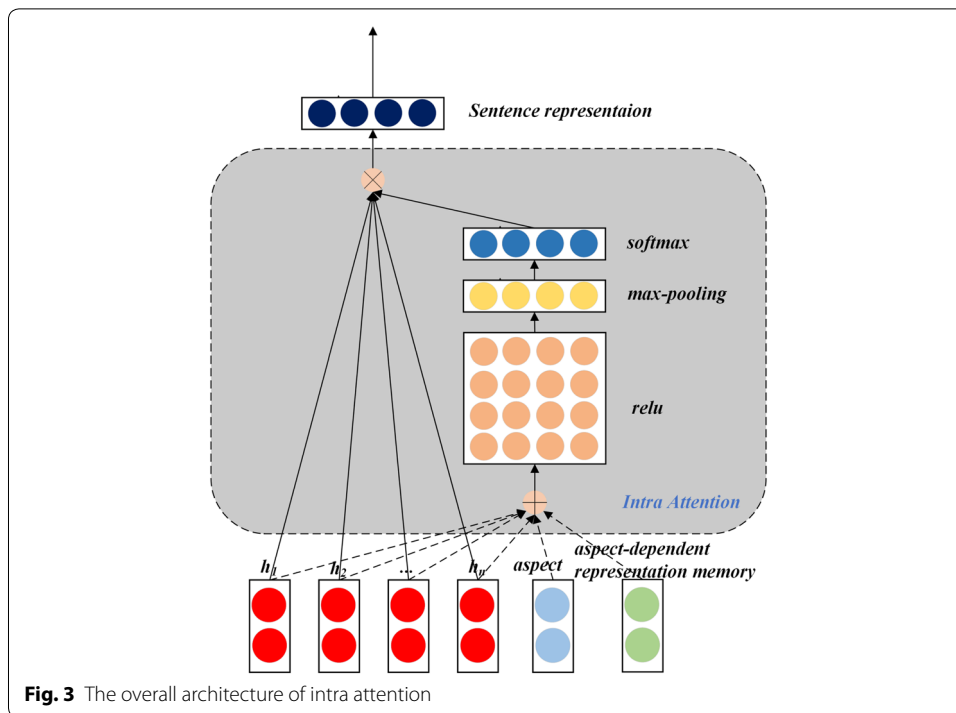
$$M = [\hat{c}_1, \hat{c}_2, \dots, \hat{c}_{n'-1}, \hat{c}_{n'}] \quad (4)$$

$$\hat{c}_j = r_{j-1}; \quad j \in [2, n'], \quad \hat{c}_j \in \mathbb{R}^{d \times 1} \quad (5)$$

where n' is the number of the layers, it is a hyper-parameter, \hat{c}_1 is initialized by the normal distribution where the mean is 0 and the stand deviation is 0.01, and, r denotes the output of the reasoner. Aspect-dependent representation memory is more like a list.

Reasoner

Based on aspect-dependent representation memory, we can store the semantics of a sentence as an object that is a highly abstract generalization. In human reading, sentiment polarity is predicted by the sentence semantics that we have processed in our brain. Additionally, the aspect information also needs to be considered in the reasoning, because the aspect is a concern in the ABSA task. As for human understanding of



text, different importance degrees for different words are assigned that correspond to a specific aspect. Thus, a better decision on meaning can be achieved by applying aspect information, and the current word and aspect-dependent representations in the attention mechanism.

In this paper, we propose two attention mechanisms to alleviate the problem of how to precisely capture sentiment expressions in ABSA to reason: intra attention and global attention.

Intra attention

Sentiment similarity between two words in a sentence can be used to identify import sentiment expressions and assign reasonable weights to different words in a specific sentence. For example, the sentence “The wine list is interesting and has many good values” contains two aspect categories. For the aspect category “DRINKS#STYLE_OPTIONS”, the model first learns that the word “interesting” implies a positive sentiment polarity. Then the model can infer that the aspect category “DRINKS#PRICES” is also positive, referring to a sentiment similarity between “interesting” and “good”. For the aspect target “wine list”, capturing the sentiment similarity between both words can confirm with increased certainty sentiment polarity.

The overall architecture of intra attention is shown in Fig. 3. Specially, attention weight α_i between each word’s hidden state of encoder h_i , and the aspect-dependent memory vector \hat{c}_i can be obtained in the following formulas:

$$u_i = \text{relu}(W_u \cdot [e_{\text{aspect}}; h_i; \hat{c}_i] + b_u); \quad i \in [1, n] \tag{6}$$

$$p_i = \text{max-pooling}(u_i) \tag{7}$$

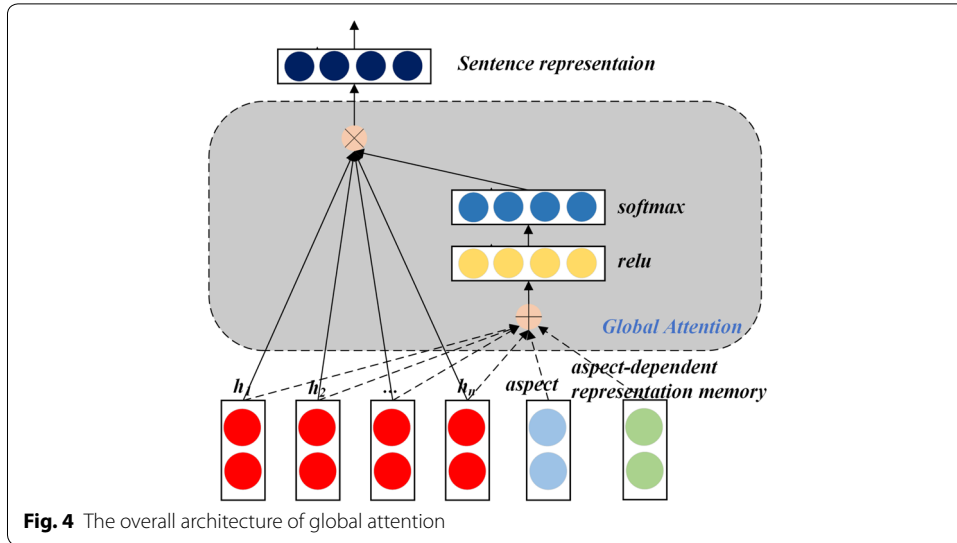


Fig. 4 The overall architecture of global attention

$$a_i = \text{softmax}(p_i) = \frac{\exp(p_i)}{\sum_{t=1}^{n'} \exp(p_t)} \tag{8}$$

where n is the number of context length, $W_u \in \mathbb{R}^{n \times 3d}$, $u_i \in \mathbb{R}^{n \times 1}$, representing the association relationship between the current word and other words in a sentence, $p_i \in \mathbb{R}$, $a_i \in \mathbb{R}$, *max-pooling* denotes a max-pooling operation which the step is 1, where the height of the window size is 1 and the weight of the window size is n . We can obtain word embedding of the most relevant word responding to the current word.

Then we compute sentence representation s_k as a weighted sum of the hidden state of the encoder on the weights, i.e.,

$$r_k = s_k = \sum_{i=1}^{n'} \alpha_{ki} \cdot h_{ki}; \quad k \in [1, n'], \quad r_k \in \mathbb{R}^{d \times 1} \tag{9}$$

where n is the number of the layer in AS-Reasoner, and r_k denotes the output vector of the k -th layer.

Global attention

Global attention captures key sentiment expressions from a global perspective within a sentence, and not just the sentiment similarity between any two word within a sentence.

The overall architecture of global attention is shown in Fig. 4. We can get attention weight α_i between each word’s hidden state of encoder h_i , and the aspect-dependent representation memory vector \hat{c}_i thru the following formulas:

$$v_i = \text{relu}(W_v \cdot [e_{\text{aspect}}; h_i; \hat{c}_i] + b_v); \quad i \in [1, n] \tag{10}$$

$$a_i = \text{softmax}(v_i) = \frac{\exp(v_i)}{\sum_{t=1}^{n'} \exp(v_t)} \tag{11}$$

where n is the number of context length, $W_v \in \mathbb{R}^{1 \times 3d}$, $v_i \in \mathbb{R}$.

After computing the global attention weights, sentence representation s_k is obtained as a weighted sum of the hidden state of the encoder on the weights, and the computational formula is the same as that in Eq. (10). The output vectors of the attention mechanism are viewed as the output representations of the reasoner. They are then used to update the aspect-dependent representation memory.

Combining both attention perspectives

Two kinds of attention are used separately in the reasoner to capture the differences in the importance of different words. We can also use both attention mechanisms to get more informative sentence representation to improve the model's performance.

After computing sentence representation by intra attention s_k^{intra} and sentence representation by global attention s_k^{global} , we can get a compact sentence representation by:

$$r_k = s_k = s_k^{intra} + s_k^{global}; \quad k \in [1, n'] \quad (12)$$

where n' is the number of the layer in AS-Reasoner, and r_k denotes the output vector of the k -th layer.

Softmax layer

We feed the sentence representation of the last layer $s_{n'}$ to a softmax function to perform the ABSA task. The sentiment polarity of the aspect can be computed as the following:

$$o = W_o \cdot s_{n'} + b_o \quad (13)$$

$$\hat{y} = \underset{k}{\operatorname{argmax}} \left(\frac{\exp(o_k)}{\sum_{t=1}^K \exp(o_t)} \right) \quad (14)$$

where $W_o \in \mathbb{R}^{K \times d}$, $b_o \in \mathbb{R}^{d \times 1}$, $o \in \mathbb{R}^{K \times 1}$, K is the number of sentiment class.

Model training

We use the cross-entropy function as the loss function to train the proposed model in an end-to-end approach. Given training data $\{x^i, e^i, y^i\}$, where x^i denotes the i -th sentence, e^i denotes the corresponding aspect and y^i denotes the one-hot representation of the ground-truth sentiment polarity for the current sentence and the current aspect. The loss function is defined as:

$$J(\theta) = - \sum_{i=1}^N y^i \log(\hat{y}^i) \quad (15)$$

where N is the number of training samples.

Experimentation

Experimental settings

We perform aspect term level ABSA task on four Chinese datasets [19] which are camera, car, notebook and phone domains and four English datasets which are restaurant and laptop domains from SemEval-2014 and two restaurant domains from SemEval-2015

Table 1 The statistical information of all datasets

Datasets	Overall	Pos.	Neg.	Neu.	Len.
Restaurant (2014)					
Train	3578	2148	800	630	78
Test	1110	721	195	194	
Laptop (2014)					
Train	2317	991	865	459	82
Test	637	341	128	168	
Restaurant (2015)					
Train	1279	963	280	36	74
Test	597	353	207	37	
Restaurant (2016)					
Train	1880	1319	489	72	78
Test	650	483	135	32	
Camera					
Train	1635	1157	478	–	26
Test	408	263	145	–	
Car					
Train	885	668	217	–	34
Test	221	166	55	–	
Notebook					
Train	485	331	154	–	18
Test	121	72	49	–	
Phone					
Train	1885	1260	625	–	39
Test	471	303	168	–	

We use “–” to indicate that the dataset doesn’t contain this kind of class. We use ‘Len.’ to represent the max length of the sentence. We respectively use “2014”, “2015” and “2016” to represent that the dataset is belong to SemEval 2014, 2015 and 2016

and SemEval-2016 respectively. We conduct aspect category level ABSA task on the restaurant domain from SemEval-2015. The four Chinese datasets contain two sentiment polarities, i.e., positive or negative. The other English datasets contain three sentiment polarities, i.e., positive, neutral or negative. The statistical information of all datasets is shown in Table 1. Glove [36] 300-dimension word embeddings are adopted as word embeddings in all English datasets. We randomly initialize word embeddings in Chinese datasets and other all parameters in our proposed model by uniform distribution in $[-0.01, 0.01]$. Adam [37] is employed as model’s optimization algorithm and the learning rate is 0.001. Dropout [38] is used to be the regularization strategy and is set to 0.5. Accuracy and Macro-F1 are employed to evaluate the performance of the model.

Experimental results

In this subsection, we compared our proposed approach with some state-of-the-art approaches to comprehensively evaluate the performance of AS-Reasoner.

1. LSTM [10]: This approach employs LSTM to encode a specific sentence to predict sentiment polarity.
2. TD-LSTM [9]: This approach employs two LSTMs to encode the sentence from two opposite directions respectively.

Table 2 Accuracy on aspect term level ABSA on four English datasets

Model	Res (2014)	Lap (2014)	Res (2015)	Res (2016)
LSTM	74.28	66.45	–	–
TD-LSTM	75.63	68.13	–	–
ATAE-LSTM	77.20	68.70	–	–
MemNN	78.38	71.11	73.36	84.15
IAN	78.60	71.78	77.55	85.84
RAM	78.93	71.81	78.56	87.08
AS-Reasoner-I	79.73	72.41	79.06	85.69
AS-Reasoner-G	79.64	72.57	78.56	86.62
AS-Reasoner	<i>80.00</i>	<i>74.77</i>	<i>79.90</i>	86.31

Italic values indicate the significance of the best result

“Res” denotes the restaurant domain, “Lap” denotes the laptop domain

3. TC-LSTM [9]: This approach extends TD-LSTM by concatenating aspect embedding with the context embeddings.
4. ATAE-LSTM [10]: This approach employs an attention mechanism and LSTM to learn sentence representation.
5. MemNN [8]: This approach regards the context as the fact and views aspect as the query to predict sentiment polarity.
6. IAN [12]: This approach employs the attention and pooling mechanisms to model the context and aspect separately.
7. RAM [11]: This approach uses multiple attention mechanisms to capture the importance of different words and combine the result in a non-linear way.
8. ATAM-S [19]: This approach encodes the sentence and aspect with word granularity.
9. ATAM-F [19]: This approach encodes the sentence and aspect in three granularities and combines them in two ways.
10. Word&Clause-Level ATT [39]: This approach employs word-level and clause-level attentions to predict sentiment.
11. AS-Reasoner-Intra: This approach is our approach which only employs intra attention.
12. AS-Reasoner-Global: This approach is our approach which only employs global attention.
13. AS-Reasoner: This approach is our approach which combines both intra attention and global attention to learn and obtain a more informative and precise sentiment representation.

The experimental results for the aspect term level ABSA task on the four English datasets are shown in Table 2. We use “–” to denote that the model didn’t report experimental results on the dataset in the original paper. From Table 2, we can see that AS-Reasoner-I and AS-Reasoner-G obtain the better results than the other state-of-the-art methods. The difference between AS-Reasoner-I and AS-Reasoner-G is that the former uses intra-attention in the reasoner module and the latter uses global attention in the reasoner module. These Results verify the effect of the reason mechanism in the proposed model and our assumption. AS-Reasoner is an ensemble model of AS-Reasoner-I and AS-Reasoner-G. AS-Reasoner performs best in three datasets and achieves a

Table 3 The experiments of state-of-the-art methods on aspect term level ABSA on all Chinese datasets

Model	Camera		Car		Notebook		Phone	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
LSTM	78.31	68.72	81.99	58.83	74.63	62.32	81.38	72.13
TD-LSTM	70.48	51.46	76.53	46.67	67.10	40.58	69.17	53.40
AT-LSTM	85.05	83.44	80.09	72.34	79.34	77.99	86.41	84.46
ATAE-LSTM	85.54	84.09	81.90	76.88	83.47	82.14	85.77	83.87
MemNN	70.59	55.13	75.55	51.01	69.10	53.51	70.29	55.93
ATAM-S	82.88	72.50	82.94	64.18	75.59	60.09	84.86	75.35
ATAM-F	88.30	–	82.94	–	77.52	–	88.46	–
AS-Reasoner	<i>89.71</i>	<i>88.66</i>	<i>85.52</i>	<i>79.88</i>	<i>85.95</i>	<i>84.41</i>	<i>89.17</i>	<i>88.02</i>

Italic values indicate the significance of the best result

The metric "Acc." means the accuracy and "F1" denotes the macro-F1

comparable result in Restaurant-2016. Compared to these state-of-the-art models, AS-Reasoner achieves an average improvement of 1.07% for restaurant-2014, 2.2% for laptop-2014, 1.34% for the restaurant-2015.

We believe the first reason why AS-Reasoner wins over other state-of-the-art methods is that we explicitly design the reason module to detect the important context words towards to the specific aspect target, which plays a decisive role in deciding the polarity of the specific aspect target. Intra and global attention in the proposed model can effectively extract informative sentiment expressions and semantically similar words between the context. The intra attention pays more attention to the association relationship between any two words in a sentence. The global attention focuses on finding beneficial sentiment expressions from a sentence towards the specific aspect target. As a result, the model can learn more precise sentence vector representations.

The second reason is that we design a multi-layered architecture. Each layer in AS-Reasoner can be seen as an encoder-reasoner structure. The reasoner of the high-level layer can make use of the generated representation of the lower-level layer to gradually learn to find semantic similarity between the words in the context and obtain the more informative aspect-dependent representation.

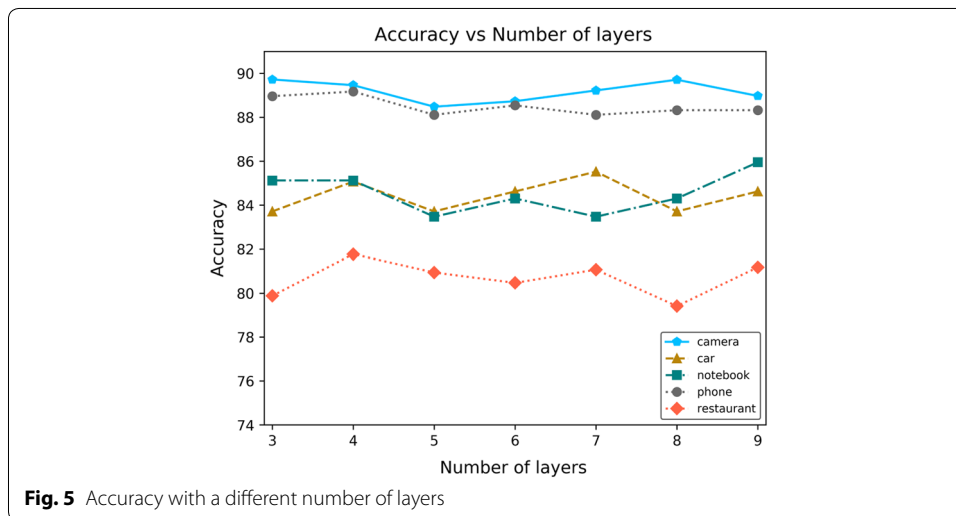
The third reason is that we propose an explicit aspect-dependent representation module, which is used to store the output of each encoder-reasoner layer and transfer these aspect-dependent representations into other layers. The specific module is essentially a residual connection between different layers, so it makes model training in the larger number of layers possible.

In order to evaluate whether our model is language-independent, we conducted extra experiments on four Chinese datasets. The results are shown in Table 3 in comparison with several of the top state-of-the-art methods, namely ATAM-S, RAM, IAN, MemNN, ATAE-LSTM, TD-LSTM and LSTM. AS-Reasoner achieves the highest accuracy and macro-F1 in all Chinese datasets by around 0.71–2.58%, 2.27–4.57%, respectively. The validity of the proposed model is further verified. The main reason is that our proposed model can continually learn the more informative aspect-dependent representation by

Table 4 Accuracy on aspect category level ABSA on the restaurant-2015 dataset

Model	Restaurant-2015
LSTM	73.50
TC-LSTM	74.70
ATAE-LSTM	75.20
IAN	75.50
RAM	76.70
Word&Clause-Level ATT	80.90
AS-Reasoner	<i>81.77</i>

Italic values indicate the significance of the best result



the multiple-layered encoder-reasoner structure and the aspect-dependent representation module.

Results of aspect category level ABSA

To further verify the generalization capacity and test whether our proposed model is effective in the related ABSA task, we conduct aspect category level ABSA experiments in SemEval 2015 restaurant dataset. The experimental results are shown in Table 4. Our proposed model achieves the best accuracy in the dataset. AS-Reasoner is 0.87% higher than the state-of-the-art approach in the aspect category level task. That indeed proves the powerful generalization capacity of the proposed model. Besides, it verifies that our assumption is correct.

Effect of layer number

In this section, we focus on the number of encoder-reasoner in the proposed model. AS-Reasoner is a multi-layered architecture, and the number of layers is a hyper-parameter in our proposed model. The accuracy of the different number of layers is illustrated in Fig. 5. In the aspect-term level, we reported the experimental results on four Chinese datasets. In the aspect-category level, we reported experimental results for the

restaurant-2015 dataset. Figure 5 shows that accuracy increases with an overall increase in the number of layers. Thus, the best accuracy is obtained when the number of layers is 9 for most scenarios.

Conclusion

This paper proposes a multiple-layered attention reasoning model, named AS-Reasoner, to alleviate the problem of how to capture precise sentiment expressions to reason in an ABSA task. AS-Reasoner is consisted of three import modules in each hop/layer: Encoder, Reasoner and Aspect-dependent Representation Memory. Encoder is designed to learn the sequence information and combine the context with an aspect. Reasoner is used to reason. Reasoner utilizes intra attention and global attention to continuously capture more informative sentiment expression words to implement reasoning. Intra attention can capture the semantic relevance of any two words in a sentence given an aspect target and then select the relevant vectors towards the specific aspect target. Global attention captures the import parts of a sentence by directly assigning attention weights, which does not consider the similarity between two words. Aspect-dependent Representation Memory is used to store these attention results and transform them into next layer for reasoning. Abundant experiments on all the datasets demonstrated that the proposed model acquires state-of-the-art results and verified that it is language-independent. In our future work, we would like to incorporate common knowledge into the proposed model. Furthermore, we would like to apply AS-Reasoner to other natural language processing tasks.

Acknowledgements

We would like to thank the reviewers for their valuable comments.

Authors' contributions

All authors read and approved the final manuscript.

Funding

This research was funded by the Fundamental Research Funds for the Central Universities grant number 2019YJS022.

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China. ² Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing, China. ³ School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China. ⁴ Beijing Huayu Software Co., Ltd, Beijing, China.

Received: 21 June 2019 Accepted: 4 September 2019

Published online: 19 September 2019

References

1. Schouten K, Flavius F (2015) Survey on aspect-level sentiment analysis. *IEEE Trans Knowl Data Eng* 28(3):813–830. <https://doi.org/10.1109/TKDE.2015.2485209>
2. Liu B (2015) Sentiment analysis: mining opinions, sentiments, and emotions. Cambridge University Press, Cambridge
3. Manek AS, Shenoy PD, Mohan MC et al (2017) Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. *World Wide Web* 20:135–154
4. Ma Y, Peng H, Cambria E (2018) Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In: Proceedings of AAAI
5. Chen T, Xu R, He Y et al (2016) Learning user and product distributed representations using a sequence model for sentiment analysis. *IEEE Comput Intell Mag* 11:34–44

6. Dong L, Wei F, Tan C et al (2014) Adaptive recursive neural network for target-dependent twitter sentiment classification. In: Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: short papers), pp 49–54
7. Zhang Z, Zou Y, Gan C (2018) Textual sentiment analysis via three different attention convolutional neural networks and cross-modality consistent regression. *Neurocomputing* 275:1407–1415
8. Tang D, Qin B, Liu T (2016) Aspect level sentiment classification with deep memory network. arXiv preprint [arXiv:160508900](https://arxiv.org/abs/160508900)
9. Tang D, Qin B, Feng X et al (2015) Effective LSTMs for target-dependent sentiment classification. arXiv preprint [arXiv:151201100](https://arxiv.org/abs/151201100)
10. Wang Y, Huang M, Zhao L (2016) Attention-based lstm for aspect-level sentiment classification. In: Proceedings of the 2016 conference on empirical methods in natural language processing, pp 606–615
11. Chen P, Sun Z, Bing L et al (2017) Recurrent attention network on memory for aspect sentiment analysis. In: Proceedings of the 2017 conference on empirical methods in natural language processing, pp 452–461
12. Ma D, Li S, Zhang X et al (2017) Interactive attention networks for aspect-level sentiment classification. arXiv preprint [arXiv:170900893](https://arxiv.org/abs/170900893)
13. Ruder S, Ghaffari P, Breslin JG (2016) A hierarchical model of reviews for aspect-based sentiment analysis. arXiv preprint [arXiv:160902745](https://arxiv.org/abs/160902745)
14. Kumar A, Irsay O, Ondruska P et al (2016) Ask me anything: dynamic memory networks for natural language processing. *Int Conf Mach Learn* 2016:1378–1387
15. Sukhbaatar S, Weston J, Fergus R (2015) End-to-end memory networks. *Adv Neural Inf Process Syst* 2015:2440–2448
16. Tay Y, Luu AT, Hui SC (2017) Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis. arXiv preprint [arXiv:171205403](https://arxiv.org/abs/171205403)
17. Tay Y, Tuan LA, Hui SC (2017) Dyadic memory networks for aspect-based sentiment analysis. In: Proceedings of the 2017 ACM conference on information and knowledge management. ACM, New York, pp 107–116
18. Li X, Bing L, Lam W et al (2018) Transformation networks for target-oriented sentiment classification. arXiv preprint [arXiv:180501086](https://arxiv.org/abs/180501086)
19. Peng H, Ma Y, Li Y et al (2018) Learning multi-grained aspect target sequence for Chinese sentiment analysis. *Knowl Based Syst* 148:167–176
20. Yang C, Zhang H, Jiang B et al (2019) Aspect-based sentiment analysis with alternating coattention networks. *Inf Process Manag* 56:463–478
21. Devlin J, Chang M-W, Lee K et al (2018) Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)
22. Sun C, Huang L, Qiu X (2019) Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. arXiv preprint [arXiv:1903.09588](https://arxiv.org/abs/1903.09588)
23. Xu H, Liu B, Shu L et al (2019) Bert post-training for review reading comprehension and aspect-based sentiment analysis. arXiv preprint [arXiv:1904.02232](https://arxiv.org/abs/1904.02232)
24. Tang F, Fu L, Yao B et al (2019) Aspect based fine-grained sentiment analysis for online reviews. *Inf Sci* 488:190–204
25. Yu J, Jiang J, Xia R (2019) Global inference for aspect and opinion terms co-extraction based on multi-task neural networks. *IEEE/ACM Trans Audio Speech Lang Process* 27:168–177
26. Wu S et al (2019) Aspect-based sentiment analysis via fusing multiple sources of textual knowledge. *Knowledge-Based.* <https://doi.org/10.1016/j.knsys.2019.104868>
27. Ma Y, Peng H, Khan T et al (2018) Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis. *Cognit Comput* 10:639–650
28. Kumar R, Pannu HS, Malhi AK (2019) Aspect-based sentiment analysis using deep networks and stochastic optimization. *Neural Comput Appl.* <https://doi.org/10.1007/s00521-019-04105-z>
29. Shuang K, Ren X, Yang Q et al (2019) AELA-DLSTMs: attention-enabled and location-aware double LSTMs for aspect-level sentiment classification. *Neurocomputing* 334:25–34
30. Zeng J, Ma X, Zhou K (2019) Enhancing attention-based LSTM With position context for aspect-level sentiment classification. *IEEE Access* 7:20462–20471
31. Zainuddin N, Selamat A, Ibrahim R (2018) Hybrid sentiment classification on twitter aspect-based sentiment analysis. *Appl Intell* 48:1218–1232
32. Ma X, Zeng J, Peng L et al (2019) Modeling multi-aspects within one opinionated sentence simultaneously for aspect-level sentiment analysis. *Future Gener Comput Syst* 93:304–311
33. Fu X, Wei Y, Xu F et al (2019) Semi-supervised aspect-level sentiment classification model based on variational autoencoder. *Knowl Based Syst* 171:81–92
34. García-Pablos A, Cuadros M, Rigau G (2018) W2V LDA: almost unsupervised system for aspect based sentiment analysis. *Expert Syst Appl* 91:127–137
35. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9:1735–1780
36. Pennington J, Socher R, Manning C (2014) Glove: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp 1532–1543
37. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
38. Srivastava N, Hinton G, Krizhevsky A et al (2014) Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 15:1929–1958
39. Wang J, Li J, Li S et al (2018) Aspect sentiment classification with both word-level and clause-level attention networks. *IJCAI* 2018:4439–4445

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.