Identifying Topic and Cause for Sarcasm: An Unsupervised Knowledge-enhanced Prompt Method

Minjie Yuan[1,2], Qiudan Li[1], Xue Mao[1,2], Daniel Dajun Zeng[1,2]
[1]State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Science
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences
Beijing, China
{yuanminjie2021, qiudan.li, maoxue2022}@ia.ac.cn, zengdaniel@outlook.com

ABSTRACT

Sarcasm is usually emotional and topical. Mining the characteristics of sarcasm semantics in different emotional tendencies and topic expressions helps gain insight into the sarcasm cause. Most of the existing work detect sarcasm or topic label based on a supervised learning framework, which requires heavy data annotation work. To overcome the above challenges, inspired by the multi-task learning framework, this paper proposes an unsupervised knowledge-enhanced prompt method. This method uses the similarity interaction mechanism to mine the hidden relationship between the sarcasm cause and topic, which integrates external knowledge, such as syntax and emotion, into the prompting and generation process. Additionally, it identifies the sarcasm cause and topic simultaneously. Experimental results on a real-world dataset verify the effectiveness of the proposed model.

CCS CONCEPTS

• Computing methodologies → Natural language processing; • Human-centered computing → Social network analysis; • Information systems → Data mining

KEYWORDS

Topic and Sarcasm Cause Identification, Unsupervised Learning, Prompt and Generation, Sarcasm Analysis

ACM Reference format:

Minjie Yuan, Qiudan Li, Xue Mao and Daniel Dajun Zeng. 2023. Identifying Topic and Cause for Sarcasm: An Unsupervised Knowledge-enhanced Prompt Method. In *Proceedings of the Web conference 2023 (WWW'23 Companion). ACM, Austin, TX, USA, 4 pages.* https://doi.org/10.1145/3543873.3587343

1 Introduction

In social media, mining and understanding sarcasm help to analyze the deep semantics of text. Sarcasm is closely related to various external factors such as emotional attitudes and topic backgrounds, which makes it difficult to mine based on single text information. Previous researches [1][2][3] indicate that sarcastic expressions contain inconsistent emotional tendencies between literal meaning and implicit semantics, and different topics provide important background information for sarcasm. Correspondingly, modeling the sarcasm cause helps to understand the emotion and deep implications of sarcastic texts, which could mine topics accurately [4]. The complex internal relationships among emotion, topic and sarcasm cause jointly promote the analysis of sarcastic expressions.

Existing studies have demonstrated the association between sarcasm and elements, such as topic, external knowledge, and sentiment. Liang et al. [3] introduced topics as the target of sarcasm and constructed templates in the pattern-exploiting form to predict sarcasm labels, demonstrating the enhancement effect of topic information on detecting sarcastic expressions. Li et al. [5] used a commonsense knowledge reasoning approach based on an attention mechanism, which considers generated knowledge information, such as sentiment and event relationships, resulting in improved performance of sarcasm detection. Part-Of-Speech (POS) information embodies linguistic knowledge in text, and integrates emotional information, which enhances the emotional representation of text, and promotes the development of sentiment analysis and related tasks [6]. Multi-task learning frameworks provide an approach to capture mutually enhanced information among different elements. Chauhan et al. [2] designed a multi-channel attention mechanism to simultaneously recognize emotion, sentiment, and sarcasm within a multi-task learning framework. The existing work focuses on supervised learning methods, which requires a large amount of annotated data and consumes high time and labor costs. Wang et al. [7] proposed an unsupervised method for detecting sarcasm, which explored a masking and generation paradigm to capture inconsistent expressions in sarcastic sentences. Based on prompt learning text classification models, Hu et al. [8] modeled text with tasks and label descriptions, and generated category labels based on pre-trained language models, which is conducive to topic mining in the low-data scenario.

Inspired by the above work, this paper proposes an Unsupervised Knowledge-enhanced Prompt method for simultaneously Identifying sarcasm Topic and Cause (UKP-ITC), which employs knowledge masking, prompt learning and

WWW'23 Companion, April 30-May 04, 2023, Austin, TX, USA

https://doi.org/10.1145/3543873.3587343

interaction similarity matching mechanism. Specifically, POS tagging and Emotion Word Lexicon [9] are first used to mask the sentences. Then, combining the topic-based prompting, a generation language model [10] is adopted to fill the masked text. Finally, the similarity matching mechanism models the internal correlation between the topic and sarcasm cause, which are predicted jointly based on the similarity matrix in an unsupervised manner. We empirically evaluate the performance of the proposed model on a Reddit social media dataset. In summary, our main contributions are as follows:

- We formulate and define a novel research question of mining topic and sarcasm cause at the same time.
- The proposed method utilizes an unsupervised learning framework that incorporates external knowledge, such as syntax and emotion, into the prompting and generation process.
- We conduct experiments on a Reddit dataset to demonstrate the effectiveness of the method.

2 Method

2.1 Problem definition and formalization

Given a text collection $D = [C_1, C_2, \dots, C_n, S]$ with n candidate sarcasm cause sentences C_i and sarcasm S, and a predefined set of topics $T = \{t_1, t_2, \dots, t_m\}$, where n and m are the numbers of sarcasm cause sentences and topics respectively. The topic and sarcasm cause identification problem aims to obtain its sarcasm topic $y_T \in T$ and sarcasm cause $y_c \in D$ jointly.

2.2 Model

The overall framework of the proposed model is shown in Figure 1.

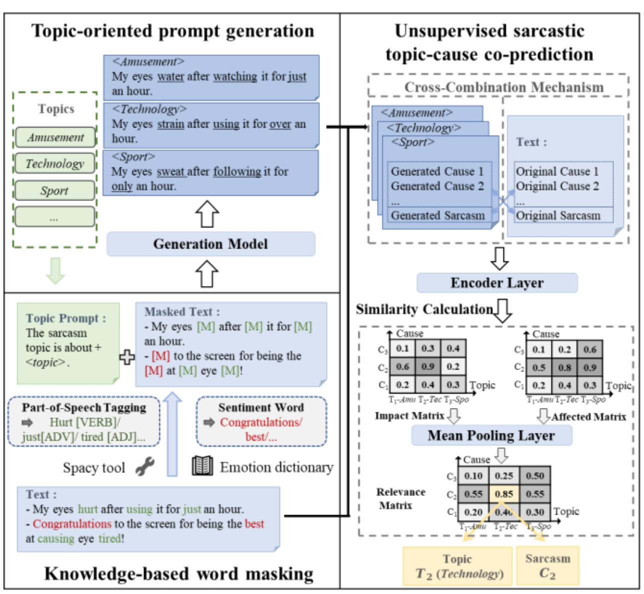

Figure 1: Overview of the proposed model

The model consists of three modules: knowledge-enhanced word masking, topic-oriented prompt generation, and sarcastic topic and cause co-prediction. Specifically, the knowledge-enhanced word masking module first uses POS tagging tool and sentiment lexicon to obtain masking words related to sarcasm expression. Then, text semantic prompt is constructed based on the candidate topic labels. The topic-oriented prompt generation module uses a generative language model to complete the masked information. Finally, the sarcastic topic and cause co-prediction module calculates the bidirectional-influence matrix of sarcasm topic-reason in an unsupervised manner. The relevance matrix obtained by mean pooling is used to simultaneously identify the corresponding topic and cause of sarcasm.

External knowledge-enhanced word masking. The use of emotion words and specific parts-of-speech (POS) words have an important impact on the semantics of text in sarcastic expression [6]. For instance, in the sarcastic text "My eyes hurt after using it for just an hour. Congratulations to the screen for being the best at causing eye tired!", the words "hurt" and "tired" convey the harm caused by the screen to the eyes, while the word "congratulations" and the emotion word "best" sarcastically express dissatisfaction with the electronic screen. Therefore, obtaining emotion words and POS words from external knowledge can help better understand sarcasm. For text $X = \{x_1, x_2, \dots, x_L\}$, the method employs the Spacy[1] tool to obtain the POS tags of the text and selects the word set IW consisting of verbs, adjectives, adverbs, and interjections. At the same time, the emotion word set SW in the NRC emotion and sentiment lexicon are obtained from the text. These two sets constitute the masking word set MW as shown in formula (1). By using the masking mechanism in (2-3), the module obtains the masked text $X^m = \{x_1, [Mask]_1, x_3, \dots, [MASK]_2, \dots, x_l\}$.

$$MW = IW \cup SW \tag{1}$$

$$x^m = MASK(x_i) = \begin{cases} [MASK], & if\ x_i \in MW \\ x_i, & if\ x_i \notin MW \end{cases} \tag{2}$$

$$X^m = MASK(X), X \in \{C_1, \dots, C_n, S\} \tag{3}$$

Topic-oriented prompt generation. Based on the guidance of different topic information, the model can generate text with diversified semantics. The topic prompt generation module assumes "The sarcasm topic is about [Topic Name]" for each candidate topic, which constructs topic-oriented prompt P_{t_j} accordingly. For each masked sarcastic and candidate cause text, this model constructs the input Z_{ij}^m for topic prompt generation sentence by sentence, formalized as formula (4).

$$Z_{ij}^m = \left[\mathrm{P}_{\mathrm{t_j}}[SEP]C_i^M[SEP]C_{i+1}[SEP]\dots \mathrm{S}\right] \tag{4}$$

When the correct candidate topic is used as a prompt, the generated words and sentences are more closely related to the original text. For example, under the "Technology" topic, the generated words "strain" and "using" are semantically similar to the original sentence, while under the "Sports" topic, the generated words "sweat" and "following" are incoherent. As shown in formula (5), this method uses generation language

[1] https://spacy.io/

models, such as BERT [10], BART [11] or T5 [12] to complete the masked text, which obtains the sarcasm and candidate cause sentence X_{ij}^g generated under each topic.

$$X_{ij}^g = Generation\_Model\left(Z_{ij}^m\right) \tag{5}$$

Unsupervised sarcastic topic-cause co-prediction. The model incorporates a cross-combination mechanism to match the generated text and the original semantic information. The text generated under each topic-oriented prompt and the original text form two types of sentence combinations, which are the sarcastic impact combination consisting of the generated cause C_{ij}^g and the original sarcasm S, the affected combination consisting of the generated sarcasm $S_{t_j}^g$ and the original cause C_i. These cross-combinations are input into an encoder to obtain the semantic representation $E_I(C_{ij})$, $E_A(C_{ij})$ of each candidate cause under each topic, and the encoding representation $E_O(C_i)$ for the original text.

The interactive similarity matching mechanism calculates the dot similarity of E_I and E_A for E_O respectively, which obtains the impact matrix M_I and the affected matrix M_A. Taking M_I as an example, the mechanism calculates similarity between E_I and E_O for each C_i and t_j sequentially, which can be expressed by the formula (6-7). Correspondingly, the affected matrix M_A is calculated by E_A and E_O.

$$Sim_I(C_{ij}) = E_I(C_{ij}) \cdot E_O(C_i) \tag{6}$$

$$M_I = \left[Sim_I\left(C_{ij}\right)\right]_{m\times n} \tag{7}$$

Finally, the model obtains the topic-cause relevance matrix M_R by mean-pooling of the impact matrix M_I and the affected matrix M_A. And the sarcastic topic and cause are jointly predicted based on the maximum value of the relevance matrix in an unsupervised manner, shown as formula (8).

$$y_C, y_T = argmax\left(Mean\_Pooling(M_I, M_A)\right)_{m,n} \tag{8}$$

3 Experimental Analysis

3.1 Dataset

To verify the effectiveness of the proposed topic and sarcasm cause mining model, we constructed a dataset based on Reddit, which is a famous social website, from 01 Jan. 2021 to 01 May. 2021. The raw data we collected contains a total of 100k sentences, including posts and replies. First, sarcasm sentences are filtered based on the "/s" tag, and we get posters and replies for sarcasm and cause sentences. Then, two annotators are required to label the topic, sarcasm cause and non-cause sentences. The Fleiss' kappa coefficient of the final annotation result is 0.78. After preprocessing and annotation, the dataset included 1495 texts containing sarcasm and candidate cause sentences, which cover the topics of technology, amusement and sports.

3.2 Evaluation metrics and baseline methods

When evaluating the performance of topic and sarcasm cause detection methods, the multi-classification task can be viewed as two independent unsupervised tasks. For the sarcasm cause recognition task, BERT [10] and Word2vec+Similarty [13] are adopted as baselines, the Precision, Recall, F1 of the cause sentences are used as evaluation indicators. For topic mining tasks, we compare our model with two baselines, NPM [14] and entailment [15] methods. And the micro-average values of Precision, Recall, F1, and Accuracy of topics are evaluation indicators.

Baseline methods for sarcasm cause detection are as follows.

- BERT [10]: This method employs a supervised BERT model to detect sarcasm cause sentences. After concatenating the embedding of the sarcasm and cause sentences, a linear classifier is used to identify the sarcasm cause.
- Word2vec+Similarity [13]: Word2vec is used to obtain the representation of the words, and the embeddings of words are averaged as the representation of the sentence. This method calculates the cosine similarity between candidate cause and sarcasm sentences, which selects the one with the largest similarity as the cause sentence.
- UKP-ITC w/o prompt: A variant of UKP-ITC model that does not consider the topic prompt in the generation phase.

Baseline methods for topic mining are as follows.

- NPM [14]: This method pre-trains a nonparametric masked language model for zero-shot text classification. To compare with our model, NPM is used without a retrieve-and-generate approach.
- Entailment [15]: Entailment approach treats zero-shot text classification as a textual entailment problem. It constructs a hypothesis by filling a candidate label into the query "the text is about [label name]", and asks the pre-trained model whether this hypothesis is true.

3.3 Experimental settings

For the supervised BERT model, bert-base-uncased is used for fine-tuning, and the result is the average of 5-fold cross-validation. For the proposed model, masked POS tags selects ['VERB', 'ADJ', 'ADV', 'INTJ'], and the algorithm during generation is a greedy search after removing stop words. The text generation model and representation encoder are pre-trained model bert-base-uncased.

3.4 Experimental results and discussion

Comparison for sarcasm cause detection. Table 1 shows the performance comparison of our proposed model with baseline methods for sarcasm cause detection. For the supervised method, the F1 of BERT reached 0.5527. The UKP-ITC method achieves the best results on P, R and F1, which are 0.5317, 0.6065 and 0.5666. The results verify that UKP-ITC method could capture the deep semantic relations between sarcasm and cause, and effectively improve the performance of sarcasm cause detection in an unsupervised manner. To verify the role of topic prompt in identifying the sarcasm cause, we apply a variant of UKP-ITC, which removes the topic-based prompt in generation stage. The P, R, and F1 values are 0.5249, 0.5855 and 0.5536 respectively. The reason may be that topic information helps to generate more coherent sentences.

Table 1: Performance of methods for sarcasm cause detection

Type	Method	P	R	F1
Supervised	BERT	0.5285	0.5800	0.5527
Unsupervised	Word2vec+sim	0.5171	0.4775	0.4965
	UKP-ITC $_{\text{w/o topic}}$	0.5249	0.5855	0.5536
	UKP-ITC	**0.5317**	**0.6065**	**0.5666**

Comparison for topic mining. Table 2 reports the experimental results of topic mining based on unsupervised methods. The NPM method demonstrates the highest precision with 71.10%. The Entailment model achieves F1 and accuracy of 0.5981 and 0.5926 by pretraining on entailment task to capture the associations between the topic label and text. Compared with baselines, the UKP-ITC model could mine topics more accurately, where the F1 and Acc values reach 60.65% and 60.00%. The possible reason is that the UKP-ITC model captures the semantic information of sarcasm topic by generating and similarity matching mechanism.

Table 2: Performance of methods for topic mining

Method	P	R	F1	Acc
NPM	**0.7110**	0.5485	0.5262	0.5485
Entailment	0.7097	0.5926	0.5981	0.5926
UKP-ITC	0.6349	**0.6000**	**0.6065**	**0.6000**

4 Conclusion

In this paper, we formulate a novel research question of detecting topic and cause for sarcasm simultaneously, and propose an unsupervised prompting and generating method, which fuses external knowledge and topic information interactively. Experimental results demonstrate the efficacy of the method. Future work can analyze fine-grained information related to sarcasm, such as the stance and users' personality.

ACKNOWLEDGMENTS

Qiudan Li is the corresponding author. This work was partially supported by the National Key Research and Development Program of China under Grant No. 2020AAA0103405, the National Natural Science Foundation of China under Grant No.62071467, 72293575, 62141608, and the Strategic Priority Research Program of Chinese Academy of Sciences under Grant No. XDA27030100.

REFERENCES

[1] Yiyi Liu, Yequan Wang, Aixin Sun, Xuying Meng, Jing Li, and Jiafeng Guo. 2022. A Dual-Channel Framework for Sarcasm Recognition by Detecting Sentiment Conflict. *In Findings of the Association for Computational Linguistics: NAACL 2022.* 1670–1680.

[2] Dushyant Singh Chauhan, Dhanush S R, Asif Ekbal, and Pushpak Bhattacharyya. 2020. Sentiment and Emotion help Sarcasm? A Multi-task Learning Framework for Multi-Modal Sarcasm, Sentiment and Emotion Analysis. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.* 4351–4360.

[3] Bin Liang, Zijie Lin, Bing Qin, and Ruifeng Xu. 2022. Topic-Oriented Sarcasm Detection: New Task, New Dataset and New Method. *In Proceedings of the 21st Chinese National Conference on Computational Linguistics*. 557–568.

[4] Hejing Liu, Qiudan Li, Zaichuan Tang, and Jie Bai. 2021. An Attention Based Multi-view Model for Sarcasm Cause Detection (Student Abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*, 15833–15834.

[5] Jiangnan Li, Hongliang Pan, Zheng Lin, Peng Fu, and Weiping Wang. 2021. Sarcasm Detection with Commonsense Knowledge. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 2, 3192–3201.

[6] Pei Ke, Haozhe Ji, Siyang Liu, Xiaoyan Zhu, and Minlie Huang. 2020. SentiLARE: Sentiment-Aware Language Representation Learning with Linguistic Knowledge. *In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP).* 6975–6988.

[7] Rui Wang, Qianlong Wang, Bin Liang, Yi Chen, Zhiyuan Wen, Bing Qin, and Ruifeng Xu. 2022. Masking and Generation: An Unsupervised Method for Sarcasm Detection *(SIGIR '22)*. 2172–2177.

[8] Shengding Hu, Ning Ding, Huadong Wang, Zhiyuan Liu, Jingang Wang, Juanzi Li, Wei Wu, and Maosong Sun. 2022. Knowledgeable Prompt-tuning: Incorporating Knowledge into Prompt Verbalizer for Text Classification. *In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics.* 2225–2240.

[9] Saif M. Mohammad and Peter D. Turney. 2013. CROWDSOURCING A WORD–EMOTION ASSOCIATION LEXICON. *Computational Intelligence* 29, 3 (2013), 436–465.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *In Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* 4171—4186.

[11] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7871–7880.

[12] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* 21, 1, Article 140, 67 pages.

[13] Tomas Mikolov, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *In International Conference on Learning Representations.*

[14] Sewon Min, Weijia Shi, Mike Lewis, Xilun Chen, Wen tau Yih, Hanna Hajishirzi, and Luke Zettlemoyer. 2022. Nonparametric Masked Language Modeling. *ArXiv abs/2212.01349 (2022)*.

[15] Wenpeng Yin, Jamaal Hay, and Dan Roth. 2019. Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach. *In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-JCNLP)*. 3914–3923.