Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 2337–2344
Marseille, 20-25 June 2022
© European Language Resources Association (ELRA), licensed under CC-BY-NC-4.0

The CRECIL Corpus: a New Dataset for Extraction of Relations between Characters in Chinese Multi-party Dialogues

Yuru Jiang*, Yang Xu*, Yuhang Zhan*, Weikai He*, Yilin Wang* , Zixuan Xi*, Meiyun Wang*, Xinyu Li*, Yu Li*, Yanchao Yu†
* Computer School, Beijing Information Science and Technology University
jiangyuru@bistu.edu.cn

† School of Computing, Edinburgh Napier University
y.yu@napier.ac.uk

Abstract

We describe a new freely available Chinese multi-party dialogue data set for automatic extraction of dialogue-based character relationships. The data has been extracted from the original TV scripts of a Chinese sitcom called "I Love My Family" with complex family-based human daily spoken conversations in Chinese. First, we introduced human annotation scheme for both global Character relationship map and character reference relationship. And then we generated the dialogue-based character relationship triples. The corpus annotates relationships between 140 entities in total. We also carried out a data exploration experiment by deploying a BERT-based model to extract character relationships on the CRECIL corpus and another existing relation extraction corpus (DialogRE (Yu et al., 2020)).The results demonstrate that extracting character relationships is more challenging in CRECIL than in DialogRE.

Keywords: Multi-Party Dialogue; Character Relation; Corpus; Annotation Scheme; Relation Extraction

1. Introduction

While bringing intelligent systems/robots out of the laboratory into the physical world, especially into the daily environment, they must become capable of understanding natural daily conversations between two or more human speakers. Among other capabilities, this involves the ability to identify/extract the character[1]/speaker relationships by analyzing the semantic meanings of conversations between different users – this is widely known as the *Relation Extraction (RE)* problem. Solving such a problem has become a critical task in Natural Language Processing, and it plays an essential role in downstream tasks.

In the past decades, the RE problem in language processing has received considerable attention in computational linguistics. On the one hand, there is work that only addresses the relation extraction problem on the document level. Some datasets focus on the RE problem between two entities in one sentence, such as the SemEval-2010 Task 8 dataset(Hendrickx et al., 2010) and the TACRED dataset(Zhang et al., 2017). And some datasets focus on the cross-sentence relationships, i.e. two entities whose relation needs to be determined are not in the same sentence or even the same paragraph, such as BC5CDR (Li et al., 2016) and DocRED dataset (Yao et al., 2019). Furthermore, different to the formal document, some RE datasets focus on annotating dialogues. There are much more complicated connections between different speakers in a dialogue, especially with more than two participants. The relationship in the dialogue is more character-related, and a complete dialogue exists between different speakers. In addition to many cross-sentence relationships, there is also a large amount of omission and co-reference information in the dialogue. So, researchers published some datasets for dialogue-based relation extraction tasks. Chen et al. (2020b) published a multi-party dialogue dataset, called MPDD, for analysing emotions and interpersonal relationships based on the dialogue in five different TV series scripts. MPDD only focus on the relationship between speaker and listener. Yu et al. (2020) constructed a multi-party dialogue relationship extraction dataset DialogRE based on the English script - Friends. Moreover, they also published a Chinese version of the same dataset using machine translation technology, but it cannot represent practical daily conversations between Chinese native speakers. The statistical analysis of the "Friends" (Chinese episode) and "I Love My Family" (Chinese episode) shows that there are differences between them in terms of the number of characters, the number of character relationships, and the number of dialogue rounds in each episode. Therefore, the DialogRE dataset's Chi-

[1]This paper is about social relationships between literary characters or protagonists in a story, rather than relations between orthographic Chinese characters.

nese version cannot cover all features/phenomena in Chinese multi-party dialogue. Furthermore, the character relationship schema defined by DialogRE is not suited for "I Love My Family".

In this paper, we present a new dialogue data set - the CRECIL corpus[2] - constructed from the authentic TV scripts of a Chinese sitcom - I Love My Family - labelled using an inter-agreed annotation scheme, for the task of dialogue-based character relationship extraction. Different to the DialogRE, which directly annotates the relationship between two arguments, our corpus annotates the global character relationships between two character entities and the referential relationships between a reference in the utterance and the character entity. We, then, automatically generate the character relationship triples by fusing the relationships of each episode, as mentioned earlier. We carry out a data experiment by investigating the performance of a BERT-based model (as baseline) (proposed by (Yu et al., 2020)) on this dialogue dataset. The result shows that the extraction of relationships in CRECIL is more challenging than DialogRE's English version.

2. Data annotation method on Multi-party Dialogue

This section describes our data annotation method and process, including the inter-agreed annotation scheme and character relationship triples generation.

The corpus is extracted from the original TV scripts of a Chinese sitcom called "I Love My Family"[3] which contains a total of 120 episodes and 679 scenes. It is not annotated directly on audio data, and the data is also not naturally occurring real-world speech but rather television scripts.

2.1. Cleaning up the data for Annotation

To annotate character relationships between speakers within the conversation in the scene, we cleaned up the original scripts as follows: 1) removing useless narrations, transitions, and text descriptions and 2) retaining only the content of the dialogue. (see dialogue example in Table 1).

Before Cleaning
（客厅，一家刚吃完饭，圆圆打开电视，志国上） (In the lecture hall, the family has just finished eating, Yuanyuan turns on the TV, Zhiguo enters) 圆圆：爸，动画片儿哪频道啊? Yuanyuan: Dad, what channel is the cartoon on? 志国：看哪门子动画片呀—-看连续剧（两人争执不下） Zhiguo: What kind of cartoon to watch —- watch a series (the two can't argue)
After Cleaning
圆圆：爸，动画片儿哪频道啊? Yuanyuan: Dad, what channel is the cartoon on? 志国：看哪门子动画片呀—-看连续剧 Zhiguo: What kind of cartoons are you watching? - TV series

Table 1: An Example of Data Cleaning

2.2. Annotation Scheme

In order to reach a new corpus with high quality but less labelling cost, we, in this paper, apply a simple annotation method[4] for labelling both the global character relationship and the referential relationship. We then automatically generate the character relationship triples using a rule-based model given both types of labelled information.

Character Relationship Triples A character relationship triple (CRT) is about two characters and their social relationship. We define the CRT as (a_i, r, a_j), where a_i and a_j represent the full name of two *character entities*, and r represents *the type of character relationship* between them. Given Chinese social background, legal relations and other actual conditions in "I Love My Family", there are more complex family characteristics. We retained 17 relationship types from the original DialogRE data set and introduced 13 new relationship types(marked with asterisks in Table 2) that apply to this dialogue and other Chinese dialogues. Hence, the corpus finally contains a total of 30 relationship types.

2.2.1. Annotation of the global character relationships

This paper labelled the relationships between character entities in the scripts as a global character relationship knowledge graph. Figure 1 shows the global character relationships between six main characters, in which each node represents a character entity's name. The arrow indicates a particular relationship type from one character entity to another. For example, in Figure 1, the double-head arrow between '贾志国 (Jia Zhiguo)' and '贾小凡 (Jia Xiaofan)' indicates that they are siblings to each other. The set of rules is listed below:

- Each character, appearing for the first time in the script, will be inserted into the graph as a new node. The node's name is the character's name. However, if the character's name is unknown when he/she first appears, the node's

[2]https://github.com/bistu-nlp-lab/CRECIL

[3]It is the first Mandarin-language sitcom using multi-camera technology in China.

[4]The annotated scheme has passed the internal agreement. After unifying the annotation rules, we adopt the mode of two people labelling back to back, then achieving statistical consistency and unified recognition, and then labelling. The consistency of the reference labels in the first eight episodes reaches more than 92%, and the average consistency is 94.15%.

Index	Main body	Relationship type	Object	Inverse relationship
1	Person	per:positive impression	Name	—
2	Person	per:negative impression	Name	—
3	Person	per:acquaintance	Name	per:acquaintance
4	Person	per:boss	Name	per:subordinate
5	Person	per:subordinate	Name	per:boss
6	Person	per:client	Name	—
7	Person	per:dates	Name	per:dates
8	Person	per:friends	Name	per:friends
9*	Person	per:classmate	Name	per:classmate
10	Person	per:neighbor	Name	per:neighbor
11	Person	per:children	Name	per:parents
12	Person	per:parents	Name	per:children
13*	Person	per:parents-in-law	Name	per:children-in-law
14	Person	per:siblings	Name	per:siblings
15*	Person	per:siblings-in-law	Name	per:siblings-in-law
16	Person	per:spouse	Name	per:spouse
17*	Person	per:relative	Name	per:relative
18*	Person	per:children-in-law	Name	per:parents-in-law
19*	Person	per:grandparents	Name	per:grandchildren
20*	Person	per:grandchildren	Name	per:grandparents
21*	Person	per:nurse	Name	—
22*	Person	per:ex-girlfriend	Name	per:ex-boyfriend
23*	Person	per:ex-boyfriend	Name	per:ex-girlfriend
24*	Person	per:teacher	Name	per:student
25*	Person	per:student	Name	per:teacher
26	Person	per:girlfriend	Name	per:boyfriend
27	Person	per:boyfriend	Name	per:girlfriend
28	Person	per:alternate_name	String	—
29*	Person	per:colleague	Name	per:colleague
30	Name	unanswerable	Name/String	—

Table 2: Chinese multi-party dialogue character relationship types

name will depend on how he/she is mentioned the first time in the script. Otherwise, the node's name will be the speaker's name of the same character.

- The character's name referring to the same character entity will be consistent for both the global character relationship labelling and the referential relationship labelling(see sec 2.2.2).
- All the relationships between each pair of character entities should be annotated. We also need to tag the time stamp when such relationships appear for the first time because the relationship between two character entities could change over time.
- When more than one name refers to the same character entity, their relationship will be alternate_name.
- Relationship types that are not obvious or cannot be answered would not be annotated.

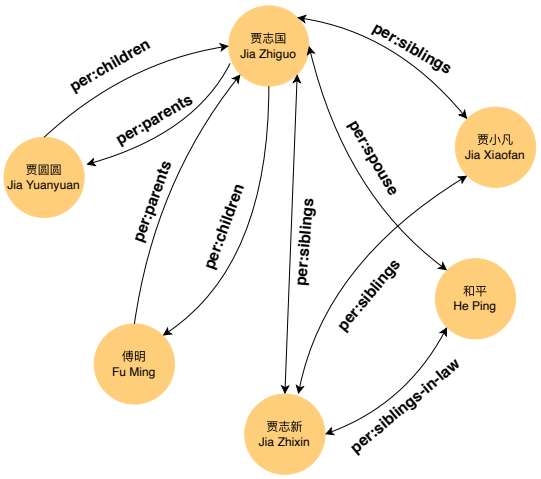


Figure 1: Global relationship diagram.

2.2.2. Annotation of the referential relationships

We also annotated two kinds of referential relationships (see the schematic diagram in Figure 2). One referential relationship is between different pronouns (referents) in the utterance and the corresponding character entities in the conversation. The other referential relationship is between the speakers and the corresponding character entities. The referential relationship triples (RRT) is defined as (a', a, p), where a' is the pronouns (referents) in the utterance $u_i (a' \in u_i)$ or the speaker. a is the character entity's name referred to by a' and p represents the position index where a' is located in u_i. (see an example of '(爸, 贾志国, 0)' in Figure 2)

Here, we also define and employ the annotation rules as follows:

- If the reference to a character entity is a single word or the entity's name, create an RRT for it.
- If the reference to a character entity has some attribute, creating an RRT for the reference and the attributive modification together.
- Create an RRT for the character's nickname.
- When there is a pronoun before or after a reference, and they refer to the same character entity, creating an RRT only for the reference itself.

2.2.3. Dialogue-based CRT Generation

- Given that p_i denote one speaker or reference to a character entity in the current scene (dialogue), if the relationship r between p_i and p_j can be found in the global CRTs, we will set their relationship as r, otherwise set as 'unanswerable'.
- Given the same dialogue, we only keep a unique relationship triple between p_i and p_j.

Figure 3 presents an example of dialogue-based CRTs were generated from one of the conversations in the CRECIL corpus.

Average turns length(in tokens)	23.8
Average dialogue length(in tokens)	707.6
Average # of turns	29.7
Average # of speakers	4.1
Average # of sentences	39.4
Average # of relational instances	57.4
Average # of no-relational instances	21.6

Table 3: Statistics per dialogue of CRECIL

3. Comparison between CRECIL and DialogRE

We investigate the similarities and differences between our CRECIL and the DialogRE corpus.

3.1. Statistics and Analysis of Corpus

Given the scripts of the "I Love My Family" Chinese sitcom, we have gathered 679 dialogues (each about one scene), with a total of 20,183 turns. Table 3 shows the distribution of dialogue length (i.e. the number of turns) in the CRECIL corpus, where the average number of turns per dialogue is 29.7 (which is significantly longer than English data in DialogRE (12.9 turns on average)). More speakers participated in a single conversation in CRECIL (4.1 speakers per dialogue on average) than those in DialogRE (3.3 speakers per dialogue on average). We show the distribution of multi-party conversations in CRECIL and DialogRE in Figure 4. The more multi-party dialogue occurs in the data, the more challenging the system can extract accurate character relationship triples.

3.2. Statistics and Analysis of Relation Types

We have identified 30 character relationship types as mentioned in table 2. Using the above annotation scheme, we have finally labeled 121 character entities with 501 global CRTs across those categories (see the distribution of global CRTs across those categories in Figure 5). We have also annotated a total of 8282 RRTs. Based on the global CRTs and RRTs, we have generated over 53,646 dialogue-based CRTs(which is directly annotated in DialogRE), which are more than the ones in DialogRE (only 10,168 in the English version and 11,365 in the Chinese-translated version). Table 4 shows the number of ways to address each character in CRECIL and DialogRE respectively[5]. Each character in the CRECIL corpus has about 71.83 different names/referents on average, significantly more than in the DialogRE corpus (only 19.67 per character on average).

On the other hand, the top three of these categories in global character relationships (excluding unan-

[5]We, here, only present the various character names for the 6 main characters in each.

CRECIL		DialogRE	
Character Name	**number**	**Character Name**	**number**
傅明 Fu Ming	102	Chandler Bing	**31**
贾志国 Jia Zhiguo	92	Ross Geller	**29**
贾志新 Jia Zhixin	81	Joey Tribiani	20
和平 He Ping	74	Rachel Karen Green-Geller	17
贾圆圆 Jia Yuanyuan	58	Monica Geller	13
贾小凡 Jia Xiaofan	24	Phoebe	8
Average	71.83	Average	19.67

Table 4: Comparison between CRECIL and DialogRE on Character Referents

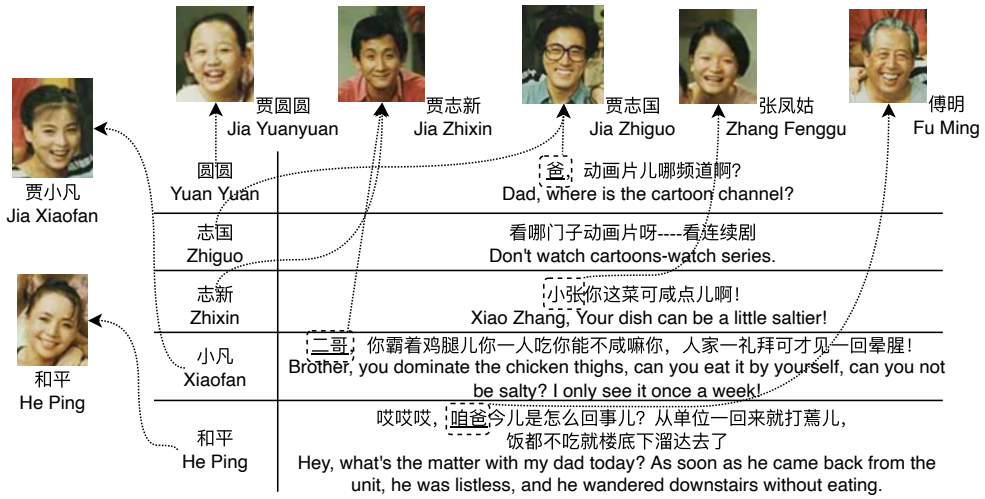


Figure 2: Schematic diagram of referential relationship labelling in the CRECIL corpus

Algorithm 1 Generating the dialogue based CRTs

function GENERATE-DCRT($D, gCRTs, dRRTs$)
  Input: dialogue D, global CRTs $gCRTs$, dialogue RRTs $dRRTs$
  Output: dialogue CRTs $dCRTs$

  Initialised the set of speakers S and references M
  S = get speakers From dialogue (D)
  M = get references from dialogue (D)
  Remove duplication $P = Set(S + M)$

  set $dCRTs = \{(a1, a2, r)|a1 \in P \& a2 \in P \& a1! = a2 \& r = `unanswerable'\}$

  # Update $dCRTs$ with $gCRTs$
  for $dcrt$ in $dCRTs$ **do**
    for $drrt$ in $dRRTs$ **do**
      if $dcrt.a1 == drrt.a$ **then**
        $a1 = drrt.a$'
      end if
      if $dcrt.a2 == drrt.a$ **then**
        $a2 = drrt.a$'
      end if
    end for
    $dcrt.r = gcrt.r$ where $gcrt.a1 = a1 \& gcrt.a2 = a2$
  end for

  Return $dCRTs$
end function

swerable) are 'neighbours' (with 97 global CRTs), 'friends' (with 58 global CRTs) and 'acquaintances' (with 34 global CRTs), which overall account for 37.7% of the total. (see more details in Figure 5). Figure 6 shows the distribution of dialogue-based CRTs across different relationship types (excluding unanswerable) in the CRECIL and DialogRE corpus. 'alternate name' (around 6700 samples) and 'neighbour' (about 5400 samples) are significantly more than other categories. Such imbalanced data sample distribution also happens in the DialogRE corpus. For example, the relationship type of 'alternate name' (containing around 2150 samples in the English corpus and about 3400 samples in the Chinese-translated one) occurs more frequently than the ones in 'girl/boyfriend', 'positive' and 'friend' separately (around 800 only).

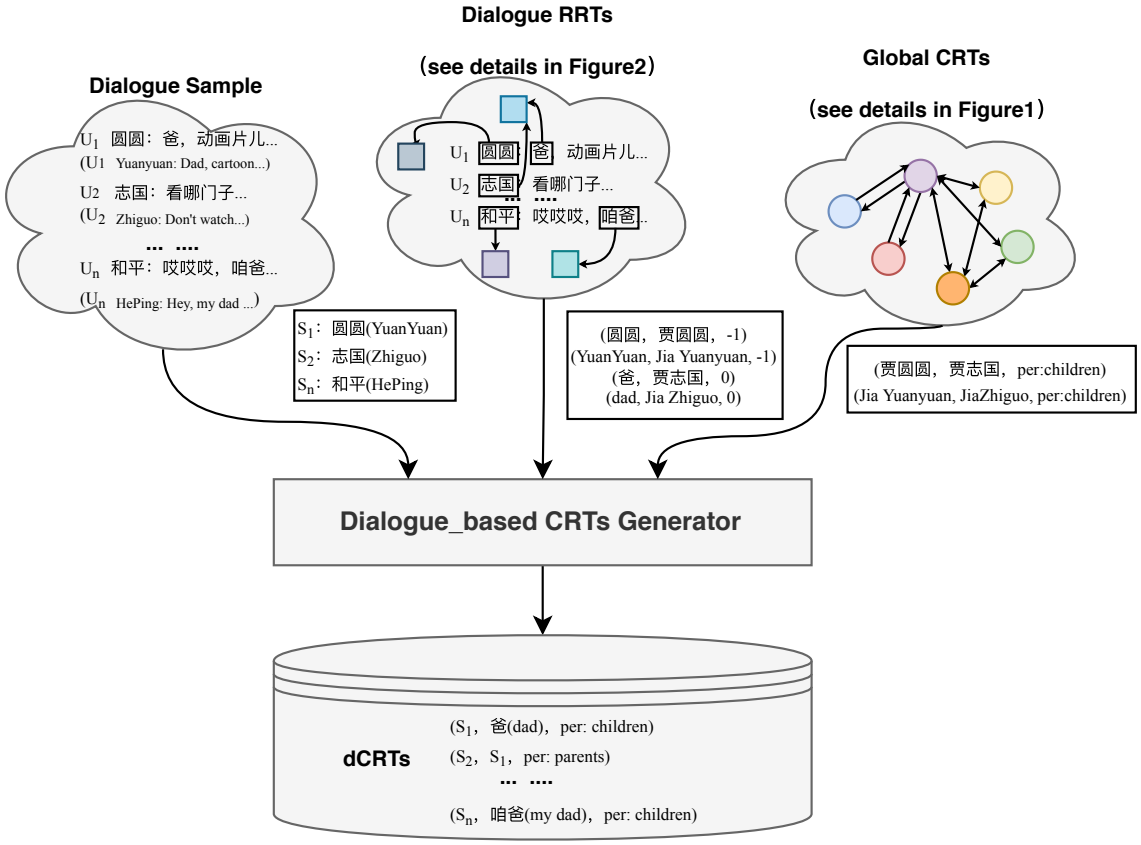


Figure 3: Example of Dialogue-based Character Relationship Triples Generator in the CRECIL corpus.

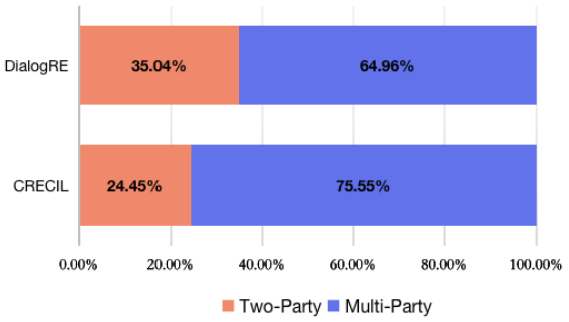


Figure 4: Multi-party dialogue Distribution in CRECIL and DialogRE corpus

4. Data Experiment

In order to demonstrate how the CRECIL corpus can be used, we deployed one of the existing approaches (BERT model) from Yu et al. (2020) to extract dialogue-based character relations. We follow the previous task and standard experiment settings (see (Yu et al., 2020)) to compare[6]. Unlike previous work, this will compare the overall performance and the performance of specific relationship types separately.

4.1. Experiment Task Formulation

Following the previous work (Yu et al., 2020; Chen et al., 2020a), we, here, redefine the dialogue-based character relation extraction (DCRE) task. Given a dialogue $D = (s_1, u_1), (s_2, u_2), ..., (s_n, u_n)$ and also a pair of arguments (a_1, a_2), where n represents the total number of utterances, and s_i denote the speaker name of utterance u_i. The task aims at employing an appropriate approach to identify/extract the relationship between a_1 and a_2 that appears in the dialogue (see more details of 'standard' experiment settings in (Yu et al., 2020)). Since such a task can be viewed as a simple multi-classification task, we employ the Macro-F1 score ($F1$), which is the harmonic mean of precision (P) and recall (R), for evaluation.

4.2. Model

This paper adopts a Bert-based extraction model proposed by Yu et al. (2020) as a baseline model. The baseline employs a pre-trained language model BERT. We deploy this model on our CRECIL corpus, and the DialogRE (Yu et al., 2020) in both English and Chinese versions. The model's input

[6]We compare both English (V2) and Chinese translated versions of DialogRE in this experiment

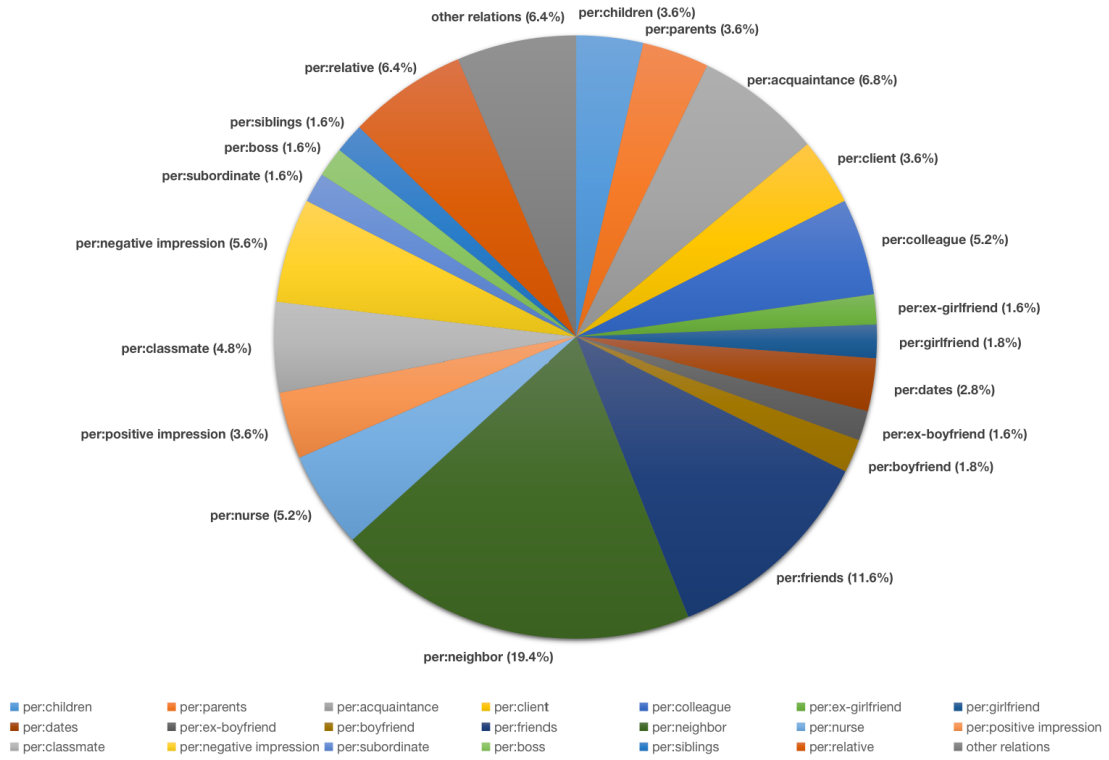


Figure 5: Global Character Relationships distribution in the CRECIL corpus

consists of a dialogue and a character entity pair to be recognized. The input to BERT has the format: $'[CLS]\ D\ [SEP]\ a_1\ [SEP]\ a_2\ [SEP]'$. The output of the BERT encoding layer contains $[CLS']$, which encodes the entire dialogue and the argument pair to be recognized. We then feed the embeddings of the $[CLS']$ into the LINEAR layer to predict the results of the relationship between a_1 and a_2.
We employed the chinese_wwm_ext[7] as the pre-trained BERT model, and the hidden layer dimension is 768. The maximum length of the dialogue is 512, and the number of gradient accumulation steps is 2. The batch size is 24, the optimizer is Adam, and the learning rate is 3e-5. The epoch is 20. The loss function is cross-entropy.

4.3. Results and Discussion

The F1 score for each relation type using the baseline BERT model on the newly created corpus is shown in Table 5. The result indicates the type of 'alternative name' (54.9% in the dev set, 55.7% in the test set) and 'neighbour' (64.2% in the dev set, 60.0% in the test set) have shown better performance than the other categories. This is because of imbalanced data across different relation types. Both 'alternative name' and 'neighbour' contain 6726 and 5464 examples respectively in CRECIL, which is almost as double as the others (see Figure 6)).

	Alternate Name	Neighbor	Children	Parents	Others
Dev	54.9	64.2	50.0	48.5	51.4
Test	55.7	60.0	47.1	48.6	46.2

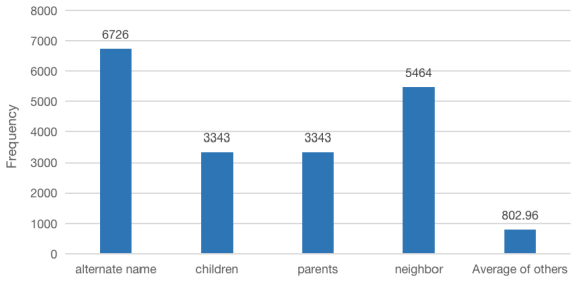
Table 5: Comparison between different categories (%) (excluding the 'unanswerable' type)

Table 6 shows the macro-F1 scores for each version of the DialogRE corpus (English vs Chinese) and our CRECIL corpus using the same baseline model. The model predicting argument relationship in DialogRE (F1-scores 59.4% in the development set and 57.9% in the test set) shows significantly better overall performance than the CRECIL data set (56.8% in the development set and 54.4% in the test set). This might be because compared with "Friends", 1) the dialogue in the Chinese episode "I Love My Family" is longer and more complex with multiple characters, and 2) there are more alternative referents in this Chinese script.

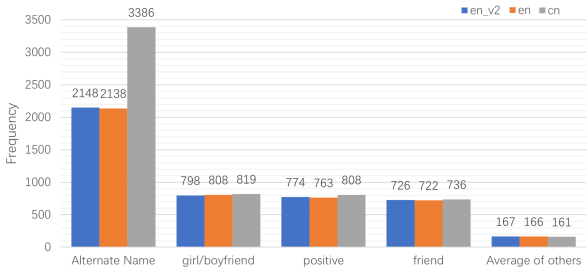
	EN-DialogRE	**CN-DialogRE**	**CRECIL**
Dev	59.4	63.7	56.8
Test	57.9	63.2	54.4

Table 6: Comparison between the CRECIL corpus and the DialogRE corpus (%)

[7]https://github.com/ymcui/Chinese-BERT-wwm



(a) CRECIL



(b) DialogRE

Figure 6: Relationship type distribution in CRECIL and DialogRE corpus (only present the top 4 relationship types per corpus, excluding the unanswerable type)

5. Conclusion

We presented a new dialogue-based relation extraction corpus (CRECIL) for multi-party conversations in Chinese. We introduced the Chinese-oriented character relationship categories and labelling rules for annotating the corpus. We also investigate the performance of a BERT-based extraction method on both CRECIL and another existing English TV-episode corpus (DialogRE). The results demonstrate that extracting character relationships is more challenging in CRECIL than in DialogRE.

Ongoing work further uses the data to explore the character relationship characteristics of Chinese multi-party dialogues and build a better-performance character relationship extraction model.

6. Acknowledgement

This work is supported by the National Natural Science Foundation of China (Grant No.61602044) and the funds for improving the quality of personnel training in 2021 of Beijing Information Science and Technology University (Grant No. 5102110805).

7. Bibliographical References

Chen, H., Hong, P., Han, W., Majumder, N., and Poria, S. (2020a). Dialogue relation extraction with document-level heterogeneous graph attention networks. *CoRR*, abs/2009.05092.

Chen, Y.-T., Huang, H.-H., and Chen, H.-H. (2020b). Mpdd: A multi-party dialogue dataset for analysis of emotions and interpersonal relationships. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 610–614.

Hendrickx, I., Kim, S. N., Kozareva, Z., Nakov, P., Ó Séaghdha, D., Padó, S., Pennacchiotti, M., Romano, L., and Szpakowicz, S. (2010). SemEval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 33–38, Uppsala, Sweden, July. Association for Computational Linguistics.

Li, J., Sun, Y., Johnson, R. J., Sciaky, D., Wei, C.-H., Leaman, R., Davis, A. P., Mattingly, C. J., Wiegers, T. C., and Lu, Z. (2016). Biocreative v cdr task corpus: a resource for chemical disease relation extraction. *Database*, 2016.

Yao, Y., Ye, D., Li, P., Han, X., Lin, Y., Liu, Z., Liu, Z., Huang, L., Zhou, J., and Sun, M. (2019). Docred: A large-scale document-level relation extraction dataset. *arXiv preprint arXiv:1906.06127*.

Yu, D., Sun, K., Cardie, C., and Yu, D. (2020). Dialogue-based relation extraction. *arXiv preprint arXiv:2004.08056*.

Zhang, Y., Zhong, V., Chen, D., Angeli, G., and Manning, C. D. (2017). Position-aware attention and supervised data improve slot filling. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*, pages 35–45.