Causality Extraction Model Based on Two-stage GCN

Guangli Zhu（✉ glzhu@aust.edu.cn ）

Anhui University of Science and Technology   https://orcid.org/0000-0002-0540-7593

Zhengyan Sun

Anhui University of Science and Technology

Shunxiang Zhang

Anhui University of Science and Technology

Subo Wei

Anhui University of Science and Technology

Kuan Ching Li

Providence University

Causality Extraction Model Based on Two-stage GCN

Guangli Zhu [1,*], Zhengyan Sun [1], Shunxiang Zhang [1], Subo Wei [1], KuanChing Li [2,*]

[1] School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, People's Republic of China

[2] Department of Computer Science and Information Engineering (CSIE), Providence University, Taizhong, 43301, Taiwan

*Corresponding author

Guangli Zhu (e-mail: glzhu@aust.edu.cn)，KuanChing Li (e-mail: kuancli@pu.edu.tw).

Abstract: In the traditional methods, the low identification accuracy of cascade implicit causalities is caused by the lack of causal inference. To solve this problem, we propose a causality extraction model based on GCN to infer the causality of the text. It can analyze the cause-effect existing in the text and realize the deep extraction under semantic enhancement. First, the data are preprocessed, and BERT is used for pre-training. In the pre-training, the candidate entities are selected by entity links. The text encoding is used context semantics and embedding location coding. Then, the semantic dependency graph is used to obtain the relationship between entities. In addition, the nodes and edges obtained by the previous step are input into the first-stage GCN to extract the whole causality. Finally, the entity relation graph obtained by the first-stage GCN is introduced into the second-stage GCN for cascade inference. The causality of the cascade is inferred and extracted by a two-stage GCN. Experiments show that the model can find implicit causality more accurately generate new implicit causal entities on the original causality.

Keywords: two-stage GCN, BERT, GCN, causality extraction.

1 Introduction

Relation extraction gives a given text, extracts entities/objects, and infer their relations in the text, forming a triple (S, R, O). S represents the subject entity, R represents the existing relationship, and O represents the object entity. At present, entity relation extraction technology impacts all aspects of natural language processing tasks and has been widely used. Causality is one of the most famous relationships. Traditional methods can quickly identify explicit causal entities, but the identification accuracy is not high because of the need for inference for implicit causal entities. At the same time, because of the uniqueness and diversity of text, the complexity of semantic structure [35], the variety of expression and other factors, it is inevitable to increase the difficulty of inference, which leads to many potential implicit causal entities are not identified.

Given the above problems in the research on causality extraction, the following points are mainly considered. (1) Entity extraction. Previous work has proposed a key sentences extraction algorithm for Chinese microblog comments [1], which considers multiple factors attributes to identify the key sentences of comments. So this algorithm can provide technical assistance for entity extraction. (2) Feature extraction and relational inference. In the microblog hot topic word extraction model [2], a feature cooccurrence method is proposed, providing a reference for causal relationship feature extraction and relational inference. (3) Causality identification, especially can effectively identify the implicit causality. Previous work has used ALN (Association link Network) [3] to achieve the hierarchical

division of associated semantics. Learning the semantic information of the text provides technical support for causality identification.

Based on the above considerations, we combine BERT technology and GCN graph neural network to propose a new model called the causality extraction model based on a two-stage GCN. Different from the traditional method, the proposed method identifies a potential causality between cascading causal entities by two-stage inference of GCN and extracts it. The motivation is that the network can learn more complex entity structures, capture more abundant local and non-local features and realize multi-hop inference of entities. The model framework is shown in Figure 1, and the main contributions of this paper can be summarized as the following two points.

- **Construction of causal entity candidate library.** We use a combination of causal prior knowledge and semantic data to extract causal entities from the corpus. Cluster analysis of the extracted entities, and the same cluster of entities randomly selected as adjacent nodes into GCN. Firstly, the review text is preprocessed, and the sentences without causality or the sentences with default components are deleted. Then, the "NLTK" word segmentation is used to segment the words in the text. Secondly, using emotional intensity to establish causality seed lexicon. Finally, the K-means clustering algorithm is used for clustering calculation to select the final causal entity library.

- **Extraction of causality tuples.** Semantic encoding by BERT combined with context. The reference nodes, entity nodes and meta-dependent path (MDP) nodes are constructed. So candidate entities are obtained. Its also are represented as node inputs, and the syntactic relations obtained from the syntactic dependency graph are introduced into the two-stage GCN network as edges. Use the full connection to capture more structural information and train deeper models.

Figure 1. the framework of the two-stage GCN model
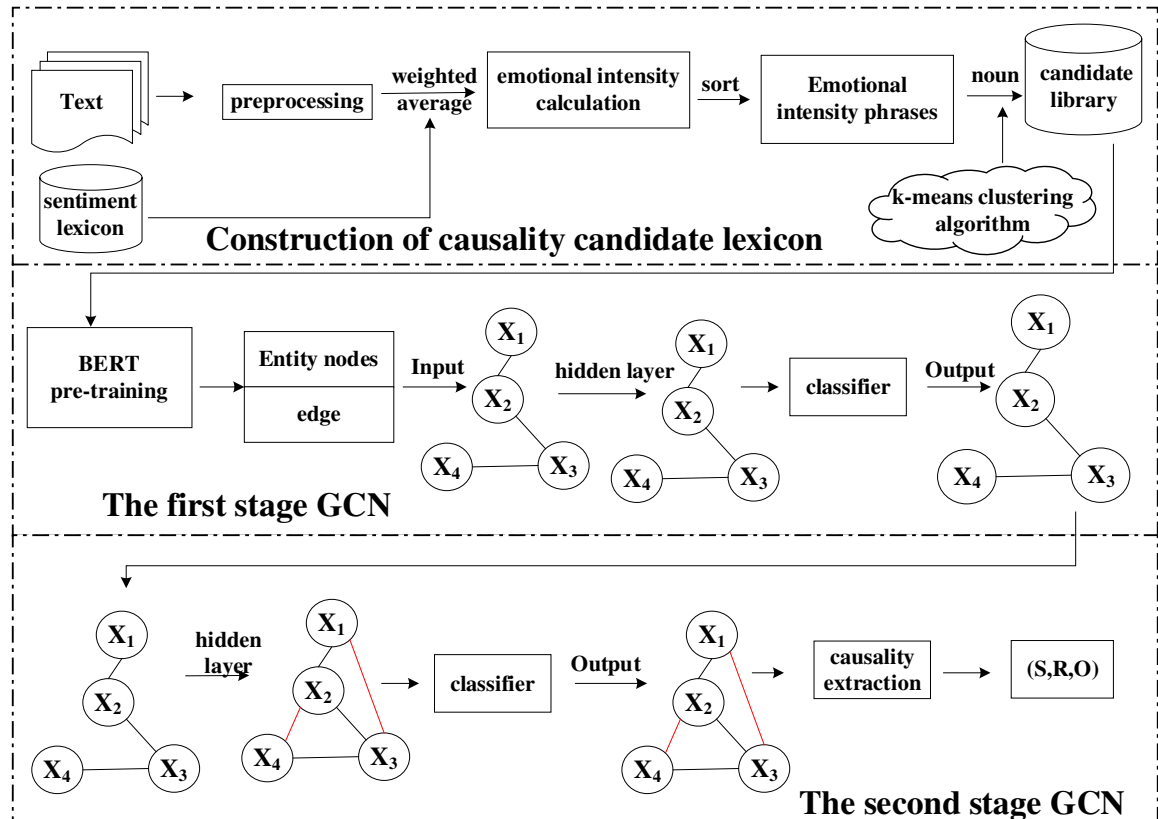
The advantage is that the two-stage GCN method can achieve better multi-hop relational inference and identify more implicit cascade causality. The main contributions of this paper can be summarized as the following two points :

- **Effective construction of causality candidate entity library.** we take into account that causal entities often have phrases with emotional polarities. Therefore, the candidate seed bank of causal entity is constructed by emotional intensity, which is convenient for constructing entity knowledge base through entity learning.
- **Strong reasoning of two-stage GCN.** When the proposed method is used to construct the causal relationship extraction model, to further strengthen the reasoning ability of the model, a two-stage GCN reasoning model is introduced.

This paper is organized as follows. Section 2 introduces the related work of relation extraction. Section 3 gives the method of building the causality entity library, and Section 4 shows the construction method of the causality extraction model. Section 5 provides the experimental analysis of the model. Section 6 summarizes this article and future work.

2 Related works

2.1 Research on causality

More and more scholars have paid attention to causality extraction in recent years. Dasgupta et al.[4] uses complex formulas to represent causality annotates the cause, result, and causal connectives in sentences through bidirectional long-term and short-term memory networks. It converts the whole sentence into a words vector sequence and adds a language layer on Bi-LSTM to achieve good results. Zhang Shu et al.[5] consider the sequence information of long sentences because of the location uncertainty of causal information. Therefore, word embedding [29] is used as the input feature to extract causal events [34] based on the Bi-LSTM method. Silva Tharini et al. [6] compared the two methods. Based on the characteristics of knowledge and deep learning, the first group of experiments trained the SVM-based causal relationship classification model on the features of semantic knowledge and selected several different CNNs for experiments.

Li et al. [7] introduced other vital features of causality to improve the performance of CNN by reducing the dimension, considering that the construction of features requires a lot of engineering. An et al. [8] extracted causality from the literature from the perspective of rules improved the syntactic pattern matching method to simplify sentences and established a verb seed set to learn the characteristics of verbs, and finally achieved good results. Chaveevan et al. [9]extracted causal relationships from web documents, mainly extracting the causal path, using the causal path can explain or express some concepts, using supervised learning method to improve the accuracy. Abbas et al. [10]aimed to extract causality from biomedical literature and implemented and evaluated several commonly used models to reduce class imbalances and improve the performance of models by random oversampling techniques. Duc et al. [11] focused on the causality extraction of document-level texts and constructed a network. Fu et al. [22] proposed an event network. Shao et al. [12] proposed a BEL method to simplify sentences and improve the accuracy of causality extraction by BERT.

2.2 Research on Graph Neural Network

Xu et al. [13] introduced the commonly used graph convolution neural network in recent years, introduced two typical graph convolution neural network methods, the general method and the spatial

method, and described these two methods in detail. They also introduced various applications of graph convolution neural networks. Bosselut et al. [14] transformed implicit knowledge into explicit knowledge through the COMET model, established a commonsense knowledge base, and produced more high-quality new knowledge. Experiments show that new knowledge can be inferred through knowledge graphs. Zhu et al. [15] proposed a GP-GCNs to generate parameters for relation extraction and described the embedding module, propagation module, and classification module in detail. GCN adjusted the hyper-parameters in the process of propagation. Finally, qualitative analysis and quantitative analysis were carried out. The model can be used for inferring relations through the multi-hop mechanism.

Fu et al. [16] and others use relational weighted GCN to extract the relationship. This paper mainly considers the text's sequential features and local features and uses a dependency structure for GCN to learn the implicit features between all words in the text. The GCN model improves the prediction results with high accuracy on public data sets. Ali et al. [17] extracted event information by dependency graph and proposed an attention-based GCN to capture potential relationships between events. MHGCN presented by Gao et al. [18] embeds entities in each view of GCN to realize cross-language entity pair and supplement knowledge map. In addition to paying attention to the characteristics of entities themselves, Gao et al. also considered the characteristics of relational semantics and entity attributes to learn the structural features of entities better. Zhou et al. [19] used grammar dependence and GCN to establish a commonsense knowledge map, and GCN could flexibly combine syntactic information into emotion analysis tasks. Zhang et al. [20] used GCN learning to extract more accurate features. [25-27] and [32] The effectiveness of convolutional neural networks is proved.

Causality extraction[21] can quickly extract causal entities for explicit causality, but it often requires inferring to extract causality for an implicit causality. The traditional GCN can learn the features of adjacent nodes through the hidden layer to infer, and further inference is needed for the implicit causality of a cascade. Therefore, this paper proposes a two-stage GCN to implement the cascade implicit causality inference.

3 Construction of causality candidate entity library

3.1 Data preprocessing

The main tasks of data preprocessing include two aspects. One is to preliminarily screen the content of the text, delete the valueless information or unify the sentence format. The second is to mark the selected sentences. For example, causality pairs use cause phrases to represent a cause, and result phrases mean an effect. Since this article involves sequence annotations, punctuation is also annotated as words (labeled ' O '). The causality trigger words are not labeled. The causality extraction in this paper is not limited to the explicit causal relationship with markers. The labeling example is shown in Figure 2. C represents the cause, O represents the other, and E represents the effect.

Figure 2. Sample causation annotation

Sentence:

| C   O   C   C   O   E   E   O   E   O |
| [**Bacterial**] and [**comedonal debris**] cause [**acne pimples**] or [**pustales**]. |

Entity1:[Bacterial]	Entity2:[acne pimples]	{[Bacterial],Cause-Effect,[acne pimples]}
Entity1:[Bacterial]	Entity2:[pustales]	{[Bacterial],Cause-Effect,[pustales]}
Entity1:[comedonal debris]	Entity2:[acne pimples]	{[Bacterial],Cause-Effect,[acne pimples]}
Entity1:[comedonal debris]	Entity2:[acne pimples]	{[Bacterial],Cause-Effect,[acne pimples]}

This paper aims to extract all the causality entity pairs, as shown in Figure 3.

Figure 3. Examples of entity extraction

{[Bacterial],Cause-Effect,[acne pimples]}	{[Bacterial],Cause-Effect,[pustales]}

[**Bacterial**] and [**comedonal debris**] cause [**acne pimples**] or [**pustales**].

{[comedonal debris],Cause-Effect,[acne pimples]}	{[comedonal debris],Cause-Effect,[pustales]}
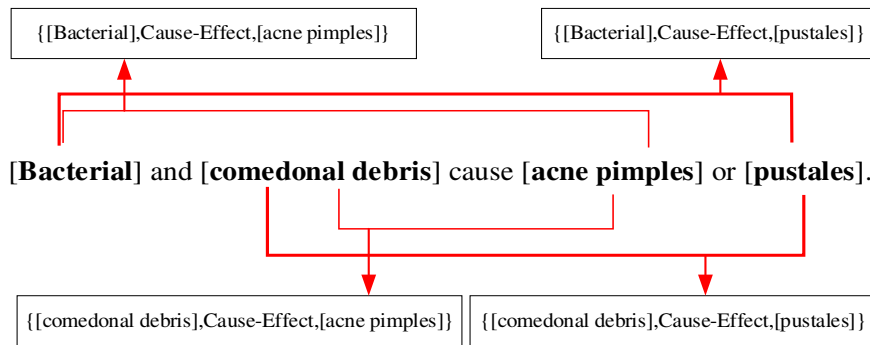
For example, in {[Bacterial], Cause-Effect,[ance pimples]}, the first entity represents the cause entity, and the second entity represents the effect entity. The relationship is Cause-Effect, which makes up a triple.

3.2 Extraction of causality candidate entities

Because causality is usually a phrase with more significant emotional tendencies, this paper will prioritize the emotional intensity of the words. The existing sentiment lexicon, such as the sentiment lexicon of the Taiwan University of China and the HowNet sentiment lexicon of CNKI, are all familiar. In the field of corpus sentiment analysis, it is impossible to judge the emotional intensity of some specific words accurately, so we need to build a corpus-oriented sentiment lexicon, as follows.

(1) **Data preprocessing, cleaning, and removing incomplete and repetitive data in the corpus.** The motivation is to ensure the corpus belongs to the same field. Stop words and special symbols are processed by word segmentation.

(2) **The construction of the word vector model.** We use the Word2Vec model in deep learning to transform words into word vectors, which build the foundation for constructing the subsequent neural network.

(3) **The construction of a neural network.** Through the dataset, we construct the corpus needed for training. At the same time, the word vector model is used to convert words into vectors for neural network training, and the emotional classifier is obtained finally.

(4) **Construction of domain sentiment lexicon.** The sentiment lexicon is mainly composed of the obtained sentiment lexicon and the candidate emotional words in the corpus field. The neural network classifier is used to judge its sentiment polarity, and the required domain sentiment lexicon is obtained.

This paper uses the neural network to construct a binary classifier of word emotion. It is known that

the training corpus is emotional words, and the corresponding label is the polarity of emotional words. Each emotional word is converted into a 100-dimensional word vector through Word2Vec, and the judgment of the sentiment polarity of words belongs to the classification problem. Therefore, we use the fully connected neural network to construct the classifier.

Step1: Determine the seed word set. According to the characteristics of the fields, the corresponding selection criteria are formulated, and the words in the corpus are extracted as seed words and added into the seed word set.

Step2: Determine the set of candidate emotion words. The seed words are converted into the word vector and calculated the vector of similarity formula (cosine similarity). N-words that are most similar to each seed word are obtained as the candidate emotion words set.

Step3: Use the trained classifier to judge the sentiment polarity of each candidate word. the candidate words with sentiment polarity are integrated and added to the sentiment lexicon for specific fields.

As for the causality in the corpus field, it may be positive or negative emotions, so we consider more about emotional intensity. Firstly, we calculate the emotional intensity of the causality, combined with the emotional intensity of manual annotation in the comment text. Whether a word frequently appears in the corpus and has strong sentiment polarity. When the conditions are met, it can be considered as a candidate seed word. The emotional intensity is divided into four levels: level-0, level-1, level-2, and level-3. The basis for the division mainly considers the following aspects: the completeness of components, the weighted average of the emotional intensity of each word, and the frequency of words that appears in the whole annotation corpus. Therefore, the following definition is given.

Definition 1: Emotional Intensity of Causality Entity (EICE)

Emotional Intensity of Causality Entity is used to measure the emotional intensity of entity i in the annotated corpus. The sentiment polarity intensity of the causality entity is calculated from the causal word. With the help of the emotional intensity characteristics, the causality entity can be extracted more accurately, as shown in Formula (1).

$$EIC_i = |F_i * I_i| * E_i * W_i \tag{1}$$

where F_i represents the frequency of causal word i in the whole annotated corpus, I_i represents the emotional intensity of causal word in the sentiment lexicon, and finally takes the absolute value. E_i represents the composition of the causality entity in the corpus (subject, predicate, object, etc.). And W_i represents the initial weight of the causality entity.

The emotional intensity values calculated by Formula (1) can be used to establish a causality seed lexicon, and appropriate weights can be given to causality entities with different emotional intensity levels. The emotional intensity level is mapped to the range of 1-4, and the integer is selected simultaneously. Therefore, the weights are 0.5, 1, 1.5, and 2, increasing by 1 for each ascending emotional intensity level. On this basis, the K-means clustering algorithm is used to cluster the candidate seed lexicon, and the candidate causality entity set is selected. The specific algorithm process is shown as follows.

Algorithm 1: Construction of the causality candidate library algorithm
Input: Preprocessed text S = { S_1, S_2,..., S_n }
Output: The candidate causality entity library L = { l_1, l_2,......, l_n }
1.for i in S_i,S = { S_1, S_2,..., S_n }:   // select each sentence after preprocessing
2.   for j in w_i: //w_i represents each word
3.   calculated EIC_i ;
4.   sentence partition Set D = { D_1, D_2,... D_n } ;
5.   calculate the *Avg (EIC_i)* of each phrase D_t ;
6.   select 2 *Max (D_t)* ;
7.for D_t in D : / / select phrases with strong polarity ;
8.   if is noun :   // judge whether this action is a noun
9.   extracts co-exist in the causal candidate library L = { l_1, l_2,......, l_n };
10.   end for;
12.end

The specific examples of the causality candidate entity library are shown in Table 1.

Table 1. examples of the causality candidate entity library

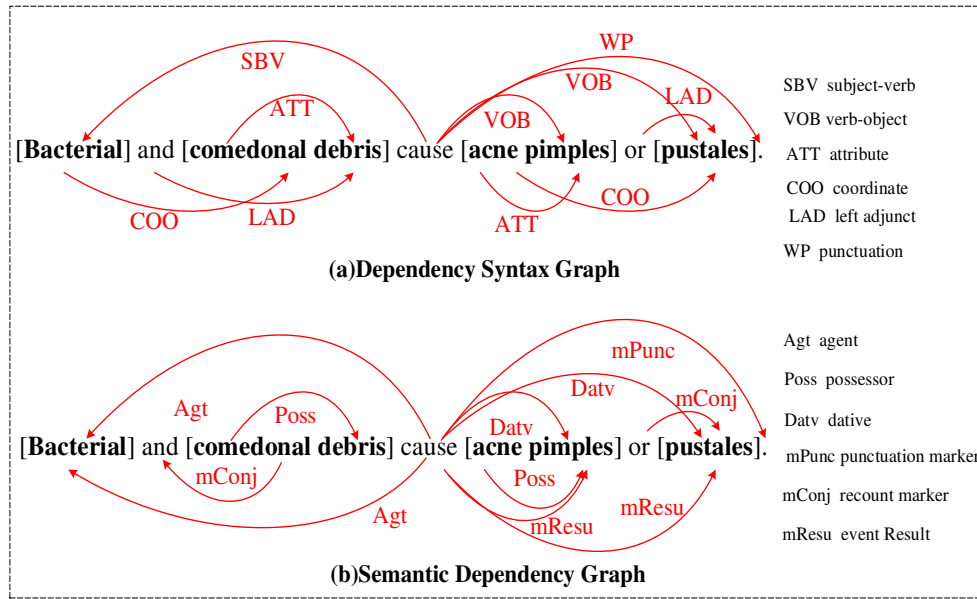| phrase	component	Weight	Emotional intensity
Account balance	subject	1	1
Accident year basis	subject	2	3
Abatement of tax	object	1.5	2

4 Construction of causality extraction model

4.1 Screening of nodes and edges

The representation of the reference node, candidate entity node, and meta-dependent path (MDP) node is constructed. MDP represents a set of shortest independent paths full reference nodes in a sentence. The representation of the candidate entity node is calculated as the average of the reference node and MDP node. At the same time, the dependency edges constructed by the syntactic dependency graph and semantic dependency graph are regarded as an adjacency matrix. The dependency syntactic graph created is centered on the core verbs in the graph. Therefore, in the input text, each sequence component is a node of the graph, and the dependency between words is the edge of the graph. In this paper, the HanLP tool is used to realize the syntactic analysis and dependency syntax analysis, and it is obtained as shown in Figure 4.

Figure 4. Semantic dependency graph and dependency syntax graph

(a)Dependency Syntax Graph

SBV subject-verb
VOB verb-object
ATT attribute
COO coordinate
LAD left adjunct
WP punctuation

(b)Semantic Dependency Graph

Agt agent
Poss possessor
Datv dative
mPunc punctuation marker
mConj recount marker
mResu event Result

4.2 Pre-training

Context word representation is usually trained on unstructured and unmarked texts that do not contain clear semantics with real-world entities. So it usually cannot remember entities other than these entities. For each sentence, use the integrated entity linker to retrieve the relevant entity embedding and then update the context word representation in the form of entity attention. The key idea is to model entities explicitly and use an entity linker to retrieve relevant entity embeddings from the constructed entity library. The motivation is to form knowledge-enhanced entity representation. The model is shown in Figure 5.

Figure 5. Pre-training process
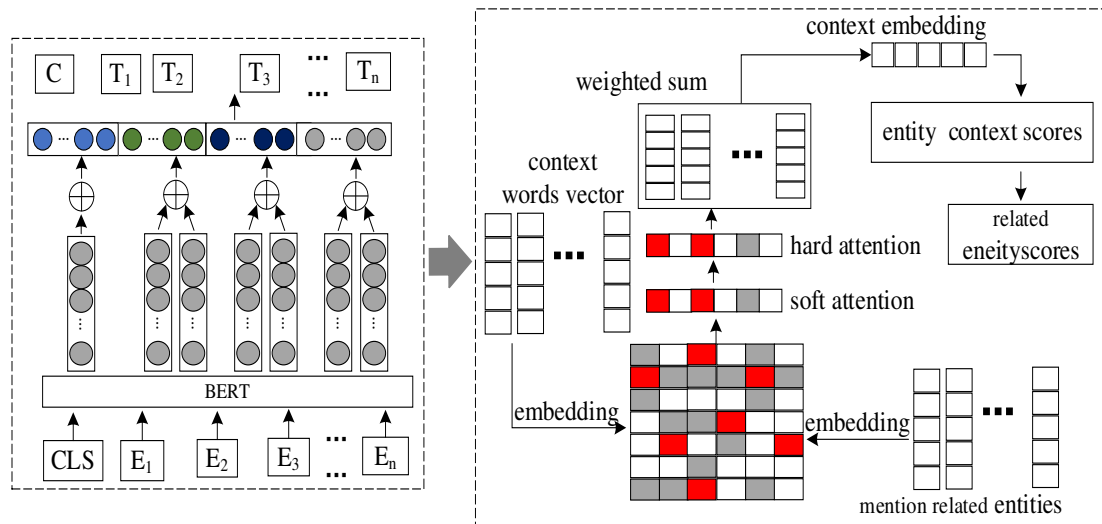
During BERT pre-training, each entity is represented by an exceptional identifier #. The task of Masked LM is described as giving a sentence, randomly erasing one or several words, and predicting what the words erased according to the remaining words are. Specifically, 15% of the words in a sentence are randomly selected for prediction. 80% of the erased words are replaced by a special symbol [MASK], 10% of the words are replaced by an arbitrary word, and 10% of the words remain unchanged to predict

a word. The model does not know whether the words that embed the corresponding position are correct. So it can force the model to rely more on context information to predict words and give the model a specific ability to correct errors.

Multiple self-attention consists of three parts: query, key, and value, allowing each vector to focus on other vectors. We train BERT to minimize the objective function that combines the next sentence prediction (NSP) with masking LM logarithmic likelihood (MLM).

$$L_{BERT}(\theta, \theta_1, \theta_2) = L_{NSP}(\theta, \theta_1) + L_{MLM}(\theta, \theta_2) \tag{2}$$

where Θ is the encoder parameter in BERT. Θ_2 is the parameter in the output layer connected to the encoder in the Mask-LM task, and Θ_1 is the classifier parameter associated with the encoder in the prediction task. Therefore, in the first part of the loss function, if the masked word set is M and its size is |V|, so it is also a multi-classification problem, then precisely loss function is formula (3).

$$L_{NSP}(\theta, \theta_1) = -\sum_{i=1}^{M} \log p(m = m_i \mid \theta, \theta_1), m_i \in [1, 2, ..., |V|] \tag{3}$$

It is also a loss function of a classification problem in the sentence prediction task.

$$L_{MLM}(\theta, \theta_2) = -\sum_{j=1}^{N} \log p(n = n_i \mid \theta, \theta_1), n_i \in [IsNext, NotNext] \tag{4}$$

Therefore, the joint learning loss function of the two tasks is :

$$L_{BERT}(\theta, \theta_1, \theta_2) = -\sum_{i=1}^{M} \log p(m = m_i \mid \theta, \theta_1) - \sum_{j=1}^{N} \log p(n = n_i \mid \theta, \theta_1) \tag{5}$$

In the pre-training process, the entity linker is used to link the causal entity library introduced in the previous section, which is mainly divided into the following three sub-modules.

(1) **The candidate entity generation module.** It is responsible for detecting the entity mention set M (including all the entities mentioned in the input text) and finding the related entity set Em corresponding to each entity mention m $\in$ M from the given causality entity library.

(2) **Related entity ranking module.** It is responsible for scoring and ranking multiple related entities in the related entity set Em (each entity mentions m) and outputs the related entity with the highest score as the entity link result of m.

(3) **Unlinkable mention prediction.** It is responsible for predicting which entities mentioned in the input text cannot be linked to the causality entity library.

4.3 Two-stage GCN

In this section, we use two-stage training GCN. GCN mainly learns entity features and corresponding relationship features in the first stage. GCN mainly considers implicit features between all entities and further infers causality in the second stage.

The first stage is GCN, and the original input is a sentence sequence. After BERT pre-training, the entity node and edge are obtained. The first stage uses GCN to extract regional dependency features. The principle of GCN is shown in Figure 6.

Figure 6. The principle of GCN
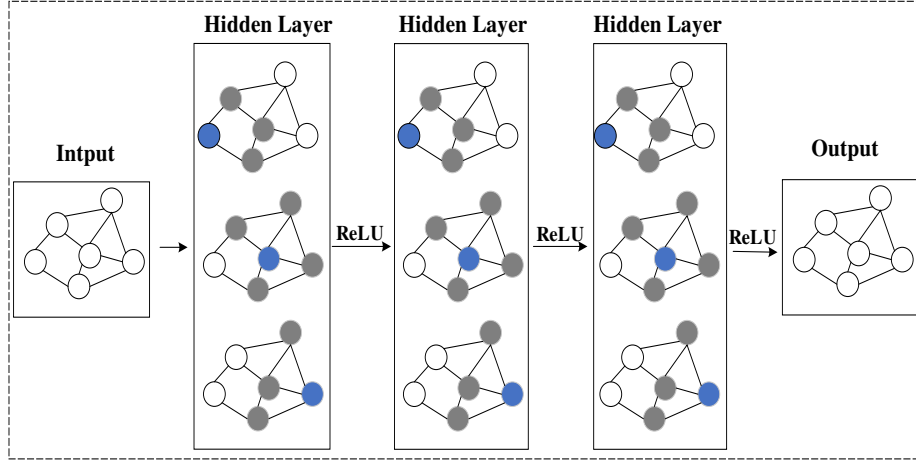
In each layer, ReLU represents the feature activation function. In this paper, three hidden layers are selected. In general, GCN accepts all the vertex feature messages transmitted from the former layer, makes corresponding transformations, and adds them together. Finally, an activation function is used as the output of this layer. For each hidden layer, there is formula (6).

$$h_u^{l+1} = \text{Re}\,LU\left(\sum_{u \in N(u)} \left(Wh_u^l + b\right)\right) \tag{6}$$

Here, h_u^l Shows the features of word u in hidden layer l, including all the words transmitted from word u and all the words introduced, including word u itself. W represents the weight. We connect the output and input word features as the last word features. Firstly, each node sends its feature information to neighbor nodes after transformation and extracts the feature information of the node. This step is to integrate the local structural information of nodes, namely the sum operation in the above formula (6) (for all neighbor nodes). And then do nonlinear transformation after gathering the previous information to increase the expression ability of the model.

The entity is predicted using the word features extracted from GCN, and the causality between entities is extracted. The dependency edges are removed and all entities are predicted. Based on the results of three hidden layers, entity causality is obtained.

$$S_{(w_1, ce, w_2)} = W_r^3 \,\text{Re}\,LU\left(W_r^1 h_{w_1} \oplus W_r^2 h_{w2}\right) \tag{7}$$

(w_1, r, w_2) denotes the scores of entity pairs (w_1, w_2) obtained under causality. When extracting causal triples, the relationship between each word pair is judged and identified as causality as possible.

In the second stage GCN, the entities and relationships extracted in the first stage are not very good for long-distance. Therefore, the second stage GCN is proposed to extract the implicit cascade causalities for long-distance multi-hop inference. The second stage considers the implicit features between all causal entities in the text so that the accuracy of the extraction is higher. In the first stage, a complete correlation weighted graph is established for each pair of causality. Where (w_1, w_2) is the weight of the edge to represent the probability that w_1 and w_2 entities are causalities. To extract the causality between each entity pair more accurately, the GCN in the second stage should carry out weighted propagation to achieve a more robust relationship prediction. The formula is shown as (8) to propagate between hidden layers.

$$h_u^{l+1} = \mathrm{Re}\,LU\left(\sum_{v \in V} \mathrm{P}_r\left(u,v\right) \times \left(W_r^l h_v^l + b_r^l\right)\right) + h_u^l \qquad (8)$$

where $P_r(u,v)$ denotes the weight with edge weight as edge, which indicates the probability that the two entities w_1 and w_2 are causal, and W_r and b_r are the weight text of layer l of GCN hidden layer, which excludes all words and all relationships.

Finally, a threshold is set to extract causality entity pairs. If $P_r(u,v) > 0.5$, the entity pair is considered to have a causality, and vice versa. The classification module takes the target entity pair (w_1, w_2) as input, and we stack the embedding of (w_1, w_2) together to infer the underlying relationship between each pair of entities. So we can obtain the causality of each pair of entities, as shown in formula 9.

$$\mathrm{P}_r\left(w_1, ce, w_2\right) = \mathrm{softmax}\left(S_{(w_1, ce, w_2)}\right) \qquad (9)$$

Here we use cross-entropy as the final classification loss function, as shown in formula 10.

$$Loss = \sum_{s \in S} \sum_{i \neq j} \log \mathrm{P}_r\left(w_i, ce, w_j\right) \qquad (10)$$

S is a set representing a collection of all entities, $P_r(w_i, w_j)$ denotes the probability that entities w_i and w_j are causality. The entire two-stage GCN algorithm process is as follows.

Algorithm 2: Two-stage GCN algorithm

Input: G=(V,E), Entity $V = \{V_1, V_2 \ldots\ldots,\ V_n\}$, dependency edge $E = \{E_1, E_2, \ldots\ldots E_n\}$

Output: Causal entity triples *(wᵢ, cause-effect, wⱼ)*

1.For i in $V = \{V_1, V_2 \ldots\ldots,\ V_n\}$:

2.　　Calculate *similarity(v_i, v_j)*,learning features;

3.　　Propagation between hidden layers: $f\left(H^l, A\right) = d\left(AH^l W^l\right)$

4.　　End for

5.If classifier $P_r(w_i, w_j) > 0.5$:

6.　　causality $(v_i, cause-effect,\ v_j)$;

7.　　Extract entities $(v_i,\ v_j)$ and relationships $cause-effect$ as a graph

8.　　Weighted graph $G = (w, e)$ as input to the second phase GCN

9.For i in $w, w = \{w_1, w_2, \ldots\ldots,\ w_n\}$:

10.　　Calculate *similarity (w_i, w_j)*,learning features;

11.　　Propagation between hidden layers $f\left(H^l, A\right) = d\left(AH^l W^l\right)$;

10.　　End for

11.Find cascaded causal entities;

12.Calculate $P_r(w_i,\ w_j)$

12.If $P_r(w_i, w_j) > 0.5$

13.　　Extraction of causality triple *(wᵢ, cause-effect ,wⱼ)*

14.End

Algorithms 2, Steps 1-6 describe the algorithm process of GCN in the first stage, mainly learning the local characteristics of nodes. Then construct a new graph. Steps 7-10 illustrate the process of the GCN algorithm in the second stage. Steps 11-14 completed the extraction of causality triples and improved extraction accuracy.

5 Experiments

5.1 Experimental setup

The datasets used are from open source databases SCIFI [28]、ECauSE Corpus2.0 [29]、CaTeRS [30]、NYT [16]、WebNLG [16]. SCIFI contains a total of 1270 valid data, 1803 sentences in the ECauSE Corpus2.0 corpus contain causality, and a third of them involve overlapping relationships. CaTeRS annotates a total of 1600 sentences in 320 five-sentence short stories extracted from the ROCStories corpus, which all contain causality. NYT and WebNLG are datasets for relation extraction, which contain overlapping entity relations. Among them, NYT contains 1230 sentences with causality, and WebNLG contains 1420 sentences. The dataset is mainly divided into a training set, test set, and verification set, which are divided according to 8: 1: 1. The construction of the dataset is shown in Table 2. We use SCIFI and NYT datasets to verify our method.

Table 2. Dataset construction

Dataset	Train	Test	Verify
SCIFI	1443	127	127
ECauSE Corpus2.0	1016	180	180
CaTeRS	1280	160	160
NYT	984	123	123
WebNLG	1136	142	142

To verify the effect of this model on causality extraction, the specific experimental operation is as follows.

Step 1: The causality dataset is obtained. The open-source databases SCIFI, ECauSE Corpus2.0, CaTeRS, NYT, and WebNLG are used for causality extraction. After denoising, 7323 sentences that are conducive to causal analysis are selected.

Step 2: Causal entity labeling. For the 7323 text sentences obtained, the causal entity labeling is unified. For the entities appearing in the sentences, the special symbol # is used to mark them, such as #entity#, and 7323 sentences with causal entity labeling are obtained.

Step 3: Candidate causality entity. Extract causality entity from a corpus and put it into entity library.

Step4: Extract the causal entity triples from the obtained review text. The BERT pre-training is used to convert the text with the semantics of each word into a word vector. At the same time, the entity linker is added to the pre-training, which can better learn the entity word vector. Then the two-stage GCN is used for feature learning. The first stage is to learn the local feature of the entity node, and the second stage is to learn the global feature. The probability of each pair of causality entities satisfies the condition for the threshold value.

To make the method more persuasive, we compare the models in many aspects. Firstly, the method is tested on multiple datasets to illustrate the effectiveness of the causality extraction model. At the same time, this paper also conducts comparative experiments on other baseline models. Including Bi-LSTM+GCN, Bi-LSTM+CRF, 3-layers CNN, GP-GCNs, CNN+RNN, CNN+BiGUR+CRF [23], Bi-LSTM+Attention.

5.2 Experimental analysis

According to the above analysis, this paper did the following experiments.

The GCN model directly extracts the causal triples. The Bi-LSTM context [24] coding is mainly to improve the pre-training process. The CNN model is often used to learn local features and three layers are selected here. GCN can learn long-distance relationship inference. The causality extraction model based on two-stage GCN, Bi-LSTM+GCN, Bi-LSTM+CRF, 3-layer GNN, GP-GCNs, CNN+RNN,

CNN+BiGUR+CRF are used. Bi-LSTM+Attention extracts causality from SCIFI and ECauSE Corpus2.0 datasets and calculates the recall. Finally, the experimental results are compared, and the specific experimental results are shown in Table 3.
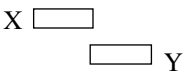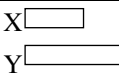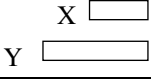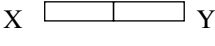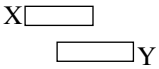
Table 3. Experimental result

Corpus	SCIFI			CaTeRS		
	P	R	F	P	R	F
Bi-LSTM+GCN	0.87	0.82	0.84	0.84	0.82	0.82
Bi-LSTM+CRF	0.85	0.81	0.83	0.82	0.80	0.81
3 layerCNN	0.88	0.84	0.86	0.83	0.81	0.82
GP-GCNs	0.90	0.88	0.89	0.85	0.83	0.84
CNN+RNN	0.80	0.77	0.78	0.77	0.75	0.76
CNN+ BiGUR+CRF	0.82	0.80	0.81	0.78	0.77	0.77
Bi-LSTM+Attention	0.81	0.80	0.80	0.76	0.74	0.75
Two-stage GCN	0.92	0.89	0.90	0.88	0.86	0.87

The results of table 3 show that based on the two-stage GCN causality extraction method for SCIFI and CaTeRS datasets, it has a higher recall score. Contrast Bi-LSTM+GCN, Bi-LSTM+CRF, 3-layer GNN, GP-GCNs, CNN+RNN, CN+BiGUR+CRF, Bi-LSTM+Attention, can be found with GP-GCNs in SCIFI dataset recall is higher than other methods.

For ECauSE Corpus2.0, NYT and WebNLG, due to many overlapping relationships, the overlapping relationships are mainly divided into the following types, as shown in Table 4.

Table 4. Relationship of overlapping entities

| Entity pair overlap	Relation	Location
X ▭ ... ▭ Y	intersection	The end of X equals the start of Y
X▭ / Y▭	overlap	XY starts the same
X ▭ / Y ▭	overlap	The end of XY is the same
X ▭▭ Y	intersection	The end of X equals the start of Y
X▭ ... ▭Y	overlap	XY overlaps at a certain position

It can be seen from Table 4 that each overlapping entity needs to be inferred, and GCN can achieve this inference. But the effect of using GCN only once is less than that of two-stage GCN, as shown in Figure 7.

Figure 7. Comparison learning ability of two-stage and one-stage GCN

One-stage GCN   Two-stage GCN
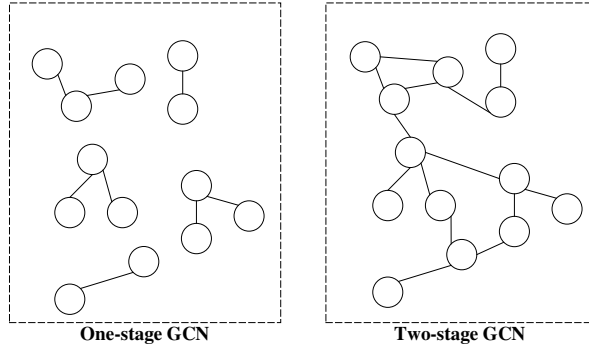
The number of layers of GCN is also an essential factor. To prove the influence of the number of layers, we also compare the models of different layers. It can be seen from Figure 8. In the two datasets, the second layer has the best effect. Although the impact of three layers is also good, the time of three layers is significantly more than two layers. It shows that more layers are considered in the inference process will lead to better performance, especially when there are more entities.

Figure 8. Comparison of Layers of GCN
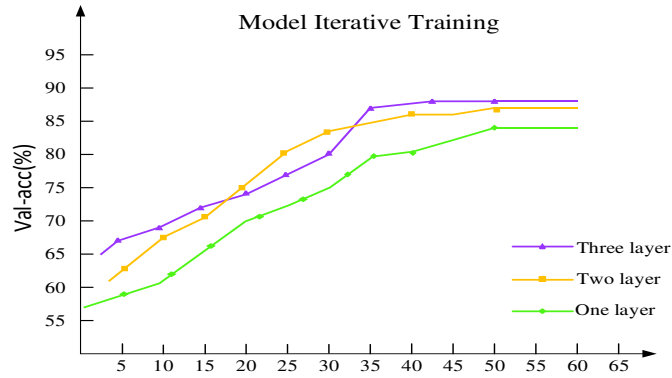
The apparent two-stage training is better than the one-stage training in the model training. Taking two datasets on SCIFI as examples, two examples of table 5.

Table 5. Two-stage GCN comparison

Sentence	The first stage GCN	The second stage GCN
Bacterial and comedonal debris cause acne pimples or pustales.	(bacterial,C-E,acne pimples) (bacterial,C-E,pustales) (comedonal debris,C-E,acnepimples) (comedonal debris,C-E,pustales)	(bacterial,C-E,acne pimples) (bacterial,C-E,pustales) (comedonal debris,C-E,acne pimples) (comedonal debris,C-E,pustales) (acne pimples,C-E,pustales)
The damages caused by mudslides, tremors, subsidence, superficial or underground water were verified, as well as swelling clay soils.	(mudslides,C-E, the damages) (tremors,C-E, the damages) (subsidence,C-E,acnepimples)	(mudslides,C-E, the damages) (tremors,C-E, the damages) (subsidence,C-E,acnepimples) (superficial,C-E,acnepimples (underground water,C-E,acnepimples)

The above experiments show that the cascade causality entity extraction relies not just on local features

but also needs an inference of global features. To get more comprehensive causal entities, each pair of causality entity are judged. The above experiments prove the effectiveness of this method.

6 Conclusions

For how to effectively extract causality, we propose a causality extraction model based on a two-stage GCN. This model mainly uses two-stage GCN to extract causality and analyzes all causal entity triples in the text. Especially the cascade implicit causality, to realize the deep extraction under semantic enhancement. The contributions of this paper mainly include the following aspects.

(1) The two-stage GCN has been proposed for entity relationship inferring. In the first stage, the local features of adjacent entity nodes are learned, and in the second stage, the features of all nodes are learned. The inferring of long-distance learning entities was carried out. Each pair of entities made causality judgment, and more cascade causality was identified.

(2) The causal entity library has been constructed. In the pre-training, the entity linker was added to find the three closest entities from the entity library so that the pre-training entity vector can learn more features. At the same time, it can better learn the features of the entity. This paper introduced the new concept of the difference coefficient of positive and negative corpus (DC-PNC) to judge the sentiment polarity of words. To a certain extent, the combination of PMI and DC-PNC improved the screening precision of new sentiment words.

In the future, the method in this paper can be considered to be applied to all relational extraction texts. The causality extraction model based on two-stage GCN can help the software platform or relevant departments to extract the causality effectively. It can carry out management measures or coping strategies, help make the best decisions, and build the foundation for subsequent emotional analysis.

Conflict of Interest:
Author Guangli Zhu declares that she has no conflict of interest.
Author Zhengyan Sun declares that she has no conflict of interest.
Author Shunxiang Zhang declares that he has no conflict of interest.
Author Kuan-Ching Li declares that he has no conflict of interest.
Author Subo Wei declares that he has no conflict of interest.

Also this manuscript is approved by all authors for publication. I (Guangli Zhu) would like to declare on behalf of all co-authors that the work described was original research that has not been published previously. All the authors listed have approved the manuscript that is enclosed.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent: No humans or any individual participants are involved in this study.

Availability of data and material: Data cannot be available for privacy reasons.

References

[1]   Zhang Shunxiang, Hu Zhaoya, Zhu Guangli, Jin Ming, Li Kuan-Ching. (2021) Sentiment classification model for Chinese micro-blog comments based on key sentences extraction. *Soft Computing*. Vol.25, pp.463-476

[2]   Liu Wenting, Zhu Guangli, Zhang Shunxiang. (2021) The method for extracting new login sentiment words from Chinese micro-blog basedf on improved mutual information. *Advances in Intelligent Systems and Computing*. Vol.1017, pp.1394-1403.

[3]   Xu Zheng, Zhang Shunxiang, Choo Kim-Kwang Raymond, Mei Lin, Wei Xiao, Luo Xiangfeng, Hu Chuanping, Liu Yunhuai. (2017) Hierarchy-cutting model based association semantic for analyzing domain topic on the web. *IEEE Transactions on Industrial Informatics*. Vol.13, No.4, pp.1941-1950.

[4]   Dasgupta Tirthankar, Saha Rupsa, Dey Lipika, Naskar Abir. (2018) Automatic extraction of causal relations from text using linguistictally informed deep neural networs. *Proceedings of the SIGDIAL 2018 Conference*, Melbourne, Australia, July, 2018, pp.306 -316.

[5]   Zhang Shu, Zheng Dequan, Hu Xinchen, Yang Ming. (2015) Idirectional long short-term memory networks for relation classification. *29th Pacific Asia Conference on Language, Information and Computation(PACLIC)*, Shanghai, China, October 30-November 1, 2015, pp.73-78.

[6]   Silva Tharini N DE, Zhibo Xiao, Rui Zhao, Kezhi Mao. (2017) Causal relation identification using convolutional neural networks and knowledge based features. *World Academy of Science*. Vol.11, No.6, pp.697-702.

[7]   Li Pengfei, Mao Kezhi.(2019) Knowledge-oriented Convolutional Neural Network for Causal Relation Extraction from Natural Language Texts. *Expert Systems With Applications*. Vol.115, pp.512-523.

[8]   An Ning, Xiao Yongbo, Yuan Jing, Yang Jiaoyun, Gil Alterovitz. (2019) Extracting causal relations from the literature with word vector mapping. *Computers in Biology and Medicine*. Online, 27 November, 2019, Vol.115, pp.103524.

[9]   Pechsiri Chaveevan, Piriyakul Rapepun. (2021) Causal pathway extraction from web-board documents. *Applied Sciences-Basel*. Online, 3 November, 2021, Vol.11, No.21, pp.10342.

[10] Abbas Akkasi, Mari-Francine Moens. (2021) Causal relationship extraction from biomedical text using deep neural models: A comprehensive survey. *Journal of Biomedical Informatics*. Online, May 24, 2021, Vol.119, pp.103820.

[11] Duc-ThuanVo, Feras Al-Obeidat, Ebrahim Bagheri. (2020) Extracting emporal and causal relations based on event networks. *Information Processing & Management*. Online, June 20, 2020, Vol.57, No.6, pp.102319.

[12] Shao Yifan, Li Haoru, Gu Jinghang, Qian Longhua, Zhou Guodong. (2021) Extraction of causal relations based on SBEL and BERT model. *Database*. Open Access, Vol.2021,pp.1-12.

[13] Xu Bingbing, Cen Keting, Huang Junjie, Shen Huawei, Cheng Xueqi. (2020) Overview of graph convolution neural network. *Journal of Computer Science*.Vol.43, No.5, pp.755–780.

[14] Bosselut Antoine, Rashkin Hannah, Sap Maarten, Malaviya Chaianya, Celikyilmaz, Choi Yejin. COMET: commonsense transformers for automatic knowledge graph construction. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 28-August 2, 2019, pp.4762-4779.

[15] Zhu Hao, Lin Yankai, Liu Zhiyuan, Fu Jie, Chua Tat-seng, Sun Maosong. Graph neural networks with generated parameters for relation extraction. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 28-August 2, 2019, pp.1331-1339.

[16] Fu Tsu-Jui, Li Peng-Hsuan, Ma Wei-Yun. GraphRel: Modeling text as relational graphs for joint entity and relation extraction. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 28-August 2, 2019, pp.1409-1419.

[17] Ali Balalia, Masoud Asadpoura, Ricardo Camposb, Adam Jatowt. (2020) Joint event extraction along shortest dependency path susing graph convolutional networks. *Knowledge-Based Systems*, Online, October 1, 2020, Vol.210, pp.106492.

[18] Gao, Jianliang, Liu, Xiangyue, Chen, Yibo, Xiong, Fan. (2022) MHGCN: Multiview highway graph convolutional network for cross-lingual entity alignment. *Tsinghua Science and Technology*.Vol.27,No.4, pp.719-728.

[19] Jie Zhou, Jimmy Xiangji Huang, QinminVivian Hu, Liang He. (2020) SK-GCN: Modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification. *Knowledge-Based Systems*. Online, July 17, 2020, Vol.205, pp.106292.

[20] Zhang Jinming, Hu Xiangyun, Dai Hengming. (2020) A graph-voxel joint convolution neural network for ALS point cloud segmentation. IEEE Access, Vol.1, No.8, pp.139781-139791.

[21] Wang Zhujun, Wang Shi, Li Xueqing, Zhu Junwu. (2021) An overview of event causality extraction based on deep learning.Computer applications. Vol.41, No.5, pp.1247-1255.

[22] Fu Jianfeng, Liu Zongtian, Liu Wei, Zhou Wen. (2011) Event causal relationship extraction based on the stratified conditional random field. *Pattern recognition and artificial intelligence*. Vol.24, No.4, pp.567–573.

[23] Zheng Qiaofen, Wu Zhendong, Zou Junying. (2021) Event causality extraction method based on double CNN-BiGRU-CRF model.*Computer engineering*. Vol.47, No.5, pp.58-64,72.

[24] Peters E. Matthew, Neumann Mark, Logan Robert, Schwartz Roy, Joshi Vidur, Singh Sameer, Smith A. Noah. (2019) Knowledge enhanced contextual word representations. *Proceeding of 9th International Joint Conference on Natural Language Processing(IJCNLP)*, Suzhou, China , December 4-7, 2019 ,pp.43-54.

[25] Zhao Haochen, Li Yaohang, Wang Jianxin. (2021) A convolutional neural network and graph convolutional network-based method for predicting the classification of anatomical therapeutic chemicals. *Bioinformatics*. Vol.37, No.18, pp.2841-2847.

[26] Xu Chang, Tao Dacheng, Zhang Xikun, Tian Xinmei. (2020) Graph edge convolutional neural networks for skeleton-based action recognition. *IEEE Transactions on Neural Networks and Learning Systems*.Vol.31, No.8, pp.3047-3060.

[27] Huan Lei, Naveed Akhtar, Ajmal Mian. (2021) Spherical kernel for efficient graph convolution on 3D point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.Vol.43, No.10, pp.3664-3680.

[28] Li Zhaoning, Li Qi, Zou Xiaotian, Ren Jiangtao. (2021)Causality Extraction based on Self-Attentive BiLSTM-CRF with Transferred Embeddings. *Neurocomputing*, Online, October 24 , 2020, Vol.423, pp.207-219.

[29] Dunietz Jesse, Levin S. Lori, Carbonell G. Jaime. (2017) The BECauSE Corpus 2.0: Annotating Causality and Overlapping Relations. *LAW 2017-11th Linguistic Annotation Workshop, Proceedings of the Workshop*, Valencia, Spain, April 3, 2017, pp.95-104.

[30] Mostafazadeh Nasrin, Grealish Alyson, Chambers Nathanael, Allen F. James, Vanderwende Lucy. (2016) CaTeRS: Causal and temporal relation scheme for semantic annotation of event structures. *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies(NAACL-HLT) San Diego*, California, June 17, 2016, pp.51-61.

[31] O'Gorman Tim, Wright-Bettner Kristin, Palmer Martha. (2016) Richer Event Description: Integrating event coreference with temporal, causal and bridging annotation. *Proceedings of the 2nd Workshop on Computing News Storylines*, Austin, TX, November 5, 2016, pp. 47-56.

[32] Suruchi Chawla. (2021) Application of convolution neural network in web query session mining for personalised web search. *International Journal of Computational Science and Engineering*. Vol.24, No.4, pp.417-428.

[33] Xu Kai, Wang Peng, Chen Xue, Luo Xiangfeng, Gao Jianqi. (2021) Causal event extraction using causal event element-oriented neural network. *International Journal of Computational Science and Engineering*. Vol.24, No.6, pp.621-628.

[34] Wu Zhihuan, Gao Yongming, Li Lei, Xue Junshi, Li Yuntao. (2019) Semantic segmentation of high-resolution remote sensing images using fully convolutional network with adaptive threshold. *Connection Science*. Vol.31, No.2, pp.169-184.

[35] Hu Junjie, Wang Zhanquan, Chen Jiequan, Dai Yonghui. (2021) A community partitioning algorithm based on network enhancement. *Connection Science*. Vol. 33, No.1, pp.42-61.