Design of an Utterance-Relation Corpus for Conversational Intention Understanding

Yuji Nakashima, Virach Sornlertlamvanich, Kyoko Kanzaki, Thatsanee Charoenporn
Asia AI Institute (AAII), Faculty Department of Data Science, Musashino University
Tokyo, Japan
{s2422018, virach, k-kan, thatsane}@musashino-u.ac.jp

Abstract

Sarcasm detection remains a challenging task in natural language processing because the intention of an utterance cannot be easily judged by only the registered word or a single sentence. Previous studies mainly focus on words individually or isolated sentences as text sequences, making it difficult to take dialogue structures into consideration. To address this, we propose a novel graph-based sarcasm corpus grounded in Rhetorical Structure Theory (RST), where utterances are modeled as nodes and their contextual dependencies as edges. We annotated the dialogues with nine relation types and four fine-grained sarcasm categories: Lexical, Propositional, Illocutionary, and Like-Prefixed Sarcasm. Designed for graph-based approaches such as Graph Neural Networks (GNNs) and Graph Attention Networks (GATs), this corpus enables models to learn the structural context of utterance relationships. To evaluate its effectiveness, we conducted classification experiments using 690 dialogues. The proposed corpus achieved 83.65% accuracy, significantly outperforming the conventional text-sequence baseline (58.65%). These results demonstrate that incorporating dialogue structures and discourse relations provides a more effective corpus for context-aware sarcasm detection.

Keywords: Sarcasm Detection, Dialogue Corpus, Rhetorical Structure Theory, Discourse Relations, Corpus Annotation

1. Introduction

Sarcasm detection is an important task in natural language processing because the intended meaning of an utterance often differs from its literal meaning, making it difficult for computational models to correctly interpret speaker intention (Joshi et al., 2017). Accurate sarcasm detection is beneficial for applications such as sentiment analysis, dialogue systems, and opinion mining. However, sarcastic expressions often depend on contextual information rather than the lexical meaning of individual words or isolated sentences.

Previous studies have mainly focused on lexical features or sentence-level representations, while more recent approaches incorporate dialogue context by representing conversations as text sequences (Joshi et al., 2017; Castro et al., 2019). Although these methods improve contextual understanding, they generally ignore discourse structures that explicitly describe relationships between utterances. Consequently, contextual dependencies essential for sarcasm interpretation cannot be effectively modeled.

The widely used MUStARD corpus provides multi-party dialogues with binary sarcasm labels and has become a standard benchmark for multimodal sarcasm detection (Castro et al., 2019). However, it does not explicitly annotate discourse relations between utterances or distinguish different sarcasm types, limiting its applicability to graph-based models that utilize structural contextual information.

To address these limitations, we propose a graph-based sarcasm corpus grounded in Rhetorical Structure Theory (RST) (Mann and Thompson, 1988). Each dialogue is represented as a graph in which utterances are modeled as nodes and discourse relations as edges. In addition to binary sarcasm labels, sarcastic utterances are annotated with four fine-grained categories—Lexical, Propositional, Illocutionary, and Like-Prefixed sarcasm—following the taxonomy proposed by Camp (2012). The resulting corpus is designed for graph-based approaches such as Graph Neural Networks (GNNs) and Graph Attention Networks (GATs), enabling models to explicitly exploit discourse structures.

The main contributions of this paper are as follows:

1. We construct a graph-based sarcasm corpus by extending MUStARD with RST-based discourse annotations.
2. We annotate dialogues with nine discourse relation types and four fine-grained sarcasm categories.
3. We demonstrate that the proposed corpus provides more effective learning resources for context-aware sarcasm detection than the existing MUStARD corpus.

The remainder of this paper is organized as follows. Section 2 reviews related work on sarcasm detection, sarcasm corpora, and discourse analysis. Section 3 describes the construction of the proposed corpus. Section 4 presents the experimental evaluation, and Section 5 concludes the paper.

2. Related Work

2.1 Sarcasm Detection

Sarcasm detection has been extensively studied in natural language processing because sarcastic expressions often convey meanings that differ from their literal interpretations. Early studies mainly relied on lexical features, sentiment incongruity, and linguistic patterns to identify sarcasm (Joshi et al., 2017). With the development of deep learning, transformer-based models such as BERT have substantially improved sarcasm detection by learning contextual semantic representations (Devlin et al., 2019). More recently, dialogue-based approaches have incorporated conversational context to better interpret speaker intention (Castro et al., 2019). However, these methods generally represent dialogues as text sequences and do not explicitly model discourse relations between utterances.

2.2 Sarcasm Corpora

Several corpora have been developed for sarcasm detection. Twitter and Reddit datasets provide large-scale collections of sarcastic expressions obtained from user-generated content, typically using hashtags or community annotations as supervision (Ptáček et al., 2014; Khodak et al., 2018). However, these corpora usually contain short texts with limited conversational context. MUStARD introduced a multimodal dialogue corpus collected from television sitcoms and has become a widely used benchmark for sarcasm detection (Castro et al., 2019). Although MUStARD includes dialogue context, it provides only binary sarcasm labels and does not explicitly annotate discourse structures or fine-grained sarcasm categories.

2.3 Discourse Analysis

Discourse analysis aims to represent semantic and pragmatic relationships between text spans. Rhetorical Structure Theory (RST) models texts using hierarchical discourse relations such as Contrast, Cause, and Elaboration (Mann and Thompson, 1988). RST has been successfully applied to various NLP tasks, including discourse parsing, summarization, and sentiment analysis, by explicitly modeling relationships between textual units. Despite its effectiveness in representing discourse structures, RST has rarely been incorporated into sarcasm corpora. To the best of our knowledge, no existing sarcasm corpus explicitly represents dialogues as graphs using RST discourse relations while simultaneously providing fine-grained sarcasm type annotations. This gap motivates the construction of the proposed corpus.

3. Corpus Construction

3.1 Source Dialogues

The proposed corpus was constructed by extending the MUStARD corpus (Castro et al., 2019), which consists of sarcastic and non-sarcastic dialogues collected from television sitcoms. In addition to the original dialogues, we collected new dialogue segments from television series, including *The Big Bang Theory*, and manually organized them into dialogue units suitable for discourse analysis.

Each dialogue consists of multiple utterances spoken by different speakers. Rather than treating a dialogue as a plain text sequence, we preserve the conversational structure by explicitly representing utterances and their relationships. As summarized in Table 1, the resulting corpus contains 1,725 dialogues, 6,258 utterances, and 4,553 discourse relations.

Metric	Value
Dialogues	1,725
Utterances	6,258
Edges	4,553
Average utterances/dialogue	3.63
Average edges/dialogue	2.64
Average node degree	1.46
Average graph density	0.31

Table 1: Corpus Overview

The corpus is balanced across five sarcasm categories, including one non-sarcastic class, enabling both binary sarcasm detection and fine-grained sarcasm classification.

3.2 Annotation Scheme

Each dialogue is represented as a directed graph. Every utterance corresponds to a node containing the speaker name, utterance text, and target information. Contextual dependencies between utterances are represented as edges labeled with discourse relations.

We annotated each dialogue with three types of information.

First, each dialogue is assigned a binary sarcasm label indicating whether the target utterance is sarcastic.

Second, sarcastic utterances are classified into four fine-grained categories following the taxonomy proposed by Camp (2012): Lexical, Propositional, Illocutionary, and Like-Prefixed sarcasm. Together with the None category, the corpus supports both binary and multi-class sarcasm classification.

Third, discourse dependencies between utterances are annotated using Rhetorical Structure Theory (RST) (Mann and Thompson, 1988). We employ nine discourse relation types: Acknowledgment, Answer, Cause, Condition, Contrast, Elaboration, Elaboration/Comment, Enablement, and Question/Evaluation.

The complete annotation scheme is summarized in Table 2.

Level	Attribute	Description
Dialogue	show	Source television program
Dialogue	target_sarcasm	Binary sarcasm label
Dialogue	sarcasm_type	Five sarcasm categories
Node	speaker	Speaker identifier
Node	text	Utterance text
Node	is_target	Target utterance indicator
Edge	source	Source utterance ID
Edge	target	Target utterance ID
Edge	rst_relation	One of nine RST discourse relation

Table 2: Annotation Labels

3.3 Corpus Statistics

The proposed corpus contains a balanced distribution of sarcasm categories while preserving diverse discourse structures.

Table 3 presents the distribution of discourse relations. Contrast, Answer, and Question/Evaluation are the three most frequent relation types, accounting for the majority of discourse dependencies. These relations frequently occur in conversational exchanges where sarcastic expressions respond to preceding utterances.

Relation	Edges	Share (%)
Contrast	1,159	25.5
Answer	1,009	22.2
Question/Evaluation	906	19.9
Elaboration/Comment	835	18.3
Acknowledgment	338	7.4
Elaboration	180	4.0
Enablement	47	1.0
Condition	46	1.0
Cause	34	0.7
Total	4,553	100.0

Table 3: Distribution of RST relations

To examine the relationship between discourse structures and sarcasm types, Table 4 summarizes the co-occurrence of RST relations and sarcasm categories. Different sarcasm types exhibit distinct discourse patterns. For example, Lexical Sarcasm frequently appears with Question/Evaluation, whereas Propositional Sarcasm is strongly associated with Contrast and Answer. These observations suggest that discourse relations provide informative contextual cues for distinguishing different forms of sarcasm.

RST Relation	Lexical	Illocutionary	Propositional	Like-Prefixed	None	Total
Contrast	143	217	331	283	185	1,159
Answer	113	164	274	44	414	1,009
Question/ Evaluation	279	208	79	193	148	907
Elaboration/Comment	144	238	95	160	197	834
Acknowledgment	84	71	17	107	59	338
Elaboration	33	17	27	28	75	180
Enablement	6	26	4	4	7	47
Condition	7	23	6	4	6	46
Cause	8	7	7	3	9	34

Table 4: Distribution of RST Relations across Sarcasm Types

Finally, Table 5 reports the average dialogue size for each sarcasm category. Non-sarcastic dialogues contain slightly more utterances and discourse relations on average than sarcastic dialogues, while the overall dialogue length remains relatively short, making the corpus suitable for graph-based models that learn local conversational structures.

Sarcasm Type	Average Nodes	Average Edges
Illocutionary Sarcasm	3.61	2.81
Lexical Sarcasm	3.28	2.37
Like-Prefixed Sarcasm	3.36	2.39
None	4.39	3.19
Propositional Sarcasm	3.50	2.43

Table 5: Average Dialogue Size by Sarcasm Type

Overall, the proposed corpus extends existing sarcasm resources by explicitly incorporating discourse structures into dialogue representations. Unlike conventional text-sequence corpora, it provides graph-structured dialogue annotations together with fine-grained sarcasm categories, enabling the investigation of discourse-aware sarcasm detection models.

4. Corpus Evaluation

To evaluate the effectiveness of the proposed corpus, we conducted two experiments. First, we compared the proposed corpus with the widely used MUStARD corpus (Castro et al., 2019) under identical experimental settings. Second, we performed an ablation study to investigate the contribution of RST annotations by comparing models with and without discourse relation information. These experiments were designed to evaluate whether explicitly modeling discourse structures improves sarcasm detection.

4.1 Experimental Setup

For the corpus comparison, both the proposed corpus and MUStARD were evaluated using 690 dialogue samples. The proposed corpus was downsampled by stratified sampling to match the size and class distribution of MUStARD. Each sample was converted into a textual representation consisting of the dialogue context and the target utterance. Sentence embeddings were extracted using all-mpnet-base-v2, followed by a logistic regression classifier. The dataset was divided into training and test sets using an 85/15 split with a fixed random seed of 42. Performance was evaluated using Accuracy, Precision, Recall, and F1-score.

To evaluate the effectiveness of the proposed RST annotations, we further conducted an ablation experiment using the same graph-based sarcasm detection model. Two inference settings were compared: With RST, in which discourse relation labels were preserved, and Without RST, in which all edge attributes were removed while keeping the graph structure and model parameters unchanged. This setting isolates the contribution of discourse relation information.

4.2 Comparison with MUStARD

Table 6 presents the corpus comparison results.

Corpus	Accuracy	Precision	Recall	F1
MUStARD	58.65	58.18	61.54	59.81
Proposed	**83.65**	**86.81**	**94.05**	**90.29**

Table 6: Performance Comparison between MUStARD and the Proposed

The proposed corpus consistently outperformed MUStARD under identical experimental conditions. The proposed corpus achieved an accuracy of 83.65%, compared with 58.65% for MUStARD, while the F1-score improved from 59.81% to 90.29%. These improvements indicate that explicitly incorporating discourse structures and fine-grained sarcasm annotations provides richer contextual information for sarcasm detection than conventional text-sequence corpora.

4.3 Effect of RST Annotations

To investigate whether the performance improvement was attributable to RST annotations, we compared the graph-based model with and without discourse relation labels.

Setting	5-class Macro-F1	Stage 1 Binary-F1	Stage2 4-class Macro-F1
RST	59.83	86.23	60.65
Without RST	37.55	86.18	39.20

Table7: Ablation Study on the Effect of RST Relations

Removing RST annotations substantially reduced the five-class Macro-F1 from 59.83% to 37.55%, while the Stage 2 four-class Macro-F1 decreased from 60.65% to 39.20%. In contrast, the binary sarcasm detection performance remained almost unchanged (86.23% vs. 86.18%). These results indicate that discourse relations contribute primarily to fine-grained sarcasm classification rather than coarse binary sarcasm detection. The explicit representation of contextual dependencies between utterances enables the model to better distinguish different types of sarcasm.

4.4 Discussion

The experimental results demonstrate that the proposed corpus is a more effective resource for context-aware sarcasm detection than the existing MUStARD corpus. The comparison with MUStARD confirms that the proposed corpus provides more informative contextual representations under identical experimental settings. Furthermore, the ablation study shows that the observed performance gain is largely attributable to the inclusion of RST discourse annotations rather than differences in corpus size or model configuration. These findings suggest that explicitly modeling discourse relations is beneficial for graph-based sarcasm detection and highlight the importance of discourse-aware corpus construction.

5. Conclusion

In this paper, we proposed a novel graph-based sarcasm corpus based on Rhetorical Structure Theory (RST). Unlike conventional sequence-based datasets with binary labels, our corpus explicitly models dialogue structures by representing utterances as nodes and discourse relations as edges, while incorporating four fine-grained sarcasm categories.

Experiments demonstrated that the proposed corpus consistently outperformed the widely used MUStARD corpus under identical settings. Furthermore, an ablation study confirmed that removing RST annotations significantly degraded multi-class classification performance, highlighting that discourse relations provide critical contextual cues for fine-grained sarcasm understanding. These findings underscore the importance of discourse-aware corpus construction for improving graph-based sarcasm detection.

In future work, we plan to expand the corpus with dialogues from more diverse domains, such as social media and online forums, to enhance its generalizability.

6. Bibliographical References

Camp, E. (2012). Sarcasm, Pretense, and the Semantics/Pragmatics Distinction. *Noûs*, 46(1):58–81.

Castro, S., Hazarika, D., Pérez-Rosas, V., Zimmermann, R., Mihalcea, R., and Poria, S. (2019).

Towards Multimodal Sarcasm Detection (MUStARD). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)*, pages 4619–4629.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*, pages 4171–4186.

Joshi, A., Bhattacharyya, P., and Carman, M. J. (2017). Automatic Sarcasm Detection: A Survey. *ACM Computing Surveys*, 50(5):1–40.

Khodak, M., Saunshi, N., and Vodrahalli, K. (2018). A Large Self-Annotated Corpus for Sarcasm. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 2312–2318.

Mann, W. C. and Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text*, 8(3):243–281.

Ptáček, T., Habernal, I., and Hong, J. (2014). Sarcasm Detection on Czech and English Twitter. In *Proceedings of COLING 2014: The 25th International Conference on Computational Linguistics*, pages 213–223.

7. Language Resource References

Castro, S., Hazarika, D., Pérez-Rosas, V., Zimmermann, R., Mihalcea, R., and Poria, S. (2019). *MUStARD: A Multimodal Sarcasm Detection Corpus.* Distributed with the ACL 2019 publication.