Majority Rules Guided Aspect-Category based Sentiment Analysis via Label Prior Knowledge

Lin Li[1], Shaopeng Tang[1], Renwei Wu[1]
[1]School of Computer Science and Artificial Intelligence, Wuhan University of Technology
Wuhan, China
{cathylilin, karitown, wrw}@whut.edu.cn

Abstract

As an important fine-grained task of sentiment analysis, Aspect-Category based Sentiment Analysis (ACSA) aims to identify the sentiment polarities of pre-defined categories in text. However, due to subjectivity, the highly semantically similar text has polysemous sentiments to different people, leading to annotation difference. To this end, we propose a **MA**jority **R**ules **G**uided (MARG) for the profound understanding of this difference. Specifically, we firstly design a rule-based prompt generation, and then label word distribution is generated through an autoregression model for token-wise semantic consistency. Last but not least, the impact to the model caused by this commonly prevailing annotation difference can be mitigated by majority rules. 1) Our local majority rule is the ensemble of label word distributions, which alleviates the influence of the difference at the distribution generation stage. And 2) our global majority rule is the refinement based on the label prior knowledge of aspect categories, which further reduces the interference of the difference at the global data level. Conducted on four benchmark datasets, our MARG outperforms the state-of-the-art models by 2.43% to 67.68% in terms of F1-score and by 1.16% to 10.22% in terms of Accuracy.

Keywords: ACSA, Majority Rules, Label Prior Knowledge

1. Introduction

Aspect-Category based Sentiment Analysis (ACSA) has gradually become more and more popular in real world applications (Brauwers and Frasincar, 2023; Fu et al., 2021). As shown in Table 1, ACSA aims to distinguish aspect categories in text modality while simultaneously predicting the sentiment polarity associated with each of these categories.

Early studies (Schmitt et al., 2018; Dai et al., 2019) were interested in the category-sentiment joint neural network model by adding the sentiment label space to indicate the occurrence of each category. Later, ACSA is transferred to the paradigm of pre-training and fine-tuning. And some works paid attention to some problems in ACSA task. Jiang et al. (2019) proposed the capsule networks to better learn some sentences where there were multiple aspects and multiple sentiments. Yin et al. (2020) designed two attention-based networks to learn the contextual sentiment for the aspect and category independently and interactively. Cai et al. (2020) proposed the Hier-GCN-BERT to capture both explicit and implicit aspects. In addition, Liang et al. (2021) explored GCN and proposed AAGCN-BERT to learn the aspect-related contextual sentiment dependencies with the aspect-aware words. In recent years, prompt-tuning has risen rapidly and has been introduced into fine-grained sentiment analysis (Li et al., 2022). For example, Gao et al. (2022) and Liu et al. (2023a) have studied the fine-grained task of sentiment analysis based on prompt-tuning, achieving promising results.

However, existing methods have not paid enough attention to annotation difference which commonly appear in sentiment analysis due to the ability differences of annotators in understandings of the labeling standards. Take the two reviews in Table 1 as a case study.

	L#P	L#G	L#D
It 's truly a great laptop for the price.	0(neutral)	1(positive)	-
This laptop is a great price and has a sleek look.	1(positive)	-	1(positive)

Table 1: Two review examples (from LAP16). L#P is Laptop#Price, L#G is Laptop#General and L#D is Laptop#Design_features.

Although the descriptions of two reviews are basically the same on Aspect#Category: Laptop#Price, the sentiment polarity annotation is different. The highly semantically similar text has polysemous sentiments to different people, which leads to annotation difference. And such cases of annotation difference for semantically similar texts are not uncommon in each category. With the help of the SBERT [1], we find that the proportion of annotation difference in some datasets is as high as 17.47% (see details in Section 2.3). Since such samples will interfere with sentiment analysis performance, how to alleviate their disturbances is a challenge.

* Corresponding author.

[1]https://www.sbert.net

LREC-COLING 2024, pages 10952–10957
20-25 May, 2024.

Input	The generated prompts
Text	**Prompt 1:** Text. The quality of the food is <mask>. **Prompt 2:** Text. So the quality of food is <mask> that refers to the taste, the freshness, the texture, the consistency, the temperature, the preparation, the authenticity, the cooking or general quality. **Prompt 3:** Text. Thus, considering the taste, freshness, texture, consistency, temperature, preparation, authenticity, cooking or general quality, I feel the quality of the food is <mask>. **Prompt 4:** The quality of food is <mask> that refers to the taste, the freshness, the texture, the consistency, the temperature, the preparation, the authenticity, the cooking or general quality. Text. **Prompt 5:** Considering the taste, freshness, texture, consistency, temperature, preparation, authenticity, cooking or general quality, I feel the quality of the food is <mask>. Text.

Table 2: The generated prompts by Ruled Prompt for Food#Quality

Motivated by this, we propose a majority rules guided framework (MARG) for ACSA. As shown in Fig. 1, our MARG mainly includes Ruled Prompt, AR-PLM and Majority Rules Module. **(1)** Inspired by the rule-based approach to generate prompts (Han et al., 2022), we design Ruled Prompt by building several sub-prompts which contain subordinate relationships and thesauri of aspect categories to generate prompts (Section 2.1). **(2)** We train an **A**uto**R**egressive **P**re-trained **L**anguage **M**odel, denoted as AR-PLM (Section 2.2), to generate sentiment label word distributions and obtain token-wise semantic consistency (Radford et al., 2019). **(3)** In our Majority Rules Module, we mitigate the impact to the performance of the model caused by the difference through two majority rules. The local majority rule alleviates the influence of the difference through ensembling label word distributions at label word distribution generation stage. Furthermore, the global majority rule reduces the interference of the difference through refining the ensembled results using Bernoulli distribution and label prior knowledge at the global data level (Section 2.3).

Experiments are conducted on four benchmark datasets from SemEval 2015 and SemEval 2016, where the laptop domain contains 198 aspect categories and the restaurant domain contains 30 aspect categories. The experimental results demonstrate that our MARG outperforms other models and achieves the most remarkable performance across all four datasets. The addition of Majority Rules Module significantly boosts our framework's performance that F1-score improvement ranges from 23.45% to 38.93% and Accuracy is improved from 7.20% to 12.79%.

2. The Proposed Framework

2.1. Ruled Prompt

Manually designed prompt might contain bias, and lack of coverage capability. To address this problem, Han et al. (2022) proposed a rule-based method to generate prompts. Each sub-prompt is generated according to the generation rules of different tasks. Inspired by this, we design our Ruled Prompt for ACSA.

We set the rule that a prompt consisting of three sub-prompts which are generated by either a unary function or a binary function. The ruled prompt can be formalized as:

$$f_{text}(\cdot) \wedge f_{relation}(\cdot,\cdot) \wedge f_{words}(\cdot,\cdot)$$

where $f_{text}(\cdot)$ is a unary function to determine the sub-prompt generated from the original input text. $f_{relation}(\cdot,\cdot)$ is a binary function to determine the sub-prompt generated from aspect and category to indicate subordinate relationship of the both. $f_{words}(\cdot,\cdot)$ is a binary function to determine the sub-prompt generated from thesaurus of aspect and category containing the prior knowledge.

Due to page limitations, we only present rule-based prompts for Food#Quality. The generated prompts of Food#Quality are shown in Table 2. Generally, each generated prompt consists of three parts, i.e., $f_{text}(\cdot)$(black font), $f_{relation}(\cdot,\cdot)$ (green font) and $f_{words}(\cdot,\cdot)$(blue font), which respectively indicate the original semantic information, the subordinate relationship of aspect and category, and the prior knowledge of specific domain.

For each aspect category, the generated prompts are more favorable to fit the contextual expression of the texts for further ensuring the semantic quality of the model input. Then, the answer space as well as the mapping to the output space are from the words with conspicuous sentiment polarities. When applying them to the mask position in the prompt, they are replaced by the token **<mask>**, as shown in the left of Fig. 1, where the token **<mask>** is the prediction object.

2.2. AR-PLM

As mentioned earlier, we propose the Autoregressive Prompting to stimulate the potential of the pre-trained language model and generate labels with token-wise semantic consistency. Generative models such as BART, XLNet, GPT can be applied to our framework, but due to the limitation of the experimental hardware, we choose the moderate XLNet to validate our idea. Pre-Trained XLNet (Yang et al., 2019) is employed as the pre-trained language model for our prompt-tuning which combines the advantages of AutoRegressive LM and AutoEncoder LM. Specifically, for training of our prompting,

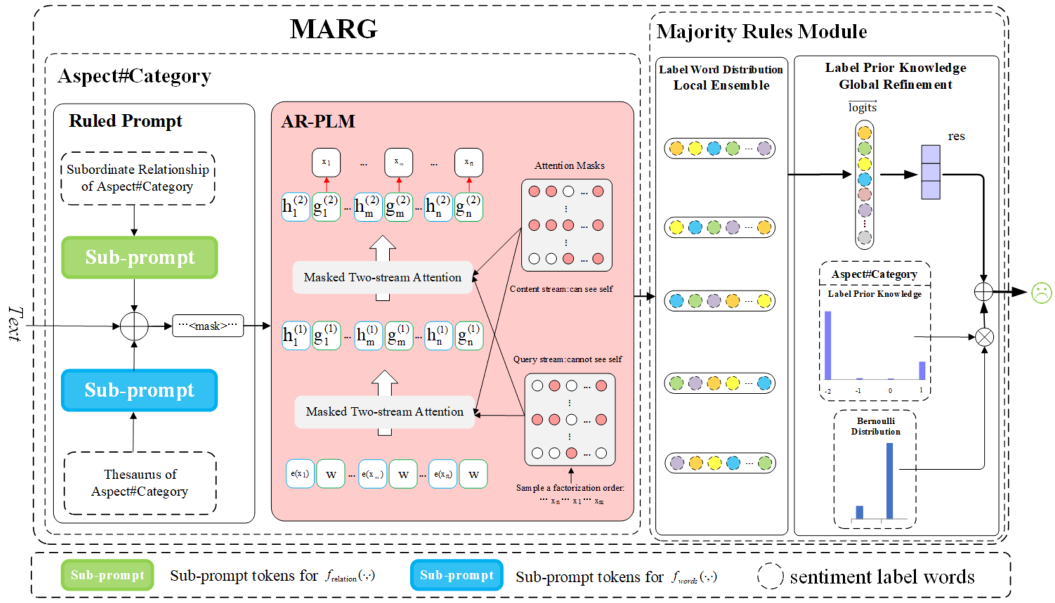

Figure 1: The framework overview of our MARG

the model input is fed to XLNetLMHeadModel [2], and the corresponding label word in the answer space is used as the training label. The loss function is the popular cross-entropy.

2.3. Majority Rules Module

In order to objectively analyze the annotation difference in the datasets, we perform statistics for SemEval 2015 and SemEval 2016 datasets using SBERT, which is the state-of-the-art model for calculating semantic similarity. In Fig. 2, the average proportion of different annotation samples with text similarity higher than 0.75 is as high as 17%, showing the difference exists objectively. Therefore, it is necessary to mitigate their interference.

To mitigate the interference of the annotation difference on the performance of ACSA, we design a majority rules module which contains two parts: 1) The local majority rule is the ensemble of label word distributions. 2) The global majority rule is the refinement based on the label prior knowledge of aspect categories.

Local Majority Rule. The commonly prevailing difference might affect the label word distribution generation, and we ensemble the label word distributions to alleviate the influence of the difference at label word distribution generation stage. Specifically, for each aspect category, the logits of the corresponding words of multiple prompts are summed to achieve better coverage capability.

Global Majority Rule. Besides using Ensemble to mitigate the impact of the difference locally, we design Refinement to contribute to the mitigation of the difference globally. We think the label distribution of each aspect category is a kind of label prior knowledge, which further reduces the impact at the global data level. Specifically, for each category, we count the frequencies of the four sentiment polarities (i.e. negative, neutral, positive and not mentioned) in a train dataset as label prior knowledge $K \in R^4$. However, this prior knowledge K is not necessarily entirely beneficial. Inspired by the BERT MLM task, we introduce the Bernoulli distribution to determine whether the label prior knowledge K is considered or not (p=0.15, following BERT). The Bernoulli distribution is expressed as follows:

$$P(x) = p^x(1-p)^{1-x} = \begin{cases} p\,, & x=1 \\ 1-p\,, & x=0 \end{cases} \quad (1)$$

where $P(x)$ denotes the probability of the random variable x, p denotes the probability of taking 1, and $(1-p)$ denotes the probability of taking 0. The random variable x can only take on 0 or 1.

We design trainable weight parameters α for K to allow the label prior knowledge to be adapted to the sentiment analysis of different aspect categories. Combining the label prior knowledge and Bernoulli distribution, the refined label prior knowledge $F(K)$

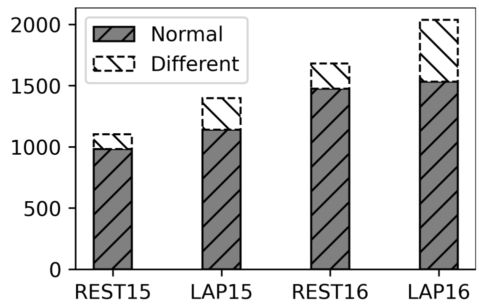

Figure 2: Annotation difference in datasets

[2]https://huggingface.co/docs/transformers/v4.17.0/en/model_doc/xlnet

is as in Equation (2):

$$F(K) = bernoulli(\alpha K, (1-p)) \quad (2)$$

where $bernoulli$ is the Bernoulli operation. If the probability (1 - p) taking x to 0, there is probability p that makes αK available. And further specifically each sentiment polarity dimension of αK is set with mutually independent probability $(1-p)$.

The label prior knowledge refinement and the final output are expressed as follows:

$$y = argmax(softmax(F(K) \bigoplus res)) \quad (3)$$

where as shown in Fig. 1, res is the set of semantic polarity logits produced by our local majority rule, and $\bigoplus$ denotes corresponding addition based on the polarities.

3. Experiments

3.1. Datasets and Experimental Settings

Our experiments are performed on public SemEval 2015 and SemEval 2016. Each of them contains both of restaurant and laptop domains.

The official training set is randomly split into the ratio of 9:1 as training, validation set. XLNetLMHeadModel is loaded with the weight of xlnet-base-cased, and refer readers to the URL [3] for the detailed setting. AdamW optimizer is adopted and its learning rate and batch size are set to 1e-5 and 6, respectively. The maximum sentence length is set to 512. The epoch of the training iteration is 100. We adopt an early stop strategy, and the training process will be ended if the performance on the validation set is not improved within 10 epochs.

To assess the generalization of MARG across different training and test sets, we execute 10 runs for each experiment. The training and validation sets are randomly divided for each run, and the test sets are the same to have a fair comparison with the state-of-the-art models.

3.2. Main Results

Overall Performance: As shown in Table 3, We divide baselines into two main types: non-LLM and LLM-based. Obviously, The non-LLM solutions are generally less effective than the LLM-based ones. Compared to AddOneDim-LSTM from scratch, MARG improves F1-score by 67.68% on REST16. Prompt_ACSA is based on pre-trained model and prompt-tuning, and MARG improves F1-score by 12.72% on LAP15 compared to it. Among other five baselines in LLM-based solutions, AAGCN-BERT achieved good performances, and our MARG further exceeds it with 4.07% improvement F1-score on REST15.

Ablation Study: To better analyze the effect of Refinement and Ensemble, we conduct two ablation experiments. First, we remove the refinement and the experimental results show a decrease of Accuracy and F1-score. For example, F1-score decreases from 84.68% to 82.12% on REST16. This suggests that the global majority rule works. Next, we additionally remove the ensemble and the experimental results decreases largely. For example, F1-score decreases from 88.94% to 80.56% on REST15. This indicates that the difference will clearly affect the performance of the model. MARG outperforms all in terms of Accuracy and F1-score, indicating that Majority Rules Module can mitigate the interference through two majority rules.

Additional Analysis: Considering the datasets are not large, we conduct a double-independent sample T-test on the corresponding prediction results (Acc. and F1-score) for MARG and some representative baselines. As shown in Table 4, the p values of MARG and CapsNet-BERT, MIMLLN-BERT, and Prompt_ACSA are 0.017, 0.028, and 0.041, respectively, which are less than 0.05.

This shows that the effect of MARG has some statistical significance and reliability. To further validate the power of Majority Rules Module on the large-scale dataset, we conduct extensive experiments on the Challenger 2018 (120K, Chinese) [4]. As shown in Table 5, the addition of the module can bring about an increase in results, which indicates it can improve the model on large-scale datasets as well.

Case Study: Table 6 shows the logits outputs for "Vanison was good but not amazing." on FOOD#QUALITY at several stages of MARG. For "Vanison was good but not amazing." on FOOD#QUALITY, its gold label is neutral. (1) After AR-PLM, AR-PLM generates five different sets of logits outputs through the five templates. We can find that the predictions of merely two templates yield correct results, which indicates that the effect is different between the different templates. (2) After Ensemble, the res logits is {"great": 21.2674, "average": 19.6598, "poor":9.1598}. It is not sufficiently effective just by local majority rule. (3) The adaptive prior knowledge on FOOD#QUALITY αK is [pos : neu : neg = 8:10:7]. After Refinement, once it is considered, the result of the global majority rule is {"great": 29.2674, " average": 29.6598, "poor":16.1598}, so this text on FOOD#QUALITY is predicted to neutral. Through the global majority rule, MARG predicts the correct result. In this case, the global majority rule plays a fundamental part.

[3] https://huggingface.co/docs/transformers/v4.17.0/en/model_doc/xlnet#transformers.XLNetLMHeadModel

[4] https://github.com/xiaoyong-z/2018-AIchallenger_emotion_analysis_27th

Solution Type	Model	REST15		LAP15		REST16		LAP16	
		ACC.	F1	ACC.	F1	ACC.	F1	ACC.	F1
non-LLM	AddOneDim-LSTM (2018, EMNLP)	-	37.32	-	-	-	50.50	-	-
	CapsNet (2019, EMNLP)	78.14	61.56	74.71	61.75	83.79	61.36	76.31	61.07
	GIN (2020, NLPCC)	81.17	62.38	75.93	63.18	87.05	65.03	78.92	62.93
	MIMLLN (2020, EMNLP)	78.27	60.59	75.30	61.39	85.76	63.52	78.57	62.63
	AAGCN (2021, EMNLP	82.79	67.43	80.02	65.87	88.32	72.55	81.76	65.96
LLM-based	CapsNet-BERT (2019, EMNLP)	81.89	61.85	82.19	59.75	86.50	62.12	80.53	61.03
	GIN-BERT (2020, NLPCC)	83.96	66.03	82.97	65.29	89.47	74.87	82.76	63.77
	MIMLLN-BERT (2020, EMNLP)	82.76	65.10	82.98	62.36	88.12	73.05	82.57	63.26
	Hier-GCN-BERT (2020, COLING)	-	64.23	-	62.13	-	74.55	-	54.15
	AAGCN-BERT (2021, EMNLP)	87.92	71.75	85.82	72.39	92.83	80.77	85.24	69.68
	Prompt_ACSA (2023b, ACM Comput. Surv.)	85.87	69.72	83.43	65.78	90.59	77.86	85.38	66.75
LLM-based	MARG (ours)	**88.94**	**74.67**	**87.05**	**74.15**	**95.34**	**84.68**	**87.89**	**71.65**
	w/o Global Refinement	87.34	73.98	85.90	72.67	93.25	82.12	87.33	71.40
	w/o Global Refinement & Local Ensemble	80.56	60.18	81.20	59.24	84.53	60.95	78.55	58.04

Table 3: Experimental results on SemEval 2015 and SemEval 2016 (%). The score marked as **bold** means the best performance among all models. The score marked with an underline means the best one among the baselines.

	MARG	
	t statistic	p value
CapsNet-BERT	-2.72	0.017
MIMLLN-BERT	-2.44	0.028
Prompt_ACSA	-2.25	0.041

Table 4: Significance test of MARG experimental results

Model	Precision	Recall	F1
MARG	**72.74**	70.15	**71.00**
w/o Refinement	72.58	**70.24**	70.77
w/o Refinement & Ensemble	69.06	66.15	68.32

Table 5: The results of extensive experiments on the Challenger 2018 (%)

Stage	Great	Average	Poor
After AR-PLM	2.1392	**3.4324**	2.2043
	5.4594	3.9346	1.9407
	4.6466	**4.9534**	1.5372
	4.1344	3.5143	1.6502
	4.8878	3.8251	1.8274
After Ensemble	**21.2674**	19.6598	9.1598
After Refinement	29.2674	**29.6598**	16.1598

Table 6: The logits outputs for a running sample, "Vanison was good but not amazing.", on FOOD#QUALITY at several stages of MARG. The logit marked as **bold** means that the prediction result of the logits set at the current stage.

4. Conclusions

This paper studies how to mitigate the impact of the annotation difference both locally and globally, and proposes a majority rules guided framework called MARG. Experiments show that it outperforms the state-of-the-art models and still works on the large-scale dataset, demonstrating its efficacy.

Limitations

Although our proposed MARG is able to achieve better results, there are some limitations. On the one hand, MARG is effective on the large-scale dataset in Chinese, and yet its effectiveness on the large dataset in English and in a real-world system is up for exploration. We might be able to enhance the ability in few-shot scenarios (Brown et al., 2020) to enhance its generalization. On the other hand, there is still a manual involvement in Ruled Prompt. PPT (Pre-trained Prompt Tuning) based on continuous prompts Gu et al. (2022) recently proposed is a promising approach. It might help to improve the performance of the model.

Acknowledgements

This work is partially supported by NSFC, China (No.62276196).

Bibliographical References

Gianni Brauwers and Flavius Frasincar. 2023. A survey on aspect-based sentiment classification. *ACM Comput. Surv.*, 55(4):65:1–65:37.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Hongjie Cai, Yaofeng Tu, Xiangsheng Zhou, Jianfei Yu, and Rui Xia. 2020. Aspect-category based sentiment analysis with hierarchical graph convolutional network. In *Proceedings of the 28th International Conference on Computational Lin-*

guistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020, pages 833–843.

Zehui Dai, Wei Dai, Zhenhua Liu, Fengyun Rao, Huajie Chen, Guangpeng Zhang, Yadong Ding, and Jiyang Liu. 2019. Multi-task multi-head attention memory network for fine-grained sentiment analysis. In *Natural Language Processing and Chinese Computing - 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9-14, 2019, Proceedings, Part I*, volume 11838 of *Lecture Notes in Computer Science*, pages 609–620.

Yujie Fu, Jian Liao, Yang Li, Suge Wang, Deyu Li, and Xiaoli Li. 2021. Multiple perspective attention based on double bilstm for aspect and sentiment pair extract. *Neurocomputing*, 438:302–311.

Tianhao Gao, Jun Fang, Hanyu Liu, Zhiyuan Liu, Chao Liu, Pengzhang Liu, Yongjun Bao, and Weipeng Yan. 2022. LEGO-ABSA: A prompt-based task assemblable unified generative framework for multi-task aspect-based sentiment analysis. In *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022*, pages 7002–7012.

Yuxian Gu, Xu Han, Zhiyuan Liu, and Minlie Huang. 2022. PPT: pre-trained prompt tuning for few-shot learning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 8410–8423.

Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, and Maosong Sun. 2022. PTR: prompt tuning with rules for text classification. *AI Open*, 3:182–192.

Qingnan Jiang, Lei Chen, Ruifeng Xu, Xiang Ao, and Min Yang. 2019. A challenge dataset and effective models for aspect-based sentiment analysis. 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 2019, Hong Kong, China, November 3-7, 2019*, pages 6279–6284.

Lin Li, Dan Liu, Lingyun Zhao, Jianwei Zhang, and Jinhang Liu. 2022. Evidence mining for interpretable charge prediction via prompt learning. *IEEE Transactions on Computational Social Systems*, pages 1–11.

Yuncong Li, Cunxiang Yin, Sheng-hua Zhong, and Xu Pan. 2020. Multi-instance multi-label learning networks for aspect-category sentiment analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 3550–3560.

Bin Liang, Hang Su, Rongdi Yin, Lin Gui, Min Yang, Qin Zhao, Xiaoqi Yu, and Ruifeng Xu. 2021. Beta distribution guided aspect-aware graph for aspect category sentiment analysis with affective knowledge. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 208–218.

Dan Liu, Lin Li, Xiaohui Tao, Jian Cui, and Qing Xie. 2023a. Descriptive prompt paraphrasing for target-oriented multimodal sentiment classification. In *Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, December 6-10, 2023*, pages 4174–4186.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023b. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Comput. Surv.*, 55(9):195:1–195:35.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Martin Schmitt, Simon Steinheber, Konrad Schreiber, and Benjamin Roth. 2018. Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 1109–1114.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime G. Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 5754–5764.

Rongdi Yin, Hang Su, Bin Liang, Jiachen Du, and Ruifeng Xu. 2020. Extracting the collaboration of entity and attribute: Gated interactive networks for aspect sentiment analysis. In *Natural Language Processing and Chinese Computing - 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14-18, 2020, Proceedings, Part I*, volume 12430 of *Lecture Notes in Computer Science*, pages 802–814.