Causality extraction model based on two-stage GCN

Guangli Zhu\textsuperscript{1,2} • Zhengyan Sun\textsuperscript{1,2} • Shunxiang Zhang\textsuperscript{1,2} • Subo Wei\textsuperscript{1,2} • KuanChing Li\textsuperscript{3}

Accepted: 4 July 2022 / Published online: 16 August 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
As one of the indirect causality, cascaded causality can be used to construct the event knowledge graph, causal inference, scenario analysis, etc. The existing GCN methods lack the mining of context information and relevant entity information, resulting in the poor ability of causality inference, which inevitably affects the extraction accuracy of cascade causality. To solve this problem, this paper proposes a causality extraction model based on a two-stage GCN to improve the extraction accuracy. To obtain rich features of entities, this work combines sentiment polarity and knowledge base to get the causality candidate entity library. Firstly, the BERT model is pre-trained using context information and relevant entity information extracted from the entity library to obtain the final entity nodes. Secondly, using the semantic dependency graph, each possible edge between any two entity nodes can be obtained, which are input into the first stage GCN to get a preliminary directed graph of causality. Finally, the directed graph of causality is input into the second stage GCN to achieve deep causality multi-hop inference. Thus, the cascade causality is inferred and extracted by the two-stage GCN model. Experiments show that the extraction accuracy of cascade causality has been further improved.

Keywords Causality extraction • Cascade causality • Two-stage GCN • BERT

1 Introduction

Relation extraction gives a given text, extracts entities/objects, and infers their relations in the text, forming a triple (S, R, O) (Bosselut et al. 2019). 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. Existing methods can quickly identify explicit causality entities, but the extraction accuracy is not high because of the need for inference for implicit causality entities. At the same time, because of the uniqueness and diversity of text, the complexity of semantic structure (Hu et al. 2021), the variety of expression and other factors, it is inevitable to increase the difficulty of inference, which leads to many potential causality entities not extracted.

Given the above problems in the research on causality extraction, the following points are mainly considered. (1) Identification of Entity nodes. Previous work has proposed a key sentences extraction algorithm for Chinese microblog comments (Zhang et al. 2021), which considers multiple factors and attributes to identify the key sentences of comments. So this algorithm can provide technical assistance for the identification of entity nodes. (2) Feature extraction and causality inference. In the microblog hot topic word extraction model (Liu et al. 2021a), a feature cooccurrence method is proposed, providing a reference for causality feature extraction and causality inference. (3) Long-distance entity causality inference. Especially, it can effectively identify the cascade causality. Previous work has used ALN (Association link Network) (Xu et al. 2017) to achieve the hierarchical division of associated...
semantics. Learning the semantic information of the text provides technical support for long-distance entity causality inference.

Based on the above considerations, we combine BERT technology and GCN (Graph Convolutional 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 extracts the potential causality among cascade entities by two-stage inference of GCN. The motivation is that the network can learn more complex entity structures, capture more abundant local and non-local entity features and achieve multi-hop inference of entities for extracting cascade causality entities. The model framework is shown in Fig. 1, and the main contents of this paper can be summarized in the following three points.

• **Construct causality candidate entity library** We use prior causality knowledge and semantic data to extract causality entities from the dataset. Firstly, the review text is preprocessed. Then, the “NLTK” word segmentation is used to segment the words in the text. Secondly, using emotional intensity to establish causality seed lexicon. Finally, combined with the knowledge base, the K-means clustering algorithm is used to extend the causality candidate entity library.

• **Build the directed graph of causality by first stage GCN** To obtain rich target entity features, the relevant entity information from the causality candidate entity library is input into BERT by entity linking. Firstly, The BERT pre-training model is used to obtain the final entity node by the relevant entity and context semantic information. Secondly, the potential edges between any entity nodes are obtained by the semantic dependency graph. Then, the final entity nodes and edges are input into the first stage GCN to get the preliminary directed graph of causality.

• **Cascaded causality extraction by second stage GCN** Firstly, the directed graph of causality by the first stage GCN is input into the second stage GCN. Secondly, find the path for all reachable entities of an entity by an entity causality directed graph. Finally, the multi-hop inference of entity nodes is carried out through the second stage GCN model to extract all potential entity relations.

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

![Fig. 1 The framework of the two-stage GCN model](image-url)
• **Effective construction of causality candidate entity library** We consider that the causality candidate entity library is constructed by emotional intensity and part of speech, which is convenient for obtaining relevant entity information to learn entity features better.

• **Strong inference of two-stage GCN** The GCN of the first stage is conducted entity causality inference to get the preliminary directed graph of causality, and the second stage GCN realizes multi-hop inference between entities. So the two-stage model proposed in this paper further strengthens the inference ability.

This paper is organized as follows. Section 2 introduces the related work of causality extraction. Section 3 gives the method of building the causality candidate entity library, and Sect. 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. (2018) uses complex formulas to represent causality and annotates the cause, effect, and causal connectives in sentences through bidirectional long-term and short-term memory networks. It converts the whole sentence into a word vector sequence and adds a language layer on Bi-LSTM to achieve good results. Zhang et al. (2015) consider the sequence information of long sentences because of the location uncertainty of causality information. Therefore, word embedding (Dunietz et al. 2017) is used as the input feature to extract causal events (Wu et al. 2019) based on the Bi-LSTM method. Silva Tharini et al. (2017) 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 and Mao (2019) 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. (2019) extracted causality from the literature from the perspective of rules, improved the syntactic pattern matching method to simplify sentences, established a verb seed set to learn the characteristics of verbs, and finally achieved good results. Pechsiri and Piriyakul (2021) extracted causality from web documents, mainly the causal path. The causal path can be used to explain or express some concepts, and the supervised learning method is used to improve the accuracy. Abbas et al. (2021) 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. Vo et al. (2020) focused on the causality extraction of document-level texts and constructed a network. Akkasi and Moens (2021) extract causality from the biomedical field. Shao et al. (2021) proposed a BEL method to simplify sentences and improve the accuracy of causality extraction by BERT.

### 2.2 Research on graph neural network

Vargas-Hakim et al. (2022) introduced the convolution neural networks (CCNs) and introduced the state-of-the-art of encodings for CNNs. GCN model is closely related to the CNN model, both of which are the operations of aggregating neighborhood information. In recent years, GCN can capture more features because of its characteristics. More and more scholars turn their attention to GCN and use effective training methods to train GCN to improve efficiency (Liu et al. 2021b). Bosselut et al. (2019) transformed implicit knowledge into explicit knowledge through the COMET model, established a common-sense knowledge base, and produced more high-quality new knowledge. Experiments show that new knowledge can be inferred through knowledge graphs. Zhu et al. (2019) 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. (2019) 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 datasets. Balalia et al. (2020) extracted event information by dependency graph and proposed an attention-based GCN to capture potential relationships between events. MHGCN model presented by Gao et al. (2022) embeds entities in
each view of GCN to realize cross-language entity pairs and supplement the knowledge map. In addition, to pay 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. (2020) used grammar dependence and GCN to establish a common-sense knowledge map, and GCN could flexibly combine syntactic information into emotion analysis tasks (Wei et al. 2021). Zhang et al. (2020a) used GCN learning to extract more accurate features. Zhao et al. (2021), Zhang et al. (2020b), Lei et al. (2021) and Chawla (2021) have proved the effectiveness of convolutional neural networks.

Causality extraction (Wang et al. 2021) can quickly extract causality 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 cascade causality. Therefore, this paper proposes a two-stage GCN to implement the cascade causality inference.

3 Construction of causality candidate entity library

3.1 Data pre-processing

The main tasks of data pre-processing 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 effect 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 causality with markers. The labeling example (Li et al. 2021) is shown in Fig. 2 and Fig. 3. C represents the cause, O represents the other, and E represents the effect.

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

For example, in [Bacterial], Cause-Effect, [ance pimple], 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 (i.e., abbreviation of China National Knowledge Internet, a paper retrieval database in China), are all familiar. In 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.

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 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 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.

Step 1: 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 which are added to the seed word set.
Step 2: Determine the set of emotional words. The seed words are converted into the word vector and calculated as the vector of similarity formula (cosine similarity). N-words that are most similar to each seed word are obtained as the emotion words set.

Step 3: Use the trained classifier to judge the sentiment polarity of each word. The words with sentiment polarity are integrated and added to the sentiment lexicon for specific fields. As the causality in the dataset field may be positive or negative emotions, so we consider more emotional intensity. Firstly, we calculate the emotional intensity of the causality entity, combined with the emotional intensity of manual annotation in the comment text. Whether a word frequently appears in the dataset and has strong sentiment polarity, when the conditions are met, it can be considered a 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 dataset. 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 the entity in the annotated dataset. 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).

\[
EIC_i = |F_i \times I_i| \times E_i \times W_i
\]

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

The emotional intensity values calculated by the formula (1) can be used to establish a causality seed entity library, 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. The K-means clustering algorithm is used to cluster the candidate seed entity library, so the causality candidate entity seed set is constructed. The specific algorithm process is shown as follows.
Algorithms 1, Steps 1–3 calculate the $EIC_i$ of each word. Then, Step 4 splits the sentence. Next, calculate $Avg\ (EIC_i)$ for the segmented clauses, and select the most emotional intensive clause in Step 6. Steps 7–8 determine the part of speech of each clause. If it is a noun, add a clause phrase to $L$ in Step 9. Steps 11–12 expand the candidate seed entity library.

Obtain seed entity library on the dataset and expand related entities with the domain knowledge base. The specific examples of the causality candidate entity library are shown in Table 1.

4 Construction of causality extraction model

4.1 Screening of nodes and edges

The representation of entity labels is constructed, and Sect. 3.1 of this paper introduces the unified label framework of entities. At the same time, the dependency edges constructed by the semantic dependency graph are regarded as an adjacency matrix input into GCN. The semantic dependency graph created is centered on the core entity in the graph. Therefore, in pre-training input, 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 semantic analysis and it is obtained as shown in Fig. 4.

4.2 Pre-training

In the causality entity extraction task, BERT mainly learns syntactic structure in the low-level network structure and learns rich semantic information features in the high-level network. However, BERT needs to be improved for the ambiguity of the entity itself. So this paper needs to learn more complex entity information by entity linking. Context word is usually trained on unstructured and unmarked texts and does not contain precise semantics for real-world entities. So it usually cannot remember entities other than these entities. For each sentence, use the integrated 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 a linker to retrieve relevant entity information embeddings from the constructed causality candidate entity library. The motivation is to form knowledge-enhanced entity representation. The model is shown in Fig. 5.

During BERT pre-training, each entity is represented by an exceptional identifier #. BERT pre-training sequence generation uses Masked LM. The task of Masked LM is described as giving a sentence, randomly erasing one or several words, and predicting what the words erased

| Phrase             | Component | Weight | Emotional intensity |
|--------------------|-----------|--------|---------------------|
| Clay soils         | Subject   | 1      | 1                   |
| Comedonal debris   | Subject   | 2      | 3                   |
| Pustales           | Object    | 1.5    | 2                   |
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], an arbitrary word replaces 10% of the words, 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).

\[
L_{\text{BERT}}(\theta, \theta_1, \theta_2) = L_{\text{NSP}}(\theta, \theta_1) + L_{\text{MLM}}(\theta, \theta_2) \quad (2)
\]

where \(\theta\) is the encoder parameter in BERT. \(\theta_2\) is the parameter in the output layer connected to the encoder in the Mask-LM task and \(\theta_1\) 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 \(|V|\), it is also a multi-classification problem, then precisely the loss function is formula (3).

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

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

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

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

\[
L_{\text{BERT}}(\theta, \theta_1, \theta_2) = - \sum_{i=1}^{M} \log p(m = m_i|\theta, \theta_1) - \sum_{j=1}^{N} \log p(n = n_j|\theta, \theta_1) \quad (5)
\]

In the pre-training process, the related entity information from the causality candidate entity library by entity linking was introduced in the previous section, which is mainly divided into the following three sub-modules.

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

(2) **Related entity ranking module** It is responsible for scoring and ranking multiple related entities in the related entity set \( E_m \) (i.e., each entity mentions \( m \)) and outputs the related entity with the four highest scores as the entity link result of \( m \).

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

### 4.3 Two-stage GCN

In this section, we use the two-stage GCN model to infer. GCN mainly learns entity node features and corresponding relationship features in the first stage to get a directed graph of causality. And the second stage GCN realizes multi-hop inference between entities. GCN mainly considers implicit features between all entities and further infers causality in the second stage. So the two-stage GCN model proposed in this paper further strengthens the inference ability.

The original input is a sentence sequence. After BERT pre-training, the entity nodes and edges are obtained. GCN can be used to extract entity regional dependency features. In this paper, the GCN model is used to realize causality inference, and the first stage GCN model and the second stage GCN model adopt the same structure. The principle of GCN is shown in Fig. 6.

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).

\[
h^{l+1}_u = \text{ReLU} \left( \sum_{v \in N(u)} (W h^l_v + b) \right) \quad (6)
\]

Here, \( h^l_u \) shows the features of the word \( u \) in a hidden layer, including all the words transmitted from the word \( u \) and all the introduced words, including the 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 (i.e., for all neighbor nodes), namely the sum operation in the above formula (6). 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 two hidden layers, entity causality is obtained.

\[
S_{(w_1, c, w_2)} = W_2^c \text{ReLU} (W_1^c h_{w_1}) \quad (7)
\]

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

In the second stage of 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 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. Here, \((w_1, w_2)\)is the weight of the edge to represent the probability that \( w_1 \) and \( w_2 \) entities are causalties. 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.

$$h_{u}^{l+1} = \text{ReLU} \left( \sum_{v \in V} P_r(u,v) \times (W^l_r{h}^l_r + b^l_r) \right) + h^l_r$$  \hspace{1cm} (8)

where $P_r(u,v)$ denotes the weight with edge weight as edge, which indicates the probability that the two entities $w_t$ and $w_j$ 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 $(w_t, w_j)$ as input, and it stack the embedding of $(w_t, w_j)$ 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).

$$P_r(w_t, ce, w_j) = \text{softmax} \left( S(w_t, ce, w_j) \right)$$  \hspace{1cm} (9)

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

$$\text{Loss} = \sum_{w \in S} \sum_{i \neq j} \log P_r(w_i, ce, w_j)$$  \hspace{1cm} (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 $V = \{V_1, V_2, \ldots, V_n\}$, dependency edge $E = \{E_1, E_2, \ldots, E_n\}$

**Output:** Causal entity triples $(w_i, cause-effect, w_j)$

1. For $i$ in $V = \{V_1, V_2, \ldots, V_n\}$:
   2. Calculate similarity $(v_i, v_j)$, learning features;
   3. Propagation between hidden layers: $f(H, A) - d(AH')$
   4. End for
5. If classifier $P_r(v_i, v_j) > 0.5$:
   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 $w, w = \{w_1, w_2, \ldots, w_k\}$:
   10. Calculate similarity $(w_i, w_j)$, learning features;
   11. Propagation between hidden layers $f(H', A) - d(AH')$
   12. End for
13. Find cascaded causal entities;
14. Calculate $P_r(w_i, w_j)$
15. If $P_r(w_i, w_j) > 0.5$
16. Extraction of causality triple $(w_i, cause-effect, w_j)$
17. End

---

Algorithms 2. Step 2 calculates the similarity between $v_i$ and $v_j$, learns features, and propagates through Step 3 hidden layer. If the $P_r(v_i, v_j)$ calculated by Steps 5–8 exceeds the set threshold, there is a causal relationship between $v_i$ and $v_j$. Steps 1–8 describe the algorithm process of GCN in the first stage, mainly learning the local characteristics of nodes. Then, construct a new graph. Steps 9–12 illustrate the process of the GCN algorithm in the second stage. Step 10 calculates the similarity between $w_i$ and $w_j$, learns features, and propagates through Step 11 hidden layer. If the $P_r(w_i, w_j)$ calculated by Step 14 exceeds the threshold set by Step 15, there is a causal relationship between $w_i$ and $w_j$. Steps 13–17 completed the extraction of causality triples and improved extraction accuracy.
5 Experiments

5.1 Experimental setup

The datasets used are from open source databases SCIFI1 (Li et al. 2021), ECauSE Corpus2.02 (Dunietz et al. 2017), CaTeRS3 (Mostafazadeh et al. 2016), NYT4 (Fu et al. 2019), WebNLG (Fu et al. 2019). 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. NYT contains 1230 sentences with causality, and WebNLG contains 1420 sentences. Mining information in large data sets requires rational allocation and composition of data (Masood et al. 2021). The dataset is mainly divided into a training set, test set, and validation 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.

The specific operation for perfect experiments 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 causality analysis are selected.

Step 2: Causality entity pre-processing For the 7323 text sentences obtained, the causality 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: Construct the causality candidate entity library Extract the potential causality entity from the dataset and put it into a candidate entity library, on this basis, using a domain knowledge base to expand the candidate entity library.

Step 4: Construction of causality extraction model The BERT pre-training is used to convert the text with the semantics of each word into a word vector. At the same time, to learn the entity features better, BERT pre-training is used to obtain context information and related entity information from the causality candidate entity library. Then, the two-stage GCN is used for causality inference. The first stage is to learn the local feature of the entity node, and the second stage is to learn the global feature.

Step 5: Extract the causality entity triples Each pair of entities meeting threshold conditions is identified as causality entities. All entity pairs satisfying causality are extracted to form a triple.

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, Bi-LSTM + Attention.

5.2 Experimental analysis

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

This paper selects different models for comparison. The Bi-LSTM context (Peters et al. 2019) 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.

Table 2 Dataset construction

| Dataset     | Train | Test | Verify | Total |
|-------------|-------|------|--------|-------|
| SCIFI       | 1016  | 127  | 127    | 1270  |
| ECauSE Corpus2.0 | 1443  | 180  | 180    | 1803  |
| CaTeRS      | 1280  | 160  | 160    | 1600  |
| NYT         | 984   | 123  | 123    | 1230  |
| WebNLG      | 1136  | 142  | 142    | 1420  |

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   |

Bold values are used to highlight the comparison model compared with the model proposed in this paper.
Bi-LSTM + Attention extracts causality from SCIFI and ECauSE Corpus2.0 datasets and calculates the recall. The proposed method has been compared with the above methods, mainly through the accuracy, recall and F1 score evaluation (Mercy Rajaselvi Beaulah et al. 2018). Finally, the experimental results are compared, and the specific experimental results are shown in Table 3.

Two-stage GCN model is more comprehensive. 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.

We consider more overlapping entity relationships. 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.

The inference ability of the two-stage GCN is stronger. 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 Fig. 7.

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 in Fig. 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 considered in the inference process will lead to better performance, especially when there are more entities.

The apparent two-stage training is better than the one-stage training in the model training. We took two datasets on SCIFI as examples, two examples in Table 5. As can be seen from Table 5, the GCN effect of two-stage is better than that of only one-stage GCN. Although the GCN of one-stage also implements inference, the inference ability of the two-stage GCN is stronger for long distances.

The above experiments show that the cascade causality entity extraction relies not just on local features but also needs of global features. It is necessary not only to judge the relationship between adjacent entities but also to infer cascade entities. So this work uses a two-stage GCN for efficient inference to extract all potential causality. Because of the need to learn each pair of causality entities more comprehensively, the candidate entity library introduced in Sect. 3 is constructed to learn entity features. And two-stage GCN is used to infer the relationship between entities. The various aspects of the experiment are discussed, and the above experiments prove the effectiveness of this method.

### 6 Conclusions

We propose a causality extraction model based on a two-stage GCN to extract causality effectively. This model mainly uses two-stage GCN to extract causality and analyzes all causality entity triples in the text—especially the cascade causality. Finally, causality extraction in deep semantics is realized. The contributions of this paper mainly include the following aspects.

1. **The causality candidate entity library has been constructed.** In the pre-training, the four closest related entities from the candidate entity library by entity linking is selected, so the pre-training entity vector could learn more features. At the same time, it can better learn the features of the entity.

2. **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 model realizes long-distance relation inference. Each pair of entities made a causality judgment to identify more cascade causality.

In the future, the method in this paper can be considered to be applied to all relational extraction texts. The method proposed in this paper can well apply cascade causality texts. But in the face of more complex multifactorial, text effect remains to be improved. In the future, the

### Table 5 Two-stage GCN comparison

| Sentence                                                                 | In the first stage GCN                      | In the second stage GCN                      |
|-------------------------------------------------------------------------|--------------------------------------------|---------------------------------------------|
| Bacterial and comedonal debris cause acne pimples or pustules           | (bacterial,C-E,acne pimples)               | (bacterial,C-E,acne pimples)                |
|                                                                         | (bacterial,C-E,pustules)                   | (bacterial,C-E,pustules)                    |
|                                                                         | (comedonal debris,C-E,acne pimples)        | (comedonal debris,C-E,acne pimples)         |
|                                                                         | (comedonal debris,C-E,pustules)            | (comedonal debris,C-E,pustules)             |
| The damages caused by mudslides, tremors, subsidence, superficial or    | (mudslides,C-E, the damages)                | (mudslides,C-E, the damages)                |
| underground water were verified, as well as swelling clay soils        | (tremors,C-E, the damages)                  | (tremors,C-E, the damages)                  |
|                                                                         | (subsidence,C-E,acne pimples)              | (subsidence,C-E,acne pimples)               |
|                                                                         | (superficial,C-E,acne pimples)             | (superficial,C-E,acne pimples)              |
|                                                                         | (underground water,C-E,acne pimples)       | (underground water,C-E,acne pimples)        |
comprehensive analysis combined with emotional analysis can be considered to broaden the scope of application, such as dialogue system and situational analysis combined with emotional analysis. The causality extraction model based on the two-stage GCN model 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.

Funding Funding was provided by the University Synergy Innovation Program of Anhui Province (Grant number GXXT-2021-008), National Natural Science Foundation of China (Grant number 62076006), Anhui 2019 Anhui Provincial Natural Science Foundation Project (Grant Number 1908085MF189), Graduate Innovation Foundation Project of Anhui University of Science and Technology (Grant Number 2021CX2110).

Data availability Data cannot be available for privacy reasons.

Declarations

Conflict of interest The authors declares that they have 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.

References

Abbás A, Moens MF (2021) Causal relationship extraction from biomedical text using deep neural models: A comprehensive survey. J Biomed Inform 119:103820

Akkasi A, Moens MF (2021) Causal relationship extraction from biomedical text using deep neural models: A comprehensive survey. J Biomed Inform 119:103820

An N, Xiao Y, Yuan J, Yang J, Alterovitz G (2019) Extracting causal relations from the literature with word vector mapping. Comput Biol Med 115:103524

Balalai A, Asadpoura M, Camposh R, Jatowt A (2020) Joint event extraction along shortest dependency path using graph convolutional networks. Knowl-Based Syst 210:106492

Bosselut A, Rashkin H, Sap M, Malaviya C, Celikyilmaz A, Choi Y (2019) COMET: Commonsense transformers for automatic knowledge graph construction. In: Proceedings of the 57th annual meeting of the association for computational linguistics, Florence, Italy, July 28-August 2, 2019, pp.4762–4779.

Chawla S (2021) Application of convolution neural network in web query session mining for personalised web search. Int J Comput Sci Eng 24(4):417–428

Dasgupta T, Saha R, Dey L, Naskar A (2018) Automatic extraction of causal relations from text using linguistically informed deep neural networks. In: Proceedings of the SIGDIAL 2018 conference, Melbourne, Australia, July, 2018, pp.306–316.

De Silva TN, Zhibo X, Rui Z, Kezhi M (2017) Causal relation identification using convolutional neural networks and knowledge-based features. Int J Comput Syst Eng 11(6):696–701

Dunietz J, Levin L, Carbonell JG (2017) The BECaUSE Corpus 2.0: annotating causality and overlapping relations. In: Proceedings of the 11th linguistic annotation workshop, Valencia, Spain, April 3, 2017, pp 95–104.

Fu TJ, Li PH, Ma WY (2019) GraphRel: Modeling text as relational graphs for joint entity and relation extraction. In: Proceedings of the 57th annual meeting of the association for computational linguistics, Florence, Italy, July 28–August 2, 2019, pp.1409–1419.

Gao J, Liu X, Chen Y, Xiong F (2022) MHGCN: Multiview highway graph convolutional network for cross-lingual entity alignment. Tsinghua Sci Technol 27(4):719–728.

Hu J, Wang Z, Chen J, Dai Y (2021) A community partitioning algorithm based on network enhancement. Connect Sci 33(1):42–61

Lei H, Akhtar N, Mian A (2021) Spherical kernel for efficient graph convolution on 3d point clouds. IEEE Trans Pattern Anal Mach Intell 43(10):3664–3680.

Li P, Mao K (2019) Knowledge-oriented convolutional neural network for causal relation extraction from natural language texts. Expert Syst Appl 115:512–523.

Li Z, Li Q, Zou X, Ren J (2021) Causality extraction based on self-attentive BiLSTM-CRF with transferred embeddings. Neurocomputing 423:207–219.

Liu W, Zhu G, Zhang S (2021a) The method for extracting new login sentiment words from Chinese micro-blog based on improved mutual information. Adv Intell Syst Comput 1017:1394–1403.

Liu X, Yan M, Deng L, Li G, Fan D (2021b) Sampling methods for efficient training of graph convolutional networks: a survey. IEEE-CAA J Autom Sinica 9(2):205–234.

Masood M, Arunarani AR, Manjula D, Sugumaravan V (2021) An efficient algorithm for identifying (f, d) motif from huge DNA datasets. J Amb Intell Hum Comput 12(1):485–495.

Mercy Rajaselvi Beaulah P, Manjula D, Sugumaravan V (2018) Categorization of images using autoencoder hashing and training of intra bin classifiers for image classification and annotation. J Med Syst 42(7):1–15.

Mostafazadeh N, Grealis A, Chambers N, Allen J, Vanderwende L (2016) CaTeRS: Causal and temporal relation scheme for semantic annotation of event structures. In: Proceedings of the fourth workshop on events, California, June 17, 2016, pp 51–61.

O’Gorman T, Wright-Bettner K, Palmer M (2016) CaTeRS: Causal and temporal relation scheme for semantic annotation of event structures. In: Proceedings of the fourth workshop on events, California, June 17, 2016, pp 51–61.

O’Gorman T, Wright-Bettner K, Palmer M (2016) Richer event description: Integrating event coreference with temporal, causal and bridging annotation. In: Proceedings of the 2nd workshop on computing news storylines, Austin, TX, November 5, 2016, pp 47–56.

Pechsiri C, Piriyakul R (2021) Causal pathway extraction from web-board documents. Appl Sci Basel 11(21):10342.

Peters ME, Neumann M, Logan IV RL, Schwartz R, Joshi V, Singh S, Smith NA. (2019) Knowledge enhanced contextual word representations. In: Proceeding of 9th international joint conference on natural language processing (IJCNLP), Suzhou, China, December 4–7, 2019, pp 43–54.

Shao Y, Li H, Gu J, Qian L, Zhou G (2021) Extraction of causal relations based on SBEL and BERT model. Database 2021:1–12.

Vargas-Hakim GA, Mezura-Montes E, Acosta-Mesa HG (2022) A review on convolutional neural network encodings for neu-rowevolution. IEEE Trans Evol Comput 26(1):12–27.
Vo DT, Al-Obeidat F, Bagheri E (2020) Extracting temporal and causal relations based on event networks. Inf Process Manage 57(6):102319

Wang Z, Wang S, Li X, Zhu J (2021) Review of event causality extraction based on deep learning. J Comput Appl 41(5):1247–1255

Wei S, Zhu G, Sun Z, Li X, Weng T (2022) GPGCN: Global features of orthogonal projection and local dependency fused graph convolutional networks for aspect-level sentiment classification. Connect Sci 34(1):1785–1806

Wu Z, Gao Y, Li L, Xue J, Li Y (2019) Semantic segmentation of high-resolution remote sensing images using fully convolutional network with adaptive threshold. Connect Sci 31(2):169–184

Xu Z, Zhang S, Choo KKR, Mei L, Wei X, Luo X, Hu C, Liu Y (2017) Hierarchy-cutting model based association semantic for analyzing domain topic on the web. IEEE Trans Industr Inf 13(4):1941–1950

Xu K, Wang P, Chen X, Luo X, Gao J (2021) Causal event extraction using causal event element-oriented neural network. Int J Comput Sci Eng 24(6):621–628

Zhang J, Hu X, Dai H (2020a) A graph-voxel joint convolution neural network for ALS point cloud segmentation. IEEE Access 1(8):139781–139791

Zhang X, Xu C, Tian X, Tao D (2020b) Graph edge convolutional neural networks for skeleton-based action recognition. IEEE Trans Neural Netw Learn Syst 31(8):3047–3060

Zhang S, Hu Z, Zhu G, Jin M, Li K (2021) Sentiment classification model for Chinese micro-blog comments based on key sentences extraction. Soft Comput 25:463–476

Zhang S, Zheng D, Hu X, Yang M. (2015) Bidirectional long short-term memory networks for relation classification. In: 29th Pacific Asia conference on language, information and computation(PACLIC), Shanghai, China, October 30–November 1, 2015, pp.73–78.

Zhao H, Li Y, Wang J (2021) A convolutional neural network and graph convolutional network-based method for predicting the classification of anatomical therapeutic chemicals. Bioinformatics 37(18):2841–2847

Zhou J, Huang JX, Hu QV, He L (2020) Sk-gcn: Modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification. Knowl-Based Syst 205:106292

Zhu H, Lin Y, Liu Z, Fu J, Chua TS, Sun M (2019) Graph neural networks with generated parameters for relation extraction. In: Proceedings of the 57th annual meeting of the association for computational linguistics, Florence, Italy, July 28-August 2, 2019, pp.1331–1339.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.