

1 Appendix

1.1 Construction of Chinese Causal Event Graph

To collect the Chinese causal event pairs, we first obtain sentences which express causality between events. Despite the pervasiveness of causality in natural language, part of them are marked by explicit causal indicators and can offer causality between events with high precision. Therefore, we recognize these sentences from raw text corpus using causal patterns. Specifically, we employ the templates proposed by Xie and Mu (2019), which are composed by two kinds of indicators: 1. the causal verbs such as “dao zhi” (lead to), “zao cheng” (result in); 2. the causal connectives such as “suo yi” (so), “yin wei” (because). Based on these causal patterns, sentences which explicitly express event causality could be recognized from the crawled Chinese web news corpus.

Although the indicators could indicate the existence of causality, the exact span of the cause and the effect event within a sentence remains undetermined. To effectively extract the causal event pairs from those sentences, we formalize the causal event pair extraction process as a sequence labeling task, and train a BERT-CRF model to conduct this task. Specifically, we randomly sample 5000 sentences from the whole sentence corpus, and manually annotate the span of the cause event and the effect event under the BIO tagging schema. Then the annotated sentences are split into subsets sized up to 4,000, 500, 500 as training, validation, and test set, respectively. After the training process, the BERT-CRF model achieves a F1 value of 71.38 on the test set. Then the BERT-CRF model is employed to automatically extract causal event pairs from the Chinese web news corpus. After the above steps, totally 2 million Chinese causal event pairs are extracted.

References

Zhipeng Xie and Feiteng Mu. 2019. Distributed representation of words in cause and effect spaces. In *IJCAI*, volume 33, pages 7330–7337.