Future Generation Computer Systems 93 (2019) 304–311

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

# Modeling multi-aspects within one opinionated sentence simultaneously for aspect-level sentiment analysis

Xiao Ma [a], Jiangfeng Zeng [b], Limei Peng [c], Giancarlo Fortino [d], Yin Zhang [a,*]

[a] School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan, China
[b] Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, China
[c] School of Computer Science and Engineering, Kyungpook National University, Daegu, Republic of Korea
[d] Department of Informatics, Modeling, Electronics and Systems, University of Calabria, Italy

## HIGHLIGHTS

- Content and position attention are involved to measure the influence of each context word on a given aspect.
- Multi-aspects are modeled within one opinionated sentence simultaneously.
- Actual experimental results empirically demonstrate the effectiveness of the proposed model for aspect-level sentiment analysis.

## ARTICLE INFO

*Article history:*
Received 8 August 2018
Received in revised form 14 September 2018
Accepted 22 October 2018
Available online 2 November 2018

*Keywords:*
Sentiment analysis
Aspect-level
Position attention
Multi-aspects

## ABSTRACT

Aspect-level sentiment analysis aims at inferring the sentiment polarity with respect to a specific aspect term in an opinionated text, and has attracted a surge of active research interest in the research community. Years of research have witnessed significant progress made in aspect-level sentiment analysis by exploiting attention mechanism to learn semantically meaningful aspect-specific representations. Although previous attention-based approaches have proven to be successful and effective for aspect-level sentiment classification, there still exist some problems not well handled in the literature. First, the explicit position context is not well explored. Second, different aspects in one opinionated sentence are processed in isolation. In other words, existing attentive methods ignore the disturbance of other aspects in the same sentence when computing the attention vector for the current aspect. Aiming to address the two issues, in this paper, we develop a two stage paradigm which can be accomplished in two steps: (1) the StageI model introduces position attention to model the explicit position context between the aspect and its context words with the goal of dealing with aspects one by one; and (2) the StageII model investigates how to model multi-aspects within one opinionated sentence all at once using the position attention mechanism. We empirically evaluate our proposed method on the SemEval 2014 datasets and encouraging experimental results turn out that the proposed approach yields a significant performance gain compared to other state-of-the-art attention-based methods.



## 1. Introduction

Aspect-level sentiment classification has garnered lots of attention in the past few years. It is a fine-grained NLP research task which aims at identifying the sentiment polarity of a specific aspect term from an opinionated text [1–3]. For instance, in the following opinionated sentence "the quality of the shoes looks nice, but the service is terrible.", the sentiment polarity of aspect "quality" is positive while the polarity of aspect "service" is negative.

* Corresponding author.
*E-mail addresses:* cindyma@zuel.edu.cn (X. Ma), jfzeng@hust.edu.cn (J. Zeng), auroraplm@knu.ac.kr (L. Peng), g.fortino@unical.it (G. Fortino), yin.zhang.cn@ieee.org (Y. Zhang).

https://doi.org/10.1016/j.future.2018.10.041
0167-739X/

Aspect-level sentiment classification has becoming a very hot topic in the research community these years, with many open problems unsolved. The traditional sentiment classification methods [4] usually assign class labels to each sentence, and ignore the different aspects that exist in the same sentence. Comparatively, aspect-level sentiment classification is much more complicated than sentence-level sentiment classification due to the difficulties of identifying the parts of sentence describing the corresponding aspects.

Although conventional machine learning methods (e.g., Support Vector Machine (SVM)) are able to solve this task very well, they heavily depend on the complicated and labor-intensive feature engineering [5,6]. As we know that, deep learning techniques are highly appreciated because they are successful and effective in automatically learning semantically meaningful representations from high dimensional raw data. In particular, recurrent neural networks (RNNs) have become the preferred choices for a variety of NLP tasks [7–9]. Researchers have investigated many RNN models for sentiment analysis including adaptive recursive neural network [10], TD-LSTM [11], etc. Recently, attention mechanism is one of the most exciting advancements in Artificial Intelligence (AI) and has been demonstrated to be very effective in plenty of applications, e.g., image generation [12,13], image caption [14], machine translation [15,16], natural language inference [17] and deep hashing [18]. Naturally, attention mechanism is introduced to deal with sentiment classification [19,20]. These attention-based models exploit attention mechanism to measure the influence of each context word on a given aspect and have proven to be successful and effective in learning semantic aspect-specific representations. In spite of the effectiveness of these methods stated above, the explicit position context between the aspect and its context words is overlooked, and different aspects in one opinionated sentence are processed in isolation. All the aforementioned facts indicate that the current state-of-the-art attention-based approaches leave space for improvement.

To address the two issues stated above, in this paper, we devise a two stage paradigm for aspect-level sentiment classification. We first assume that the context words neighboring to the aspect should be given higher attention since they are intuitively more valuable than those far away. Based on this assumption, we introduce position attention to model the position context between the aspect and the context words. On the other hand, the more aspect terms one opinionated sentence has, the more disturbance exists. It encourages us to investigate how to model multi-aspects within one opinionated sentence simultaneously on the basis of the position attention mechanism. Inspired by the work of [21], the Frobenius norm of a matrix is adopted to regularize the attention weights of all aspects in the same sentence. We first train our StageI model to process each aspect one by one. Then we train the proposed StageII model to predict more than one aspects of one review at once. As a result, our StageI model is utilized in terms of those reviews which only contain one aspect term, and the StageII model is exploited for those reviews which contain more than one aspect terms.

To summarize, the contributions of our paper can be listed as follows.

- In order to enhance the aspect-specific attentive representations, the utilization of both content attention and position attention better measures the influence of each context word on a given aspect than simply exploiting content attention.
- To further improve the performance, we model multi-aspects within one opinionated sentence simultaneously.
- Experiments on several real-world datasets are conducted to empirically demonstrate the effectiveness of our two stage approach for aspect-level sentiment analysis.

The remainder of this paper is organized as follows. In Section 2, we survey the field of aspect-level sentiment classification. Afterwards, we present the proposed two-stage framework in Section 3. In Section 4, extensive experiments are conducted to demonstrate the superiority of the proposed algorithm. Finally, we draw a conclusion in Section 5.

## 2. Related work

Sentiment analysis is a big suitcase and has been investigated for a long time. Apart from the sentence level sentiment classification, target dependent sentiment analysis (subject-based sentiment analysis) and aspect based sentiment analysis (ABSA) are emerged as fine-grained branches of sentiment analysis. Aspect based sentiment analysis (ABSA) infers sentiments in terms of the aspects of a given entity, while target-dependent sentiment analysis aims to predict the sentiment polarity of a given entity. Generally, methods suitable for target dependent sentiment analysis can be naturally applied to ABSA [11,20,22]. There are two paradigms to tackle sentiment analysis: traditional machine learning based approaches and neural network model based approaches.

Traditional machine learning based methods generally solve the problem in two steps. First, they manually extract a number of carefully handcrafted features like bag-of-words and sentiment lexicons. Then, sentiment classifiers like SVMs [5,6,23–25] are used to compute the probability distribution over all classes. Therefore, the quality of the handcrafted feature engineering decides the performance of these methods. However, these methods are limited in their ability to extract high-quality handcrafted features which are labor-intensive and complicated. By contrast, neural networks models, characterized by automatically learning semantic representations from high dimensional raw data, get rid of the complicated handcrafted feature engineering [7–10,26]. However, both the neural network based methods and traditional machine learning based approaches turn a blind eye on the objects or targets depicted in the sentences. The work of Jiang et al. [27] find that about 40% of the sentiment classification errors are caused by neglecting the importance of targets. Following his work, Poria et al. [28] extract aspects for opinion mining with a deep convolutional neural network. To address aspect based sentiment analysis, Akhtar et al. [29] extract aspect terms in original sentences before final sentiment classification.

Taking into consideration the targets described in the sentences, aspect-level sentiment analysis formally enters the academic arena. Tang et al. [11] propose TD-LSTM which first separates the whole context into two parts, i.e., left part with target and right part with target, then models the two parts with two long short-term memory networks respectively, and finally concatenates the two target-specific representations for sentiment classification. In the work of [29], aspect based sentiment analysis is performed in two steps, i.e., aspect term extraction and sentiment classification. Peng et al. [3] propose ATSM to explicitly model the aspect and conduct sentiment classification at the aspect level via three granularities: radical, character and word. Recently, assisted by attention mechanism, some models enhance the aspect-specific representations via measuring the relatedness between the aspect term and its context words. Tang et al. [22] devise MemNet which learns the impact of each context word on the given aspect via deep memory networks with multiple computational layers. Wang et al. [19] propose attention-based LSTM networks with aspect embeddings for aspect-level sentiment classification, the key idea of which is to first learn an embedding vector for each aspect and then generating aspect-specific representations for final classification via computing attention weights between an aspect and its context words. Considering that previous approaches ignore the separate modeling of aspects, recently, Ma et al. [20] put forward the interactive attention networks (IAN) which uses two attention networks to model the aspect and its context interactively, i.e., obtaining aspect-specific representations for context while learning better representations for aspects. Then, the classifier takes as input the concatenation of the two representations and achieves the state-of-the-art results. In the work of [30], Chen et al. propose a framework which adopts multiple attention mechanism to capture sentiment features separated by a long distance. Moreover, Saeidi et al. [31] and Ma et al. [32] attempt to tackle the challenges of both aspect based sentiment analysis and target dependent sentiment analysis. The latter incorporates affective commonsense knowledge into a hierarchical attention model. However, much remains to be done to obtain satisfying classification performance, e.g., taking into account the explicit position context between the aspect and its context words, and processing multi-aspects within one opinionated sentence simultaneously.

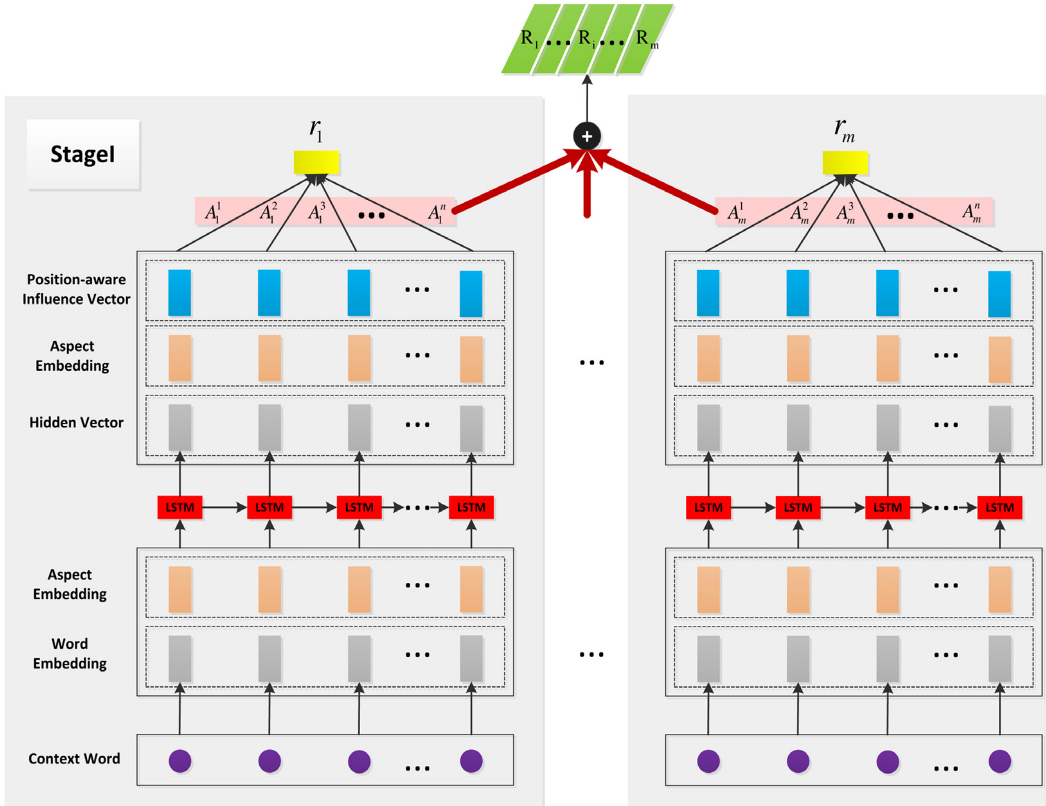


**Fig. 1.** Overview of the proposed two stage approach for aspect-level sentiment classification. $A_1 = \{A_1^1, A_1^2, \cdots, A_1^n\}$ means the attention weights for the first aspect in the sentence computed by our position attention. Note that StageI model and StageII model share almost the network architecture except that the attention weights of StageII are regulated by using a new penalization term whose purpose is to diminish the disturbance of other aspects in the same sentence when addressing all the aspects at one time.

## 3. The proposed approach

The proposed two stage approach for aspect-level sentiment classification is depicted in this section. Given one sentence including $m$ aspects, a high-level illustration of our proposed two stage framework for the $i$th aspect of the sentence is shown in Fig. 1. Note that our StageI model and StageII model share almost the network architecture. Due to the use of a new penalization term to remove interruptions from other aspects in the same sentence, StageII model can generate better representations for aspects. First, we give a formal definition of the research problem. Then, we briefly provide background on Long Short-Term Memory networks (LSTMs), and adopt the long short-term memory network for sentence modeling. Afterwards, two unsolved challenging issues are tackled: (1) modeling the explicit position context between the aspect and its context words, and (2) processing multi-aspects within one opinionated sentence simultaneously. Finally, we present the training details.

### 3.1. Task definition

Suppose a review $S$ has $n$ words and $m$ aspect terms. The task of aspect-level sentiment classification aims to predict a sentiment category for a (aspect, sentence) pair. For example, the sentence "Our [waiter] was friendly, but it is a shame that he did not have a supportive [staff] to work with." contains two aspect terms: [waiter] and [staff]. The expected outputs for [waiter] and [stuff] are **positive** and **negative**, respectively. Generally, according to this review two (aspect, sentence) pairs are obtained as independent instances, and then the classifier tries to predict a sentiment category for each instance in isolation. Our StageI model processes a review with multi-aspects one by one While the proposed StageII model attempts to regard the (aspect, sentence) pairs as one instance and predicts the sentiment polarities for the two aspect terms simultaneously.

### 3.2. Long Short-Term Memory (LSTM)

Recurrent neural networks (RNNs) [7,33,34] have become a cornerstone to process sequential data for many natural language processing (NLP) tasks. With the emergence of Long Short-Term Memory networks (LSTMs), the long-term dependencies are captured so that the gradient vanishing/exploding problem faced by RNNs is effectively handled [35]. The standard LSTM consists of three gates (i.e., input gate, forget gate and output gate) which are designed to let valuable information flow into the cell and discard the useless information in every timestep. Suppose we have a sentence which has $n$ words. For simplicity in notation, the current input word vector $w_t \in R^d$ vectorizes the $t$th word of the sentence. Then the current cell state $c_t$ and current hidden value $h_t$ of the standard LSTM can be updated as follows:

$$i_t = \sigma(W_i \cdot [w_t, h_{t-1}] + b_i) \quad (1)$$

$$f_t = \sigma(W_f \cdot [w_t, h_{t-1}] + b_f) \quad (2)$$

$$o_t = \sigma(W_o \cdot [w_t, h_{t-1}] + b_o) \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c \cdot [w_t, h_{t-1}] + b_c) \quad (4)$$

$$h_t = o_t \odot \tanh(c_t) \quad (5)$$

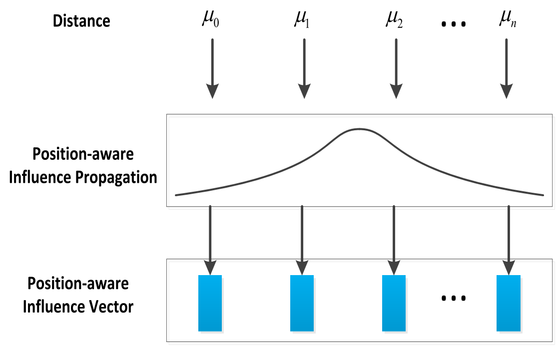


**Fig. 2.** The generation of position-aware influence vectors using the Gaussian kernel.

where $i$, $f$ and $o$ represent input gate, forget gate and output gate respectively, $\sigma$ in the equations is the sigmoid function used to control the in and out of information in each iteration, and $\{W_i, W_f, W_o, W_c, b_i, b_f, b_o, b_c\}$ are the parameters to be learned during training. The dimension of hidden values are equal to that of word vector.

Some LSTM-based approaches consist in creating a simple vector representation for the sentence by using the final hidden vector of the LSTM or the max (or average) pooling from all the LSTM hidden states. It is worth noting that the comparison model only using LSTM in Section 4 regards the last hidden vector as the sentence representation which is input into a multi-class classifier.

Most previous studies [11,22] treat the average of all the aspect or target word representations as the representation of the aspect or target, which fails to represent the aspects accurately and results in generating bad aspect-specific representations via attention mechanism for final sentiment classification. Motivated by word embedding, Wang et al. [19] devise aspect embedding to automatically learn representations for all aspect terms by the neural network. In other words, aspect embeddings are regarded as model parameters to be learned during training. In our case, LSTM takes both the context word vector and the aspect embedding as input, which has empirically proven to be effective by [19]. Thus, in this paper aspect embeddings are utilized to guide the attention mechanism to find more information related to the aspect. The equations computing $i_t, f_t, o_t, c_t$ are replaced as follows:

$$i_t = \sigma(W_i \cdot [w_t, v_a, h_{t-1}] + b_i) \tag{6}$$

$$f_t = \sigma(W_f \cdot [w_t, v_a, h_{t-1}] + b_f) \tag{7}$$

$$o_t = \sigma(W_o \cdot [w_t, v_a, h_{t-1}] + b_o) \tag{8}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c \cdot [w_t, v_a, h_{t-1}] + b_c) \tag{9}$$

where $v_a \in R^{d_a}$ comes from the aspect embeddings matrix $M$. $\Theta^{lstm} = \{W_i \in R^{d\times(2d+d_a)}, W_f \in R^{d\times(2d+d_a)}, W_o \in R^{d\times(2d+d_a)}, W_c \in R^{d\times(2d+d_a)}, b_i \in R^d, b_f \in R^d, b_o \in R^d, b_o \in R^d\}$ are denoted as the model parameters.

### 3.3. StageI: Position attention

In the first stage, we introduce our position attention based on the intuition that the polarity of a given aspect is decided to a large extent by its neighboring context words and influenced little by the faraway context words. This hypothesis conforms to the normal rules of criticism in natural language. The StageI model is designed to address aspects contained in the sentence one by one. Inspired by the remarkable improvements achieved by using the position context in information retrieval (IR) [36] and question answering (QA) [37], the Gaussian kernel is used to compute the position-aware influence propagation:

$$Kernel(\mu) = \exp(\frac{-\mu^2}{2\gamma^2}) \tag{10}$$

where $\mu$ means the distance between the current context word and the given aspect, and $\gamma$ represents the propagation scope. Obviously, the farther the distance is, the faster the position-aware influence diminishes. It is assumed that the influence of a specific distance obeys the Gaussian distribution over each dimension. Thus the influence is extended into a matrix noted as $P$. The influence in the $i$th dimension as to the distance of $\mu$ is computed as:

$$P(i, \mu) \sim N(Kernel(\mu), \sigma') \tag{11}$$

where $P(i, \mu)$ follows the normal distribution with mean value of $Kernel(\mu)$ and standard deviation of $\sigma'$. Note that each column of $P$ denotes the influence vector corresponding to a specific distance. For simplicity in notation, we denote $p_j \in R^{d_p}$ from $P$ as the influence vector for the $j$th context word with a distance of $\mu$, and $d_p$ is the dimension of position vector. Fig. 2 clearly shows the generation of position-aware influence vectors.

After we obtain the position-aware influence vectors, position attention weights are computed on top of the LSTM. Specifically, in order to infer for the $i$th aspect term of the review, by offering its aspect embedding $v_a^i$, the position-aware vector $p_j$ and the hidden vectors $[h_1, h_2, \ldots, h_n]$ of the context words, the attention distribution corresponding to the context word at position $j$ in the sentence can be denoted as:

$$A_i^j = \frac{\exp(e(h_j, v_a^i, p_j))}{\sum_{k=1}^{n} \exp(e(h_k, v_a^i, p_k))} \tag{12}$$

$$e(h_j, v_a^i, p_j) = \eta^T \tanh(W_h h_j + W_p p_j + W_a v_a^i + b) \tag{13}$$

where $e(h_j, v_a^i, p_j)$ is a score function that calculates the semantic relatedness between the $j$th context word and the given aspect. $\Theta^{att} = \{W_h \in R^{d\times d}, W_p \in R^{d\times d_p}, W_a \in R^{d\times d_a}, b \in R^d, \eta \in R^d\}$ are defined as the training parameters. Therefore, the aspect-specific sentence representation is obtained:

$$r_i = \sum_{j=1}^{n} A_i^j h_j \tag{14}$$

Finally, the aspect-specific attentive representation $r$ is mapped into the target space of $C$ classes:

$$\hat{r}_i = \tanh(W_r r_i + b_r) \tag{15}$$

where $\Theta^{(classifier)} = \{W_r \in R^{d\times C}, b_r \in R^C\}$ are the model parameters. Afterwards, a softmax layer is used to compute the sentiment distribution:

$$g_c = \frac{\exp(\hat{r}_i^c)}{\sum_{z=1}^{C} \exp(\hat{r}_i^z)} \tag{16}$$

### 3.4. StageII: Modeling multi-aspects simultaneously

Considering the disturbance among the multiple aspects, in this stage, we try to model multi-aspects all at once. Drawing inspirations from [21], we adopt the Frobenius norm to construct the penalization term which is similar to the $L_2$ regularization term. $A = \{A_1, A_2, \cdots, A_m\}$ is a matrix constructed by the attention weights of $m$ aspects in the sentence, and $A_i = \{A_i^1, A_i^2, \cdots, A_i^n\}$ means the attention probability distribution for the $i$th aspect in the sentence computed by our position attention aforementioned.

**Table 1**
Statistics of SemEval 2014 Datasets.

| Dataset | Positive | | Negative | | Neural | |
|---|---|---|---|---|---|---|
| | Train | Test | Train | Test | Train | Test |
| Laptop | 994 | 341 | 870 | 128 | 464 | 169 |
| Restaurant | 2164 | 728 | 807 | 196 | 637 | 196 |

**Table 2**
Statistics of aspect number in each sentence on SemEval 2014 Datasets.

| Dataset | 1 | 2 | 3 | More |
|---|---|---|---|---|
| Laptop-Train | 887 | 288 | 93 | 35 |
| Laptop-Test | 240 | 71 | 23 | 8 |
| Restaurant-Train | 978 | 485 | 221 | 99 |
| Restaurant-Test | 286 | 163 | 61 | 31 |

In order to make each aspect focus on different parts of the sentence and diminish the disturbance among all the aspects, the penalization term used here is denoted as:

$$P = \left\|(AA^T - I)\right\|_F^2 \tag{17}$$

where F means the Frobenius norm and I is the identity matrix. In this stage, we continue to finetune the pre-trained model in the first stage with this penalization term. Consequently, the $i$th aspect-specific attentive representation is computed as:

$$R_i = \sum_{j=1}^{n} A_i^j h_j \tag{18}$$

where $A_i^j$ is computed similar to $\alpha_j$. Then the aspect-specific attentive representations are input into the classifier.

### 3.5. Model training

In our implementation, all the parameters of our approach are notated as $\Theta = \{\Theta^{lstm}, \Theta^{att}, \Theta^{(classifier)}, M\}$. Note that $M$ is the aspect embeddings to be learned. We choose cross entropy with $L_2$ regularization as the loss function to optimize the model:

$$L = -\sum_{e \in E} \sum_{c=1}^{C} y_c(e) \cdot \log(g_c(e)) + \lambda L_2(\Theta) \tag{19}$$

where $E$ is the dataset, $e$ is one sample, $y_c(e)$ is the golden sentiment distribution and $\lambda$ is the coefficient for $L_2$ regularization. It is worth noting that the first stage model and the second stage model share the network architecture in common but the penalization term using the Frobenius norm is exploited to finetune the second stage model.

## 4. Experiments

### 4.1. Experimental settings

To evaluate our proposed method, we conduct experiments on two open datasets from SemEval 2014,[1] i.e., Laptop and Restaurant. Each dataset is split into train and test set. Table 1 summarizes the statistics of the SemEval 2014 Datasets, and the number in table means the number of reviews for training and test in each sentiment category. For example, the Laptop training dataset has 994 reviews labeled with "Positive", 870 reviews labeled with "Negative" and 464 reviews labeled with "Neural". To demonstrate the necessity of modeling more than one aspect, we calculate the aspect number in each sentence on SemEval 2014 Datasets, which is shown in Table 2. It is observed that nearly half of the sentences include more than one aspect and the number of reviews which contain more than 3 aspects in one sentence is very low. According to this observation, the second stage model is designed to address two aspects one time. By contrast, the first stage model regards multiple aspects in one sentence as independent instances just as done by all previous approaches.

In our implementation, we initialize the words using the GloVe 300-dimensional word embeddings [38], and initialize each aspect to be a 300-dimensional zero vector. The dimension of position-aware influence vectors is also set to be 300. It is empirically suggested from the work of [37] that the propagation scope $\gamma$ in Eq. (10) should be assigned to be 25, and the standard deviation $\sigma'$ in Eq. (11) should be assigned to be 0.1. The first stage model, PosATT-LSTM, adopts a batch size of 25 to model instances one by one, while the second stage model use a batch size of 12 to model two aspects one time. In addition, we employ $L_2$-regularization weight of 0.00001 and initial learning rate of 0.05 for AdaDelta.

Our work choose classification accuracy to be the evaluation metric the same as other studies. *Accuracy* measures the overall sentiment classification performance, and can be formalized as:

$$Accuracy = \frac{T}{N} \tag{20}$$

where $T$ is the number of samples correctly predicted and $N$ is the total number of test dataset.

### 4.2. Baselines

The following baseline methods are listed to compare with our two stage approach.

- **Majority:** The sentiment labels of the test dataset are simply decided by the majority sentiment polarity in the training dataset.
- **LSTM:** Regardless of the importance of the aspects, [19] models sentences using the LSTM network, and regards the last hidden vector as the representation of a sentence which is input to a classifier for final classification.
- **TD-LSTM:** [11] develops a left-directed LSTM and a right-directed LSTM, each of which models the left/right context with aspect respectively. Taking as input the concatenation of the left and right aspect-dependent representations, the classifier predicts the sentiment polarity of the aspect.
- **TC-LSTM** [11] is structurally similar to TD-LSTM [11] but takes as input the combination of aspect vector (average over multiple word vectors) and input word embedding. Then the classifier predicts the sentiment polarity of the aspect relying on the concatenation of the left and right aspect-dependent representations.
- **AE-LSTM:** [19] introduces the aspect embeddings which are served as the training parameters to be learned. It combines the hidden vectors produced by the LSTM with aspect embeddings for classification.
- **ATAE-LSTM:** [19] adds the aspect embedding into each word embedding to further strengthen the effect of aspect information. Then on top of the LSTM layer, the relevance between the aspect and its context words is measured using content attention.
- **MemNet** [22] captures importances of context words with respect to the given aspect term with multiple computational hops, each of which is a neural attention model over an external memory.
- **IAN:** Considering the aspect terms often contains multiple words, it is necessary to generate semantical representations for these aspect terms. Thus, [20] not only exploits an attention-based LSTM network to generate representations

[1] alt.qcri.org/semeval2014/task4/index.php?id=data-and-tools.


**Fig. 3.** Case Study: Heatmaps of reviews for aspect level sentiment classification. The deeper the color is, the higher the weight scores.

**Table 3**
Sentiment classification accuracy results of our models against competitor models on Laptop and Restaurant. The results of ours are in bold.

| Methods | Laptop | Restaurant |
|---|---|---|
| Majority | 0.650 | 0.535 |
| LSTM | 0.665 | 0.743 |
| TD-LSTM | 0.681 | 0.756 |
| TC-LSTM | 0.682 | 0.760 |
| AE-LSTM | 0.689 | 0.762 |
| ATAE-LSTM | 0.687 | 0.772 |
| MemNet | 0.703 | 0.781 |
| IAN | 0.721 | 0.786 |
| StageI | **0.728** | **0.794** |
| StageI + StageII | **0.731** | **0.801** |

for the context, but also uses another attention-based LSTM network to generate representations for aspects. The output of each LSTM network is obtained via interactive attention. Then the aspect representation and its corresponding context representation are concatenated to infer the polarity of the aspect.

### 4.3. Model comparisons

In order to show the effectiveness of our position attention, we first conduct experiments using our stageI model which processes aspects in one sentence one by one. Then to demonstrate that the disturbance among the multiple aspects is diminished, we conduct experiments using our two stage framework which first pre-trains the StageI model to infer polarities for those test cases where one sentence has one aspect, and then finetunes the parameters of stageI to process two aspects at a time with the goal of inferring polarities for those test cases where one sentence contains more than one aspects.

Table 3 shows the classification accuracy results of our models compared with other competitive models. We can clearly concluded from the results that: (1) LSTM based approaches have shown huge superiority over Majority because they are good at generating semantical feature representations without handcrafted feature engineering; (2) Because of taking into the importance of aspects, TD-LSTM, TC-LSTM, AE-LSTM, ATAE-LSTM, MemNet, IAN and our proposed methods outperform the basic LSTM approach largely; (3) Attention-based methods beat non-attentive methods for attention mechanism is exploited to obtain semantically meaningful aspect-specific representations for final classification; (4) Compared with the state-of-the-art IAN, our position attention brings absolute increments of 0.7% and 0.8% on the Laptop dataset and Restaurant dataset respectively; (5) Compared with the state-of-the-art IAN, our two stage approach achieves absolute increments of 1.0% and 1.5% on the Laptop dataset and Restaurant dataset respectively, which proves that the scheme modeling two aspects at one time takes effects.

### 4.4. Case study

To have an intuitive understanding of our proposed approach, we draw heat maps for one sentence with three aspects, i.e., “service”, “food”, “price” in the opinionated sentence “although the restaurant is blamed for its awful service, the reason why I always come to this restaurant is that the food are clearly among the best in the city and the price is affordable.” Fig. 3 presents the attention weights measured by our two stage approach. It can be observed that different from the attention distributions of other methods, our two stage method not only concerns much more around the given aspect term but also forces the attention mechanism to focus on several words. For example, the polarity of aspect “food” is mainly decided by only two words (“clearly” and “best”).

## 5. Conclusion

In this paper, we develop a two stage paradigm which first introduces position attention to model the position context between the aspect and its context words and then investigates how to model multi-aspects within one opinionated sentence simultaneously using the position attention mechanism. To be specific, the first stage model is used to deal with aspects one by one while the second stage model is used to process multi-aspects within the sentence all at once. Finally, extensive experiments conducted on well-known datasets have demonstrated that our two stage approach which models multi-aspects simultaneously is capable of diminishing the disturbance among the multi-aspects and making different aspects attend to different parts of the sentence.

## Acknowledgments

We are grateful to the anonymous reviewers for their helpful and thought-provoking suggestions and feedbacks. This work was supported in part by the National Natural Science Foundation of China under grants No.61802440, No.61232004, No.61502189 and No.61672254, and the National Key Research and Development Program of China No.2016YFB0800402.

## References

[1] B. Pang, L. Lee, Opinion mining and sentiment analysis, Found. Trends Inf. Retr. 2 (1–2) (2007) 1–135.
[2] B. Liu, Sentiment Analysis and Opinion Mining, in: Synthesis Lectures on Human Language Technologies, vol. 5, Morgan & Claypool Publishers, 2012.
[3] H. Peng, Y. Ma, Y. Li, E. Cambria, Learning multi-grained aspect target sequence for chinese sentiment analysis, Knowl.-Based Syst. 148 (2018) 167–176.
[4] K. Ravi, V. Ravi, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, Knowl.-Based Syst. 89 (2015) 14–46.
[5] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, in: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, vol. 10, 2002, pp. 79–86.
[6] T. Mullen, N. Collier, Sentiment analysis using support vector machines with diverse information sources, in: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 2004, pp. 412–418.
[7] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, S. Khudanpur, Recurrent neural network based language model, in: INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, 2010, pp. 1045–1048.
[8] R. Socher, A. Perelygin, J.Y. Wu, J. Chuang, C.D. Manning, A.Y. Ng, C. Potts, Recursive deep models for semantic compositionality over a sentiment treebank, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 2013, pp. 1631–1642.

[9] K.S. Tai, R. Socher, C.D. Manning, Improved semantic representations from tree-structured long short-term memory networks, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, 2015, pp. 1556–1566.
[10] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, K. Xu, Adaptive recursive neural network for target-dependent twitter sentiment classification, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014, pp. 49–54.
[11] D. Tang, B. Qin, X. Feng, T. Liu, Effective lstms for target-dependent sentiment classification, in: COLING 2016, 26th International Conference on Computational Linguistics, 2016, pp. 3298–3307.
[12] V. Mnih, N. Heess, A. Graves, K. Kavukcuoglu, Recurrent models of visual attention, in: Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, 2014, pp. 2204–2212.
[13] K. Gregor, I. Danihelka, A. Graves, D.J. Rezende, D. Wierstra, DRAW: A recurrent neural network for image generation, in: Proceedings of the 32nd International Conference on Machine Learning, 2015, 1462–1471.
[14] K. Xu, J. Ba, R. Kiros, K. Cho, A.C. Courville, R. Salakhutdinov, R.S. Zemel, Y. Bengio, Show, attend and tell: Neural image caption generation with visual attention, in: Proceedings of the 32nd International Conference on Machine Learning, 2015, pp. 2048–2057.
[15] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, CoRR abs/1409.0473.
[16] T. Luong, H. Pham, C.D. Manning, Effective approaches to attention-based neural machine translation, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 1412–1421.
[17] T. Rocktäschel, E. Grefenstette, K.M. Hermann, T. Kociský, P. Blunsom, Reasoning about entailment with neural attention, CoRR abs/1509.06664.
[18] K. Zhou, J. Zeng, Y. Liu, F. Zou, Deep sentiment hashing for text retrieval in social ciot, Future Gener. Comput. Syst. 86 (2018) 362–371.
[19] Y. Wang, M. Huang, X. Zhu, L. Zhao, Attention-based LSTM for aspect-level sentiment classification, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 606–615.
[20] D. Ma, S. Li, X. Zhang, H. Wang, Interactive attention networks for aspect-level sentiment classification, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017, pp. 4068–4074.
[21] Z. Lin, M. Feng, C.N. dos Santos, M. Yu, B. Xiang, B. Zhou, Y. Bengio, A structured self-attentive sentence embedding, CoRR abs/1703.03130.
[22] D. Tang, B. Qin, T. Liu, Aspect level sentiment classification with deep memory network, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 214–224.
[23] N. Kaji, M. Kitsuregawa, Building lexicon for sentiment analysis from massive collection of HTML documents, in: Proceedings of the 2007 Conference on Empirical Methods in Natural Language Processing, 2007, pp. 1075–1083.
[24] D. Rao, D. Ravichandran, Semi-supervised polarity lexicon induction, in: 12th Conference of the European Chapter of the Association for Computational Linguistics, 2009, pp. 675–682.
[25] V. Pérez-Rosas, C. Banea, R. Mihalcea, Learning sentiment lexicons in spanish, in: Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, 2012, pp. 3077–3081.
[26] Q. Qian, B. Tian, M. Huang, Y. Liu, X. Zhu, X. Zhu, Learning tag embeddings and tag-specific composition functions in recursive neural network, in: Proceedings of the 53th Annual Meeting of the Association for Computational Linguistics, 2015, pp. 1365–1374.
[27] L. Jiang, M. Yu, M. Zhou, X. Liu, T. Zhao, Target-dependent twitter sentiment classification, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, 2011, pp. 151–160.
[28] S. Poria, E. Cambria, A.F. Gelbukh, Aspect extraction for opinion mining with a deep convolutional neural network, Knowl.-Based Syst. 108 (2016) 42–49.
[29] M.S. Akhtar, D.K. Gupta, A. Ekbal, P. Bhattacharyya, Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis, Knowl.-Based Syst. 125 (2017) 116–135.
[30] P. Chen, Z. Sun, L. Bing, W. Yang, Recurrent attention network on memory for aspect sentiment analysis, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 452–461.
[31] M. Saeidi, G. Bouchard, M. Liakata, S. Riedel, Sentihood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods, in: COLING 2016, 26th International Conference on Computational Linguistics, 2016, pp. 1546–1556.
[32] Y. Ma, H. Peng, E. Cambria, Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
[33] J.L. Elman, Finding structure in time, Cognitive Sci. 14 (2) (1990) 179–211.
[34] M. Schuster, K.K. Paliwal, Bidirectional recurrent neural networks, IEEE Trans. Signal Process. 45 (11) (1997) 2673–2681.
[35] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[36] B. Liu, X. An, J.X. Huang, Using term location information to enhance probabilistic information retrieval, in: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2015, pp. 883–886.
[37] Q. Chen, Q. Hu, J.X. Huang, L. He, W. An, Enhancing recurrent neural networks with positional attention for question answering, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, pp. 993–996.
[38] J. Pennington, R. Socher, C.D. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1532–1543.

**Xiao Ma** is an Assistant Professor of the School of Information and Safety Engineering, Zhongnan University of Economics and Law (ZUEL), China. She got her Ph.D. from Huazhong University of Science and Technology in 2017. During the period 2015.10–2017.3, she was visiting the University of Illinois at Urbana-Champaign. She has published more than 10 prestigious conference and journal papers. Her research interests include Recommendation Systems, Data Mining, Machine Learning et al.

**Jiangfeng Zeng** is currently pursuing the Ph.D. degree at Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, China. His current research interests include machine learning, natural language processing, sentiment analysis, deep hashing and image generation.

**Limei Peng** received the M.S. and Ph.D. degrees from the Chonbuk National University, South Korea, in 2006 and 2010, respectively. In 2011, she was a Research Professor with Grid Middleware Research Center, Korea Advanced Institute of Science and Technology, South Korea. She has been an Associate Professor with the School of Electronic and Information Engineering, Soochow University, China, for more than two years. She has been an Assistant Professor with the Department of Industrial Engineering, Ajou University, South Korea. She is now an assistant professor with the School of Computer and Engineering, Kyungpook National University, South Korea. Her research interests fall in optical communication networks and protocols, datacenter networks, software de_ned networks, and cloud computing networks.

**Giancarlo Fortino** received the Laurea (B.S and M.S) and Ph.D. in computer engineering from the University of Calabria, Italy, in 1995 and 2000, respectively. He is currently an Associate Professor of Computer Engineering (since 2006) with the Department of Informatics, Modeling, Electronics and Systems (DIMES), University of Calabria. He holds the scientific national Italian habilitation for Full Professor and he is also Adjunct Full Professor of Computer Engineering at Wuhan University of Technology in the framework of High-End Foreign Experts in China and adjunct senior researcher at Italian National Research Council. He authored about 300 publications in journals, conferences, and books. His research interests include Internet of Things, body area networks, wireless sensor networks, software agents, and distributed and cloud computing. He is currently an associate editor of IEEE Trans. on Affective Computing, IEEE Trans. on Human-Machine Systems, Information Fusion, Engineering Application of Artificial Intelligence, Journal of Network and Computer Applications. He is the chair of the Italian chapter of IEEE SMC society and founder co-chair of the IEEE SMC TC on Interactive and Wearable Computing and Devices. He is co-founder and CEO of SenSysCal S.r.l., a spin-off of University of Calabria, engaged in advanced applied research and development of innovative IoT systems.

**Yin Zhang** is an Associate Professor of the School of Information and Safety Engineering, Zhongnan University of Economics and Law (ZUEL), China. He is a Wenlan Distinguished Scholar at ZUEL and a Chutian Distinguished Scholar, Hubei, China. He is Vice-chair of IEEE Computer Society Big Data STC. He serves as editor or associate editor for IEEE Access, IEEE Sensors Journal, Journal of Information Processing Systems, etc. He is a Guest Editor for Future Generation Computer Systems, IEEE IoT Journal, Mobile Networks and Applications, Sensors, Sensors, Multimedia Tools and Applications, Wireless Communications and Mobile Computing, Electronic Markets, Journal of Medical Systems, New Review of Hypermedia and Multimedia, etc. He also served as Track Chair of IEEE CSCN 2017, TPC Co-Chair of CloudComp 2015 and TRIDENTCOM 2017, etc. He has published more than 80 prestigious conference and journal papers, including 8 ESI Highly Cited Papers. He is an IEEE Senior Member since 2016. He got the Systems Journal Best Paper Award of the IEEE Systems Council in 2018. His research interests include intelligent service computing, big data, social network, etc.