6V-02

Multilingual Complementation of Causality Property on Wikidata Based on ALBERT with Causality Tagging Scheme

JIN Yuxi† Shun SHIRAMATSU †

Graduate School of Engineering, Nagoya Institute of Technology†

1 Introduction

We aim to develop an agent for understanding distribution of public opinions and preferences. For this development, the agent needs to have causality knowledge. At the same time, when discussing social issues, Wikidata, as a knowledge base, can provide linked data and play an important role in analyzing the content of the discussion. However, there is a lack of causal content on Wikidata, and even some content has errors. Therefore, how to extract knowledge from news automatically and add it to Wikidata is a worthy of attention.

In this paper, we propose a method of automatically judging causality and directly extracting cause or effect from Wikipedia. We collected sentences in Wikipedia and used ALBERT to infer whether the article is causally related to the entity and further infer the cause or effect of the entity by Causality Tagging Scheme. Moreover, we are trying to increase the reliability of extracted causality knowledge by dealing with multilingual texts.

2 Experimental Evaluation

Extract Japanese articles $W = \{W_1, W_2, \dots, T_1, T_2, \dots, T_e, \dots, W_m\}$ containing the entity $T = \{T_1, T_2, \dots, T_n\}$ from Wikipedia, determine whether this article contains causality related to T, and further infer the cause of T or the effect of T. We use "BIO" to represent the position information of “begin”, “inside”, and “other” respectively, and use C and E to represent the semantic roles of “cause” and “effect” causal events.

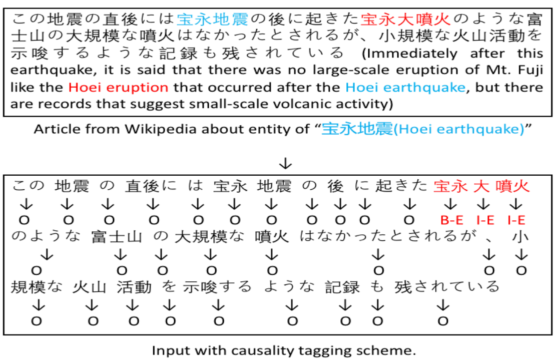

Figure 1: Input of Extraction Effect with *Causality Tagging Scheme*

Figures 1 show examples of such causal sequence labeling. In the extraction reason model, the label “B-C” stands for “beginning of cause”, the label “I-C” stands for “internal cause”, in the extraction effect model, the label “B-E’ stands for “start of effect”, and the label “I-E” stands for “internal effect”. If there is no causal relationship, all labels are represented by "O". Therefore, the total number of tags in each model is 3.

We used the published ALBERT Japanese pre-learning model in HuggingFace Transformers package. “Combination models” and “Separate model” two sets of experiments were conducted using the method of Causality Tagging Scheme.

2.1 Combination models. The first set of experiments is to use the two models of Binary classification and Sequence labeling to conduct experiments. The experimental steps are as follows: The first set of models is to use the AlbertForSequenceClassification model to classify the article to determine whether there is a causal relationship about entity in the article. If there is a causal relationship, the article label is "1", if it does not exist the article label is "0". Then add Causality Tagging to the article whose prediction label is “1” in first model. In the input part, we separate entity T from article W with [SEP], add [CLS] at the beginning of the sentence, and add [SEP] at the end of the sentence as shown in Figure2.

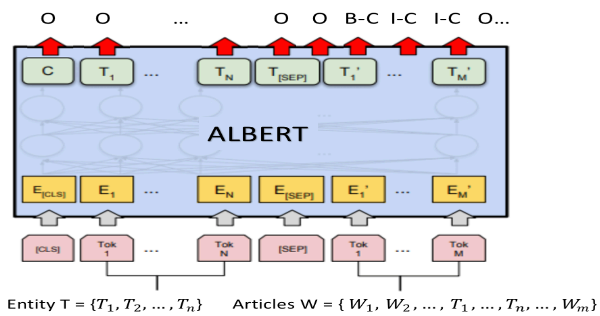

Figure 2: The framework of ALBERT

In the output part, each token will be classified in the form of label classification. If the cause for the entity is speculated, it will be classified as B-C or I-C, and if the effect of the speculation is caused by the entity, it will be classified as B-E or I-E.

2.2 Separate model. The single model only uses the Causality Tagging Scheme to simultaneously judge the causality and automatically extract the cause or effect. The input part is the same as the second part of 2.1. Enter the entity and the article together into the AlbertForTokenClassification model. In the output part, the tags B-C, I-C and O are utilized if the article has the cause of entity, the tags B-E, I-E and O are utilized if the article has the effect of entity, and all tokens are tagged as O tags if the article does not contain any causal relationship of entity.

3 Results and Analysis

A total of 974 data, that the data has a causal relationship has 611, there is no causal relationship data is 363. 90% of the data is used for training and 10% of the data is used for testing. We trained all models using the following hyperparameters: batch-size of 32, sequence length of 128, Adam with a learning rate of 2e-5 as optimization and epoch of 5. When predicting the cause or effect, it is too difficult to be completely consistent, so two methods are used to judge whether the inferred result is correct. (1) Due to the morphological characteristics of which the last token of the entity is the most important in Japanese, if the model can successfully infer the last token, the calculation is correct as "last". (2) If the overlap between the predicted token and the real token exceeds 50%, 70%, 100% it will be judged as correct as "overlap50", "overlap70", "overlap100".

The performance of combination models and separate model on the extraction of cause-effect relationships is shown in Table1.

model	Combination models				Separate model			
property	non-causal	causal			non-causal	causal		
			cause	effect			cause	effect
Precision	0.78	last	0.48	0.64	0.69	last	0.52	0.64
		overlap 50	0.62	0.69		overlap 50	0.58	0.66
		overlap 70	0.38	0.58		overlap 70	0.55	0.64
		overlap 100	0.33	0.49		overlap 100	0.49	0.57
Recall	0.68	last	0.35	0.39	0.73	last	0.51	0.59
		overlap 50	0.45	0.42		overlap 50	0.56	0.61
		overlap 70	0.28	0.35		overlap 70	0.53	0.59
		overlap 100	0.24	0.30		overlap 100	0.47	0.53
F1-Score	0.72	last	0.40	0.49	0.71	last	0.52	0.61
		overlap 50	0.52	0.52		overlap 50	0.57	0.63
		overlap 70	0.37	0.44		overlap 70	0.54	0.61
		overlap 100	0.32	0.37		overlap 100	0.48	0.55

Table 1: Evaluation results on test dataset

3.1 Analysis of No-causal: Compared with the Combination model, the Separate model uses Causality Tagging for prediction. It requires all tokens in the article to be correctly predicted to be calculated as correct, which should be more difficult. Therefore, we originally predicted that its performance would be far less than the Combination model. However, the results of the two models are very similar, which proves that when the Causality Tagging scheme model is used to predict causality, a single model already has good performance, and there is no need to superimpose multiple models.

3.2 Analysis of Causal: Combination models only perform Causality Tagging on data that has been judged to have a causal relationship, so the prediction result is the product of the two results of binary classification and Sequence labeling. Therefore, for data with causal relationships, the performance of the Separate model is much higher than that of the Combination model. Especially for the improvement of the accuracy requirements of the prediction results, the Separate model can still maintain good performance. This proves once again the simplicity and effectiveness of extracting the relationship using the Causality Tagging method.

4 Conclusion and Future Work

In this paper, we propose Causality Tagging Scheme to extract the causality of a specific entity in article. And through experiments, it is proved that Separate model can achieve good prediction results.

However, the causal relationship extracted in this experiment is dependent on the sentence itself, that is, the extracted result is a singular causal relation, and the result of this entity linking cannot be added to Wikidata. So, in order to add it to Wikidata, we should automatically extract the general causal relation. In the future work, we will collect social news items from English, Japanese and Chinese News and train them separately to calculate the precision of each language model to infer causality. After that, in order to extract the general causality of a Wikidata entity, we will collect news data of the three languages containing this Wikidata entity and predict causality of this Wikidata entity. The predicted results of this Wikidata entity were also grouped using similarity scores and the weight of each group of results in the overall results was calculated. After that, the weight of the same results and precision of model in the three languages is superimposed to calculate the reliability scores. Finally, the results with high reliability scores are subjected to entity linking and added to Wikidata.

ACKNOWLEDGMENTS

This study was partially supported by JST CREST (JPMJCR15E1) and NEDO (JPNP20006).

REFERENCES

[1] Khoo, C. S., Kornfilt, J., Oddy, R. N., & Myaeng, S. H. (1998). Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. Literary and Linguistic Computing, 13(4), 177-186.
[2] Khoo, C. S., Chan, S., & Niu, Y. (2000, October). Extracting causal knowledge from a medical database using graphical patterns. In Proceedings of the 38th annual meeting of the association for computational linguistics (pp. 336-343).
[3] Girju, R., & Moldovan, D. I. (2002, May). Text mining for causal relations. In FLAIRS conference (pp. 360-364).