Modeling Inter-Aspect Dependencies for Aspect-Based Sentiment Analysis

Devamanyu Hazarika*
School of Computing,
National University of Singapore
devamanyu@comp.nus.edu.sg

Soujanya Poria*
Artificial Intelligence Initiative,
A*STAR, Singapore
sporia@ihpc.a-star.edu.sg

Prateek Vij
Department of Computer Science
and Engineering, IIT, Guwahati
v.prateek@iitg.ernet.in

Gangeshwar Krishnamurthy
Artificial Intelligence Initiative,
A*STAR, Singapore
gangeshwark@ihpc.a-star.edu.sg

Erik Cambria
School of Computer Science and
Engineering, NTU, Singapore
cambria@ntu.edu.sg

Roger Zimmermann
School of Computing,
National University of Singapore
rogerz@comp.nus.edu.sg

Abstract

Aspect-based Sentiment Analysis is a fine-grained task of sentiment classification for multiple aspects in a sentence. Present neural-based models exploit aspect and its contextual information in the sentence but largely ignore the inter-aspect dependencies. In this paper, we incorporate this pattern by simultaneous classification of all aspects in a sentence along with temporal dependency processing of their corresponding sentence representations using recurrent networks. Results on the benchmark SemEval 2014 dataset suggest the effectiveness of our proposed approach.

1 Introduction

Aspect-based Sentiment Analysis (ABSA) is a fine-grained task of sentiment classification. Sentimentally involved sentences in reviews, debates, etc., often comprise of multiple aspects that have varied sentiment polarities. An important subtask of ABSA is aspect or aspect-term classification which involves predicting sentiment of aspects embodied in a sentence (Young et al., 2017). Present works in the literature approach this task by analyzing associations between aspects and their contexts provided in the sentence. In this work, we argue that to classify an aspect into sentiment categories, knowledge of surrounding aspects, their sentiment orientation, and resulting inter-dependencies, is beneficial.

Inter-aspect dependencies abound in sentences with multiple aspects. Largely ignored in present literature, these dependencies may reveal themselves in many forms, such as a) *Incomplete information*, where a certain aspect does not contain enough contextual information to convey the sentiment. In such cases, the surrounding aspects and their sentiment tone become crucial to fill the contextual gap. As an example, in the sentence *The menu is very limited - I think we counted 4 or 5 entries.*, the subsentence *I think ... entries* containing aspect *entries* does not provide the required sentiment unless considered with the aspect *menu*. Here, the negative sentiment of *menu* induces *entries* to have the same sentiment. b) *Sentiment influence in conjunctions*, in which, the sentiment of an aspect in a sentence influences the succeeding aspects due to the presence of conjunctions. In particular, for sentences containing conjunctions like *and, not only, also, but, however, though,* etc., aspects tend to share/contrast their sentiments. In the sentence *Food is usually very good, though I wonder about freshness of raw vegetables*, the aspect *raw vegetables* does not have any sentiment marker linked to it. However, the positive sentiment of *food* due to the word *good* and presence of conjunction *though* determines the sentiment of *raw vegetables* to be negative. Thus, aspects when arranged as a sequence, reveal high correlation and interplay of sentiments.

In this paper, we facilitate such phenomena by proposing a neural network where the information is shared among the aspects by means of a Long Short-Term Memory (LSTM) network (Hochreiter and Schmidhuber, 1997). In other words, we model the sequential relationship between the aspects as per their occurrence in the sentence.

* means authors contributed equally.

Proceedings of NAACL-HLT 2018, pages 266–270
New Orleans, Louisiana, June 1 - 6, 2018.

Specifically, our model first takes a sentence along with all of its aspect-terms and then generates the sentential representations relative to each aspect to get better aspect-oriented features (Tang et al., 2016a). This is done using an attention-based LSTM network, where the attention mechanism enables the model to focus on key parts of the sentence that modulate the sentiment of the aspects. To further guide the attention process the model incorporates aspect information at the word-level by concatenating aspect representations with each word (Wang et al., 2016). Finally, to capture the inter-aspect dependencies, the aspect-based sentential representations are ordered as a sequence and temporally modeled using another LSTM. Each timestep of this LSTM corresponds to a particular aspect. The hidden state output for each timestep is then projected to a dense layer and fed to a softmax classifier to predict the polarities of the corresponding aspect. To the best of our knowledge, use of inter-aspect dependencies in neural models is unprecedented and fills a significant gap in the literature.

In the remaining paper, Section 2 first provides a summary of existing works; Section 3 then describes the proposed approach in detail; Section 4 gives training and dataset details followed by results and a qualitative case study. Finally, Section 5 concludes the paper.

2 Related Works

Traditional methods in this field leveraged sentiment lexicons to solve this task (Rao and Ravichandran, 2009; Perez-Rosas et al., 2012) whereas present methods have transitioned to neural-based approaches. Tang et al. 2016a introduced the idea of aspect-based sentential representations which generates a custom representation of the sentence based on the aspect. This approach has been heavily adapted by modern works. Wang et al. 2016 built on this framework and introduced attention mechanism for generating these sentential features. They also incorporated aspect information into the attention module by concatenating them with the words. More recently, Ma et al. 2017 proposed a model where both context and aspect representations interact with each other's attention mechanism to generate the overall representation. Tay et al. 2017 proposed word-aspect associations using circular correlation as an improvement over Wang et al.'s work. ABSA has also been approached from a question-answering perspective where memory networks have played a major role (Tang et al., 2016b; Li et al., 2017). Our work is different from all these works since we train all aspects of a particular sentence together and capitalize on inter-aspect dependency modeling which they ignore.

3 Proposed Approach

Let us take a sentence $S = [w^1, ..., w^n]$ having n words. Each word is represented as a low-dimensional real-valued vector of size d_{em}, called word embedding. To get the embeddings, we use the pre-trained Glove vectors (Pennington et al., 2014) having $d_{em} = 300$. We can thus represent S as a matrix of dimensions $\mathbb{R}^{d_{em} \times n}$.

The sentence S also contains m aspect-terms (or aspects), where for each $i \in [1, m]$, aspect A_i is a multi-word subsequence of S, i.e., $\exists j \in [1, n]$, such that, $A_i = [w^j, ..., w^{j+|A_i|-1}] \in \mathbb{R}^{d_{em} \times |A_i|}$. All the aspects $A_1, ..., A_m$ are enumerated as per their order of occurrence in the sentence. The goal is to determine the sentiment label for each of these m aspects belonging to S.

The proposed model comprises two distinct phases (Figure 1). The first phase involves the generation of aspect-based sentential representations $s_1, ..., s_m$, where, vector s_i is created by coupling aspect A_i with sentence S. The second phase models the inter-aspect dependencies in a sentence using an LSTM which is followed by the sentiment prediction for all the aspects.

3.1 Phase 1: Aspect-based sentential representations

Below, we describe the methodology to generate the i^{th} aspect-based sentential representation s_i for aspect A_i and sentence S.

Given sentence S and aspect-term A_i, the model first generates the aspect representation t_i. This is done by passing A_i through an LSTM, named *LSTM*$_a$, having internal dimension d_a. *LSTM*$_a$'s final hidden state vector $h_a^{|A_i|} \in \mathbb{R}^{d_a}$ is taken to be this representation, i.e., $t_i = h_a^{|A_i|}$.

Following this, an attention-based LSTM model is used to create s_i using S and t_i (Wang et al., 2016). First, each word vector w^j in S is concatenated with aspect t_i to create a comprehensive feature vector $x_i^j = (w^j; t_i) \in \mathbb{R}^{(d_{em}+d_a)}$, where ; is the concatenation operator. We then take this new sequence representation $X_i = [x_i^1, ..., x_i^n]$ and

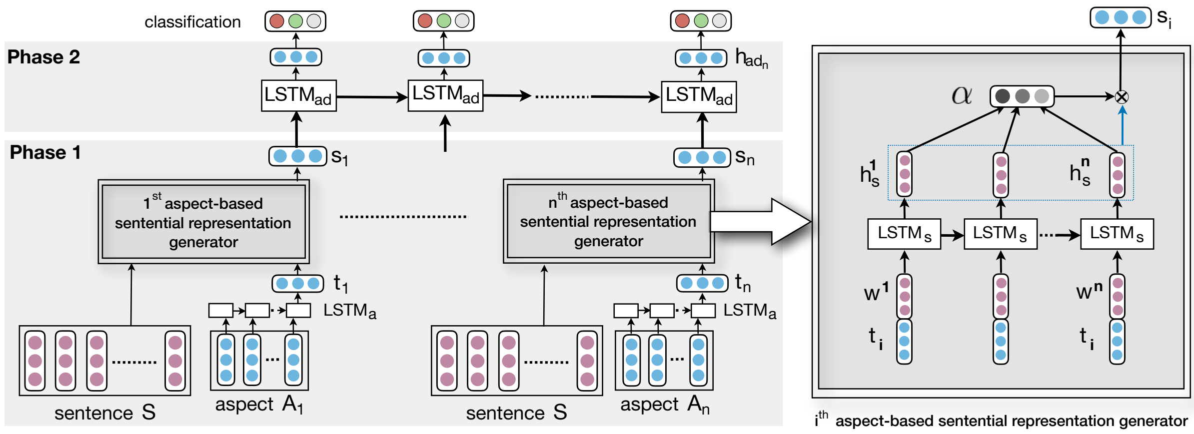

Figure 1: Overall architecture of the proposed method. The *aspect-based sentential representation generator* is described in the right end of the figure.

apply an LSTM, named $LSTM_s$ with dimension d_s, to model the long-term temporal dependencies within the sentence. The hidden state memory vectors across all n timesteps result in matrix $H_i = [h_s^1, ..., h_s^n] \in \mathbb{R}^{d_s \times n}$.

Attention: Attention mechanism is applied on H_i to get an attention vector α, which is in turn used to generate a weighted representation of H_i. We use this weighted representation to be the i^{th} aspect's sentential representation s_i. Previous concatenations of words with the aspect-representations infuse aspect information into the attention process. This enables the attention mechanism to focus on relevant segments in the sentence with respect to the aspect. The overall attention mechanism to generate s_i is summarized as:

$$M = tanh(H_i.W_h) \tag{1}$$

$$\alpha = softmax(M^T.W_b) \tag{2}$$

$$s_i = H.\alpha^T \tag{3}$$

where, $W_h \in \mathbb{R}^{n \times 1}$ and $W_b \in \mathbb{R}^{d_s \times n}$ are projection parameters to be learnt during training and d_s is the dimension of the final sentence vector, i.e., $s_i \in \mathbb{R}^{d_s}$.

The overall process described above is individually applied to all m aspects to get sentential representations $s_1, ..., s_m$.

3.2 Phase 2: Inter-aspect relationship

To capture the implicit inter-aspect dependencies, we model the sentential representations as a sequence $[s_1, ..., s_m]$, following the order of occurrence of their corresponding aspect-terms in sentence S. An LSTM, named LSTM$_{ad}$ with dimension d_{ad} is then applied on this sequence and at each of the i^{th} timestep, its hidden state is projected to another vector having dimensions equal to the number of classes to predict. Finally, softmax operation is applied on this vector to get the prediction probabilities for the sentiment of this i^{th} aspect-term for sentence S. The transitions are as follows:

$$[h_{ad_1}, ..., h_{ad_m}] = LSTM_{ad}([s_1, ..., s_m]) \tag{4}$$

$$\hat{y}_i = softmax(W_{ad}.h_{ad_i}) \tag{5}$$

Here, $\hat{y}_i \in \mathbb{R}^C$ is the predicted probability distribution for the i^{th} aspect of sentence S where C is the number of sentiment classes. $W_{ad} \in \mathbb{R}^{C \times d_{ad}}$ is a parameter and $softmax(x_i) = e^{x_i} / \sum_j e^{x_j}$.

Loss Function: We use categorical cross-entropy as the loss function which is averaged over all aspects for a sentence. Thus, stochastic loss for sentence S is calculated as:

$$Loss = \frac{-1}{m} \sum_{i=1}^{m} \sum_{j=1}^{C} y_{i,j} \; log_2(\hat{y}_{i,j}) + \lambda \| \theta \|^2 \tag{6}$$

Here, m is the number of aspects for a sentence and C is the number of sentiment categories. y_i is the one-hot vector ground truth of i^{th} aspect of sentence S and $\hat{y}_{i,j}$ is its predicted probability of belonging to sentiment class j. λ is the L$_2$ - regularization term and θ is the parameter set, i.e., $\theta = \{W_{[h,b,ad]}, LSTM_{[t,s,ad]}\}$, where $LSTM_{[]}$ represents the internal parameters of that LSTM.

4 Experimentation

Training details: To perform experiments and subsequent hyperparameter tuning, we first split the training set randomly in the ratio $9:1$ to get a held-out validation set. For optimization, we use the Adam optimizer (Kingma and Ba, 2014) having learning rate 0.01. Embedding dimensions are set as follows, $d_a = 100, d_s$ and $d_{ad} = 300$. To facilitate batch processing, we attach dummy aspects in sentences with lesser aspects and also provide masking schemes. For termination, we use the early-stopping procedure with a patience value of 10 that is monitored on the validation loss.

Dataset: We conduct our experiments using the dataset for SemEval 2014 Task 4 containing customer reviews on restaurants and laptops. Each review has one or more aspects with their corresponding polarities. The polarity of an aspect can be positive, negative, neutral or conflict; however, we consider the first three labels for classification. Table 1 contains the statistics for the dataset.

4.1 Results

Table 2 presents the results of our proposed model along with state-of-the-art methods. Our model significantly surpasses the performance of ATAE-LSTM (Wang et al., 2016). Given that ATAE's architecture has a strong correlation to our aspect-based sentential generator (see Figure 1), their work can be categorized as a baseline to our model. This reinforces our hypothesis that a model capable of capturing inter-aspect dependencies indeed performs better. We also compare our model to the recently proposed IAN (Ma et al., 2017). On both datasets, our model performs competitively with IAN and produces nominal improvement. Given that IAN explores the inter-dependencies of aspects with their contexts, while we try to model inter-dependencies between aspects, an interesting direction would be to explore the IAN modeled in our proposed setting (Phase 2 of Figure 1). We set this path as an option for future research.

Table 1 also presents variations of our proposed model. Specifically, we try out variants (a) *Without attention:* in this setting, we omit the attention mechanism while generating aspect-based sentential representation s_i (Equation 1-3). Instead, we define s_i to be h_s^n, i.e., the last hidden state vector of LSTM$_s$ with input S and A_i. However, removing attention brings degradation in the performance of our model on the Restaurant and Laptop dataset by 4% and 3%, respectively. This signifies the importance of an attention mechanism to derive the aspect-based sentential representations. (b) *With hadamard fusion:* instead of concatenation of w^j and t_i, we use the hadamard product which is the element wise multiplication of the vectors. Although this variation reduces the total parameter sizes of the network, it still does not benefit the model and gives a poorer performance to simple concatenation. Numerous other fusion methods such as tensor fusion (Zadeh et al., 2017), compact bilinear pooling (Gao et al., 2016), attention-based fusion (Poria et al., 2017; Hazarika et al., 2018), etc. are applicable, whose analyses, however, is not the focus of this paper.

Data		Aspect Labels			No. of reviews
		Positive	Negative	Neutral	
Rest.	Train	2148	790	628	1977
	Test	725	195	196	600
Laptop	Train	974	839	450	1462
	Test	340	125	169	411

* Rest. = Restaurant

Table 1: Labels and review statistics for the dataset SemEval 2014.

Models	Attn.	Fusion	3-way classification	
			Rest.	Laptop
LSTM	✗	-	74.3	66.5
AE-LSTM	✓	Concat	76.6	68.9
ATAE-LSTM	✓	Concat	77.2	68.7
IAN	✓	-	78.6	72.1
Our Model	✓	Hadamard	73.42	63.7
Our Model	✗	Concat	74.5	69.6
Our Model	✓	Concat	**79.0**	**72.5**

* Attn. = Attention, Rest. = Restaurant

Table 2: Accuracies for three-way classification on the Restaurant and Laptop SemEval 2014 dataset.

4.2 Case Study

A qualitative study on the test set classifications by our model reveals its capability to learn inter-aspect dependencies (Section 1). For the sentence *I love the keyboard and the screen*, the model correctly identifies the sentiment of *screen* as positive which is hinted by positive aspect *keyboard* and conjunction *and*. In another case, for the sentence *The best thing about this laptop is the price along with some of the newer features*, aspect *features* is correctly classified as positive which is influenced by aspect *price* and positive word *best*. This shows that our model is performing well in classifying joint aspects having conjunctions. For the slightly harder case of tackling incomplete information, our model fares well in sentences having this pattern. For example, one of the sentence *Boot up slowed significantly after all windows updates were installed* has aspect *windows update* which does not have a clear sentiment orientation but is implicitly dependent on the aspect *boot up* having a negative sentiment. This was also correctly classified by our model. Moreover, the above examples were incorrectly classified by ATAE. This reaffirms our hypothesis that the ability to learn inter-aspect dependencies is a crucial factor in the task of ABSA.

5 Conclusion

In this paper, we present a way to incorporate inter-aspect dependencies in the task of Aspect-based Sentiment Analysis. Our results suggest that capturing such information indeed improves the task of prediction. Through this work, we hope that future attempts by researchers include this idea in their methods.

6 Acknowledgement

This research was supported in part by the National Natural Science Foundation of China under Grant no. 61472266 and by the National University of Singapore (Suzhou) Research Institute, 377 Lin Quan Street, Suzhou Industrial Park, Jiang Su, People's Republic of China, 215123. We would also like to thank the anonymous reviewers for their valuable feedback.

References

Yang Gao, Oscar Beijbom, Ning Zhang, and Trevor Darrell. 2016. Compact bilinear pooling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 317–326.

Devamanyu Hazarika, Sruthi Gorantla, Soujanya Poria, and Roger Zimmermann. 2018. Self-attentive feature-level fusion for multimodal emotion detection.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Cheng Li, Xiaoxiao Guo, and Qiaozhu Mei. 2017. Deep memory networks for attitude identification. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 671–680. ACM.

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Veronica Perez-Rosas, Carmen Banea, and Rada Mihalcea. 2012. Learning sentiment lexicons in spanish. In *LREC*, volume 12, page 73.

Soujanya Poria, Erik Cambria, Devamanyu Hazarika, Navonil Mazumder, Amir Zadeh, and Louis-Philippe Morency. 2017. Multi-level multiple attentions for contextual multimodal sentiment analysis. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 1033–1038. IEEE.

Delip Rao and Deepak Ravichandran. 2009. Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 675–682. Association for Computational Linguistics.

Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. 2016a. Effective lstms for target-dependent sentiment classification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3298–3307.

Duyu Tang, Bing Qin, and Ting Liu. 2016b. Aspect level sentiment classification with deep memory network. *arXiv preprint arXiv:1605.08900*.

Yi Tay, Anh Tuan Luu, and Siu Cheung Hui. 2017. Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis. *arXiv preprint arXiv:1712.05403*.

Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. Attention-based lstm for aspect-level sentiment classification. In *EMNLP*, pages 606–615.

Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. 2017. Recent trends in deep learning based natural language processing. *arXiv preprint arXiv:1708.02709*.

Amir Zadeh, Minghai Chen, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. 2017. Tensor fusion network for multimodal sentiment analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1103–1114.