Multi-Task Multi-Head Attention Memory Network for Fine-Grained Sentiment Analysis

Zehui Dai(✉), Wei Dai, Zhenhua Liu, Fengyun Rao, Huajie Chen, Guangpeng Zhang, Yadong Ding, and Jiyang Liu

NLP Group, Gridsum, Beijing, China
{daizehui,daiwei,liuzhenhua,raofengyun,chenhuajie,
zhangguangpeng,dingyadong,jliu}@gridsum.com

Abstract. Sentiment analysis is widely applied in personalized recommendation, business reputation monitoring, and consumer-driven product design and quality improvement. Fine-grained sentiment analysis, aimed at directly predicting sentiment polarity for multiple pre-defined fine-grained categories in an end-to-end way without having to identify aspect words, is more flexible and effective for real world applications. Constructing high performance fine-grained sentiment analysis models requires the effective use of both shared document level features and category-specific features, which most existing multi-task models fail to accomplish. In this paper, we propose an effective multi-task neural network for fine-grained sentiment analysis, Multi-Task Multi-Head Attention Memory Network (**MMAM**). To make full use of the shared document level features and category-specific features, our framework adopts a multi-head document attention mechanism as the memory to encode shared document features, and a multi-task attention mechanism to extract category-specific features. Experiments on two Chinese language fine-grained sentiment analysis datasets in the Restaurant-domain and Automotive-domain demonstrate that our model consistently outperforms other compared fine-grained sentiment analysis models. We believe extracting and fully utilizing document level features to establish category-specific features is an effective approach to fine-grained sentiment analysis.

Keywords: Fine-grained sentiment analysis · Multi-head Attention Memory · Multi-task learning

1 Introduction

The main purpose for sentiment analysis is to identify the sentiment polarity (i.e. positive, neutral, and negative) from input documents. Most existing sentiment analysis tasks are carried out at document level [1–3] or aspect level [4–7]. Document level sentiment analysis outputs the general sentiment polarity

J. Tang et al. (Eds.): NLPCC 2019, LNAI 11838, pp. 609–620, 2019.
https://doi.org/10.1007/978-3-030-32233-5_47

of the whole document, while aspect level sentiment analysis predicts sentiment for an aspect. Aspect sentiment analysis is a two-step process, i.e. aspect word extraction and sentiment analysis. Fine-grained sentiment analysis [8,9] is an approach that directly analyzes sentiment polarity (positive, neutral, negative or not mentioned) for multiple pre-defined fine-grained categories in a specific domain. Take the Restaurant domain as an example, pre-defined categories such as *ease of transportation, price level, cost effectiveness, discounts, taste, overall experience* and so on should be analyzed collectively to provide a fine-grained sentiment analysis approach to document understanding. Fine-grained sentiment analysis is also able to predict sentiment polarity from implicit expressions in the absence of aspect words. It is more suitable in real world applications, especially for documents containing oral expressions. For example, the user review snippet "*The food is expensive but the taste is delicious*" contains two categories of sentiment, i.e. the price is negative while the taste is positive. The negative comment for price is "*expensive*", which is expressed implicitly without an aspect word.

In order to analyze these categories collectively, multi-task learning has been suggested for fine-grained sentiment analysis. For example, [10] proposed a multi-task learning framework with an individual attention for each category on the shared LSTM encoding layer. However, these models perform poorly for categories which rely on multiple document level features, especially on conflicting features. These approaches tend to obscure the characteristics of each attended word by forcing multiple words into one attention or one pooling for each category [5]. For example, the sentiment expressed in the review snippet "*Although tables on the top floor of the restaurant are visible from the road crossing, it is still a long way from there, and the restaurant sign is not as clear as others*" relies on multiple words with conflicting expressions. The negative sentiment on category "*easy to find*" is influenced more strongly by the expression "*the restaurant sign is not as clear as others*". Models with only one attention or one pooling for each category are not able to provide appropriate weights for these features. On the other hand, certain sentiments are synthetic in nature. For example, the category "*overall experience*" should be synthesized from the combination of the sentiment polarities from all other categories, especially if no explicit expression is provided. Therefore, in addition to individual category-specific features, obtaining document level features in a shared way and making full use of them is necessary for effective fine-grained sentiment analysis.

In order to capture multiple shared features of a document, as well as category-specific features for fine-grained sentiment classification, we propose an effective multi-task learning framework, i.e. Multi-Task Multi-Head Attention Memory Network (**MMAM**), for fine-grained sentiment analysis. With the document tokens as input, our model adopts an embedding look-up layer to generate the document embedding matrix, a Bi-LSTM layer for document encoding, and a document attention memory layer with multiple attention heads to capture features of different expressions. All the above layers are trained with shared parameters. Subsequently, a fine-grained attention layer is adopted on the multi-head document attention memory layer by paying specific attention to each fine-grained category. The final output of each category consists of an individual fully connected layer and an individual softmax layer.

In summary, our contributions are two-fold: (i) We proposed an effective approach to making full use of document level features and category-specific features for fine-grained sentiment analysis. (ii) We developed a multi-task framework with multi-head attention layer to capture shared document level features, and a fine-grained attention layer to make full use of these document level features for fine-grained sentiment analysis.

Our framework outperforms other compared fine-grained sentiment analysis models on two Chinese language fine-grained sentiment analysis datasets, i.e., the Fine-grained Sentiment Analysis of Online User Reviews dataset 2018 (AI Challenger 2018)[1] in the Restaurant-domain with 20 categories, and the Fine-grained Sentiment Analysis of User Reviews in Automotive Industry (DataFountain 2018)[2] containing reviews in the Automotive-domain with 10 categories, as shown in Table 1.

Table 1. Details of the experiment datasets

Datasets	Training	Validation	Test	Number of categories
Restaurant-domain reviews	100k	10k	10k	20
Automotive-domain reviews	6632	829	829	10

2 Related Work

Fine-grained sentiment analysis is to analyze sentiment polarity on multiple pre-defined categories in an end-to-end way. It is able to predict sentiment from implicit expressions in the absence of aspect words [8,9]. For example, [11] applied structured features for fine-grained sentiment analysis. [12] proposed a multi-layer perceptron model for multi-task emotion classification and regression. [13] combines the final states of a bi-LSTM neural network with additional features for fine-grained emotion analyses. [14] applied multi-task framework with shared CNN or LSTM encoder and task-specific softmax mechanism for fine-grained sentiment analysis.

Common to these approaches to fine-grained sentiment analysis is the use of **multi-task learning** (MTL). MTL based on neural networks has proven to be effective in many NLP tasks, such as information retrieval [15], machine translation [16], part-of-speech tagging and semantic role labeling [17]. MTL utilizes both the commonalities in the document features and the differences in each task to perform multiple learning tasks collectively. Therefore, MTL can strengthen the training data by transferring useful information from one task to another. For example, [18] used shared CRFs and domain projections for multi-domain multi-task sequence tagging. [16] and [19] shared encoders or decoders in one to

[1] https://challenger.ai/dataset/fsaouord2018.

[2] https://www.datafountain.cn/competitions/329.

many or many to many neural machine translations. [20] used multiple shared LSTM layers with a separate softmax layer for each semantic sequence labeling task. Multi-task learning has also been applied to multi-aspect sentiment analysis tasks [21]. However, existing approaches in fine-grained sentiment analysis are not so effective because document level features are not fully utilized since only word encoding layers are shared in these models.

3 Approach

Our framework consists of five layered modules (Fig. 1), the word embedding layer, the Bi-LSTM encoding layer, the document attention layer, the fine-grained attention layer, and the output layers for each category consisting of a fully connected layer and a softmax layer.

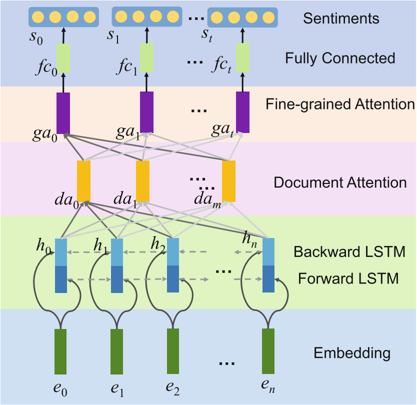


Fig. 1. MMAM model Framework

3.1 Input Embedding Layer

An embedding lookup matrix $\mathbb{L} \in \mathbb{R}^{d \times |V|}$ is generated by concatenating all the word vectors from pre-trained models, such as word2vec and ELMo, in which d is the dimension of the embedding vector and $|V|$ is the size of the vocabulary. In forward-propagation, $\mathbf{E} = \{e_0, e_1, \ldots, e_n\}$ is generated by retrieving the matrix $\mathbb{L}$ from the input words w, where $e_i \in \mathbb{R}^d$ is the embedding vector for each word.

3.2 Bi-LSTM Layer

A Bi-LSTM layer is applied for encoding the embedded words to form sequential features. 1-layer Bi-LSTM is applied in this research. The inputs for the forward LSTM encoder and backward LSTM encoder are both the embedded word vectors $\mathbf{E}$, while the outputs are the encoded forward and backward vectors. The Bi-LSTM layer produces the concatenated vectors $\mathbf{H} = \{h_0, h_1, \ldots, h_n\}$ as the output, where h_i is the concatenation of the hidden states in the i-th forward LSTM cell and the i-th backward LSTM cell.

3.3 Multi-Head Document Attention Memory Layer

Attention is applied for document encoding. Different from previous researches, [22,23], for this fine-grained multi-task learning, multiple attention heads are applied as memory on the output of Bi-LSTM layer to capture shared features in the document. For each attention head, the forward-propagation is listed as follows:

$$\alpha_{da_i} = softmax(\overrightarrow{w}_{da2,i} tanh(\mathbf{W}_{da1,i}\mathbf{H}^\top)) \tag{1}$$

$$da_i = \alpha_{da_i}\mathbf{H} \tag{2}$$

where da_i is the output of the i-th document attention head vector, $\mathbf{W}_{da1,i} \in \mathbb{R}^{dim_{da} \times 2dim_h}$ is a dense transformation matrix for hidden states $\mathbf{H}$, dim_{da} is the document attention dimension, dim_h is the hidden states size for the LSTM cell, and vector $\overrightarrow{w}_{da2,i} \in \mathbb{R}^{dim_{da}}$ is the query vector for each document attention query head. Supposing there are m document attention features, the output of document attention is a matrix $\mathbf{DA} \in \mathbb{R}^{2dim_h \times m}$, generated by the concatenation of the m attention heads, i.e. $\mathbf{DA} = \{da_0, da_1, \ldots, da_m\}$.

3.4 Fine-Grained Attention Layer

While the document multi-head attention memory layer captures shared features from the document, the fine-grained attention layer is employed on the output of document attention memory layer in order to obtain the category-specific features. For each category, the calculation in the forward propagation for a fine-grained attention vector is given as follows:

$$\alpha_{ga_i} = softmax(\overrightarrow{w}_{ga2,i} tanh(\mathbf{W}_{ga1,i}\mathbf{DA}^\top)) \tag{3}$$

$$ga_i = \alpha_{ga_i}\mathbf{DA} \tag{4}$$

where ga_i is the output of the i-th fine-grained attention vector, $\mathbf{W}_{ga1,i} \in \mathbb{R}^{dim_{ga} \times 2dim_h}$ is a dense transformation matrix for document attention matrix $\mathbf{DA}$, dim_{ga} is the fine-grained attention dimension, dim_h is the hidden states size for the LSTM cell, and vector $\overrightarrow{w}_{ga2,i} \in \mathbb{R}^{dim_{ga}}$ is a specific query vector for each category.

3.5 Output Layers and Multi-task Learning

The output layers consist of a fully connected layer and a 4-class softmax layer (positive, neutral, negative, and not-mentioned) for each category. Both layers are trained with category-specific parameters. The forward-propagation for each category is listed as follows:

$$fc_i = dense(ga_i) \tag{5}$$

$$p_i = softmax(fc_i) \tag{6}$$

where $fc_i \in \mathbb{R}^{dim_{fc}}$ is the output of a fully connected layer, and $p_i \in \mathbb{R}^4$ is the output probability for each class in the i-th category.

The model is trained by minimizing the sum of cross-entropy loss in each fine-grained category. L_2 regularization is employed in all the attentions and dense layers to ease over-fitting. The loss function of this model is given as follows:

$$L = \sum_{x,y \in D} \sum_{i=0}^{t} \sum_{c \in C} y_i^C \cdot log f_i^C(x;\theta) + \lambda ||\theta||_2 \tag{7}$$

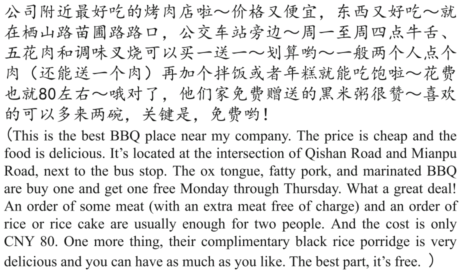
where D is the training dataset, C is the sentiment classes including positive, neutral, negative, and not-mentioned, $y_i^C \in \mathbb{R}^4$ is the one-hot label vector for the i-th category with true label marked as 1 and others marked as 0, $f_i^C(x;\theta)$ is the probability result for the i-th category, and λ is the L_2 regularization weight. Besides L_2 regularization, we also employed dropout and early stopping to ease overfitting.

4 Experiments

4.1 Experiment Settings

The effectiveness of the model was tested on two Chinese language fine-grained sentiment analysis datasets, as shown in Table 1. The original Restaurant-domain dataset with 120k labeled data was split into training, validation, and test datasets, containing 100k, 10k, and 10k samples respectively. The positive, neutral, and negative classes are labeled as 1, 0, and –1 respectively, while the not-mentioned class is labeled as –2. There are 20 categories in Restaurant-domain, *ease of transportation, distance from business location, ease of finding, waiting duration, waiters' attitude, ease of parking, serving duration, price level, cost effectiveness, discount, decoration, noise, space, cleanness, portion, taste, look, recommendation, overall experience*, and *willingness to return*, while the 10 categories in Automotive-domain are *price level, engine power, comfort, configuration, appearance, fuel consumption, space, safety, ease of control*, and *trim.* All these categories are predefined by the datasets providers. A user review example is given in Fig. 2. In this case, the *ease of transportation, price level, cost effectiveness, discounts, taste*, and *overall experience* categories are labeled as 1 (positive). The others are labeled as −2 since they were not mentioned in this review, while no category is labeled as 0 or −1. The original Automotive-domain reviews dataset with 8290 labeled data was also split into training, validation, and test datasets, containing 6632, 829, and 829 samples respectively. The original labels were transformed to −2, −1, 0 and 1, similar to the Restaurant-domain.

We used a concatenation of a 300-dimension word2vec [24] and a 1024-dimension Embedding Language Model (ELMo) [25] as input features for the Restaurant-domain dataset. Both word2vec and ELMo embedding are



公司附近最好吃的烤肉店啦～价格又便宜，东西又好吃～就在栖山路苗圃路路口，公交车站旁边～周一至周四点牛舌、五花肉和调味叉烧可以买一送一～划算哟～一般两个人点个肉（还能送一个肉）再加个拌饭或者年糕就能吃饱啦～花费也就80左右～哦对了，他们家免费赠送的黑米粥很赞～喜欢的可以多来两碗，关键是，免费哟！

（This is the best BBQ place near my company. The price is cheap and the food is delicious. It's located at the intersection of Qishan Road and Mianpu Road, next to the bus stop. The ox tongue, fatty pork, and marinated BBQ are buy one and get one free Monday through Thursday. What a great deal! An order of some meat (with an extra meat free of charge) and an order of rice or rice cake are usually enough for two people. And the cost is only CNY 80. One more thing, their complimentary black rice porridge is very delicious and you can have as much as you like. The best part, it's free. ）

Fig. 2. A sample review of Restaurant-domain from AI Challenger 2018 (Fine-grained sentiment analysis)

pre-trained on a large Dianping corpus[3] for the Restaurant-domain dataset. The codes we used for ELMo model pre-training were released by the authors[4]. For the Automotive-domain dataset, a 300-dimension word2vec pre-trained on an Automotive-domain corpus was used as network input embedding features.

4.2 Compared Methods

The Multi-Task Multi-Head Attention Memory (**MMAM**) model was compared with the following models. All comparisons were conducted by augmenting a multi-task fine-grained sentiment analysis layer on top of the existing networks in order to achieve comparable results.

- SVM [26]: A traditional support vector machine classification model with extensive feature engineering.
- multi-task CNN-attention and CNN-pooling networks [2]: A multi-task framework with an attention or a max-pooling layer is applied on the concatenation of the output of CNN kernels with various kernel sizes.
- multi-task LSTM-attention and LSTM-pooling networks [10]: A multi-task framework with an attention or a max-pooling layer is applied on the concatenation of the output of a forward LSTM layer and a backward LSTM layer.
- multi-task Recurrent Attention network on Memory (RAM) [5]: RAM model adopts a multiple attention layer combined with a recurrent neural network. The final state of the recurrent attention network is used for classification in the original RAM network. We applied multi-task RAM by adding an individual softmax layer on the final states for each category.
- Multi-Head single-task model: A set of single-task models (**MAM**-single) that is trained for each specific category.

[3] https://github.com/SophonPlus/ChineseNlpCorpus.

[4] https://github.com/allenai/bilm-tf.

Table 2. Fine-grained sentiment prediction results

Multi-task models	Restaurant-domain		Automotive-domain	
	Macro-F1	Acc	Macro-F1	Acc
SVM	.5244	.7171	.5371	.8680
CNN-pooling	.6997	.8748	.5540	.9285
LSTM-pooling	.7171	.8774	.5591	.9299
CNN-attention	.7170	.8784	.5574	.9309
LSTM-attention	.7199	.8787	.5588	.9274
RAM	.7170	.8770	.5597	.9274
MAM-single	.7195	.8788	.5636	.9300
MMAM	**.7229**	**.8799**	**.5852**	**.9355**

4.3 Main Results

We evaluated the models with two metrics. The first metric is Accuracy [5,6,27], the average accuracy across all categories. We also used the Macro-Averaged F-measure (Macro-F1) [5,6,27] calculated by averaging the Macro-F1 across all categories as the sentiment is polarized in some categories.

As shown in Table 2, our **MMAM** model consistently outperforms all other models on both metrics. SVM model performs the worst because it takes n-gram words directly as input without any embedding. CNN based multi-task models perform poorly both with attention and with max-pooling feature extractor. This is because CNN models are efficient in capturing the informative n-gram features, but are likely to fail when reviews of multiple categories are expressed in one document due to the loss of sequential features. Multi-task LSTM based models perform better than CNN since they may extract some sequential features. However, LSTM does not perform as well as our **MMAM** model since they only apply one pooling or attention layer for each fine-grained classification task, and lack shared document level attention memory features. Comparison with multi-task LSTM model confirms that the multi-head document attention is necessary to capture multiple document level features.

Our **MMAM** model also performs better than multi-task RAM model. For Automotive-domain dataset, the Macro-F1 of **MMAM** is 0.0255 higher than RAM, a **4.6%** improvement. Multi-task RAM model also adopts multiple attentions after the LSTM encoder layer combined together by a GRU layer. However, the nonlinear recurrent attention concentrates on the sentiment transition of one category, rather than capturing category-specific features. This confirms the effectiveness of the collective extraction of document level features in our framework.

To validate the effectiveness of the multi-task learning structure, we tested our **MMAM** against a set of single-task models (**MAM**-single), where one model is trained for each specific category. As expected, **MAM**-single did not perform as well as our **MMAM** model. This is because the **MAM**-single model does not utilize any encoding information from other categories.

Table 3. Fine-grained sentiment prediction results for **MMAM** with various number of document attention heads

Document attention heads	Restaurant-domain		Automotive-domain	
	Macro-F1	Acc	Macro-F1	Acc
2	.7171	.8778	.5549	.9299
4	.7181	.8781	.5649	.9331
6	.7224	.8791	.5714	.9325
8	.7224	.8795	.5811	.9329
10	.7227	.8797	**.5852**	**.9355**
15	**.7229**	**.8799**	.5834	.9353
20	.7227	.8799	.5847	.9355

4.4 Effect of Document Attention Memory Heads

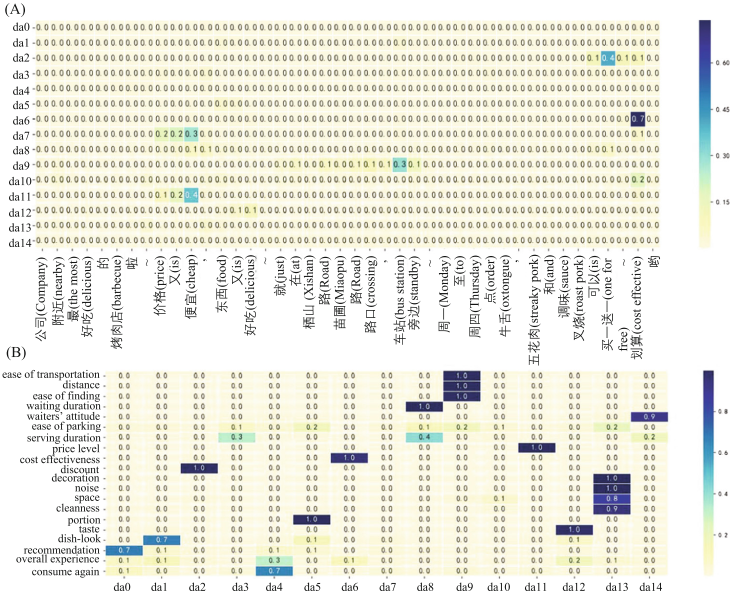


Fig. 3. Visualization of document attention with 15 heads for document from AI Challenger 2018 (fine-grained sentiment analysis). (A) document attention plots of sub-sentences, and (B) fine-grained attention plot

We tested our model with various number of document attention memory heads, as it is a crucial setting that affects the performance of **MMAM** model. The results are shown in Table 3. With only 2 attention memory heads, **MMAM** performs worse than multi-task LSTM-attention model. This is because the features learned from 2 attention memory heads are quite limited for fine-grained classification tasks. The performance of our **MMAM** model improves as the number of document attention memory heads increases until it reaches 10 when the performance begins to level off for both datasets. The optimal performance is obtained with 15 attention heads for the Restaurant-domain dataset, and with 10 attention heads for the Automotive-domain dataset. More attention heads were needed for Restaurant-domain dataset to reach optimal performance because the Restaurant-domain dataset contains more categories, requiring more shared features for classification.

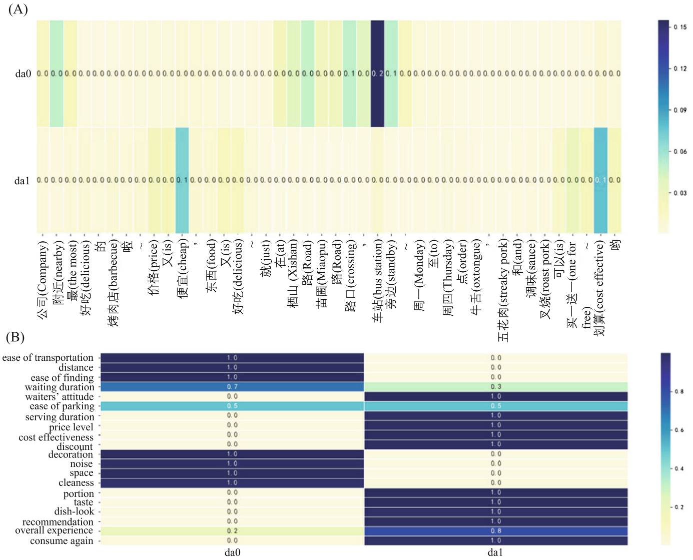


Fig. 4. Visualization of document attention with only 2 heads for document from AI Challenger 2018 (fine-grained sentiment analysis). (A) document attention plots of sub-sentences, and (B) fine-grained attention plot

4.5 Case Study

To directly understand the information flow in the **MMAM** model, we visualized the attention results in the multiple attention heads from the document attention layer and the attention results in the fine-grained attention layer. The Bi-LSTM encoding layer was removed in the visualization plots in order for the attention plots to reveal the words on which each document attention head focused. The visualization results shown in Figs. 3 and 4 are attention plots for some sentences in the sample document in Fig. 2.

Figure 3 presents the attention results of document attention layer with 15 heads. These attention memory heads focus on different word-level features for fine-grained sentiment classification. For example, document attention head 6 strongly focuses on the word "*cost-effective*" in Fig. 3(A), which dominantly contributes to the feature in category "*cost-effective*" in Fig. 3(B). The document attention head 9, which is focused on the words "near the bus station", is the sole contributor to the "*ease of transportation*" category. Category "*overall experience*" relies on 4 document attention heads, i.e. head 1, 4, 6, 12 that focus on different document level features, to predict the sentiment polarity. The multiple document attention heads provide the fine-grained attention layer with the ability to combine features from multiple categories. For comparison, the visualization plots in Fig. 4 present the attention results of **MMAM** model with only 2 attention heads in the document attention layer. The location items, such as *ease of transportation, distance*, and *easy to find* are all supported by attention head 0. Other categories of positive sentiments, such as *food taste, portion, prices*, and *cost-effectiveness* are all contributed by attention head 1. Therefore, the attention in each attention head is distributed across multiple categories, preventing the multi-task model from achieving optimal performance.

5 Conclusions and Future Work

In this paper, we proposed an effective neural network framework for fine-grained sentiment analysis. This model employs a shared multi-head attention layer to capture document level features, followed by an individual fine-grained attention layer to capture category-specific features. We evaluated the performance of our model on two datasets and demonstrated that it outperforms other fine-grained sentiment analysis models we tested.

The performance of fine-grained sentiment analysis can be further improved in many ways. One approach is to combine domain knowledge with machine learning for sentiment analysis to provide additional features to the neural network. For example, the knowledge that Wudaokou and Wangfujing are popular business locations can be very useful in predicting sentiment polarity for the category "*distance from business location*". Therefore, we believe that the learning framework enhanced with domain-knowledge may perform even more effectively in fine-grained sentiment analysis systems.

References

1. Qian, Q., Tian, B., Huang, M., Liu, Y., Zhu, X., Zhu, X.: Learning tag embeddings and tag-specific composition functions in recursive neural network. In: Proceedings of ACL, vol. 1, pp. 1365–1374 (2015)
2. Kim, Y.: Convolutional neural networks for sentence classification. In: Proceedings of EMNLP, pp. 1746–1751 (2014)
3. Shi, B., Fu, Z., Bing, L., Lam, W.: Learning domain-sensitive and sentiment-aware word embeddings. In: Proceedings of ACL, vol. 1, pp. 2494–2504 (2018)
4. Wang, Y., Huang, M., Zhu, X., Zhao, L.: Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of EMNLP, pp. 606–615 (2016)
5. Chen, P., Sun, Z., Bing, L., Yang, W.: Recurrent attention network on memory for aspect sentiment analysis. In: Proceedings of EMNLP, pp. 452–461 (2017)

6. Tang, D., Qin, B., Feng, X., Liu, T.: Target-dependent sentiment classification with long short term memory. CoRR, abs/1512.01100
7. Tang, D., Qin, B., Liu, T.: Aspect level sentiment classification with deep memory network. In: Proceedings of EMNLP, pp. 214–224 (2016)
8. Mohammad, S., Bravo-Marquez, F., Salameh, M., Kiritchenko, S.: Semeval-2018 task 1: affect in tweets. In: Proceedings of The 12th International Workshop on Semantic Evaluation, SemEval, pp. 1–17 (2018)
9. Wu, C., Wu F., Liu, J., Yuan, Z., Wu, S., Huang, Y.: Thu_ngn at semeval-2018 task 1: Fine-grained tweet sentiment intensity analysis with attention CNN-LSTM. In: Proceedings of The 12th International Workshop on Semantic Evaluation, SemEval, pp. 186–192 (2018)
10. Meisheri, H., Khadilkar, H.: Learning representations for sentiment classification using multi-task framework. In: Proceedings of EMNLP, pp. 299–308 (2018)
11. Zirn, C., Niepert, M., Stuckenschmidt, H., Strube, M.: Fine-grained sentiment analysis with structural features. In: Proceedings of IJCNLP, pp. 336–344 (2011)
12. Akhtar, M., Ghosal, D., Ekbal, A., Bhattacharyya, P.: A multi-task ensemble framework for emotion, sentiment and intensity prediction. CoRR, abs/1808.01216
13. Balikas, G., Moura, S., Amini, M.: Multitask learning for fine-grained twitter sentiment analysis. CoRR, abs/1707.03569
14. Schmitt, M., Steinheber, S., Schreiber, K., Roth, B.: Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks. In: Proceedings of EMNLP, pp. 1109–1114 (2018)
15. Liu, X., G., Gao, J., He, X., Deng, L., Duh, K., Wang, Y.: Representation learning using multi-task deep neural networks for semantic classification and information retrieval. In: Proceedings of NAACL, pp. 912–921 (2015)
16. Dong, D., Wu, H., He, W., Yu, D., Wang, H.: Multi-task learning for multiple language translation. In: Proceedings of ACL, vol. 1, pp. 1723–1732 (2015)
17. Collobert, R., Weston, J.: A unified architecture for natural language processing: deep neural networks with multitask learning. In: Proceedings of ICML, pp. 160–167 (2008)
18. Peng, N., Dredze, M.: Multi-task multi-domain representation learning for sequence tagging. CoRR, abs/1608.02689
19. Luong, M., Le, Q., Sutskever, I., Vinyals, O., Kaiser, L.: Multi-task sequence to sequence learning. CoRR, abs/1511.06114
20. Alonso, H., Plank, B.: Multitask learning for semantic sequence prediction under varying data conditions. CoRR, abs/1612.02251
21. Hu, M., et al.: CAN: constrained attention networks for multi-aspect sentiment analysis. CoRR, abs/1812.10735
22. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. CoRR, abs/1409.0473
23. Santos, C., Gatti, M.: Deep convolutional neural networks for sentiment analysis of short texts. In: Proceedings of COLING, pp. 69–78 (2014)
24. Mikolov, T., Chen, K., Corrado, G., Deam, J.: Efficient estimation of word representations in vector space. CoRR, abs/1301.3781
25. Peters, M., et al.: Deep contextualized word representations. In: Proceedings of NAACL, vol. 1, pp. 2227–2237 (2018)
26. Kiritchenko, S., Zhu, X., Cherry, C., Mohammad, S.: NRC-Canada-2014: detecting aspects and sentiment in customer reviews. In: Proceedings of SemEval@COLING 2014, vol. 1, pp. 437–442 (2014)
27. Li, X., Bing, L., Lam, W., Shi, B.: Transformation networks for target-oriented sentiment classification. In: Proceedings of ACL, vol. 1, pp. 946–956 (2018)