Opinion Knowledge Injection Network for Aspect Extraction

Shaolei Zhang[1(✉)], Gang Lu[2], and Kai Shuang[3]

[1] School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China
zhangshaolei@bupt.edu.cn
[2] Intelligent Networks and Devices Research Institute, China Telecom Corporation Limited, Guangzhou, China
[3] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China

Abstract. Aspect term extraction (ATE) is to extract explicit aspect expressions from online reviews. This paper focused on the supervised extraction of aspect term. Previous models for ATE either ignored the opinion information or improperly utilized the opinion information with a high-coupling method. We proposed a model to perform ATE with the assistance of opinion knowledge, called opinion knowledge injection network. Specifically, the proposed model distills the opinion knowledge through the attention mechanism and joins it into each word to assist aspect extraction. The proposed model achieved surprisingly good results, improving 1.34% and 1.23% than the best results before respectively on the laptop and restaurant datasets, and reached state-of-the-art.

Keywords: Aspect extraction · Opinion knowledge · Unidirectional injection · Attention mechanism

1 Introduction

The aspect-based sentiment analysis (ABSA) task is to identify opinions expressed towards specific entities or attributes of entities [1]. The first step of the ABSA is the aspect term extraction (ATE), which is the key to sentiment analysis. The goal of ATE is to find the phrase that is evaluated in the sentence. For example, in the sentence "The food is simply unforgettable!", "*food*" should be chosen as the aspect term because it is commented with "*unforgettable*".

Aspect term extraction has been performed with supervised approaches [2–4,8] and unsupervised approaches [5–7,9,10]. This paper focuses on supervised approaches which usually perform better than unsupervised approaches.

S. Zhang–This work was supported in part by the National key research and development program of China(2016QY01W0200). Research Innovation Fund for College Students of Beijing University of Posts and Telecommunications. National Key Research and Development Program of China (2017YFB1400603).

T. Gedeon et al. (Eds.): ICONIP 2019, LNCS 11954, pp. 669–681, 2019.
https://doi.org/10.1007/978-3-030-36711-4_56

In supervised aspect extraction, many of the early models only focus on the modeling of the aspect term, ignoring the contribution of the opinion word. According to the definition of the task, if a word/phrase is treated as an aspect term, it should be commented on by some opinion words which indicate the emotional polarity of it. The neglect of the opinion word will result in the word which was not commented in the sentence being incorrectly treated as the aspect term. Take the sentence "The service is fantastic at this fun place." as an example, where "*service*" and "*place*" are aspect terms. Without the assistant of opinion words, the previous models are difficult to extract "*place*" as an aspect term, but as an adverbial more likely.

The opinion word has been exploited in a few models, such as RNCRF [11], CMLA [12], and MIN [16]. Although these methods achieve better performance than the previous models, there are still two shortcomings. (1) These previous models that utilize the opinion word either rely on dependency phase tree [11] or have a high-coupling network architecture [12,16]. Depending on dependency parsing will cause errors in informal comments, and high-coupling network cause that the inaccurate prediction will interfere with each other and amplify. (2) Most previous models are based on RNN and LSTM, and the models that perform both aspect extraction and opinion extraction often require two or more LSTMs and even interact between each layer. This will result in a lower speed, which needs to be considered when the model is actually deployed.

To solve the first problem, making more rational use of opinion information, we proposed a model called Opinion Knowledge Injection Network (OKIN). For each word, OKIN distills the opinion knowledge of it from the whole sentence and injects the opinion knowledge into the representation of the word. To be more reasonable in the representation of opinion knowledge, OKIN computes the relevance between the aspect word and opinion words as association score, and then the opinion knowledge is weighted aggregated by opinion word according to the association score. To avoid the noise being continuously amplified during the coupling of aspect and opinion extraction, we adopt a unidirectional injection method, which only appears on the last layer of the network. Unidirectional injection guarantees that there is only information flow from opinion extraction to aspect extraction without reverse. Unidirectional injection effectively aggregates opinion knowledge into each word while avoiding the inaccurate aspect extraction to generate noise on the opinion extraction and preventing two tasks from interfering with each other. Take the sentence "The service is fantastic at this fun place." as an example as well, where the opinion terms are "*fantastic*" and "*fun*". Previously highly coupled models were prone to errors in dealing with "*place*", but still used "*place*" to associate with opinion words layer by layer, making the "*place*" and "*fun*" both difficult to be predicted accurately. OKIN computes the association score, where "*fantastic*" $\rightarrow$ "*service*" and "*fun*" $\rightarrow$ "*place*" will get a higher score. Then, the opinion knowledge of "*fun*" will be aggregated into "*place*" unidirectionally. This approach ensures that the opinion extraction is as accurate as possible and then uses it to assist aspect extraction. Thus, it is much easier to extract the "*place*" as an aspect accurately. To solve the second problem, improving the speed of processing, our model uses CNN instead of LSTM. CNN has been proved is effective to ATE [13]. OKIN uses two CNNs to extract the aspect and the opinion individually and only perform unidirectional injection on the last layer of the network, which is parallelizable.

In summary, our contribution is four-fold: (1) To the best of our knowledge, we are the first to use double CNNs to extract aspect and opinion simultaneously. (2) We proposed unidirectional injection, a more reasonable low-coupling method, to control the spread of noise when utilizing the opinion word in ATE. (3) We provided an acceleration method for aspect and opinion extraction with CNN. (4) We conducted an experiment on two datasets to verify that our model achieves state-of-the-art for aspect term extraction.

2 Related Work

Reference [5] put forward aspect level sentiment analysis in which aspect term extraction is an important task. In recent research, aspect term extraction has been abstracted into sequence labeling tasks.

Aspect term extraction is mainly divided into unsupervised and supervised approaches. The unsupervised approach includes methods such as topic modeling [15,20], syntactic rules-based extraction [8,15], frequent pattern mining [5,14], word alignment [21] and label propagation [22].

The earliest supervised approach was using Conditional Random Fields (CRF) [17]. Recently, more methods are using neural network for aspect term extraction., e.g., using LSTM [18],using CNN [19] and attention mechanism [12,16]. Further, the opinion word is gradually paid more attention to the aspect term extraction. Some models [11,12,16] directly performed aspect term and opinion term co-extraction through dependency parsing or high-coupling network. RNCRF [11] is a novel joint model that integrates recursive neural networks and conditional random fields. For RNCRF, it tends to suffer from parsing errors since the structure of the recursive network hinges on the dependency parse tree. When applied to informal comments, it is easier to make mistakes. CMLA [12] consists of coupled attentions to exploit the correlations between aspect and opinion terms through tensor operators. Similarly, MIN [16] employs three LSTMs, using memory interactions and Sentimental sentence constraints for the aspect and opinion co-extraction. For CMLA and MIN, when coupling the extraction of aspect and opinion on each layer, noise and error are continuously amplified layer by layer. In addition, most of the previous models for aspect extraction were based on RNN and LSTM, until [13] proposed a CNN-based model and proved to be effective.

Existing methods used highly-coupled LSTM-based models, resulting in bad interference between aspect and opinion extraction, and slow processing speed. Therefore, we proposed OKIN with double CNNs and unidirectional injection, reducing the degree of coupling to avoid bad interference and using two parallel CNNs to speed up the processing.

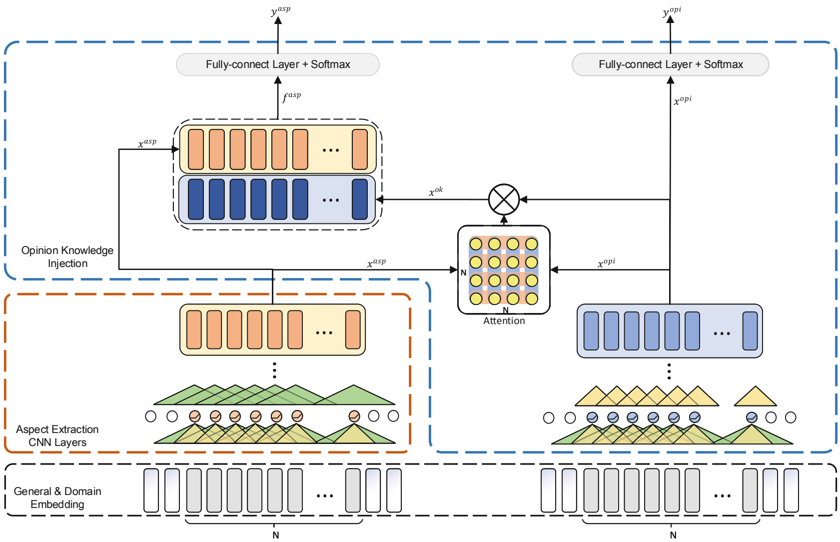


Fig. 1. Proposed model architecture. Gray rectangles, red rectangles, and blue rectangles represent a vector in the embedding layer, the aspect extraction CNN layers, and the opinion extraction CNN layers, respectively. The white rectangles represent zero vector padded. The triangles represent the 1D-CNN filters. (Color figure online)

3 Model

To reduce the coupling degree of the model, OKIN uses two independent CNNs to extract the aspect and opinion feature respectively, and then, connects the two parts together by unidirectional injection. Unidirectional injection limits the direction of possible noise propagation. OKIN first ensures that opinion knowledge is accurate and then use it to assist in aspect extraction. In addition, two independent CNNs can be processed in parallel, which improves processing speed compared to LSTM-based models. The proposed model is depicted in Fig. 1.

Assume the input is a sequence of word indexes $x = \{x_1, \cdots, x_i, \cdots, x_n\}$. The output of the model is $y = \{y_1, \cdots, y_i, \cdots, y_n\}$, $y_i \in \{B, I, O\}$, where 'B' represents the starting position of the aspect term, 'I' represents the subsequent position of the aspect term, and 'O' representing the position that is not the aspect term.

3.1 Embedding Layers

The embedding module of our model also uses double embedding [13]. Let $\omega^{general} \in \mathbb{R}^{d_1 \times |V_1|}$ and $\omega^{domain} \in \mathbb{R}^{d_2 \times |V_2|}$ be the embedding matrix of general embedding and domain embedding, where d_1 and d_2 are the dimension of word vectors and $|V_1|$ and $|V_2|$ are the vocabulary size.

Through matrix $\omega^{general}$ and ω^{domain}, the embedding layer transforms the input x into a list of vectors $x^{general} = \{x_1, \cdots, x_i, \cdots, x_n\}$, where $x_i \in \mathbb{R}^{d_1}$, and $x^{domain} = \{x_1, \cdots, x_j, \cdots, x_n\}$, where $x_j \in \mathbb{R}^{d_2}$.Then we concatenate two embeddings:

$$x^0 = \left[x^{general} : x^{domain}\right], \tag{1}$$

and input x^0 into aspect extraction CNN layers and opinion knowledge injection module.

3.2 Aspect Extraction CNN Layers

Aspect extraction CNN layers contain many 1D-convolution filters and each filter has a fixed kernel size $k = 2c+1$ and step size $s = 1$. We perform convolution on c words before and after the center word $[i-c, i+c]$. For the case of less than c, we pad the zero vector to make up. For the r^{th} convolution filter in l^{th} CNN layer, it performs the following convolution operation and ReLU activation:

$$x_i^{(l+1,r)} = ReLu\left(\left(\sum_{j=-c}^{c} \omega_j^r \times x_{i+j}^{(l)}\right) + b^r\right), \tag{2}$$

where $\omega \in \mathbb{R}^{k \times d_2 \times d_1}$, $b \in \mathbb{R}^{d_2 \times 1}$ are the 1D-convolution parameters. d_1 and d_2 are the number of input channels (word vector dimension) and output channels.

Aspect Extraction CNN layers contain l_{asp} layers. The output of Aspect Extraction CNN layers is $x^{asp} = \{x_1, \cdots, x_i, \cdots, x_n\}$, where $x_i \in \mathbb{R}^{d_{asp} \times 1}$ and d_{asp} is the number of output channel of aspect extraction CNN layers.

3.3 Opinion Knowledge Injection

This module firstly uses CNN layers to extract the opinion feature, similar to the aspect extraction CNN layers. However, the difference is the number of layers and the kernel size of 1D-convolution. Opinion Extraction CNN layers contain l_{opi} layers. The output of Aspect extraction CNN layers is $x^{opi} = \{x_1, \cdots, x_i, \cdots, x_n\}$, where $x_i \in \mathbb{R}^{d_{opi} \times 1}$ and d_{opi} is the number of output channel of opinion extraction CNN layers.

In order to reasonably incorporate the opinion knowledge into each word, OKIN employs the co-attention mechanism between the output of the aspect extraction CNN layers x^{asp} and opinion extraction CNN layers x^{opi}. We apply co-attention as follows:

$$S = ReLU\left(\left(x^{asp}\right)^T \omega x^{opi}\right), \tag{3}$$

Where $\omega \in \mathbb{R}^{d_{asp} \times d_{opi}}$is the parameter being trained and $S \in \mathbb{R}^{n \times n}$is the correlation matrix. $S_{i,j}$ represents the association score of the i^{th} word as aspect word and the j^{th} word as opinion word. Then, we perform $softmax$ on each row of the correlation matrix S to obtain the attention weight between the current aspect word and each opinion word:

$$W = softmax\left(S, dim = 1\right), \tag{4}$$

where $\sum_{j=1}^{n} W_{i,j} = 1, 1 \leq i \leq n$.The opinion knowledge of the current aspect word X_i is obtained by weighting all the partial opinion according to the correlation matrix W:

$$x_i^{ok} = \sum_{j=1}^{n} W_{i,j} \times x_j^{opi}, \tag{5}$$

where $x_i^{ok} \in \mathbb{R}^{d_{opi} \times 1}$, $1 \leq i \leq n$, is the opinion knowledge of x_i. Finally, we connect the aspect feature x_{asp} with its opinion knowledge x^{ok} to get the vector containing both the the aspect information and sentimental information,

$$f^{asp} = \left[x^{asp} : x^{ok}\right]. \tag{6}$$

3.4 Loss Function

Since aspect extraction is the main part and opinion extraction is the auxiliary part, we proposed a new loss function $\jmath$, which consists of two parts. The first part is computed by the aspect word with opinion knowledge f^{asp}, and the second part is only computed by the opinion information x^{opi}.

For both f^{asp} and x^{opi}, we applied a fully-connected layer and a $softmax$ to calculate the label distribution probability for each word. Dimension of the output of two fully-connected layers $|\nu| = 3$.

$$P\left(y^{asp} \mid f^{asp}\right) = softmax\left(\omega^{asp} f^{asp} + b^{asp}\right), \tag{7}$$

$$P\left(y^{opi} \mid x^{opi}\right) = softmax\left(\omega^{opi} x^{opi} + b^{opi}\right), \tag{8}$$

where w^{asp}, b^{asp}, w^{opi} and b^{opi} are parameters of fully-connected layer.

During the training process, for both aspect extraction and opinion extraction, we use the token-level cross-entropy error between the predicted distribution $P\left(y_i^{\tau} \mid x_i\right)$, $\tau \in \{asp, opi\}$, and the gold distribution $P\left(y_i^{\tau,g} \mid x_i\right)$ as the loss function:

$$L_{\tau} = -\sum_{i=1}^{N} P\left(y_i^{\tau,g} \mid x_i\right) \odot log\left[P\left(y_i^{\tau} \mid x_i\right)\right], \tag{9}$$

Then, the proposed loss functions of both tasks are combined to form the training objective of the entire model:

$$\jmath = L_{asp} + \alpha L_{opi}, \tag{10}$$

Where α is the hyperparameter we set.

4 Experience

To evaluate the proposed model for aspect extraction, we compared OKIN with the previous model in terms of both aspect and opinion extraction and conducted two sets of containment experiments. In addition, we visualized the attention mechanism and made some case studies on the proposed model and the baseline. All the experimental results prove that OKIN is effective for ATE.

4.1 Datasets

The statistics of two benchmark datasets from SemEval challenges [6] was shown in "Table 1". The first dataset is from the laptop domain on subtask 1 of SemEval-2014 Task 4. The second dataset is from the restaurant domain on subtask 1 of SemEval-2016 Task 5. The aspect terms and opinion terms of these two datasets were labeled as spans of characters.

Table 1. Statistics of datasets. S., A., O. represent sentence, aspect, and opinion.

		#S.	#A.	#O.	#S. w/ A.	#S. w/ O.
SemEval-14	TRAIN	3045	2358	4979	1484	2499
Laptop	TEST	800	654	1229	422	652
SemEval-16	TRAIN	2000	1743	2758	1233	1590
Restaurant	TEST	676	622	874	420	522

4.2 Comparison Models

We compare our model with the following four parts: simple original models, models that use the opinion word to assist the aspect extraction, CNN-based models, and the variations of OKIN.

- **CRF:** Conditional Random Fields with basic feature templates.
- **LSTM:** Vanilla bi-directional LSTM with pre-trained word embeddings.
- **IHS RD** [3], **DLIREC** [23], **NLANGP** [24]**:** The winning systems in the ATE subtask in SemEval ABSA challenge.
- **WDEmb** [9]**:** Enhanced CRF with word embeddings, dependency path embeddings and linear context embeddings.
- **RNCRF** [11]**:** A model that integrates recursive neural networks and conditional random fields.
- **CMLA** [12]**:** A network with coupled multilayer attentions, which propose coupled attentions to exploit the correlations among input tokens.
- **MIN** [16]**:** A model that jointly handles the extraction tasks of aspects and opinions via memory interactions and Sentimental sentence constraint.

- **DECNN** [13]**:** A CNN-based aspect extraction model with a double embeddings mechanism without extra supervision.
- **OKIN w/o** L_{opi}**:** OKIN removes the loss of opinion extraction and does not perform any supervised learning on opinion extraction, which only trains the model with the loss of aspect extraction.
- **OKIN w/o ATT:** OKIN removes the co-attention mechanism and directly stitches the output of the opinion extraction CNN layers to the output of the aspect extraction CNN layers.

4.3 Experiment Setting

We use NLTK to tokenize each sentence into a sequence of words and do not make special treatments for punctuation marks that are different from words. To ensure the fairness of the comparison, our Embedding layers and aspect extraction CNN layers are the same as DECNN [13]. The opinion extraction CNN contains four layers. The first layer ($l = 1$) is the same as the first layer in aspect extraction CNN layers. The next four layers ($l = 2, 3, 4, 5$) each contain a convolution filter ($kernelsize : 3; step : 1; padding : 1$).

We divided 150 samples from the training set as the validation set. The hyperparameter α in loss function is set to 0.7. We apply dropout after each embedding layer and CNN layer, and dropout is set to 0.6. Due to the instability of CNN training, the learning rate was set to 0.0001.

Table 2. Experimental results of aspect extraction(F1 score, %). The first ten results are copied from their papers, '-' indicates the results were not available in their papers.

Model	Laptop	Restaurant
CRF	72.77	66.96
LSTM	75.71	70.35
IHS_RD	74.55	-
DLIREC	73.78	-
NLANGP	-	72.34
WDEmb	75.16	-
RNCRF	78.42	-
CMLA	77.80	-
MIN	77.58	73.44
DECNN	81.59	74.37
OKIN w/o L_{opi}	82.11	74.17
OKIN w/o ATT	81.16	73.58
OKIN	**82.93**	**75.60**

4.4 Results and Analysis

"Table 2" shows that OKIN exceeds all previous models and achieves the best F1 values on the restaurant and laptop dataset. Compared with the winning systems of SemEval ABSA, our model has an absolute advantage of 8.38% and 3.26% on the Laptop and Restaurant datasets, respectively.

Compared to RNCRF, OKIN does not rely on dependency parsing, but the score on the Laptop dataset does exceed 4.51%. OKIN could predict more accurately when dealing with unofficial comments or comments with confusing sentence structures, which perform better than RNCRF. CMLA and MIN emphasize the interaction between aspect and opinion with high-coupling. OKIN proposes unidirectional injection to control the spread of noise, so that OKIN works better than MIN and CMLA on both datasets. To verify the effect of opinion extraction, we evaluated the accuracy of opinion extraction on RNCRF, CMLA and OKIN. As shown in "Table 3", OKIN largely outperforms the RNCRF and CMLA in the accuracy of opinion extraction. This is the effect of using unidirectional injection, we do not allow any aspect information to interfere with the training of the opinion, and directly output the result of the opinion extraction CNN layers to the fully-connected layer and the softmax.

Table 3. Experimental results of opinion extraction(F1 score, %).

Model	Laptop	Restaurant
RNCRF	79.44	-
CMLA	80.17	-
OKIN	**90.56**	**83.22**

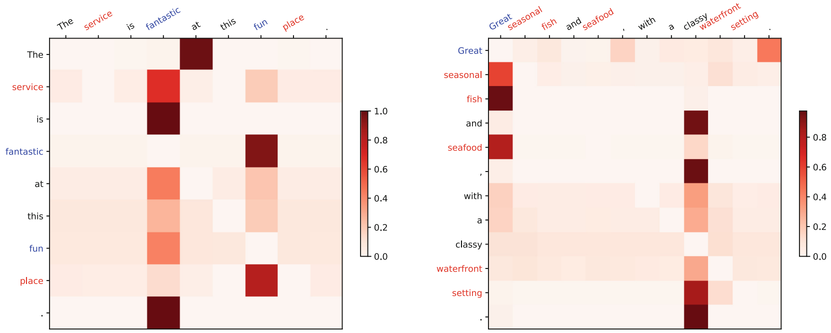
One important baseline is DECNN, which is the first model using CNN on the ATE. Our model has an accuracy of 1.34% and 1.23% higher than the DECNN on the datasets of Laptop and Restaurant, which proves that the injection of opinion knowledge can effectively assist in aspect extraction.

In addition, we conducted two sets of ablation study. When we remove the L_{opi} or attention mechanism, the effect is significantly reduced, even lower than the baseline. This shows that the addition of opinion extraction leads to the complexity of the model, which will bring noise. If the proposed model does not use opinion information reasonably, it will achieve the opposite effect, which proves the necessity of L_{opi} and attention mechanisms in the proposed model.

4.5 Attention Visualization and Case Study

As shown in Fig. 2, we visualize the association score in two difficult example sentences.

The first sentence is "The service is fantastic at this fun place.". From Fig. 2(a), we can see that when the aspect word is "*service*" (the second row),



(a) "The service is fantastic at this fun place."

(b) "Great seasonal fish and seafood, with a classy waterfront setting."

Fig. 2. Visualization of the attention. The vertical and horizontal axis represent the word as aspect and opinion, respectively. The deeper the color is, the higher the attention score is, the more the word's opinion knowledge will be injected into the aspect. (Color figure online)

"*fantastic*" is most concerned, followed by "*fun*". Similarly, when the aspect word is "*service*" (the eighth row), "*fun*" is most concerned. It proves that the attention mechanism can effectively associate the aspect word with the opinion word. Almost all words pay more attention to the opinion word. However, the aspect word focuses on the opinion word more closely to one-hot, while other non-aspect word's attention is more dispersed. Therefore, it is more beneficial for the aspect to get more opinion knowledge.

The second sentence is "Great seasonal fish and seafood, with a classy waterfront setting.". The aspect terms for this sentence are "*seasonal fish*" and "*seafood*", which are connected by "*and*". For aspect extraction, if "*A*" is an aspect and when "*A* and *B*" appears, "*B*" should also be extracted as an aspect. One shortage of DECNN [13] is that it can't predict correctly when requiring the semantics of the conjunction word, while OKIN can avoid this error with the help of opinion words. For example, when "*A* and *B*" appears, and the opinion word "*C*" exists in the sentence to comment "*A* and *B*". OKIN can extract comment relationship "*C*" → "*A*" and "*C*" → "*B*" through attention, and inject knowledge of "*C*" into "*A*" and "*B*", thus it is more easier to obtain "*A*" and "*B*" both as aspect term. As shown in Fig. 2(b), we can see that the "*seasonal*" (the second row) and the "*fish*" (the third row) pay more attention to "*great*", and the "*seafood*" (the fifth row) also pays attention to "*great*". So "*seasonal fish*" and "*seafood*" are both correctly identified as aspect terms. By using the transitivity of the comment relationship on conjunctions, our model subtly addresses the errors caused by conjunctions.

As mentioned before, DECNN [13] is the baseline of our model. OKIN introduces opinion knowledge injection to assist aspect extraction and achieve better results. To show that, we pick a few example reviews from the test datasets

Table 4. Prediction comparison between our model and DECNN. The red label is the aspect term, and the blue label is the opinion term.

Input sentences	Output of DECNN	Output of OKIN
1. Seattle's BEST Winelist	*Seattle, Winelist*	*Winelist*
2. DONOT GO!	*DONOT*	*None*
3. Great seasonal fish and seafood, with a classy waterfront setting	*Fish, waterfront setting*	*Seasonal fish, seafood, waterfront setting*
4. I had to buy a wireless mouse to go with it, as I am old school and hate the pad, but knew that before I bought it, now it works great, need to get adjusted to the key board, as I am used to a bigger one and pounding	*Wireless mouse, pad, key board*	*Wireless mouse, pad, works, key board*

as presented in "Table 4". The first two examples illustrate that OKIN avoids no-commented words being identified as aspect terms, and the latter two examples illustrate that OKIN can extract difficult-to-recognize aspect terms with the help of opinion extraction. OKIN also has a good performance in short sentence (i.e., the first and the second), long sentence (i.e., the forth), the sentence with many aspect terms (i.e., the third and the forth), and sentence containing conjunctions (i.e., the third). All these examples prove that OKIN can handle all kinds of online reviews and achieve good results.

5 Conclusion

We considered the weaknesses of complex networks on ATE and provided an effective way to reduce the coupling degree of network and speed up the processing. For models that handle multitasks, we provided a unidirectional injection method that effectively limits the propagation of noise. Finally, we proposed an opinion knowledge injection network for aspect term extraction, which exceeds all existing models.

References

1. Liu, K., Xu, L., Zhao, J.: Opinion target extraction using word-based translation model. In: EMNLP-CoNLL 2012, pp. 1346–1356 (2012)
2. Jakob, N., Gurevych, I.: Extracting opinion targets in a single- and cross-domain setting with conditional random fields. In: EMNLP 2010, pp. 1035–1045 (2010)
3. Chernyshevich, M.: IHS R&D Belarus: crossdomain extraction of product features using CRF. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp. 309–313 (2014)
4. Shu, L., Xu, H., Liu, B.: Lifelong learning CRF for supervised aspect extraction. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), vol. 2, pp. 148–154 (2017)
5. Minqing, H., Liu, B.: Mining and summarizing customer reviews. In: KDD 2004, pp. 168–177 (2004)
6. Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S.: Semeval-2014 task 4: aspect based sentiment analysis. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pp. 27–35 (2014)
7. Zhang, Z., Singh, M.: Limbic: author-based sentiment aspect modeling regularized with word embeddings and discourse relations. In: EMNLP 2018, pp. 3412–3422 (2018)
8. Huang, M., Rao, Y., Liu, Y., Xie, H., Wang, F.L.: Siamese network-based supervised topic modeling. In: EMNLP 2018, vol. 37, no. 1, pp. 4652–4662 (2018)
9. Yin, Y., Wei, F., Dong, L., Xu, K., Zhang, M., Zhou, M.: Unsupervised word and dependency path embeddings for aspect term extraction (2016). arXiv preprint arXiv:1605.07843
10. He, R., Lee, W.S., Ng, H.T., Dahlmeier, D.: An unsupervised neural attention model for aspect extraction. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, pp. 388–397 (2017)
11. Wang, W., Pan, S.J., Dahlmeier, D., Xiao, X.: Recursive neural conditional random fields for aspect-based sentiment analysis (2016). arXiv preprint arXiv:1603.06679
12. Wang, W., Pan, S.J., Dahlmeier, D., Xiao. X.: Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In: AAAI, pp. 3316–3322 (2017)
13. Xu, H.. Liu, B., Shu, L., Yu, P.S.: Double embeddings and CNN-based sequence labeling for aspect extraction. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 592–598 (2018)
14. Popescu, A.-M., Etzioni, O.: Extracting product features and opinions from reviews. In: HLT-EMNLP, vol. 05, pp. 339–346 (2005)
15. Giannakopoulos, A., Musat, C., Hossmann, A., Baeriswyl, M.: Unsupervised aspect term extraction with B-LSTM & CRF using automatically labelled datasets. In: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 180–188 (2017)
16. Li, X., Lam, W.: Deep multi-task learning for aspect term extraction with memory interaction. In: EMNLP, vol. 17, pp. 2886–2892 (2017)
17. Lafferty, J., McCallum, A., Pereira, F.C.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: ICML 2001, pp. 282–289 (2001)

18. Liu, P., Joty, S., Meng, H.: Finegrained opinion mining with recurrent neural networks and word embeddings. In: EMNLP 2015, pp. 1433–1443 (2015)
19. Poria, S., Cambria, E., Gelbukh, A.: Aspect extraction for opinion mining with a deep convolutional neural network. In: Knowledge-Based Systems, pp. 42–49 (2016)
20. Moghaddam, S., Ester, M.: ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. In: SIGIR, vol. 11, pp. 665–674 (2011)
21. Liu, K., Liheng, X., Liu, Y., Zhao, J.: Opinion target extraction using partially-supervised word alignment model. In: IJCAI, vol. 13, pp. 2134–2140 (2013)
22. Zhou, X., Wan, X., Xiao, J.: Collective opinion target extraction in Chinese microblogs. In: EMNLP, vol. 13, pp. 1840–1850 (2013)
23. Toh, Z., Wang, W.: DLIREC: aspect term extraction and term polarity classification system. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval-2014), pp. 235–240 (2014)
24. Toh, Z., Su, J.: NLANGP at semeval-2016 task 5: improving aspect based sentiment analysis using neural network features. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pp. 282–288 (2016)