Detecting Emotional Incongruity of Sarcasm by Commonsense Reasoning

Ziqi Qiu[†], Jianxing Yu[*†], Yufeng Zhang, Hanjiang Lai, Yanghui Rao, Qinliang Su, Jian Yin

School of Artificial Intelligence, Sun Yat-sen University, Zhuhai, 519082, China

School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, 510006, China

Key Laboratory of Sustainable Tourism Smart Assessment Technology, Ministry of Culture and Tourism

Pazhou Lab, Guangzhou, 510330, China

{qiuzq25,zhangyf283}@mail2.sysu.edu.cn

{yujx26,laihanj3,raoyangh,suqliang,issjyin}@mail.sysu.edu.cn

Abstract

This paper focuses on sarcasm detection, which aims to identify whether given statements convey criticism, mockery, or other negative sentiment opposite to the literal meaning. To detect sarcasm, humans often require a comprehensive understanding of the semantics in the statement and even resort to external commonsense to infer the fine-grained incongruity. However, existing methods lack commonsense inferential ability when they face complex real-world scenarios, leading to unsatisfactory performance. To address this problem, we propose a novel framework for sarcasm detection, which conducts incongruity reasoning based on commonsense augmentation, called *EICR*. Concretely, we first employ retrieval-augmented large language models to supplement the missing but indispensable commonsense background knowledge. To capture complex contextual associations, we construct a dependency graph and obtain the optimized topology via graph refinement. We further introduce an adaptive reasoning skeleton that integrates prior rules to extract sentiment-inconsistent subgraphs explicitly. To eliminate the possible spurious relations between words and labels, we employ adversarial contrastive learning to enhance the robustness of the detector. Experiments conducted on five datasets demonstrate the effectiveness of *EICR*.

1 Introduction

Sarcasm detection aims to endow machines with the ability to identify the emotional reversal between the literal word and its true intention about the given statement. This task holds significant potential to benefit various real-world applications, such as sentiment analysis (Chen et al., 2024b), social opinion analysis (Okawa and Iwata, 2022), and political intent identification (Bülow and Johann, 2023). According to Joshi et al. (2017), the key to sarcasm detection is to discern subtle incongruity in the statement. The simple cases of sarcasm typically involve only the shallow meaning of the text. For example, an obvious sarcastic statement like "*I am so happy that the car broke down*" uses the word '*happy*', but the author is unhappy because the car '*breakdown*'. However, sarcasm on social platforms often arises from complex real-world scenarios, making it less manifest and more difficult to detect (Oprea and Magdy, 2020). It involves complex background knowledge and social relationships, where simple word matching is insufficient to grasp the full semantic meaning (Li et al., 2021). As shown in Figure 1, *Trump's* sarcastic intentions need to be deduced from multiple relevant clues in the contexts, i.e., implicit commonsense knowledge "*Trump aborted PA*" and "*Trump and Biden are hostile*". Here, commonsense knowledge refers to the well-established fact and emotional causality that people are familiar with. Although they do not appear in the statement, it is hard to find the incongruity without them due to the incomplete context. Detecting such complex sarcasm requires a comprehensive understanding of the semantics in the statement and even resorting to external commonsense to make inferences. However, research

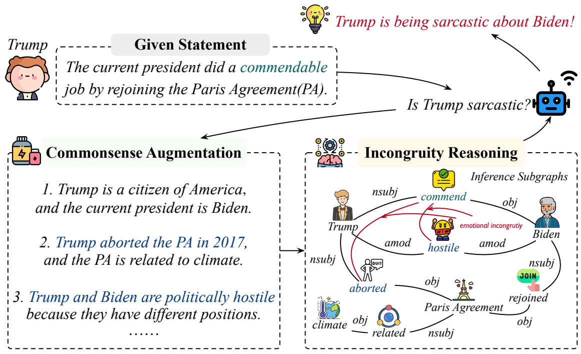


Figure 1: A sarcasm detection example that needs a good commonsense reasoning ability to identify the sarcastic intention of *Trump*.

[*]Corresponding author.

[†]These authors have contributed equally to this work.

Proceedings of the 31st International Conference on Computational Linguistics, pages 9062–9073

on commonsense reasoning in such complex scenarios is still less investigated, so we propose a new research topic to fill this gap in sarcasm detection.

Through our investigation, we discover that it requires two steps to identify complex incongruity in sarcasm detection, including retrieving reliable commonsense and conducting the accurate reasoning process. For the first step, earlier rule-based and attention-based (Babanejad et al., 2020) approaches struggle to generalize effectively across diverse sarcasm patterns due to the incomplete context. To provide the essential commonsense, existing knowledge-enhanced (Min et al., 2023) methods utilize knowledge graphs *KGs* (Liu et al., 2022) or pre-trained language models *PLMs* (Yao et al., 2024) as augmented tools. However, *KGs* are usually constrained by their fixed knowledge scope and lacked adaptability to evolving contexts, while *PLMs* suffer from hallucinations (Wei et al., 2024) due to their reliance on implicitly parametric knowledge. As a result, they can not always provide suitable commonsense for social instances. For the second step, existing methods often employ graph structures (Lou et al., 2021), which are effective at capturing long-range dependencies and facilitating intricate information interactions. However, these methods typically rely on coarse-grained, global graph representations extracted by *GCNs* (Yu et al., 2023b), but overlooking the fine-grained reasoning necessary for identifying subtle incongruity. For instance, fine-grained incongruity reasoning needs to capture the *Trump's* inconsistent attitude towards *Biden* and *PA*. Essentially, their detection results depend solely on uninterpretable high-dimensional vectors without explicitly identifying where the incongruity occurs within the graph. Besides, lack of reasoning makes models vulnerable to spurious bias, such as the positive word '*commend*' is usually related to non-sarcasm, leading to unpredictable results and deteriorating their robustness.

To address the aforementioned challenges, we propose a novel framework for sarcasm detection called *ECIR*, which conducts incongruity reasoning based on the commonsense-augmented graph. Concretely, we first retrieve input-related passages from the search engine *Bing* for each statement. To filter noise introduced in passages, we employ a hybrid method that considers word matching and semantic similarity. The filtered content is utilized to construct the retrieval-augmented prompt, which improves the quality of knowledge provided to the *LLMs*. This retrieval-augmented strategy enhances the generative capabilities of *LLMs*. We then construct the dependency graph to capture the complex correlations across multiple texts follow (Lou et al., 2021). To learn the optimized graph topology, we employ the innovative graph enrichment strategy to capture semantic relations and the graph pruning method to remove irrelevant edges based on meta-path. We further utilize the practical reasoning skeleton with per-defined rules to infer the fine-gained incongruity, which is observed in the suspicious incongruity subgraphs. A fusion module is employed to integrate multi-grained comprehensive features from the refined graph and subgraphs. We finally devise adversarial contrastive learning in feature space to mitigate the word biases. Extensive experimental results from five datasets demonstrate the effectiveness of our approach.

The main contribution of this paper includes,

- We point out the challenges of providing reliable commonsense knowledge and performing sufficient emotional incongruity inference when facing complex instances in sarcasm detection, which are new for this task.
- We propose a novel framework for sarcasm detection called *ECIR*, which detects emotional incongruity by commonsense reasoning.
- We conduct extensive experiments on the five public datasets to evaluate the rationality and effectiveness of our proposed method.

2 Methodology

We first give the problem definition of sarcasm detection. Suppose we have a set of training samples, where each sample consists of a given statement $\mathcal{S}$ and its corresponding label y. Specifically, $y = 1$ if the statement is sarcastic, and $y = 0$ represents non-sarcastic. To enhance $\mathcal{S}$ with relevant context, we utilize a retrieval-augmented *GPT-4o* to provide the commonsense knowledge C. We aim to train a novel sarcasm detection model $\mathcal{F}$ that can precisely identify incongruity across texts as follows,

$$\hat{y} = \mathcal{F}(\mathcal{S}, C|\theta), \tag{1}$$

where θ denotes all the trainable parameters of $\mathcal{F}$ and $\hat{y}$ is the probability distribution. Subsequently, we present the commonsense reasoning model for detecting emotional incongruity, as depicted in Figure 2. This model integrates *RAG-based* commonsense augmentation, graph-based incongruity reasoning, and adversarial contrastive learning.

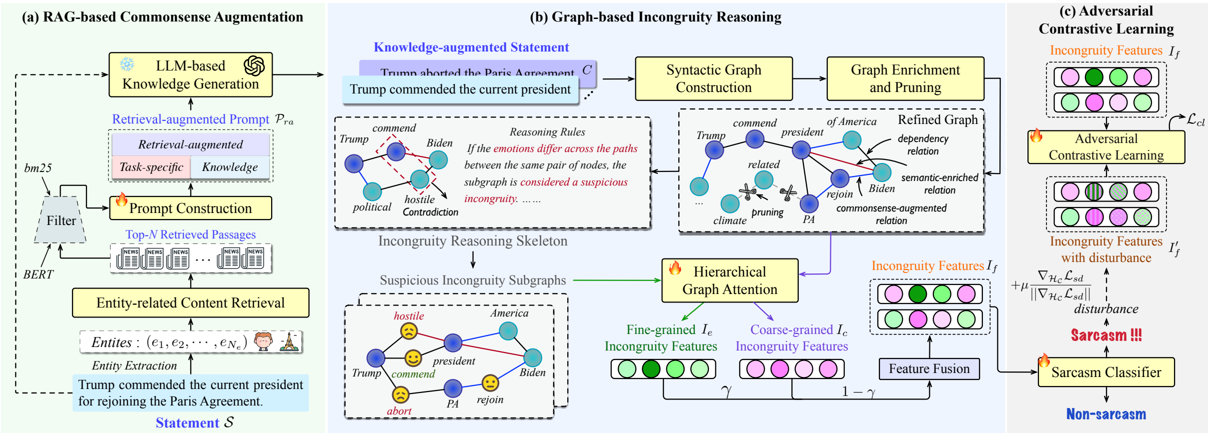


Figure 2: Overall framework of the commonsense reasoning model *EICR* for sarcasm detection

2.1 RAG-based Commonsense Augmentation

To supplement the indispensable commonsense required for sarcasm detection, we resort to retrieval-augmented *LLMs*, which has demonstrated the capability to generate reliable knowledge. In particular, we first employ *NLTK* to extract the entities $E_s = \{e_1, e_2, \ldots, e_{N_e}\}$ from the given statement $\mathcal{S}$, where N_e denotes the number of entities. We then employ the general web search engine *Bing* to retrieve entity-related content and select the *Top-N* relevant passages $P = \{p_1, p_2, \ldots, p_n\}$ as candidate corpus for *LLMs*. However, the retrieved passages might inevitably introduce noise harmful to *LLMs*. To alleviate this problem, inspired by Zhang et al. (2023), we devise a hybrid noise filter to improve the quality of generated commonsense knowledge. We employ the *BERT-BM25 Hybrid* method to calculate the *BM25* keyword matching score f_{bm25} and the *BERT*-based semantic similarity score f_{bert} between each passage and statement $\mathcal{S}$. The final score is obtained by weighted computation of the two scores as follows,

$$f_{bebm} = \frac{\alpha}{1 + e^{-f_{bm25}}} + (1-\alpha) f_{bert}, \qquad (2)$$

where α is a hyper-parameter and passages will be filtered if the f_{bebm} is less than the predefined threshold ϵ. After that, remained passages P_f will be fed into the *BART Encoder* (Lewis et al., 2019) incorporating an additional multi-head pooling layer to get the representation of passages R_p.

To ensure the efficiency of training, we freeze the parameters in *LLMs* and adopt prompt-tuning (Radford et al., 2021) to generate statement-related commonsense knowledge C. Thereby, based on R_p, we employ a prompt builder which consists of a cross-attention network and a feed-forward network to create the retrieval-augmented prompt as follows,

$$\begin{aligned} \mathcal{P}'_{ra} &= \mathrm{MHA}(\mathcal{D}M_q; R_pM_k; R_pM_v), \\ \mathcal{P}_{ra} &= \mathrm{LN}(\mathcal{P}_{ra'} + \mathrm{FFN}(\mathcal{P}'_{ra})), \end{aligned} \qquad (3)$$

Where $\mathcal{D}$ is a trainable length controller with a fixed length $l_{\mathcal{D}}$, each M represents a projection matrix, *MHA* denotes multi-head attention, and *LN* stands for layer normalization. The *LLMs* take as input the concatenation of the retrieval-augmented prompt $\mathcal{P}_{ra}$, the task-specific prompt $\mathcal{P}_{ts}$, and the entities embedding $\mathcal{P}_E$, collectively denoted as $\mathcal{P}_{lm} = [\mathcal{P}_{ra} \| \mathcal{P}_{ts} \| \mathcal{P}_E]$. The task-specific prompt $\mathcal{P}_{ts}$ is a predefined instruction: "*Please provide reliable commonsense knowledge related to this statement and its relevant passages.*" To control irrelevant noise, the maximum number of generated commonsense C is constrained to N_c, a hyper-parameter. Since lacking corresponding labels, we manually review commonsense quality and analyze the effectiveness of commonsense in section 3.3.

2.2 Graph-based Incongruity Reasoning

To capture the complex correlations across multiple texts, we employ graph structures, which are well-suited for modelling long-range dependencies. However, existing methods fail to find fine-grained incongruity due to the incomplete graph topology and poor inferential ability. To address this problem, we introduce a graph refinement strategy to capture robust associations and utilize an inference module to retrieve the incongruity subgraphs. We further supplement the overall context features to prevent over-reliance on the inference skeleton.

(1) Graph Construction. To capture the syntactic dependency relations between the statement and commonsense, we first transform each piece of text into an undirected graph using *Spacy*, denoted as $\mathcal{G}_i = \{V_i, E_i\}$ (Lou et al., 2021). Here, V_i represents the meaningful concepts in the text, while E_i denotes the set of syntactic dependency edges. We then construct a commonsense-augmented graph $\mathcal{G}$ through the off-the-shelf technique by (Zhu et al., 2021) for effectively aligning the concepts. To better learn the topology, we define $\mathcal{A}^d_{i,j} = 1$ if there is an edge between nodes v_i and v_j, where $\mathcal{A}^d$ is the dependency adjacency matrix. Additionally, each node includes a self-loop, denoted as $\mathcal{A}^d_{i,i} = 1$.

(2) Graph Enrichment. To augment the semantic associations in the dependency graph, inspired by Zhao et al. (2021), we design a graph enrichment mechanism by employing the metric learning to obtain the semantic matrix $\mathcal{A}^{se}$. This promotes interactions among sparsely connected long-tail nodes, which enhances the capabilities of *EICR* to discover subtle incongruities. The learning rule is formulated as Eq.(4),

$$\mathcal{A}^{se}_{i,j} = \begin{cases} \psi^{se}(h_i, h_j), & \psi^{se}(h_i, h_j) \geq \delta, \\ 0, & \psi^{se}(h_i, h_j) < \delta, \end{cases} \tag{4}$$

where h_i is the node embedding encoded by *Bi-LSTM*, δ is a threshold to control sparsity and $\psi^{se}(\cdot)$ is the β-head weighted cosine similarity as follows,

$$\psi^{se}(h_i, h_j) = \frac{1}{\beta} \sum_{i=1}^{\beta} \cos(w^{se} \odot h_i, w^{se} \odot h_j), \tag{5}$$

where $\odot$ denotes *Hadamard* product and w^{se} is a trainable parameter. The message propagation matrix is computed as $\mathcal{A}^p = \mathcal{A}^{se}\mathcal{A}^d$. The semantic-enriched matrix $\tilde{\mathcal{A}}$ is then formed by combining $\mathcal{A}^p$ and $\mathcal{A}^d$, represented as $\tilde{\mathcal{A}} = \mathcal{A}^p \oplus \mathcal{A}^d$.

(3) Graph Pruning. To learn the task-optimal graph topology, we adopt a meta-path-based edge pruning strategy, which efficiently removes noisy edges. An optimal graph topology should retain the relevant edges and be conducive to achieving effective sarcasm detection. Thus, we focus on sampling *Top-L* meta-path neighbors based on the enriched graph, subject to certain constraints. In particular, we first define the edge weight coefficient $\omega^{\Phi}_{v,j}$ of node j to the center node v, as follows,

$$\omega^{\Phi}_{v,j} = \frac{\exp(\sigma(M_\omega[h_v \| h_j]))}{\sum_{j \in \mathcal{N}^{\Phi}_v} \exp(\sigma(M_\omega[h_v \| h_j]))}, \tag{6}$$

where $\mathcal{N}^{\Phi}_v$ denotes the nodes connected with v through meta-path Φ and M_ω is the learnable matrix. Then we employ a linear programming strategy to perform the sampling process, formulated as $\text{argmax}_{\mathcal{O}\in\mathcal{C}}\langle \mathcal{O} \cdot \mathcal{M}^L_v \rangle$. Here $\mathcal{O}$ is the one-hot sampling pointers, $\mathcal{M}^L_v$ is the weight matrix constructed by L interactions of computation. To ensure the one-hot property and sorting reasonableness of *Top-L*, the constraint $\mathcal{C}$ is denoted as Eq.(7),

$$\mathcal{C} = \begin{cases} \sum_{i=1}^{D} \mathcal{O}_{ij} = 1, & \forall j \leq L, \\ \mathcal{O}_{ij} \geq \mathcal{O}_{i(j+1)} \geq 0, & \forall i < D, \end{cases} \tag{7}$$

where D represents the number of meta-path neighbors. Subsequently, we merge the sampling pointers from each Φ to form the pruned matrix $\hat{\mathcal{A}}$. Due to the non-differentiability of one-hot and *Top-L* operations, the perturbed maximum strategy (Berthet et al., 2020) is employed during the training stage, as Eq.(8). This strategy allows for gradient-based optimization by introducing controlled randomness, enabling effective learning in scenarios where traditional backpropagation would fail.

$$\mathcal{O}_\vartheta = \mathbb{E}_{\mathbb{U}}[\text{argmax}_{\mathcal{O}\in\mathcal{C}}\langle \mathcal{O}, \mathcal{M}^L_v + \vartheta U \rangle], \tag{8}$$

where ϑ controls the proportion of injected *Gaussian Noise* $\mathbb{U}$. Finally, we incorporate $\mathcal{O}_\vartheta$ into $\hat{\mathcal{A}}$ to get the matrix $\mathcal{A}^{opt}$, which represents the task-optimal graph topology of the refined graph $\tilde{\mathcal{G}}$.

(4) Incongruity Reasoning. To grasp the fine-grained incongruity in the refined graph $\tilde{\mathcal{G}}$, we propose a novel reasoning skeleton, which incorporates inference rules as prior knowledge. These inference rules come from multiple disciplines such as psychology (Michel et al., 2022) and sociology (Klomberg et al., 2024), helping the model better understand the diverse sarcasm patterns. Nevertheless, these rules are not scalable and are difficult to integrate into *EICR*. To address this problem, we leverage these rules to fine-tune the entailment checker based on pre-trained *RoBERTa-large* (Liu et al., 2019). In the skeleton, we first separate multiple paths between the same node pair with special tokens <s><s> to a single input sequence. We then input them into the entailment checker as conditions and conclusions, respectively, to generate predicted logits for each group of paths. Here,

Datasets	#Train	#Test	#Avg.Len	%Sarcasm
Ghosh	33,373	4,121	13.0	45.13%
Reddit	21434	5210	16.0	49.99%
IAC-V2	5,216	1,043	65.0	50.01%
iSarcasm	3,116	887	27.0	18.01%
SemEval2018	3,398	780	14.5	49.35%

Table 1: Statistics of datasets. #Avg.Len denotes the average length of the texts.

the logits correspond to three results: contradiction, neutral, and entailment. We sample the node pairs with the contradiction logit as the suspicious incongruity subgraphs $\tilde{\mathcal{G}}^s$. Relying solely on extracted subgraphs may lose some important context. Thus $\tilde{\mathcal{G}}$ and $\tilde{\mathcal{G}}^s$ are fed into hierarchical graph attention*(HGAs)* (Wang et al., 2019) simultaneously to learn multi-grained incongruity features as follows,

$$\begin{aligned} I_e &= \mathcal{M}_{mpl}(\mathcal{M}^l_{sem}(\mathcal{M}^l_{nod}(I_e^{l-1}, \tilde{\mathcal{G}}^s, \mathcal{N}^{\Phi_s}))), \\ I_c &= \mathcal{M}_{mpl}(\mathcal{M}^l_{sem}(\mathcal{M}^l_{nod}(I_c^{l-1}, \tilde{\mathcal{G}}, \mathcal{N}^{\Phi}))), \end{aligned} \tag{9}$$

where $\mathcal{M}^l_{nod}$ and $\mathcal{M}^l_{sem}$ represent the hierarchical attention mechanisms in the l-th layer of *HGAs*, while $\mathcal{M}_{mpl}$ and $\mathcal{N}^{\Phi_s}$ denote max-pooling and the meta-path neighbor sets in $\tilde{\mathcal{G}}^s$, respectively. To balance multi-grained features, a gating strategy combines global context and local incongruity into the final feature I_f, formulated as $I_f = \gamma I_c + (1-\gamma) I_e$, where γ is a hyper-parameter. This enables robust performance by integrating both perspectives. The feature I_f is then fed into a single-layer *MLP* to predict its probability distribution $\hat{y}$ as $\hat{y} = \mathrm{Softmax}(\mathrm{MLP}(I_f))$. Finally, cross-entropy loss $\mathcal{L}_{sd}$ is used for classification as Eq.(10),

$$\mathcal{L}_{sd} = -(y \log(\hat{y}) + (1-y)\log(1-\hat{y})). \tag{10}$$

2.3 Adversarial Contrastive Learning

To learn robust representations, we employ supervised contrastive learning (Kim et al., 2020), which aims to pull together samples with the same label and push apart those with different labels. Further, to mitigate the spurious correlation and reduce sensitivity to local knowledge, we introduce adversarial gradient perturbations in the feature space to construct augment instances in the feature space. The perturbation process is applied to the commonsense node embeddings $\mathcal{H}_c$ as Eq.(11),

$$\mathcal{H}'_c = \mathcal{H}_c + \mu \frac{\nabla_{\mathcal{H}_c} \mathcal{L}_{sd}}{||\nabla_{\mathcal{H}_c} \mathcal{L}_{sd}||}, \tag{11}$$

where ∇ is the derivation operator, μ is a validation controller and $\mathcal{H}'_c$ is perturbed commonsense. Compared with direct delete or substitute words, the gradient-based perturbation can maintain the original semantics of commonsense. Thereafter, $\mathcal{H}'_c$ along with statement representations are fed into the *Graph-based Incongruity Reasoning Network* to get the perturbed incongruity feature I'_f. So far, the sarcastic set Λ_s contains I_f and I'_f in the same batch. The same is true in the non-sarcastic set Λ_n. Finally, the supervised contrastive learning loss (Khosla et al., 2020) is computed as Eq.(12),

$$\mathcal{L}_{cl} = \frac{-1}{|\Lambda_s|} \sum_{I_f^+ \in \Lambda_s} \log \frac{\exp(\mathrm{sim}(I_f, I_f^+)/\tau)}{\sum_{I_f^- \in \Lambda_n} \exp(\mathrm{sim}(I_f, I_f^-)/\tau)}, \tag{12}$$

where I_f^+ is the sarcastic samples and I_f^- is the non-sarcastic samples. τ is a temperature parameter to adjust the smoothness of the shrinkage distribution and $\mathrm{sim}(\cdot)$ denotes the cosine similarity function.

The learning objective is to train the framework by jointly minimizing the two losses derived from sarcasm detection and adversarial supervised contrastive learning. The overall loss is written as,

$$\mathcal{L} = \mathcal{L}_{sd} + \lambda \mathcal{L}_{cl}, \tag{13}$$

where λ is a hyper-parameter to control the weight of contrastive learning in the overall loss function.

3 Evaluations

We conducted experiments with qualitative and quantitative analyses to evaluate our approach.

3.1 Data and Implementation Details

We conducted our experiments on five benchmark datasets, including *Ghosh* (Ghosh and Veale, 2017) which was collect from *Twitter* and annotated automatically, *Reddit* (Khodak et al., 2018) that only contained political content, *IAC-V2* (Oraby et al., 2017) which was obtained from *Internet Argument Corpus* and designed for analyzing sarcastic remarks, *iSarcasm* (Oprea and Magdy, 2020) that encompassed tweets which were written by online users and *SemEval2018* (Van Hee et al., 2018) collected using hashtags from *SemEval 2018 Task 3*. The statistics of datasets are shown in Table 1.

Our experiments were conducted using *Pytorch* and ran on four NVIDIA GeForce 3090 GPUs. For the passages filtering, we utilized the embedding from the *BERT-based* model and set the balanced hyper-parameter α to 0.4 and the threshold ϵ to 0.95. To learn the semantic relationships, we set the sparse-controlled hyper-parameter δ to 0.95. The number of β-head in Eq.(5) was set to 3. For the adversarial contrastive loss, the temperature parameter τ was set to 0.07. The training batch size was fixed at 8, and we took *Adam* as the optimizer with a learning rate of 0.001.

3.2 Model Comparison

We compared our method against eight mainstream models, including (1) *MIARN* (Tay et al., 2018)

Method	Ghosh		Reddit		IAC-V2		iSarcasm		SemEval2018	
	Acc	Ma-F1	Acc	Ma-F1	Acc	Ma-F1	Acc	Ma-F1	Acc	Ma-F1
MIARN	79.1	78.6	70.4	69.2	75.6	75.7	79.4	57.3	68.5	67.8
RoBERTa	72.2	72.4	66.7	66.7	76.6	76.7	78.6	56.8	70.2	69.1
SAWS	78.8	78.5	70.8	71.9	76.2	76.2	76.8	57.5	69.9	68.9
SarDeCK	83.4	83.0	73.5	73.1	77.5	77.5	78.1	59.6	71.7	70.2
ADGCN	79.7	79.5	73.2	72.3	78.0	78.0	79.2	58.5	71.7	70.1
DC-Net	80.2	78.6	72.9	72.4	78.0	77.9	78.8	58.7	70.8	69.6
SD-APRR	82.6	82.3	-	-	78.8	78.8	<u>80.3</u>	<u>61.2</u>	72.2	70.7
SensoryT5	<u>86.1</u>	**86.1**	-	-	<u>83.0</u>	<u>83.0</u>	-	-	<u>77.7</u>	<u>77.9</u>
SarcasmCue	83.0	82.9	<u>74.1</u>	<u>73.7</u>	73.4	72.3	79.4	60.3	74.0	74.0
EICR	**86.2**	<u>84.3</u>	**77.2**	**75.3**	**84.5**	**83.8**	**83.3**	**70.4**	**80.1**	**80.3**

Table 2: Performance comparison of different methods on datasets. Acc denotes Accuracy and Ma-F1 denotes Macro-F1. The best results are represented in **bold**. The second-best results are <u>underlined</u>.

Method	Ghosh		Reddit		IAC-V2		iSarcasm		SemEval2018	
	Acc	Ma-F1	Acc	Ma-F1	Acc	Ma-F1	Acc	Ma-F1	Acc	Ma-F1
EICR-BLOOM 3B	85.2	83.1	75.7	73.6	83.1	82.9	81.4	66.9	78.5	78.1
EICR-Qwen 2 7B	85.3	83.1	75.9	73.7	83.3	83.0	81.9	67.4	78.9	78.4
EICR-Llama 3 8B	85.3	83.2	76.0	73.7	83.6	83.2	82.2	67.8	78.9	78.5
EICR-T5 11B	85.9	83.8	76.6	74.0	83.8	83.2	82.6	68.6	79.3	79.2
EICR-GPT4o	**86.2**	**84.3**	**77.2**	**75.3**	**84.5**	**83.8**	**83.3**	**70.4**	**80.1**	**80.3**

Table 3: Performance comparison of *LLMs* of different scales **(without fine-tune)** as the generative model on datasets. The best results were represented in **bold**. Improvements were statistically significant ($p < 0.05$).

and *SAWS* (Pan et al., 2020), which were representative attention-based methods and identified sarcasm based on the input text features; (2) *RoBERTa* (Liu et al., 2019), which served as a strong baseline by capturing nuanced contextual and linguistic features; (3) *SarDeCK* (Li et al., 2021), a competitive *BERT-based* method that utilized *COMET* to supplement commonsense knowledge and enriched context via attention networks; (4) *ADGCN* (Lou et al., 2021), a *GCN-based* model that fused features from the dependency graph and sentiment graph to discern incongruity; (5) *DC-Net* (Liu et al., 2022), which modeled literal and implied sentiments separately to recognize sentiment conflict; (6) *SD-APRR* (Min et al., 2023), an incongruity reasoning model that employed a denoising module based on a commonsense-augmented dependency graph; (7) *SensoryT5* (Zhao et al., 2025), which integrated sensory knowledge into the *T5* framework's attention mechanism to facilitate sensory-emotional interactions; (8) *SarcasmCue* (Yao et al., 2024), which introduced a prompting framework that elicited *LLMs* to detect sarcasm by considering sequential and non-sequential prompting methods.

Our approach was evaluated against the baseline methods illustrated in Table 2. We observed that (1) *EICR* achieved state-of-the-art performance across five public benchmarks in most settings, demonstrating its effectiveness in sarcasm detection; (2) Baseline methods exhibited relatively low *Ma-F1* scores on the *iSarcasm* dataset, likely due to the imbalanced label distribution. In contrast, *EICR* demonstrated strong robustness in handling such imbalances; (3) Compared with *RoBERTa* and *SarDeCK*, our proposed commonsense reasoning perspective fully captured the sentiment incongruity across contexts; (4) The performance gains observed over *SarcasmCue* suggested that solely prompting *LLMs* led to hallucinations. Our retrieval-augmented strategy effectively mitigated this issue, as further confirmed by the superior results compared to *SensoryT5*, which highlighted the combined impact of noise filtering; (5) When compared to *ADGCN*, *EICR* highlighted the supplementation of comprehensive commonsense and the ability for commonsense inference instead of directly representing the entire graph for detection; (6) As opposed to *DC-Net* and *SD-APRR*, the novel

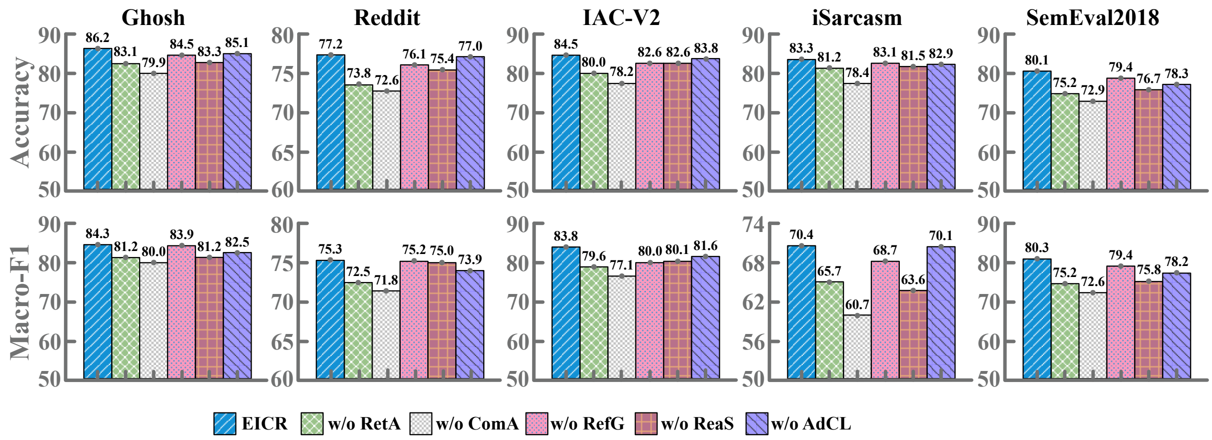
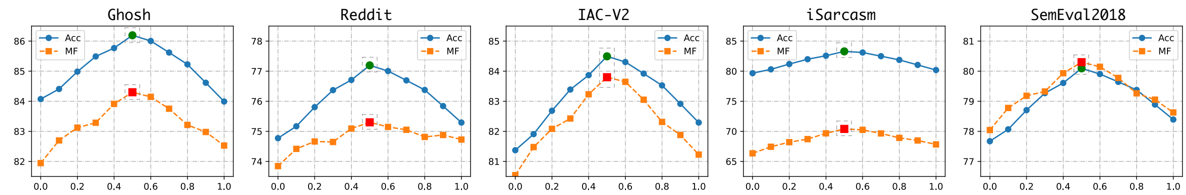
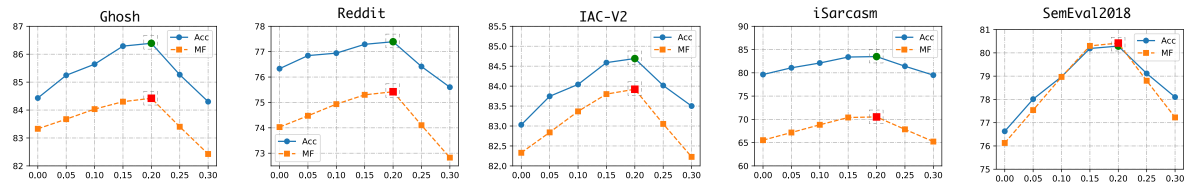


Figure 3: Experiment results of ablation studies.



(a) Variation of the trade-off parameter α when fixed $\epsilon = 0.95$



(b) Variation of the trade-off parameter λ

Figure 4: Hyper-parameter sensitivity analysis on score balance parameter α and contrastive learning coefficient λ. The best results were shown by green point and red square.

reasoning skeleton was helpful in explicitly finding fine-grained incongruity.

3.3 Ablation Studies

We performed extensive ablation studies to verify the effectiveness of our proposed key modules, including (1) *RetA* that omitted retrieval-augment strategy and solely relied on the *GPT-4o*; (2) *ComA* that removed the entire commonsense-augmented module and only based on the statement; (3) *RefG* that dropped the graph refinement strategy; (4) *ReaS* that discarded the reasoning skeleton and depended on the features extracted by *GCNs*; (5) *AdCL* threw away the adversarial contrastive loss.

As depicted in Figure 3, the ablation studies provided meaningful insights into the effectiveness of various components in *EICR*, including (1) *RetA* performed worse than *EICR*. This indicated solely *LLMs* may be influenced by hallucinations, which will generate noise and make reasoning difficult; (2) *ComA* caused the terrible performance underscored the critical role of external commonsense. Without external knowledge to contextualize and facilitate reasoning, the model struggled to detect sarcasm; (3) *RefG* led to the noticeable degradation demonstrated that the optimized graph topology was conducive to our model discovery of implicit emotional incongruity in the graph; (4) *EICR* outperformed the *ReaS* proving the necessity of reasoning about fine-grained incongruity, which improved the robustness and generalization of incongruity features; (5) *AdCL* had a suboptimal performance confirmed adversarial contrastive learning helped us mitigate biases and learn robust representations.

3.4 Model Analysis

(1) Hyper-parameter Sensitivity Analysis. We conducted a detailed analysis of the trade-off parameters employed in our proposed model, including the score balance parameter α in Eq.(2) when

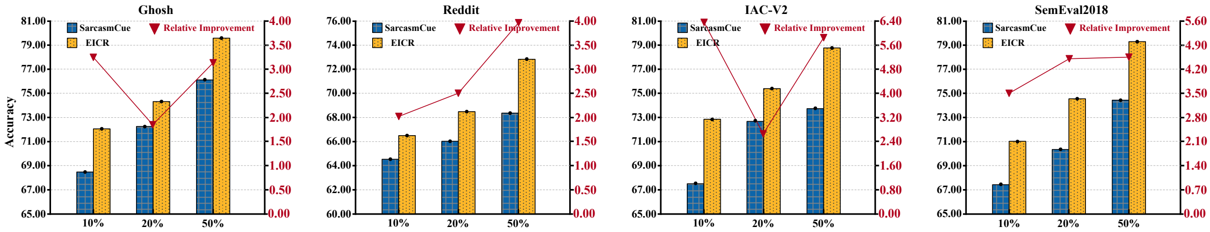


Figure 5: Low-resource performance on four benchmark datasets. The left axis represents the accuracy, while the right axis illustrates our improvement over the baseline.

threshold $\epsilon = 0.95$ and the contrastive learning coefficient λ in Eq.(13). Specifically, we adjusted α within [0, 1], while λ varied within [0, 0.3]. To better observe the tendency, we employed 0.1 and 0.05 as intervals for α and λ, respectively. Figure 4a showed that performance initially improved as α increased but subsequently declined. The best results were achieved when $\alpha = 0.5$. This indicated that an inappropriate value could lead to an imbalance in the filtering process, which may introduce irrelevant noise and perturb the quality of the retrieval-augmented prompt. In Figure 4b, as λ increased, the curve showed an upward trend followed by a decline. This suggested that too small λ prevented the model from mitigating the biases and learning robust representations, while a larger λ diverted attention away from sarcasm detection.

(2) Evaluations on Language Models. To assess the performance of our method across different large language models, we conducted experiments on *BLOOM 3B*, *Qwen 2 7B*, *Llama 3 8B*, *T5 11B*, and *GPT-4o*. As presented in Table 3, the results exhibited only minor fluctuations and were still better than the baselines. This demonstrated that our method was largely insensitive to the parameter scale of the *LLMs*, which was attributed to the proposed retrieval-augmented strategy. That effectively supplemented the models with relevant external commonsense, and the carefully designed prompts used this external knowledge and the parameterized knowledge embedded within the *LLMs* to generate the emotional commonsense required for this task. This highlighted the strong generalization ability of our model across different *LLMs*.

(3) Robustness under Low-resource Scenario. To further evaluate the effectiveness of *EICR* in low-resource scenarios (Liu et al., 2023), we implemented experiments using various proportions of training samples-10%, 20%, and 50%-from four benchmarks. Since the *iSarcasm* contained a small number of sarcastic instances, it was difficult for the model to capture the true incongruous features in low-resource scenarios. The experimental results were too random to discuss. For a fair comparison, we selected the best results from various prompting sub-methods in the *GPT-4o-based SarcasmCue* (Yao et al., 2024) as the baseline. The observed improvements shown in Figure 5 can be attributed to the integration of reliable external commonsense, which enriched the contextual content. Moreover, explicit reasoning about fine-grained incongruities between the knowledge and the given statements further enhanced the performance of the sarcasm detector. These findings validated the robustness of *EICR* in addressing distributional incongruities between training and testing samples, demonstrating its effectiveness in low-resource settings.

3.5 Case Study

To gain deeper insights into *EICR*, we conducted case studies focusing on complex sarcastic statements. As illustrated in Figure 6, *EICR* identified the sarcastic intent in *Hillary's* statement, as shown in the first case example. *EICR* effectively retrieved relevant knowledge, including the well-established fact "*political opponents*" and sentiment-related cues "*rarely praise each other publicly*." This underscored the value of the retrieval-augmented strategy in grounding sarcasm detection with commonsense knowledge. Further, the extracted subgraphs precisely captured the emotional incongruity without introducing extraneous information. In addition, *EICR* exhibited strong generalization capabilities across various domains, adapting to diverse sarcastic contexts. These findings demonstrated the outstanding commonsense reasoning ability of *EICR* neglected by traditional methods.

4 Related Work

Sarcasm Detection (SD) has attracted extensive attention in recent years. Early efforts were mainly based on formulated rules (Riloff et al., 2013). For

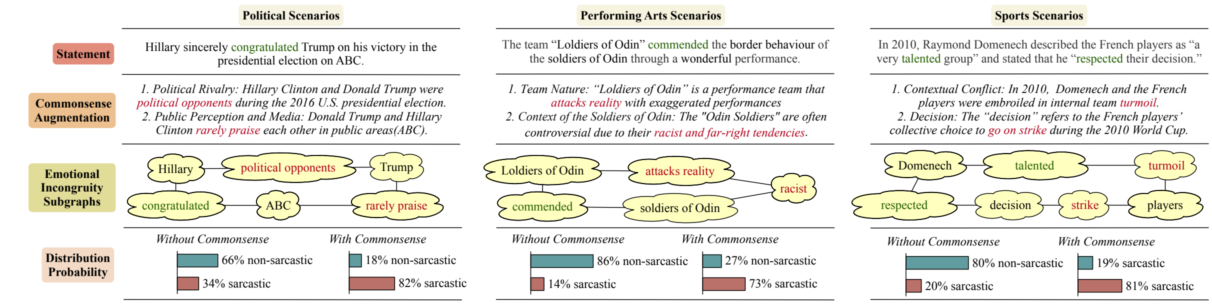


Figure 6: Case studies on the complex sarcastic instance. Red highlights the key commonsense conductive to find incongruity. Green denotes the positive words in the given statement.

instance, (Maynard and Greenwood, 2014) considered that hashtags might contain sarcastic features. Bamman and Smith (2015) found that the incongruity between positive verbs and negative situations indicated sarcasm. With the advancement of neural networks, co-attention tricks (Pan et al., 2020) and *BERT* (Babanejad et al., 2020) were applied to capture incongruity patterns. Recent methods employed knowledge graphs like *SenticNet* (Liu et al., 2022) or *COMET* (Li et al., 2021) (Yu et al., 2023b) to introduce external commonsense knowledge for daily sarcastic instances. To capture the intricate relations between multi-source knowledge and the statement, Lou et al. (2021) injected emotional commonsense into the dependency graph for *SD*. Additionally, Min et al. (2023) augmented potential results and reactions from *COMET* to mimic the way humans judge sarcasm. Liu et al. (2023) and Yao et al. (2024) both employed the *LLMs* with a prompting framework in *SD*. Besides, other works focused on behavior-level (Zhou et al., 2024) and deep text representation and learning (Gedela et al., 2024). However, these methods lack commonsense inferential ability when they face complex real-world scenarios, which might lead to unsatisfactory performance.

Commonsense reasoning has become a prominent focus in natural language processing (Guo et al., 2023) and computer vision (Liu et al., 2024). Many knowledge-intensive tasks usually require associating with commonsense knowledge, such as question answering (Zhang et al., 2024), question generation (Yu et al., 2021), and sarcasm detection (Yue et al., 2023). Effective commonsense reasoning often requires external knowledge resources, including *KGs* such as *SenticNet*(Cambria et al., 2020) and *ConceptNet*(Speer et al., 2017) and *PLMs* like *COMET*(Hwang et al., 2021) and *GPT-4*(Achiam et al., 2023). While *KGs* typically aggregate knowledge through structured matching processes (Ye et al., 2022), *PLMs* generate knowledge dynamically via prompting (Radford et al., 2021). Commonsense reasoning for detecting incongruity features generally follows two paradigms (Chen et al., 2024a). The first leverages neural networks to encode contextual clues (Tay et al., 2018), yet these methods often struggle with capturing complex relations (Lou et al., 2021). The second approach involves constructing graphs and incorporating commonsense knowledge directly into the graph structures (Qiao et al., 2023). However, existing *GNN-based* methods are limited in their ability to explain the reasoning process, focusing primarily on the final detection outcomes (Yu et al., 2023a). In contrast, our proposed reasoning skeleton integrates prior rules creating a more interpretable and effective framework for sarcasm detection.

5 Conclusion

This paper tackled the challenges of supplementing commonsense knowledge and inferring fine-grained incongruity when facing complex instances in sarcasm detection. To address these issues, we proposed a novel commonsense reasoning framework for sarcasm detection named *EICR*. Concretely, we first devised retrieval-augmented *LLMs* to provide the essential commonsense knowledge. To capture sophisticated contextual associations, we constructed a dependency graph and obtained the optimized topology through graph refinement. We further introduced an adaptive reasoning skeleton that integrated prior rules to extract emotional-incongruity subgraphs explicitly. To eliminate the possible spurious relations between words and labels, we employed adversarial contrastive learning to enhance the robustness of the detector. Experiments conducted on five popular datasets demonstrate the effectiveness of our proposed method.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (62276279, 62372483, 62276280, U2001211, U22B2060), Guangdong Basic and Applied Basic Research Foundation (2024B1515020032), and Research Foundation of Science and Technology Plan Project of Guangzhou City (2023B01J0001, 2024B01W0004).

Limitations

In this paper, we focused on text-based sarcasm detection. While *EICR* demonstrated its effectiveness, there was room for improvement. For example, sarcastic content often involved multiple modalities, such as images and videos. Next, we planned to expand our uni-modal model to a multi-modal one to handle more complex cases. In addition, our method did not provide concrete explanations for why a statement was identified as sarcastic, which was another important area we intended to explore in future research.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Nastaran Babanejad, Heidar Davoudi, Aijun An, and Manos Papagelis. 2020. Affective and contextual embedding for sarcasm detection. In *Proceedings of the 28th international conference on computational linguistics*, pages 225–243.

David Bamman and Noah Smith. 2015. Contextualized sarcasm detection on twitter. In *proceedings of the international AAAI conference on web and social media*, volume 9, pages 574–577.

Quentin Berthet, Mathieu Blondel, Olivier Teboul, Marco Cuturi, Jean-Philippe Vert, and Francis Bach. 2020. Learning with differentiable pertubed optimizers. *Advances in neural information processing systems*, 33:9508–9519.

Lars Bülow and Michael Johann. 2023. Effects and perception of multimodal recontextualization in political internet memes. evidence from two online experiments in austria. *Frontiers in Communication*, 7:1027014.

Erik Cambria, Yang Li, Frank Z Xing, Soujanya Poria, and Kenneth Kwok. 2020. Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 105–114.

Wangqun Chen, Fuqiang Lin, Guowei Li, and Bo Liu. 2024a. A survey of automatic sarcasm detection: Fundamental theories, formulation, datasets, detection methods, and opportunities. *Neurocomputing*, 578:127428.

Yifan Chen, Kuntao Li, Weixing Mai, Qiaofeng Wu, Yun Xue, and Fenghuan Li. 2024b. D2r: Dual-branch dynamic routing network for multimodal sentiment detection. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 3536–3547.

Ravi Teja Gedela, Ujwala Baruah, and Badal Soni. 2024. Deep contextualised text representation and learning for sarcasm detection. *Arabian Journal for Science and Engineering*, 49(3):3719–3734.

Aniruddha Ghosh and Tony Veale. 2017. Magnets for sarcasm: Making sarcasm detection timely, contextual and very personal. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 482–491.

Qiutong Guo, Jianxing Yu, Yufeng Zhang, Haowei Jiang, Wei Liu, and Jian Yin. 2023. Discovery of emotion implicit causes in products based on commonsense reasoning. In *International Conference on Advanced Data Mining and Applications*, pages 277–292.

Jena D Hwang, Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, and Yejin Choi. 2021. (comet-) atomic 2020: On symbolic and neural commonsense knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6384–6392.

Aditya Joshi, Pushpak Bhattacharyya, and Mark J Carman. 2017. Automatic sarcasm detection: A survey. *Journal of ACM Computing Surveys*, 50(5):1–22.

Mikhail Khodak, Nikunj Saunshi, and Kiran Vodrahalli. 2018. A large self-annotated corpus for sarcasm. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673.

Minseon Kim, Jihoon Tack, and Sung Ju Hwang. 2020. Adversarial self-supervised contrastive learning. *Advances in neural information processing systems*, 33:2983–2994.

Bien Klomberg, Joost Schilperoord, and Neil Cohn. 2024. Constructing domains in visual narratives: Structural patterns of incongruity resolution. *Journal of Comparative Literature and Aesthetics Vol*, 47(3):37–55.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

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*, 29:3192–3201.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

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*, pages 1670–1680.

Yiyi Liu, Ruqing Zhang, Yixing Fan, Jiafeng Guo, and Xueqi Cheng. 2023. Prompt tuning with contradictory intentions for sarcasm recognition. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 328–339.

Yuting Liu, Liu Yang, and Yu Wang. 2024. Hierarchical fine-grained visual classification leveraging consistent hierarchical knowledge. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 279–295. Springer.

Chenwei Lou, Bin Liang, Lin Gui, Yulan He, Yixue Dang, and Ruifeng Xu. 2021. Affective dependency graph for sarcasm detection. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 1844–1849.

Diana G Maynard and Mark A Greenwood. 2014. Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. In *Lrec 2014 proceedings*. ELRA.

Géraldine Michel, Carlos J Torelli, Nathalie Fleck, and Benoit Hubert. 2022. Self-brand values congruity and incongruity: Their impacts on self-expansion and consumers' responses to brands. *Journal of Business Research*, 142:301–316.

Changrong Min, Ximing Li, Liang Yang, Zhilin Wang, Bo Xu, and Hongfei Lin. 2023. Just like a human would, direct access to sarcasm augmented with potential result and reaction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 10172–10183.

Maya Okawa and Tomoharu Iwata. 2022. Predicting opinion dynamics via sociologically-informed neural networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1306–1316.

Silviu Oprea and Walid Magdy. 2020. isarcasm: A dataset of intended sarcasm. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1279–1289.

Shereen Oraby, Vrindavan Harrison, Lena Reed, Ernesto Hernandez, Ellen Riloff, and Marilyn Walker. 2017. Creating and characterizing a diverse corpus of sarcasm in dialogue. *arXiv preprint arXiv:1709.05404*.

Hongliang Pan, Zheng Lin, Peng Fu, and Weiping Wang. 2020. Modeling the incongruity between sentence snippets for sarcasm detection. In *ECAI 2020*, pages 2132–2139. IOS Press.

Yang Qiao, Liqiang Jing, Xuemeng Song, Xiaolin Chen, Lei Zhu, and Liqiang Nie. 2023. Mutual-enhanced incongruity learning network for multi-modal sarcasm detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 9507–9515.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR.

Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, and Ruihong Huang. 2013. Sarcasm as contrast between a positive sentiment and negative situation. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 704–714.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Yi Tay, Luu Anh Tuan, Siu Cheung Hui, and Jian Su. 2018. Reasoning with sarcasm by reading inbetween. *arXiv preprint arXiv:1805.02856*.

Cynthia Van Hee, Els Lefever, and Véronique Hoste. 2018. Semeval-2018 task 3: Irony detection in english tweets. In *Proceedings of the 12th international workshop on semantic evaluation*, pages 39–50.

Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous graph attention network. In *The world wide web conference*, pages 2022–2032.

Jiaheng Wei, Yuanshun Yao, Jean-Francois Ton, Hongyi Guo, Andrew Estornell, and Yang Liu. 2024. Measuring and reducing llm hallucination without goldstandard answers via expertise-weighting. *arXiv preprint arXiv:2402.10412*.

Ben Yao, Yazhou Zhang, Qiuchi Li, and Jing Qin. 2024. Is sarcasm detection a step-by-step reasoning process in large language models? *arXiv preprint arXiv:2407.12725*.

Hongbin Ye, Ningyu Zhang, Shumin Deng, Xiang Chen, Hui Chen, Feiyu Xiong, Xi Chen, and Huajun Chen. 2022. Ontology-enhanced prompt-tuning for few-shot learning. In *Proceedings of the ACM Web Conference 2022*, pages 778–787.

Jianxing Yu, Qinliang Su, Xiaojun Quan, and Jian Yin. 2021. Multi-hop reasoning question generation and its application. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):725–740.

Jianxing Yu, Shiqi Wang, Libin Zheng, Qinliang Su, Wei Liu, Baoquan Zhao, and Jian Yin. 2023a. Generating deep questions with commonsense reasoning ability from the text by disentangled adversarial inference. In *Findings of the Association for Computational Linguistics*, pages 470–486.

Zhe Yu, Di Jin, Xiaobao Wang, Yawen Li, Longbiao Wang, and Jianwu Dang. 2023b. Commonsense knowledge enhanced sentiment dependency graph for sarcasm detection. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 2423–2431.

Tan Yue, Rui Mao, Heng Wang, Zonghai Hu, and Erik Cambria. 2023. Knowlenet: Knowledge fusion network for multimodal sarcasm detection. *Information Fusion*, 100:101921.

Yufeng Zhang, Meng-Xiang Wang, and Jianxing Yu. 2023. Answering subjective induction questions on products by summarizing multi-sources multi-viewpoints knowledge. In *2023 IEEE International Conference on Data Mining*, pages 848–857. IEEE.

Yufeng Zhang, Jianxing Yu, Yanghui Rao, Libin Zheng, Qinliang Su, Huaijie Zhu, and Jian Yin. 2024. Domain adaptation for subjective induction questions answering on products by adversarial disentangled learning. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, pages 9074–9089.

Jianan Zhao, Xiao Wang, Chuan Shi, Binbin Hu, Guojie Song, and Yanfang Ye. 2021. Heterogeneous graph structure learning for graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 4697–4705.

Qingqing Zhao, Yuhan Xia, Yunfei Long, Ge Xu, and Jia Wang. 2025. Leveraging sensory knowledge into text-to-text transfer transformer for enhanced emotion analysis. *Information Processing & Management*, 62(1):103876.

Liming Zhou, Xiaowei Xu, and Xiaodong Wang. 2024. Bns-net: A dual-channel sarcasm detection method considering behavior-level and sentence-level conflicts. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE.

Renbo Zhu, Meng Ma, and Ping Wang. 2021. Raga: relation-aware graph attention networks for global entity alignment. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 501–513. Springer.