

Causal relationship recognition algorithm based on external semantic and contextual structural features

Jiaxin Li ^{1, a}, Kai Shuang ^{1, b}

¹School of Computing, State Key Laboratory of Networking & Switching Technology Beijing
University of Posts & Telecommunications (BUPT), Beijing, 100000, China;

^a jiaxinli@bupt.edu.cn, ^b shuangk@bupt.edu.cn

Abstract. Identifying the causal relationship of events plays an important role in determining the development of known events and evaluating the possible outcomes of different decisions. At present, neural network models are widely used to identify the relationships between events. Based on obtaining events, researchers distinguish the relationships between events by mining their semantics. However, due to the complexity of events and the dynamic changes in relationships between events, models often cannot fully meet the needs of accurately identifying causal relationships by only learning simple event descriptions in sentences; Moreover, focusing too much on the events themselves often leads to neglecting the structural features of statements and neglecting the impact of specific structural patterns on the relationships between events. In this article, we propose an event masking algorithm that combines external semantics to address the aforementioned issues. In this algorithm, external semantics are first introduced into the statement to enrich the information behind the event, allowing the model to mine the deep connections between events through a wider range of background knowledge; Then, the event masking module is used to enhance the model's extraction of sentence structured features, mining specific contextual representations that are unrelated to the event. The results show that compared to existing neural network algorithms, the algorithm proposed in this paper improves the F1 value of predictions on publicly available datasets by more than 4%.

Keywords: Natural language understanding; external knowledge Introduction; Introduction of external knowledge; causal relationships.

1. Introduction

Event causal relationship extraction refers to the use of natural language processing techniques and extraction methods to identify causal relationships between events from specified texts. This process is beneficial for us to deepen our understanding of the content of the text and analyze its inherent logical connections. By extracting causal relationships from events, we can enhance our ability to interpret and analyze the meaning of the text, assisting us in discovering valuable information from a large amount of text.

The causal relationship of an event is mainly constructed by two components: the triggering event and the resulting event. The task of extracting causal relationships from events is to extract the triggering and outcome events from the text that describes the event information, and present them in an organized manner such as "cause->result". The current research on causal relationships usually extracts causal relationships of events in a pipeline manner: based on obtaining event pairs, semantic features of events are extracted through deep learning to identify relationships between events. This method requires the model to be able to fully recognize and extract the semantic features of events and the relationship features between events. However, the same event may have multiple different interpretations, and the relationship between the same two events may be different in different contextual environments. The description of an event presented in a statement often only describes the appearance of an event in a certain time series, without delving into the background of the event and its related case relationships. This makes the model only able to learn surface features that exist in the statement when learning events, and unable to mine deep connections. This greatly affects the model's judgment of deep relationships between different events, thereby reducing the accuracy of identifying causal relationships between events. In addition,

some sentences have special sentence structures that can assist the model in judging the relationships between events, such as "because, so", "so", "for", etc. These special sentence structures not only have a positive impact on the judgment of relationships between known events, but even when we encounter events that we have never encountered before, we can identify causal relationships between events in the sentence based on these sentence structures. However, existing research methods mainly focus on extracting semantic features of events, ignoring the importance of special sentence structures. This situation reduces our ability to identify causal relationships between events and further reduces our ability to discover causal relationships between new events to a certain extent. Due to these issues, existing causal relationship extraction models have insufficient research on the causal relationships between events, which requires us to further improve our understanding of event causal relationships.

In order to solve the above problems and consider the valuable clues of causal relationships between events contained in the event structure, this article also uses a pipeline approach. Based on the obtained event pairs, we design and implement an event masking recognition model that combines the external semantics of event associations and sentence structures. This model retrieves the background related to events in the corpus and supplements them in the sentences, Enriches the semantic representation of events, allowing the model to learn the deep connections behind the events. At the same time, algorithms are used to mask events in sentences through special identifier words, allowing the model to learn sentence structures, mine special contextual structures unrelated to events, and improve the model's ability to recognize causal relationships. This article demonstrates the effectiveness of the model through experiments. Compared with existing neural network models, the model predicts an F1 value increase of more than 4% on the dataset.

2. Related work

At present, there are mainly the following methods to identify the relationship between events.

The pattern based method mainly relies on manual writing rules such as lexical, syntactic and semantic analysis, and realizes the identification of causal relationships through pattern matching. Khoo et al. ^[1] extracted causal knowledge from the Wall Street Journal by analyzing the semantic characteristics of news clues and established language patterns; Khoo et al. ^[2] summarized the graphical features in the medical field and extracted the causal relationship of medical events from the medical database using the general graphical pattern; Girju et al. ^[3] summarized the syntactic patterns of causal sentences and used semantic constraints to extract candidate event pairs and identify causal relationships; Ittoo et al. ^[4] proposed a causal pair extraction method that takes both part of speech and syntactic analysis into account and integrates them into a causal template. The causal relationship recognition method that completely depends on pattern matching requires a lot of manual work and expert domain knowledge. Although it does not need training and a lot of data, it requires rich pre experience and has poor cross domain generalization ability.

With the development of technology and computing resources, event causality identification has gradually developed to classify causality into two tasks through pipeline, namely, the extraction of candidate event pairs and the classification of event pair relationships.

The pipeline method can be realized by combining pattern based and machine learning technology. This method first extracts event pairs that may have causal relationship according to pattern matching or keyword matching, and then classifies candidate event pairs according to characteristics. Girju et al. ^[5] first constrained English events by constraint words, and then identified them by grammatical features; Luo et al. ^[6] summarized the causal terms based on the analysis of a large number of text corpora, and then measured the causal strength between network text corpora based on the statistical measurement of pointwise mutual information through machine learning, so as to determine whether there is a causal relationship between event pairs.

Based on the strong feature learning ability of deep neural network, it can effectively reveal the hidden and fuzzy causal relationship. Therefore, using deep learning technology to extract causality

has become more and more popular in recent years. De Silva et al. [7] used the convolution neural network (CNN) to identify the events in the text, and used CNN to further classify the causal relationship of the identified events; Kruengkrai et al. [8] used CNN to not only learn the event itself, but also learn the context information of the event as background knowledge to classify the causality between events in the common sense causality; Li et al. [9] proposed knowledge oriented CNN, which classifies the relationship between events by adding prior knowledge in the knowledge base in the network; Li et al. [10] proposed a causal relationship extractor based on the bidirectional long short term memory (LSTM) network and the conditional random field model, which extracts the features of events and contexts between texts through the bidirectional LSTM, and uses the conditional random field to identify the causal relationship between events; Zeng et al. [11] proposed an end-to-end joint learning model using a two-way LSTM structure, which uses two strategies to encode and decode the corpus respectively to extract triplet events and the causal relationship between events.

3. Event masking algorithm based on joint external semantics

3.1 Section Headingsdescription of relevant basic concepts

3.1.1 Event

An event is a specific behavior or situation that occurs at a specific time and place, usually related to a subject (individual, group, organization or object). The event mentioned in this paper is a continuous description consisting of several or all of the subject, predicate, object, attribute and adverbial. For example, "a strong earthquake occurred in northern Japan and led to a tsunami on the east coast of South Korea." in the sentence, "a strong earthquake occurred in northern Japan" and "a tsunami occurred on the east coast of South Korea" are two independent events respectively. Since the focus of this paper is not event extraction, the event extraction method mentioned in this paper is completed through the event extraction framework [17] based on query and extraction paradigm.

3.1.2 External semantic knowledge base

The external semantic knowledge base mentioned in this paper is conceptnet. The knowledge base is a large corpus that classifies the knowledge structure into a graph. Each node corresponds to a common sense knowledge (entity or event), and each edge corresponds to a semantic relationship.

The conceptnet library originated from MIT Laboratory and was initiated by the crowdsourcing project open mind common sense. During the duration of the project, it continuously expanded other crowdsourcing projects and resources in the field of experts. So far, the library has been updated to version 5.7, and the semantic relationship has been expanded to more than 28million, and it is still being updated. The conceptnet library is completely free and open source, with multiple language versions such as Chinese and more than 50 relationship categories. The initiator of the project provides users with three different ways of using: Online API interface, offline application of library version and offline data download.

3.2 Algorithm framework

The algorithm mainly has three modules: knowledge integration module, event shielding reasoning module and discrimination module. The overall algorithm framework is shown in Figure 1.

(1) Knowledge integration module: first, get the semantic relationship of events from concept net, enrich the event representation, and then use Bert to code to get the knowledge representation of events;

Some deep learning research methods rely on the model (neural network) to learn the semantic features of events when identifying the relationship between existing event pairs, and identify the relationship between events according to the semantic features mined, which leads to the model greatly relying on the semantic features contained in events. The current event description sentences

mostly strive to describe the real scene with simple and clear words and sentences, which leads to the fact that the event semantics contained in the sentences are too few, and only include the characteristics of the event, not including the background and deep connection of the event, which often fails to meet the requirements of the model. This requires us to adopt a reasonable way to expand the semantic features of events, so that the model can learn more semantic features to the maximum extent in the process of sentence coding.

In the introduction of chapter 2.1.2, we can learn that the conceptnet corpus is a constantly updated corpus of common sense knowledge from multiple sources, and the examples involved and the relationships between them cover most industries.

In this module, we select relevant examples in the corpus that are closely related to the event part in the sentence to supplement the event relationship, expand the semantic characteristics of the event in the sentence, so that the event description contains more background and development, and then input the expanded sufficient semantic and context information into the model, so that the output of the model contains more knowledge related to the event.

(2) Event masking processing module: replace the event pair with [MASK], and then encode it with Bert to get the context representation of the event pair;

In the existing research work, the judgment of the relationship between events is more focused on the semantic characteristics of events; There have also been research methods that put the whole sentence with context information into model training, but because some specific context patterns (so, for, because, etc.) that can show the causal relationship between events account for a relatively small proportion in the sentence, the model cannot give sufficient proportion to its impact.

In this module, in order to enable the model to give sufficient attention to the context information and to mine the specific context pattern that is separated from the event, the masking algorithm is used to replace the event with the special identifier [MASK], and the replaced text is sent to the model training as input.

(3) Discrimination module: the attention mechanism is used to splice the knowledge representation and context representation of events together for prediction.

In the first two modules, we have expanded the event semantics of the original text and covered the event. In the actual training and prediction process, the model needs to consider the impact score given by the above algorithm. In this module, the attention mechanism is used to integrate the two scores.

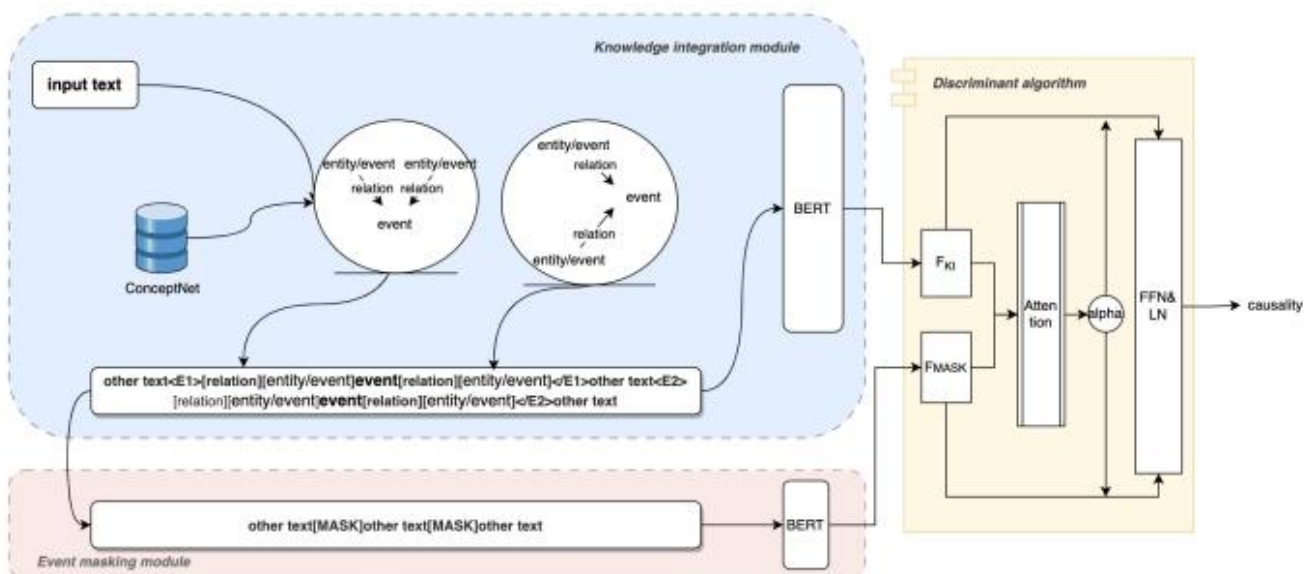


Fig. 1 knowledge semantics

3.2.1 module algorithm Introduction

(1) knowledge integration module

The event pairs (expressed as E1 and E2) are fused with external semantics to form new context information. The events mentioned in this module are those defined in Chapter events. Retrieve the relevant knowledge in the conceptnet corpus for the instances in E1 and E2 events, then connect the retrieved knowledge with the original event, and put it into the Bert pre training model for coding.

For the input text, first locate the event location in the text that needs relationship recognition, and segment it with `<en>`, `</en>`. The event in this section is a continuous description obtained by the way mentioned in Chapter 3.1.1.

In the actual process of knowledge fusion, it is necessary to analyze the semantic roles of the two events to form the form of "predicate (action agent, action patient)", then abstract the predicate, action agent and action patient into nodes, and then retrieve the edges (semantic relations) that may affect the causal relationship in the conceptnet corpus by string matching: capableof, ISA, hasproperty, causes, manner of Causesdesire, usedfor, hassubevent, hasprerequisite, notdesires, partof, hasa, entitles, receivesaction, usedfor, createdby, madeof and desires, and then extract the corresponding instances to supplement the instances and semantic relationships in the statements.

Take "a strong earthquake occurred in northern Japan and led to a tsunami on the east coast of South Korea" as an example. According to the framework mentioned in Chapter 3.1.1, two consecutive events "a strong earthquake occurred in northern Japan" and "a tsunami occurred on the east coast of South Korea" in the sentence were identified, and then semantic role analysis was carried out for the two events respectively. Two groups of different semantic role tuples are obtained: occurrence (northern Japan, earthquake) and occurrence (east coast of Korea, tsunami). Then the results of semantic role analysis are retrieved in the corpus, and the retrieved results are incorporated into the original sentence. Figure 2 shows the whole process of inputting statements to obtain background knowledge, and sending them into coding to get the final result of the module.

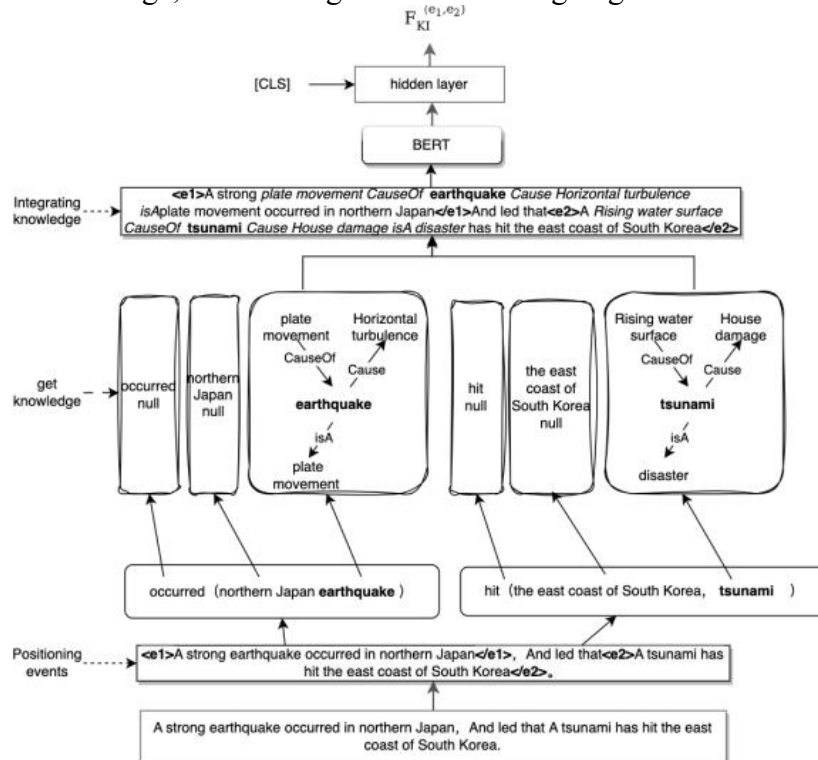


Fig. 2 Knowledge Integration Structure Diagram

Due to the combination of a large amount of corpus in the pre training stage, BERT can more comprehensively obtain the relevant features of the input text. This algorithm uses the BERT model to encode events and their contextual information. After obtaining background knowledge and structuring it into new sentences, the entire sentence is input into the pre trained model BERT to obtain the output encoding of the hidden layer, and then special encoding [CLS] is added to obtain the context information encoding representation F_{KI} with knowledge perception.

$$F_{KI}^{(e_1, e_2)} = h_{[CLS]} \oplus h_{<E1>} \oplus h_{<E2>}. \quad (1)$$

(2) Event Masking Processor

The event masking processor mechanism aims to separate from events and mine context specific causal sentence expression patterns based on sentence structure itself. The overall algorithm structure is shown in Figure 3.

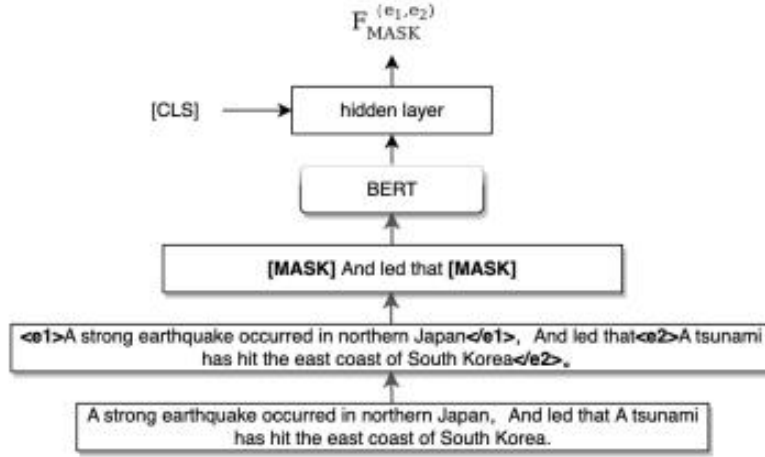


Fig. 3 Event Masking Processor Architecture Diagram

Firstly, events e_1 and e_2 are replaced by a special flag sequence "[MASK]", which can mask the specific information of the event. Then, the pre trained model BERT is used to encode the masked sentences, while adding a [CLS] flag sequence. The following characterization was obtained:

$$F_{MASK}^{(e_1, e_2)} = h_{[CLS]} \oplus h_{[MASK]}^{e_1} \oplus h_{[MASK]}^{e_2}. \quad (2)$$

When there is a causal relationship between two events in a sentence, it indicates that the themes expressed by these two events in the context structure of the sentence are similar. Therefore, when we respectively mask the two events, the expression of the other event in that context should have thematic similarity. Based on the above analysis, we can convert the probability of whether the masked event has a causal relationship to the probability that the masked events e_1 and e_2 are similar, which can be expressed as:

$$p(l = 1|e_1, e_2) = \frac{1}{1 + \exp(F_{MASK}^{e_1} - F_{MASK}^{e_2})}. \quad (3)$$

The BERT model used in this module needs to be fine tuned to obtain the fine tuned hidden layer output encoding as context encoding. During the training process, the loss function of this module is calculated as follows:

$$L = -\delta_{e_1, e_2} * \log(p(l = 1|e_1, e_2)) + (1 - \delta_{e_1, e_2}) * \log(1 - p(l = 1|e_1, e_2)). \quad (4)$$

(3) Discriminant algorithm

In order to learn the weights between FKI and FMASK in the knowledge integration module and event masking processing module of the two components mentioned above, attention mechanism is used to balance the two components for the final prediction. The overall algorithm framework is shown in Figure 4.

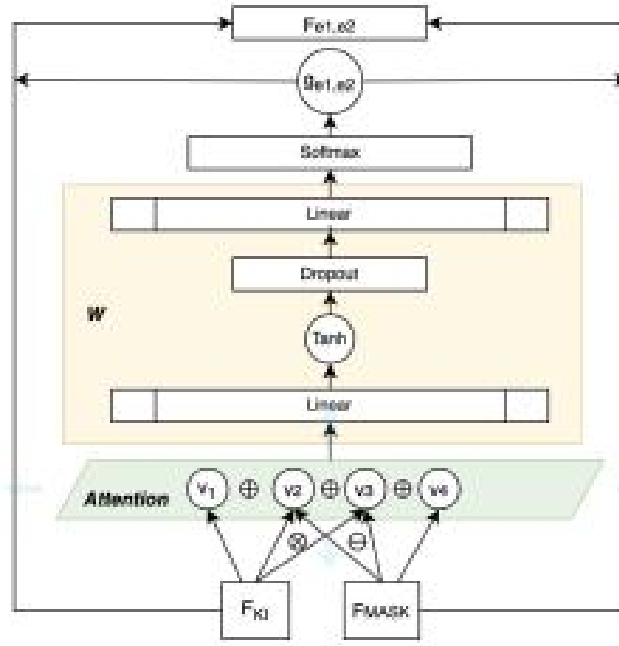


Fig. 4 Discriminant Algorithm

In order to compare and integrate the semantic structural features of events with the contextual structural features after masking events, an attention mechanism is adopted in this module. The two features obtained from the above modules are used for difference and vector product calculation to capture the differences and similar features between the two features. Then, the obtained differentiated and similar features are fused with the original features, making the model learn more rich and complex feature representations. The signal of reinforcement model learning allows the model to adjust the learned feature weights based on different data characteristics. Then, non-linear relationships are captured through fully connected layers and activation functions, and finally, Softmax is used to convert them into weights $g_{e1,e2}$:

$$g_{e1,e2} = \sigma(W(F_{KI}^{(e1,e2)} \oplus F_{MASK}^{(e1,e2)} \oplus (F_{KI}^{(e1,e2)} * F_{MASK}^{(e1,e2)}) \oplus (F_{KI}^{(e1,e2)} - F_{MASK}^{(e1,e2)}))). \quad (4)$$

Among them, W represents the two fully connected layers and the activation function, and σ represents the Softmax function.

After fusing the previously obtained features, a representation that integrates all features and focuses on key parts is obtained $g_{e1,e2}$. Then, the key parts are represented by fusing the two features separately to obtain new features that focus on the key parts. The final representation between events $e1$ and $e2$ is as follows:

$$F_{e1,e2} = g_{e1,e2} * F_{KI}^{(e1,e2)} + g_{e1,e2} * F_{MASK}^{(e1,e2)}. \quad (5)$$

3.2.2 Training and prediction

In the actual training process, the task of extracting causal relationships is transformed into a binary classification task to determine whether events $e1$ and $e2$ have causal relationships

$$o_{e1,e2} = \sigma(W_o F_{e1,e2} + b_o). \quad (6)$$

This module uses cross entropy as the loss function:

$$J(\theta) = - \sum_s \sum_{\substack{e_i, e_j \in E_s \\ e_i \neq e_j}} y_{e_i, e_j} \log(o_{e_i, e_j}) + (1 - y_{e_i, e_j}) \log(1 - o_{e_i, e_j}). \quad (7)$$

4. experiments

4.1 Evaluation indicators

The algorithm proposed in this article regards event causal relationship recognition as a classification task. Adopt accuracy (P), recall (R), and F1 score (F1) evaluation metrics in this task. Among them, accuracy refers to the proportion of correctly extracted causal relationship event pairs to the total number of event causal relationship pairs extracted by the model; Recall rate refers to the proportion of correctly extracted causal event pairs to the total number of truly causal event pairs. The F1 score is a weighted value of accuracy and recall.

4.2 Datasets

This algorithm has been validated on the following datasets: Chinese medical causal relationship extraction dataset - MedCasual.

This dataset consists of a large number of diagnostic results based on clinical cases or diseases as the core, including laboratory test results of various medical device examinations. There are three types of event relationships in the dataset: causal relationships, conditional relationships, and hierarchical relationships. The difference between conditional relationship and causal relationship is that it weakens the degree to which conditions lead to results, and is not distinguished in this task.

4.3 Experimental Results

4.3.1 Baseline Model Comparison Experiment

To verify the effectiveness of the algorithm experiment, publicly available models with good performance in recent years and the algorithm proposed in this chapter were selected for evaluation. Experiments were conducted on both MedCasual and FinCausal datasets, and the main experimental results are shown in Table 1 and Table 2. The BERT model used in each module of the algorithm in this experiment is the Bert base model, which consists of 12 individual Transformer blocks with 768 hidden layer units and 12 attention heads. All parameters are 110M in total.

Table 1. Experimental results on the MedCasual

| | P (%) | R (%) | F (%) |
|-----------------------------------|---------|---------|---------|
| BERT-softmax,2018 ^[12] | 75.66 | 75.11 | 75.38 |
| GNN,2020 ^[13] | 51.90 | 72.54 | 60.51 |
| BiLSTM-CRF,2021 ^[14] | 68.78 | 63.23 | 65.89 |
| GMTL,2022 ^[15] | 82.79 | 79.57 | 81.15 |
| CEPN,2022 ^[16] | 83.72 | 81.43 | 82.56 |
| Our model | 88.98 | 83.26 | 86.03 |

Table 2. Experimental results on the FinCausal

| | P (%) | R (%) | F (%) |
|-----------------------------------|---------|---------|---------|
| BERT-softmax,2018 ^[12] | 79.63 | 75.31 | 77.41 |
| GNN,2020 ^[13] | 56.36 | 66.89 | 61.18 |
| BiLSTM-CRF,2021 ^[14] | 71.20 | 68.34 | 69.74 |
| GMTL,2022 ^[15] | 80.49 | 79.59 | 80.04 |
| CEPN,2022 ^[16] | 89.01 | 68.91 | 77.67 |
| Our model | 89.02 | 82.34 | 85.55 |

When conducting model evaluation, all models are trained using the same training set, with model training iterations set to 20 epochs and batch sizes set to 32. It can be seen that in the comparison process with different baseline models, the algorithm proposed in this chapter performs well compared to the previous model on both datasets.

4.3.2 Ablation comparative experiment

In order to verify the effectiveness of each module in this algorithm, ablation tests were conducted on the following variants on two datasets, MedCasual and FinCausal, targeting the key modules of the algorithm:

(1) Remove the external semantic embedding module of events: do not fuse the external semantics of events, and directly identify causal relationships between events.

(2) Remove [MASK] operation: Remove the event masking processor module and do not mask events in sentences

(3) Remove discrimination module: Remove the dynamic parameter learning of the Attention mechanism in the discrimination module, and directly adjust the parameters to 0.5.

Table 3. Results of ablation comparison experiments on the MedCasual

| | <i>P (%)</i> | <i>R (%)</i> | <i>F (%)</i> |
|-------------------------------------------------------------|--------------|--------------|--------------|
| Our Model | 88.98 | 83.26 | 86.03 |
| Remove the external semantic embedding module of events (1) | 80.18 | 77.37 | 78.75 |
| Remove [MASK] operation (2) | 88.03 | 81.17 | 84.46 |
| Remove discrimination module (3) | 86.99 | 81.46 | 84.13 |

Table 4. Results of ablation comparison experiments on the FinCausal

| | <i>P (%)</i> | <i>R (%)</i> | <i>F (%)</i> |
|-------------------------------------------------------------|--------------|--------------|--------------|
| Our Model | 89.02 | 82.34 | 85.55 |
| Remove the external semantic embedding module of events (1) | 83.97 | 77.06 | 80.37 |
| Remove [MASK] operation (2) | 87.10 | 81.23 | 84.06 |
| Remove discrimination module (3) | 87.57 | 80.99 | 84.15 |

From the testing of the above two datasets, it can be seen that after removing the external semantic embedding module, the accuracy, recall, and F-value of both datasets have significantly decreased, indicating that the innovation and effectiveness of the joint external semantic module in this algorithm have been confirmed. In addition, after removing the masking mechanism and the attention mechanism of the discrimination module, the performance of the two datasets also showed varying degrees of decline. The above results validate the effectiveness of the event masking processor algorithm proposed in this chapter, which combines external knowledge semantics.

4.3.3 Experimental analysis

Through baseline model comparison experiments and ablation comparison experiments, it can be seen that the proposed fusion external semantic module in this algorithm greatly enriches the semantic structure of events while integrating external semantics into sentences, and to some extent alleviates the drawbacks caused by dynamic changes in relationships between events; The introduction of event masking mechanism can fully consider the importance of event structure in sentences, enabling the model to learn context specific information and internal correlations between causal events, greatly enhancing the model's ability to discover new events. The discrimination module based on Attention mechanism introduced in the final model can balance the impact of causal trigger words and external semantics on the final result, increasing the robustness of the model.

5. Conclusion

This article mainly studies the issue of causal relationships between events and proposes an event masking algorithm that combines external semantics. This algorithm considers the semantic features of events, the relationship features between events, and the structural features of contextual

information, and infers the causal relationship of event pairs in sentences based on the above information. After the implementation of the overall algorithm, baseline comparison model experiments and ablation comparison experiments were carried out in the MedCasual dataset, and the experimental results proved the effectiveness and progressiveness of the algorithm. Possible improvement directions in the future: (1) Replace the BERT model with a pre trained model with more parameters and expected pre training, such as ELECTRA. (2) Overlay a layer of network, such as CNN, on the external semantics of event pairs for interactive information processing.

Acknowledgment

This work is supported by Beijing Natural Science Foundation (Grant No. 4222032).

References

- [1] Khoo C S G, Kornfilt J, Oddy R N, et al. Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing [J]. *Literary and Linguistic Computing*, 1998, 13 (4) : 177-186.
- [2] Khoo C S G, Chan S, Niu Y. Extracting causal knowledge from a medical database using graphical patterns [C] // *Proceedings of 38th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics, 2000: 336-343.
- [3] Girju R, Moldovan D. Text mining for causal relations [C] // *Proceedings of the 15th International Florida Artificial Intelligence Research Society Conference*. Palo Alto, CA: AAAI Press, 2002: 360-364.
- [4] Ittoo A, Bouma G. Extracting explicit and implicit causal relations from sparse, domain-specific texts [C] // *Proceedings of 16th International Conference on Applications of Natural Language to Information Systems*, LNCS 6716. Berlin: Springer, 2011: 52-63.
- [5] Girju R. Automatic detection of causal relations for question answering [C] // *Proceedings of the ACL 2003 Workshop on Multilingual Summarization and Question Answering*. Stroudsburg, PA: Association for Computational Linguistics, 2003: 76-83.
- [6] Luo Z, Sha Y, Zhu K Q, et al. Commonsense causal reasoning between short texts [C] // *Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning*. Palo Alto, CA: AAAI Press, 2016: 421-430.
- [7] De Silva T N, Xiao Z, Zhao R, et al. Causal relation identification using convolutional neural networks and knowledge based features [J]. *World Academy of Science, Engineering and Technology: International Journal of Mechanical and Mechatronics Engineering*, 2017, 11 (6) : 703-708.
- [8] Kruengkrai C, Torisawa K, Hashimoto C, et al. Improving event causality recognition with multiple background knowledge sources using multi-column convolutional neural networks [C] // *Proceedings of the 31st AAAI Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press, 2017: 3466-3473.
- [9] Li P, Mao K. Knowledge-oriented convolutional neural network for causal relation extraction from natural language texts [J]. *Expert Systems with Applications*, 2019, 115: 512-523.
- [10] Dai D, Xiao X, Lyu Y, et al. Joint extraction of entities and overlapping relations using position-attentive sequence labeling [C] // *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press, 2019: 6300-6308.
- [11] Zhu H, Lin Y, Liu Z, et al. Graph neural networks with generated parameters for relation extraction [C] // *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics, 2019: 1331-1339.

- [12] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [13] Chen Y, Wang K, Yang W, et al. A multi-channel deep neural network for relation extraction [J] . IEEE Access, 2020, 8: 13195-13203.
- [14] Zhang S , Zheng D , Hu X , et al. Bidirectional long short-term memory networks for relation classification [C/OL] // Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation. [2021-08-05] . <https://www.aclweb.org/anthology/Y15-1009.pdf>.
- [15] Makino, T., Geras, K., & Cho, K. (2022). Generative multitask learning mitigates target-causing confounding. *Advances in Neural Information Processing Systems*, 35, 36546-36558.
- [16] Nayak, T. , Sharma, S. , Butala, Y. , Dasgupta, K. , Goyal, P. , & Ganguly, N. . (2022). A generative approach for financial causality extraction.
- [17] Wang, S., Yu, M., Chang, S., Sun, L., & Huang, L. (2021). Query and extract: Refining event extraction as type-oriented binary decoding. arXiv preprint arXiv:2110.07476.