Zero-Shot Dialogue Relation Extraction by Relating Explainable Triggers and Relation Names

Ze-Song Xu Yun-Nung Chen
National Taiwan University, Taipei, Taiwan
r10922a07@csie.ntu.edu.tw y.v.chen@ieee.org

Abstract

Developing dialogue relation extraction (DRE) systems often requires a large amount of labeled data, which can be costly and time-consuming to annotate. In order to improve scalability and support diverse, unseen relation extraction, this paper proposes a method for leveraging the ability to capture triggers and relate them to previously unseen relation names. Specifically, we introduce a model that enables zero-shot dialogue relation extraction by utilizing trigger-capturing capabilities. Our experiments on a benchmark DialogRE dataset demonstrate that the proposed model achieves significant improvements for both seen and unseen relations. Notably, this is the first attempt at zero-shot dialogue relation extraction using trigger-capturing capabilities, and our results suggest that this approach is effective for inferring previously unseen relation types. Overall, our findings highlight the potential for this method to enhance the scalability and practicality of DRE systems.[1]

1 Introduction

Relation extraction (RE) is a key natural language processing (NLP) task that identifies the semantic relationships between arguments in various types of text data. It involves extracting relevant information and representing it in a structured form for downstream applications (Zhang et al., 2017; Cohen et al., 2020; Zhou and Chen, 2021; Huguet Cabot and Navigli, 2021). Dialogue relation extraction (DRE) is a specialized area of RE that focuses on identifying semantic relationships between arguments in conversations. Recent DRE research has used diverse methods to improve relation extraction performance, including constructing dialogue graphs (Lee and Choi, 2021), identifying explicit triggers (Albalak et al., 2022; Lin et al., 2022), and using prompt-based fine-tuning approaches (Son et al., 2022).

Supervised training for RE tasks can be time-consuming and expensive due to the requirement for a large amount of labeled data. Models trained on limited data can only predict the relations they have been trained on, making it challenging to identify similar but unseen relations. Hence, recent research has explored methods that require only a few labeled examples or no labeled examples at all, such as prompt-based fine-tuning (Schick and Schütze, 2020; Puri and Catanzaro, 2019). Additionally, Sainz et al. (2021) improved zero-shot performance by transforming the RE task into an entailment task. However, this approach has not yet been applied to DRE due to the challenge of converting long conversations into NLI format.

In this work, we observe that different relations may be dependent on each other, such as the *parent-child* relationship listed in Table 1. Prior work has treated all relations independently and modeled different labels in a multi-class scenario, making it impossible for models to handle unseen relations even if they are relevant to previously seen relations. Therefore, this paper focuses on enabling zero-shot relation prediction. Specifically, if we encounter an unseen relation during testing but have previously seen a similar relation, we can relate them through explicitly mentioned trigger words, such as per:children (seen relation) → “mom” (trigger) → per:parents (unseen relation).

To achieve this, we need to identify the key information of the relation as a tool for relation reasoning during inference. We adopt the approach proposed in Lin et al. (2022), which achieves remarkable results in DRE by capturing explainable keywords in a dialogue for guiding relation extraction. By leveraging such trigger-capturing capabilities, our proposed model can better deduce unseen relations from known relations and associated triggers. Therefore, the proposed DRE model is more practical, as it can generalize to unseen relations.

[1] Code: https://github.com/MiuLab/UnseenDRE.

DialogRE Relation	Similar DialogRE Relation
per:positive_impression	per:negative_impression
per:boss	per:subordinate
per:children	per:parents
gpe:residents_of_place	per:place_of_residence
per:place_of_birth	gpe:births_in_place
org:students	per:schools_attended
per:visited_place	gpe:visitors_of_place
per:employee_or_member_of	org:employees_or_members

Table 1: Similar relation examples in DialogRE.

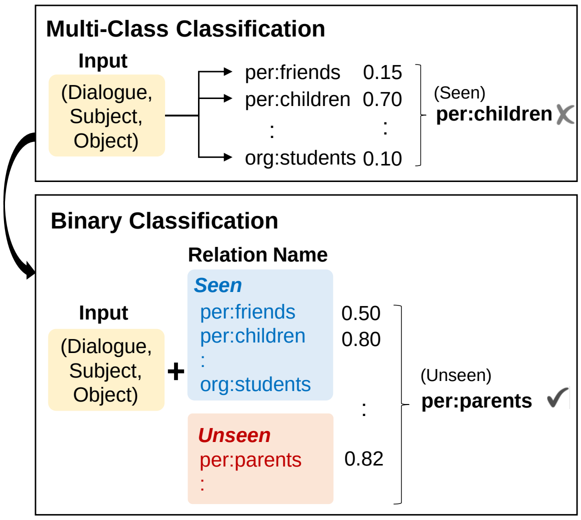

Figure 1: The illustration of our proposed zero-shot relation extraction model.

2 Proposed Approach

Prior work on classical DRE has treated it as a multi-class classification problem, which makes it challenging to scale to unseen relation scenarios. To enable a zero-shot setting, we reformulate the multi-class classification task into multiple binary classification tasks by adding each relation name as input, as illustrated in Figure 1. The binary classification task predicts whether the subject and object in the dialogue belong to the given relation. This approach is equivalent to predicting whether a set of subject-object relations is established, which can estimate any relations based only on their names (or natural language descriptions).

2.1 Model Architecture

Our model is illustrated in Figure 2, where there are three components in our architecture.

Trigger Prediction Inspired by Lin et al. (2022), we incorporate a trigger predictor into our model, allowing us to employ explicit cues for identifying subject-object relationships within a dialogue. Specifically, we adapt techniques from question-answering models to predict the start and end positions of the trigger span. By detecting these triggers, our model not only reasons the potential unseen relations but also enhances the interpretability of the task, making it more practical for real-world applications. To identify the keywords associated with `(Subject, Object, RelationType)` in a dialogue, we formulate the task as an extractive question-answering problem (Rajpurkar et al., 2016). In this setting, the dialogue can be viewed as a document, where the subject-object pair represents the question, and the answer corresponds to the span of keywords that explain the associated relation, i.e., the triggers.

Relation Name Injection In contrast to most prior work (Lee and Choi, 2021; Lin et al., 2022; Albalak et al., 2022), our input format includes the relation name after `[CLS]`, and we use the `[CLS]`-associated embeddings as relation name embeddings shown in Figure 2. By doing so, the model has access to *natural language descriptions* of the given relation, which facilitates more accurate capture of trigger words and further enables the zero-shot capability of the proposed model.

Binary Relation Prediction In our model, the relation predictor takes as input the learned relation name embedding and a predicted trigger span, as illustrated in the upper part of Figure 2. To establish the relationship between the relation name and its associated trigger words, we employ a general attention mechanism, where the relation name embedding serves as the query, while the trigger words are encoded by BERT and used as keys and values. The resulting features are then concatenated and fed through a fully connected layer, which generates the final prediction indicating whether the input subject and object have the given relation as

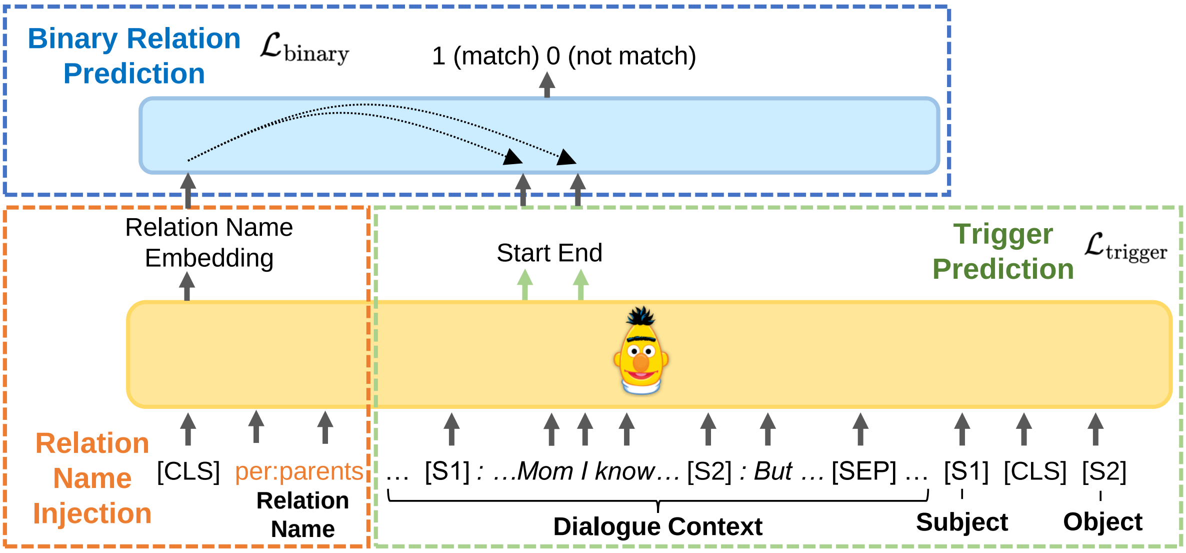

Figure 2: The illustration of our proposed model architecture.

expressed in the dialogue.

2.2 Training

As depicted in Figure 2, the input (`Dialogue`, `Subject`, `Oubject`, `RelationType`) will be initially expanded into a sequence resembling BERT's input format. The model is trained to perform two tasks. Firstly, it learns the ability to find the trigger span, and secondly, it learns to incorporate the triggers into the relation prediction.

Negative Sampling In accordance with Mikolov et al. (2013), we have adopted the negative sampling method in our training process. Specifically, we randomly select some relations from the set of previously observed relations that do not correspond to the given subject-object pair to create negative samples. Notably, the trigger spans of these negative samples remain unchanged.

Multi-Task Learning The trigger prediction task involves identifying the most likely trigger positions, and is treated as a single-label classification problem using cross-entropy loss $\mathcal{L}_{Trigger}$. On the other hand, the relation prediction task employs binary cross-entropy loss $\mathcal{L}_{Binary}$ to compute the prediction loss. To train the model simultaneously on both tasks, we employ multi-task learning. We use a linear combination of the two losses as the objective function. This enables us to train the entire model in an end-to-end fashion.

2.3 Inference

During inference, our model follows a similar setting to the one used during training. However, we have observed that the model tends to predict the seen relation when the captured trigger words are present in the training data. To prevent the model from overfitting to the seen relations, we replace the trigger span with a general embedding (the embedding of `[CLS]`), which is assumed to carry the information of the entire sentence. This embedding is used as the input for our relation prediction. By doing so, our model can better generalize to unseen scenarios and can avoid the tendency to predict the seen relation when capturing seen trigger words. This approach enhances the model's ability to handle diverse unseen relations during inference.

3 Experiments

We conducted experiments using the DialogRE dataset, which is widely used as a benchmark in the field. To assess our model's zero-shot capability, we divided the total of 36 relations into 20 seen and 16 unseen types detailed in the Appendix. We only train our model on data related to seen relation types. During training, we set the learning rate to 3e-5 and used a GeForce RTX 2080 Ti. The training process involves 10 epochs without early stopping[2], and the number of negative samples was 3. To ensure a fair comparison with prior work (Lin et al., 2022; Yu et al., 2020), we use the same testing set for evaluation.

3.1 Evaluation Metric

After performing multiple binary classification tasks, our model can rank the relation candidates

[2] The models with early stopping achieve similar performance.

Model	Unseen		Seen		Overall[2]	
	Top 1	Top 2	Top 1	Top 2	Top 1	Top 2
Multi-class BERT	0.0	0.0	60.6	-	48.5	-
TUCORE-GCN (Lee and Choi, 2021)	0.0	0.0	65.5[1]	-	48.4[1]	-
TREND (Lin et al., 2022)	0.0	0.0	**66.8**[1]	-	53.4[1]	-
Binary-Reformulated BERT	24.5	28.9	57.0	45.5	50.5	42.2
Proposed (with predicted triggers)	23.5	**34.8**	66.7	**51.5**	58.0	**48.2**
Proposed (with relation name embeddings)	**32.5**	**34.8**	65.6	51.0	**60.0**	47.8
Proposed with gold triggers	35.6	40.4	70.4	53.2	63.4	50.6

Table 2: The micro-F1 performance of DialogRE in terms of unseen, seen, and overall settings (%).

based on their predicted scores. Typically, the model outputs the relation with the highest score, as done in prior work, and micro-F score is calculated for evaluation. However, since our task is focused on zero-shot performance, we are also interested in whether our model can correctly rank the unseen relations, even if the top-ranked relation is incorrect. To better understand how our model estimates all relation candidates, we evaluate our model not only on the top-ranked relation but also on the top-2 ranked relations in our experiments. This allows us to gain insight into how well our model can rank the correct relations, even if they are not the top-ranked ones.

3.2 Model Setting

We perform different model settings on BERT-Base for fair comparison.

- **Multi-class BERT** is a baseline, where BERT-Base (Devlin et al., 2019) is adopted and treated DRE as multi-class classification.
- **TUCORE-GCN** construct a dialogue graph to utilize the graph strucutre for prediction (Lee and Choi, 2021).
- **TREND** proposed to capture explicit triggers for better performance (Lin et al., 2022).[3]
- **Binary-reformulated BERT** performs binary classification shown in Figure 1, which is a proper baseline for zero-shot settings.
- **Proposed** has three settings in binary relation prediction during inference: 1) based on predicted triggers, 2) based on relation name embddings, 3) based on gold triggers. The third is listed as an upper bound for reference.[4]

3.3 Results

Table 2 presents our results. Prior work achieves micro-F scores above 60% for seen relations but cannot predict unseen relations (0%) due to their multi-class formulation. The reformulated BERT serves as the baseline for zero-shot settings, achieving 24.9% and 28.9% for top 1 and top 2 ranked relations, respectively.

Our proposed method of inputting predicted triggers for relation prediction did not rank correct unseen relations as top 1 (23.5% vs. 24.5%). However, the performance of top 2 ranked relations significantly improved (from 28.9% to 34.8%), suggesting that trigger prediction is indeed useful. The lower top 1 relations score can be attributed to similar triggers for relevant relations, which easily favor seen relations. An example of incorrect prediction is provided in Table 3.

Replacing predicted triggers with relation name embeddings, our proposed model achieves the best performance for unseen relations (32.5% for top 1 and 34.8% for top 2). This indicates that this setting avoids overfitting to seen relations and allows prediction to better generalize to unseen scenarios.

Moreover, feeding gold triggers into relation extraction during inference yields the best results, indicating the potential for improvement with the proposed trigger mechanism. In sum, the experiments demonstrate that our proposed model can connect trigger words with relation names and enables zero-shot relation extraction.

In terms of performance on seen data, our proposed models outperform the reformulated BERT baseline by a significant margin. Moreover, our models achieve comparable scores to previous work (66.7% vs. 66.8% in top 1 scores), even though we consider more candidates. These results further validate the effectiveness of our model and its superior generalization capability.

[3] The scores are reported from the prior work for reference, which cannot be directly compared with our scores.

[4] Overall performance is estimated based on data size.

S1: What about Ben? We can't bring a baby to a hospital. S2: We'll watch him. S1: I don't think so. S3: What? I have seven Catholic sisters. I've taken care of hundreds of kids. Come on, we wanna do it, don't we? S2: I was looking forward to playing basketball, but I guess that's out the window. S1: Ok, well, if you do take him out for his walk, you might wanna bring his hat, and there's extra milk in the fridge, and there's extra diapers in the bag. S3: Hat, milk, got it. S1: ??? Thro up a thro thro–a thro thro! S3: Consider it done. S2: You understood that? S3: Yeah, my uncle Sal has a really big tongue. S2: Is he the one with the beautiful wife?
(Subject, Object) : (Sal, S3) Predicted trigger: uncle Gold trigger: uncle Predicted relation: per:children Gold relation: per:other_family

Table 3: An incorrectly-predicted example.

After comprehensive analysis, we found that our proposed method incorporating a general context embedding not only leverages the trigger capturing capability but also assists the DRE task indirectly, leading to the best overall performance among all proposed models. The ability to relate trigger keywords to relation names enables the model to generalize better to unseen relations and overcome the limitations of relying on specific trigger words. The results of our experiments demonstrate the effectiveness of our proposed method and its potential for real-world applications.

3.4 Qualitative Study

Table 3 showcases an example about the predicted triggers and relations for the DialogRE dataset. As an instance, Sal is the uncle of Speaker 3, so the relation between them should be "other_family". Although the trigger word mechanism accurately captures the crucial keyword "uncle", the model still outputs the "children" relation from the seen relation category rather than the "other_family" relation from the unseen relation category. This suggests that while capturing significant subject and object information through trigger words, the model tends to prioritize predicting relations from the seen relation category.

4 Conclusion

This paper introduces a novel approach for zero-shot dialogue relation extraction by relating explainable trigger words and relation names. Our proposed method effectively utilizes trigger-capturing capability and demonstrates a significant improvement in inferring unseen relations. The experimental results on benchmark data show that our approach achieves better generalization and practicality, making it a promising solution for real-world applications.

Acknowledgements

We thank the reviewers for their insightful comments. This work was financially supported by the Young Scholar Fellowship Program by the National Science and Technology Council (NSTC) in Taiwan, under Grants 111-2222-E-002-013-MY3 and 111-2628-E-002-016.

References

Alon Albalak, Varun Embar, Yi-Lin Tuan, Lise Getoor, and William Yang Wang. 2022. D-REX: Dialogue relation extraction with explanations. In *Proceedings of the 4th Workshop on NLP for Conversational AI*, pages 34–46.

Amir D. N. Cohen, Shachar Rosenman, and Yoav Goldberg. 2020. Relation extraction as two-way span-prediction. *CoRR*, abs/2010.04829.

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 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Pere-Lluís Huguet Cabot and Roberto Navigli. 2021. REBEL: Relation extraction by end-to-end language generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2370–2381, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Bongseok Lee and Yong Suk Choi. 2021. Graph based network with contextualized representations of turns in dialogue. In *EMNLP 2021-2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, pages 443–455. Association for Computational Linguistics (ACL).

Po-Wei Lin, Shang-Yu Su, and Yun-Nung Chen. 2022. TREND: Trigger-enhanced relation-extraction network for dialogues. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 623–629, Edinburgh, UK. Association for Computational Linguistics.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space.

Raul Puri and Bryan Catanzaro. 2019. Zero-shot text classification with generative language models. *CoRR*, abs/1912.10165.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. SQuAD: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Austin, Texas. Association for Computational Linguistics.

Oscar Sainz, Oier Lopez de Lacalle, Gorka Labaka, Ander Barrena, and Eneko Agirre. 2021. Label verbalization and entailment for effective zero- and few-shot relation extraction. *CoRR*, abs/2109.03659.

Timo Schick and Hinrich Schütze. 2020. Exploiting cloze questions for few-shot text classification and natural language inference. *CoRR*, abs/2001.07676.

Junyoung Son, Jinsung Kim, Jungwoo Lim, and Heuiseok Lim. 2022. GRASP: Guiding model with RelAtional semantics using prompt for dialogue relation extraction. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 412–423, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Dian Yu, Kai Sun, Claire Cardie, and Dong Yu. 2020. Dialogue-based relation extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4927–4940, Online. Association for Computational Linguistics.

Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning. 2017. Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 35–45, Copenhagen, Denmark. Association for Computational Linguistics.

Wenxuan Zhou and Muhao Chen. 2021. An improved baseline for sentence-level relation extraction. *CoRR*, abs/2102.01373.

A Criteria for Relation Dividing

We categorized the relations into two sets, namely, seen and unseen, as presented in Table 4. Our categorization was based on the similarity of relations, where dependent ones are assigned to different categories. For those not related, we assigned them randomly to either category. This categorization aims to train the model on seen relations to enhance its ability to predict unseen relations during testing.

B Prediction Distribution Comparison

We analyze the distribution of correctly predicted top 1 unseen relations for two models, one with predicted triggers and the other with relation name embeddings, and present the results in Table 5. We observe that the two methods exhibit a similar pattern of correctly predicted relations, with a concentration on particular unseen relations such as siblings and spouses, among others. However, the proposed method with the relation name embeddings significantly outperforms the one with the predicted triggers method in this aspect.

Seen Relations	Unseen Relations
per:positive_impression	per:subordinate
per:client	gpe:visitors_of_place
per:origin	per:place_of_residence
per:works	per:schools_attended
per:place_of_work	per:parents
per:title	gpe:births_in_place
per:alternate_names	org:employees/members
per:acquaintance	per:dates
per:alumni	per:other_family
per:friends	per:siblings
per:girl/boyfriend	per:spouse
per:neighbor	per:negative_impression
per:roommate	per:age
per:boss	per:date_of_birth
per:children	per:major
gpe:residents_of_place	per:pet
per:place_of_birth	
per:visited_place	
per:employee/member_of	
org:students	

Table 4: Seen and unseen relations in our experiments.

Unseen Relation	Unseen	
	Predict	CLS
per:siblings	26	42
per:spouse	21	30
per:negative_impression	4	11
per:parents	5	9
per:dates	0	4
per:major	2	2
per:age	1	1
gpe:births_in_place	0	0
org:employees/members	0	0
per:other_family	0	0
per:date_of_birth	0	0
per:pet	0	0
per:subordinate	0	0
gpe:visitors_of_place	0	0
per:place_of_residence	0	0
per:schools_attended	0	0

Table 5: The distribution of correct predictions in the predict trigger method and cls trigger method.