Enhance Weakly-Supervised Aspect Detection with External Knowledge (Student Abstract)

Zhuoming Zheng[1,2], **Yi Cai** [1,2*], **Liuwu Li** [1,2]

[1] School of Software and Engineering, South China University of Technology, Guangzhou, China
[2] Key Laboratory of Big Data and Intelligent Robot (SCUT), Ministry of Education, China
202021045449@mail.scut.edu.cn, ycai@scut.edu.cn, liuwu.li@outlook.com

Abstract

Aspect detection aims to identify aspects of reviews and is an essential up-stream task of opinion mining and so on. However, existing weakly-supervised methods suffer from lacking the ability of identifying implicit aspects with infrequent aspect terms and "*Misc*" aspects. To tackle these problems, we propose to enhance the representation of segment with external knowledge by a weakly-supervised method. Experiments demonstrate the effectiveness of our model and the improvement by incorporating external knowledge.

Introduction

Aspect detection seeks to identify the aspect categories (e.g. *Size* , *Price*) that a review segment depicts explicitly or implicitly. Recent methods focus more on unsupervised or weakly-supervised methods which do not need human labelling or just leverage a few human-annotated texts.

Existing weakly-supervised methods known as "seed-driven" methods (Angelidis and Lapata 2018) have achieved great improvements in this task. These methods aim to identify indicative aspect by choosing a few seed-words to represent each aspect. However, most of them can not identify implicit aspects effectively because of the low frequency of aspect terms (the core words utilized to identify aspect in review segments). For an example in Laptop domain, "This laptop is wonderful with its touchscreen and AHCI technology", this segment depicts aspect "*Hardware*". However, unlike the frequent words "laptop" and "technology" in this domain, the aspect terms "touchscreen" and "AHCI" may not be frequent enough to be captured. Because seed-words set is usually extracted from a very few annotated segments and most aspect terms are infrequent actually (Luo et al. 2019). Another challenge is to detect "*Misc*" aspect, which depicts no aspect and is very common in corpora. However, existing seed-driven methods are usually inapplicable because it is difficult to nominate seed-words for "*Misc*" aspect, even for language experts.

To address the above challenges, we propose to leverage external knowledge graph to infer aspect effectively with intent domain features. First, commonsense knowledge

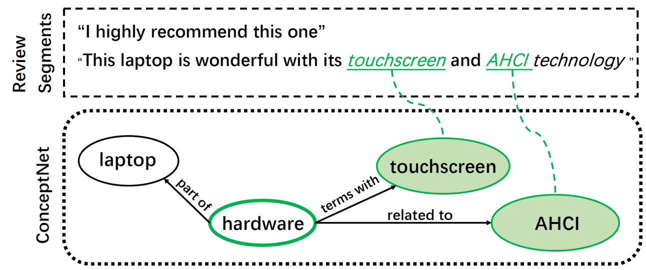

Figure 1: The first segment is labeled "*Misc*" with no domain knowledge in knowledge graph. The second segment describes "*Hardware*" aspect with meaningful domain knowledge about the aspect terms "touchscreen", "AHCI"

graph like ConceptNet has abundant knowledge which is also domain-specific. For the aspect terms "touchscreen" and "AHCI" (Figure 1), there exists their relations to corresponding aspect "hardware" in knowledge graph. It could be effective to infer the real aspect even with infrequent aspect terms by utilizing domain knowledge. Second, segments with "*Misc*" aspect label, e.g. "I highly recommend this one", usually contain less or none domain knowledge which could distinguish from not "*Misc*" ones.

Based on the analysis above, in this paper, we leverage external knowledge from knowledge graph and introduce a aspect knowledge-enhanced model. Our method is weakly-supervised and infer the aspect effectively by incorporating corresponding sub-graph . The experimental results show the effectiveness of the proposed model.

Proposed Model

Graph Construction and Self-Training

ConceptNet is a set of triples $(e_i, r_{i,j}, e_j)$ and is represented as a directed labeled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$. First, we utilize Spacy to choose nouns, adjectives, and adverbs in our review corpora as seeds set $w_i^* \in C$ and extract one-hop entity and relation with it $(w_i^*, r_{i,j}, e_j)$ to get the final sub-graph of ConceptNet for self-training, which contains domain knowledge of the whole corpora. And then We use R-GCN to get

*Corresponding author: Yi Cai, ycai@scut.edu.cn

each entity's latent representation $\mathbf{h}_i^{(l+1)}$:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_i^r}\frac{1}{c_{i,r}}\mathbf{W}_r^{(l)}\mathbf{h}_j^{(l)} + \mathbf{W}_0^{(l)}\mathbf{h}_i^{(l)}\right) \quad (1)$$

DistMult factorization is adopted for scoring where each relation r is associated with a diagonal $\mathbf{R}_r$ and then a triple $(e_i, r_{i,j}, e_j)$ is scored as :

$$f(e_i, r_{i,j}, e_j) = \mathbf{h}_i^T\mathbf{R}_{r_{i,j}}\mathbf{h}_j \quad (2)$$

Segment Reconstruction

Based on seed-driven methods, we set a matrix $\mathbf{A}$ to represent each aspect. Combined with latent domain features, each aspect is represented as $\mathbf{a}_i = \mathbf{z}_i \oplus \mathbf{h}_i$, where $\oplus$ here means "concat", $\mathbf{z}_i$ is the seed-words' weighted sum embedding while $\mathbf{h}_i$ is the average embedding of the seed-corresponding sub-graph. We also set a few learn-able vectors to represent "*Misc*" aspect. Each word $\mathbf{w}_i = \mathbf{c}_i \oplus \mathbf{h}_s$ in segment s makes attention to aspects $\mathbf{v}_s = \sum_i^n \beta_i\mathbf{w}_i$, where $\mathbf{c}_i$ is word embedding while $\mathbf{h}_s$ is the average embedding of the segment-corresponding sub-graph. Then each segment is mapped to the possibility of each aspect and reconstruct to the linear combination of aspects:

$$\mathbf{p}_s^{asp} = softmax\left(\mathbf{W}\mathbf{v}_s + \mathbf{b}\right) \quad (3)$$

$$\mathbf{r}_s = \mathbf{A}^\top\mathbf{p}_s^{asp} \quad (4)$$

Contrastive Learning

After reconstruction, we utilize contrastive loss to make each reconstructed segment similar to its original representation:

$$L_s = -\log\frac{\exp\left(sim\left(\mathbf{r}_s, \mathbf{v}_s\right)\right)}{\sum_j I_{[j\neq s]}\exp\left(sim\left(\mathbf{r}_s, \mathbf{v}_j\right)\right)} \quad (5)$$

Knowledge Distillation

After getting trained model, we obtain student BERT model by adding an attention layer and a softmax layer through knowledge distillation.

Experiment

Datasets and Baseslines

OpoSum. A review dataset from Amazon website proposed by (Angelidis and Lapata 2018) to verify the effectiveness of our method. We compare our proposed method with serveral weakly-supervised methods, including MATE (Angelidis and Lapata 2018), BERT-Stu(Karamanolakis, Hsu, and Gravano 2019), AspMem (Zhao and Chaturvedi 2020).

Experiment Results

We use the micro-f1 score as evaluation indicator, experimental results are showed in Table 1. By incorporating external knowledge, our model performs better than previous methods. It may be because with external knowledge, our model can infer the true aspect better than than other models in such highly specialized domain.

Model	TV	Keyboards	Avg.
MATE(2018)	48.8	43.5	46.2
MATE-MT(2018)	51.8	45.3	48.6
BERT(2019)	63.0	57.5	60.1
AspMem(2020)	60.0	61.8	61.0
Ours(Stu.)	**65.3**	**64.2**	**64.8**

Table 1: Experiments on dataset of different domain.

Model	TV	Keyboards	Avg.
Base	61.3	62	61.7
Base+KG (Tea.)	62.3	63.5	62.9
Base+KG+BERT (Stu.)	65.3	64.2	64.8

Table 2: Ablation experiments of proposed model.

In ablation experiments (Table 2), we denote our teacher model without external knowledge graph (KG) as base model and then compare it with teacher model (Tea.) and student (Stu.) model. Our teacher model can infer aspects better than base model without knowledge. By using knowledge distillation, student model can achieve better results than teacher which may benefit from the effectiveness of pre-trained language models for semantic understanding.

Conclusion

In this paper, we proposed a knowledge enhanced aspect-aware memory neural network. Our model incorporates external knowledge to infer aspect effectively. Experiments show that our model outperforms other models.

Acknowledgments

This work was supported by National Natural Science Foundation of China (62076100), and Fundamental Research Funds for the Central Universities, SCUT (D2210010,D2200150,and D2201300), the Science and Technology Planning Project of Guangdong Province (2020B0101100002)

References

Angelidis, S.; and Lapata, M. 2018. Summarizing Opinions: Aspect Extraction Meets Sentiment Prediction and They Are Both Weakly Supervised. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3675–3686.

Karamanolakis, G.; Hsu, D.; and Gravano, L. 2019. Leveraging Just a Few Keywords for Fine-Grained Aspect Detection Through Weakly Supervised Co-Training. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 4611–4621.

Luo, L.; Ao, X.; Song, Y.; Li, J.; Yang, X.; He, Q.; and Yu, D. 2019. Unsupervised Neural Aspect Extraction with Sememes. In *IJCAI*, 5123–5129.

Zhao, C.; and Chaturvedi, S. 2020. Weakly-supervised opinion summarization by leveraging external information. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 9644–9651.