The Journal of Supercomputing (2022) 78:6503–6527
https://doi.org/10.1007/s11227-021-04097-5

Financial causal sentence recognition based on BERT-CNN text classification

Chang-Xuan Wan[1,3] · Bo Li[1,2,3]

Accepted: 20 September 2021 / Published online: 20 October 2021

Abstract
By studying the causality contained in financial texts, we can further reveal more potential laws of economic activities, such as "factors promoting stable and healthy economic development," "The central bank's use of the loan window to issue money will increase the probability of inflation," "The consequence of overcapacity is a decline in product prices," and so on. Causal sentence recognition usually includes two sub-tasks: one is to design rules or templates to find candidate causal sentences; the other is to design a classifier to sort candidate causal sentences to finally identify the causal sentence. This article first focuses on the characteristics of complex sentence patterns of multiple causes and one effect, multiple effects and one cause, and multiple causes and multiple effects in financial review texts, and provides a relatively complete candidate causal sentence identification rules, which can identify both simple causal sentences and complex causal sentences. A BERT-CNN (Bidirectional Encoder Representations from Transformers-Convolutional Neural Networks) combination model is proposed for the classification of candidate causal sentences. On the one hand, by adding a CNN (Convolutional Neural Networks) structure to the specific task layer of the BERT (Bidirectional Encoder Representations from Transformers) model to capture important local information in the text. On the other hand, in order to make better use of the self-attention mechanism, the local text representation and the output of the BERT are input together in the multi-layer transformer encoder. A complete representation of the text is finally obtained through a single-layer transformer encoder. Experimental results show that our model is significantly better than the most advanced baseline model, with a 5.31 pts improvement in F1 over previous analyzers.

Keywords Text classification · Recognition of causality · BERT model

✉ Chang-Xuan Wan
wanchangxuan@263.net

Extended author information available on the last page of the article

1 Introduction

Causality is an important type of relationship [1]. People obtain the background of an event and the evolutionary rules between events by identifying the causality of the event and further predict the changes that may occur in the future [2]. The study found that the Chinese financial text contains causal event pairs. For example:"原材料价格上涨导致企业生产成本的增加。"("Raw material prices have led to an increase in the production cost of a company.") It shows that the event "原材料价格上涨"("Raw material prices") is the cause of the event "企业生产成本的增加"("increase in production cost of a company."). Event causality extraction is also one of the key steps in natural language processing applications such as event prediction [3], future scene generation [4], automatic question answering systems, and text entailment. Hashimoto et al. [4] proposed a supervision method based on the open text of Web pages, which uses semantic relations (between nouns), context, and correlation features to extract event causality such as degradation continues-global warming worsens and use its output to generate future scenarios. The causal sentence defined in this article refers to the sentence that describes the causal relationship. The causal sentence contains the cause part and the effect part (collectively referred to as the causal part). The causal part contains cause and effect events (collectively referred to as the causal event). A causal event is a semantic triple (*subj*, *pred*, *obj*) represented by a subject-predicate-object structure, where *subj* represents the subject of the event, *pred* represents the predicate of the event, and *obj* represents the object of the event.

Causal sentence recognition usually includes two sub-tasks: one is to design rules or templates to find candidate causal sentences; the other is to design a classifier to sort candidate causal sentences in order to finally identify the causal sentence. Hashimoto [5] studies the causal relationship between noun phrase pairs, and Mehwish and Roxana [6] study the causal relationship between verb phrase pairs. However, the causality of events in Chinese financial texts is expressed in sentence form. For the discovery of candidate causal sentences, most of the current related work is based on English texts. There are huge differences in the structure of causal sentence groups between Chinese and English. First, Chinese causality prompt words are abundant, especially verb causal prompt words. Second, Chinese financial texts often have multiple branches and complex sentence patterns. The author expresses that the occurrence of one because such an event will cause multiple effect events, or the occurrence of an effect event is the effect of multiple cause events. These all increase the difficulty of identifying Chinese candidate causal sentences. Therefore, this article combines causal prompt words with punctuation marks and conjunction to analyze the structure of Chinese sentence groups, which is not only conducive to determining the scope of causal sentences, but also helps to determine the boundary between the cause part and the effect part. For the classification of candidate causal sentences, many classifiers have been proposed, but language model pre-training has shown the effectiveness of learning common language representations using large amounts of unlabeled data. Representative models are ELMo (Embeddings from Language Models) [6], compared with the traditional word vector of Word2Vec, the model proposed in Peters et al. [7] is a dynamic model. In the previous word vector representation, words are in a static form, and the same vector is used in any context. In this case, it is difficult to express the polysemous phenomenon of a word, and ELMo can dynamically generate word vectors through the context. In theory, it will be a better model. From the actual test results, it has reached the time in many tasks. SOTA results. GPT (Improving language understanding by generative pre-training) [7], ULMFiT (Universal Language Model Fine-tuning) [8], and BERT [9, 10]. These neural network language models use unsupervised methods to train text data. For example, the BERT model is based on a multi-layer two-way transformer mechanism, and two tasks of next sentence prediction and mask prediction are added to the pre-training for training. In order to apply the pre-trained model to a specific task, we need to fine-tune it with the training data of the specific, and design additional task-specific layers after the pre-training. For example, in order to perform text classification tasks, BERT adds a simple softmax layer after the pre-trained model and can be fine-tuned in this way, so that better models can be created in text classification tasks based on specific data sets.

Through research, it is found that although the mask prediction mechanism of the BERT model can achieve the training of the two-way language model, it will lead to inconsistencies in the text representation during pre-training and fine-tuning on downstream tasks during sentence recognition. This paper proposes a combined model of BERT-CNN for the problem that text representation is inconsistent when pre-training and downstream task fine-tuning affect the classification result. The combined model has two advantages: One is the convolutional neural network CNN that is applied to text classification tasks, and multiple kernels of different sizes are used to extract key information in sentences (similar to n-grams with multiple window sizes), so as to better capture local correlations, The second is to input partial segment (phrase) information and BERT pre-training results into the transformer structure and use the self-attention mechanism to make the final text representation focus on the important part of the text. Pota et al. [11] recommended a Bidirectional Encoder Representations from Transformers (BERT)-based pipeline for Twitter sentiment analysis.

The approach used combines the knowledge embedded in pre-trained deep bidirectional transformer BERT [9] with Convolutional Neural Networks (CNN) for text [12], which is one of the most utilized approaches for text classification tasks. This combination of models has been shown to yield better results than using BERT or CNN on their own, as was shown in Li et al. [13], and shown in this paper.

The main contributions of this article include:

1) Aiming at the complex sentence patterns of multiple causes and one effect, multiple effects and one cause, and multiple causes and multiple effects in financial review texts, combined with causal prompt word expansion technology, Chinese lexical features and grammatical features, a relatively complete candidate causal sentence recognition rules are given, can recognize both simple causal sentences and complex causal sentences.

2) The combined model of BERT-CNN is proposed for the task of candidate causal sentence classification. The BERT-CNN model efficiently obtains the local segment information through the CNN structure on the specific task layer, and then inputs it into the transformer structure together with the BERT pre-training results, and uses the self-attention mechanism to make the final text representation focus on the importance of the text section.

Section 2 of this article introduces related research progress in candidate causal sentence recognition and candidate causal sentence classification. Section 3 describes the candidate causal sentence recognition rules in detail. Section 4 describes the candidate causal sentence classification model in detail. Section 5 describes the experimental results and analysis. Section 6 summarizes the full text.

2 Related work

2.1 Candidate causal sentences recognition

The key to discovering candidate causal sentences is how to obtain causal clues. Pattern matching is usually used to identify candidate causal sentences. Chung et al. [14] derive candidate causal sentence discovery rules on the basis of vocabulary and syntax. First, define words (or phrases) that can express causality, simple causal verbs (such as generate, trigger, make.), phrasal verbs (such as result in.), noun phrases with nouns + prepositions (such as cause of.), causal verb phrases in the passive voice (such as caused by, triggered by.), prepositions that can express causality (such as from, after.); then, based on the dependency syntax, four candidate causal sentence discovery rules are obtained.

Garcia [15] analyzed the language patterns and clues associated with causal verbs for French and used them as the basis for discovering candidate causal sentences. There are similar works: Khoo et al. [16] constructed a set of linguistic patterns that usually indicate the presence of a causal relation and used for the pattern matching. Khoo et al. [17] develop a method to identify and extract cause–effect information that is explicitly expressed in medical abstracts in the Medline database and constructed a set of graphical patterns to indicate the presence of a causal relation in sentences, and which part of the sentence represents the cause and which part represents the effect. Shen et al. [18, 19] put forward the concept of predicate meaning polarity. Predicate meaning polarity can be divided into positive, negative, and neutral. The positive direction means that an action increases or improves the object of an action (such as increase, preserve), negative direction means that a certain action can reduce its action object (such as reduce, reduce); use the predicate and its action object to form a polarity template (such as increase X, reduce Y), and assign a value to the template according to the polarity of the predicate, positive assignment 1, negative assignment −1, neutral assignment 0; by calculating the polarity values of the cause event predicate and the effect event predicate in the known causal sentence, 46 seed causal templates are obtained, and the seeds are used the causal template discovers new candidate causal sentences.

Radinsky et al. [3] use causal conjunctions such as "because," "as," or "after," causal prepositions such as "due to" or "because of," and causal verbs such as "cause," or "lead to." In order to constitute the template of causal prompt word, each type of template sets a certain constraint and a certain priority, <Pattern, Constraint, Priority>, the template of a candidate causal sentence is found based on this definition. When the candidate causal sentence matches multiple causal prompt word templates, the template with the highest priority is selected. There is also Yang et al. [20] that considers the priority of matching templates.

Jiang [21] studied the method of extracting the causal relationship of emergencies in Chinese text. Starting from the causal prompt words, combined with the part of speech of the prompt words, a syntactic pattern for identifying clear causal sentences and finding candidate causal sentences is designed, and at the same time, the rules for defining the cause part and the result part are given. Hashimoto et al. [22] used Wikipedia's parallel corpus to identify new causal phrase tags, created a training set through a remote supervision method and used open class tag features and contextual semantic features to train candidate causal sentence recognition classifiers. Ittoo and Bouma [23] acquire a set of explicit and implicit causal patterns from Wikipedia which as a knowledge base. Then, they use these patterns to extract causal relations from domain specific documents.

In Hidey and McKeown [24], in order to solve the problem of data sparseness caused by the lack of knowledge base in a specific field; firstly, many seed templates for finding candidate causal sentences were obtained from Wikipedia, and the k templates with the highest confidence were selected by calculating the template confidence to find the candidate causal sentence; then calculate the confidence of the causal entity pairs contained in the candidate causal sentence, and select the m causal entity pairs with the highest confidence; finally, match the selected causal entity pairs to the text on the Internet to obtain a new one candidate causal sentence discovery templates; through continuous iteration, more templates for discovering candidate causal sentences can be obtained, and the semantic pattern dictionary is expanded at the same time.

The above literature mainly studies the recognition patterns of English causal sentences. Due to the complexity of Chinese sentence patterns, there are multi-layer or nested sentence patterns, which brings some difficulties to the pattern extraction of causal sentence recognition. English causal sentence recognition model cannot be directly applied to Chinese causal sentence recognition. Therefore, this paper designs the corresponding recognition patterns according to the different hierarchical structures of Chinese causal sentences, so as to improve the recognition accuracy of causal sentences.

2.2 Candidate causal sentence classification

Candidate causal sentence classification is a typical text classification problem [25]. Compared with the problem of emotion classification, this paper uses the text two-classification method to deal with the problem of causal sentence discrimination. Text classification has methods based on traditional machine learning and methods based on deep learning [26–31]. The text classification method based on traditional machine learning splits the text classification problem into two parts: feature engineering and classifier. Typical traditional classifiers include Naïve Bayes, KNN (K-Nearest Neighbor) [32], SVM (Support Vector Machine), ME (maximum entropy) [33], and NN (neural networks) [34–41]. Traditional text classification needs to rely on many lexical and syntactic-related features. These features do not have strong versatility and often require an understanding of special tasks.

With the rapid development of deep learning technology, especially RNN (Recurrent Neural Network) [42] and CNN [43] have gradually been widely used in the field of NLP, making text classification tasks easier and increasing accuracy. The RNN recurrent neural network model includes two parts: LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Units), which is suitable for processing word sequence classification tasks. Several variants have also been proposed, such as Tree-LSTM [44] and TG-LSTM (Tag Weighted-Long Short-Term Memory) [45]. Priyadarshini and Cotton [46] propose a novel long short-term memory (LSTM)–convolutional neural networks (CNN)–grid search-based deep neural network model for sentiment analysis.

CNN is also one of the popular deep neural networks [47–50]. VDCNN (Very Deep Convolutional Neural Networks) [51] attempts to build a deeper text classification CNN model. DCNN (Dynamic Convolutional Neural Networks) [52] uses a dynamic *k*-max pool mechanism. The goal of DPCNN (Deep Pyramid Convolutional Neural Networks) [53] is to deepen CNNs without increasing computational cost. LK-MTL (Leaky Multi-task Learning) [54] is a multi-task convolutional neural network with leaky units, which filters the feature flow between tasks through memory and forgetting mechanisms. Different from the above methods, char-CNN [55] is a character-level model that encodes text characters as input.

Tang et al. [56] represent document with convolutional-gated recurrent neural network, which adaptively encodes semantics of sentences and their relations. The model first maps the text to a vector and then uses CNN/LSTM (CNN with 3 filters in the paper) for sentence representation.

Some studies combine CNN and RNN models to make full use of the advantages of both. C-LSTM [57] first uses CNN to capture the local information and then uses the LSTM network to encode each output of the convolution kernel to obtain global information. CNN-RNN [58] also uses a similar structure, but the difference between the two models is the way the CNN layer and the RNN layer. DRNN (Disconnected Recurrent Neural Networks) [59] uses RNN units instead of convolution kernels. In fact, it uses the structure of CNN and the encoding of RNN. Qian et al. [60] propose a self-attentive convolutional neural networks (SACNNs) trained on top of pre-trained word vectors for emotion detection on user-generated contents (UGC), which reserves different kinds of emotion information, avoids the loss of emotional aspects in the pooling process of CNNs structure and increases the interpretability of the model by visualizing the extraction process of features.

Recently, pre-trained models have achieved good results in multiple tasks of natural language processing. They usually learn common language representations by utilizing large amounts of unlabeled data and adopt other task-specific layers after pre-training modules for different tasks. ELMo [6] is dedicated to extracting context-related features from language models and has a good effect on problem processing including question answering systems [61], sentiment analysis [, 62, 63], and named entity recognition [, 63, 64]. GPT [7] and ULMFiT [8] first pre-trained some model structures and then fine-tuned the model for specific downstream tasks. BERT [9] is based on a multi-layer two-way transformer structure and performs two prediction tasks: next sentence prediction and mask prediction in the text. Jain et al. [65] proposed a Sparse Self-Attention Network (SpSAN) improves the fine-tuning performance of the Bidirectional Encoder Representations from Transformers (BERT) model by introducing sparsity into the self-attention procedure to predict Consumer Recommendation Decisions. It empowers us to understand sparse attention distribution with a more intelligible representation of the complete input data.

Studies have found that there are some causal prompt words in Chinese texts, such as "because," "caused," which can directly identify causal sentences with causal relationships and find out the cause or effect parts; however, some causal prompt words can only identify candidate causal sentences, and it is necessary to combine relevant background knowledge, semantics, and contextual analysis to determine whether they express a true causal relationship [44]. For this reason, further processing of candidate causal sentences is needed to identify causal sentences and non-causal sentences. This paper selects causal sentences from Chinese financial texts by using deep learning model.

3 Candidate causal sentence finding

Conneau et al. [12], Joulin et al. [66] pointed out that in Chinese causal sentences, causal sentences guided by single conjunctions or matching conjunctions usually express the true causal relationship. The sentences guided by verbs, prepositions or adverbs contain some sentences that do not have causality. It is necessary to judge whether they express real causality in combination with contextual semantic analysis. As shown in Table 1, the same connective "促成"("facilitates") leads to two sentences. Through contextual semantic analysis, it is concluded that example sentence 1 expresses true causal relationship, and example sentence 2 does not express true causal relationship.

3.1 Causal prompt words structure

Causal prompt words in Chinese mainly include conjunctions, prepositions, adverbs, verbs that can express causal relations, as well as phrases that express causal relations and super-words of super-grammatical units. The causal prompt words in this article mainly include three types.

(1) Causal prompt words such as conjunctions, prepositions, and adverbs. According to the research results of causal prompt words in Peng [67, 68], conjuncts, prepositions and adverbs, and other prompt words are summarized. Among them, there are two kinds of conjunctions: single conjunction and matching conjunction. Xing [69] divides the conjunction causal prompt words into three types:

Table 1 Candidate causal sentence

No	Example	Causal Sentence (Yes/No)
1	短期资本净流出重新小于基础国际收支顺差, [促成]外汇储备资产止跌回升。 Net short-term capital outflows were once again smaller than the underlying balance of payments surplus, [facilitating] the decline of foreign exchange reserve assets and the rebound	Yes, true causal relationship: (短期资本净流出(Net short-term capital outflows), 小于(smaller), 基础国际收支顺差(underlying balance of payments surplus))(外汇储备资产(foreign exchange reserve assets), 止跌回升(rebound),)
2	简单地上市或引入多个国有法人股东虽然可以[促成]股权多元化, 但达不到改善治理结构的目的。 Although simply listing or introducing multiple state-owned legal person shareholders can [promote] equity diversification, it fails to achieve the goal of improving the governance structure	No

conjunction cause prompt words, conjunction effect prompt words, and matching conjunction causal prompt words composed of conjunction cause prompt words and conjunction effect prompt words. Common prompting words for the cause of conjunctions include “because,” “because of,” “since,” “if,” “as long as,” and so on. Common prompting words for the effect of conjunctions include “so,” “therefore,” “from this,” and so on. There are two collocations for the matching conjunction causal prompt words: cause–effect matching conjunction causal prompt words and effect–cause matching conjunction causal prompt words. See Table 2 for details.

(2) Verbs causal prompt words. HowNet is a large language knowledge base marked by Mr. Dong Zhen dong and Mr. Dong Qiang, mainly for Chinese (including English) vocabulary and concepts [67]. HowNet adheres to the idea of reductionism and believes that vocabulary/word meaning can be described in smaller semantic units. This kind of semantic unit is called “Sememe,” as the name implies, it is atomic semantics, that is, the most basic and smallest semantic unit that should not be divided. In the process of continuous labeling, HowNet has gradually constructed a sophisticated system of Sememe (about 2000 Sememes). HowNet accumulatively annotates the semantic information of hundreds of thousands of vocabulary/words meaning based on the original system. The meaning of causality in HowNet is CauseEffect (causality) and MakeAct (make it move). This paper summarizes 47 seed verbs causal prompt words from the meaning of causality in HowNet. See Table 3 for details.

(3) Phrases and super words that express causality. In the analysis of financial corpus, this article found that some phrases or super words can also express causality. They are not a word in themselves, but some are habitual phrases, such as “the result” and “the reason.”

Table 2 Causal prompt words such as conjunctions, prepositions, and adverbs

Part of Speech	Causal Prompt Words
Single conjunction	**Conjunction cause prompt words:** 既[然]、因[为]、如果、由于、只要、因为、由于 already, if, because, as long as **Conjunction effect prompt words:** 因而、于是、所以、故、致使、以致[于]、因此、以至[于]、从而、因而 therefore, so, cause, therefore, even, thus
Matching conjunction	**Cause–effect matching conjunction causal prompt words:** (因为, 从而)、(因为, 为此)、(既[然], 所以)、(因为, 为此)、(由于, 为此)、(只有|除非, 才)、(由于, 以至[于])、(既[然], 却)、(如果, 那么|则)、(由于, 从而)、(既[然], 就)、(既[然], 因此)、(如果, 就)、(只要, 就)、(因为, 所以)、(由于, 于是)、(因为, 因此)、(由于, 故)、(因为, 以致[于])、(因为, 因而)、(由于, 因此)、(因为, 于是)、(由于, 致使)、(因为, 致使)、(由于, 以致[于])、(因为, 故)、(因[为], 以至[于])、(由于, 所以)、(因为, 故而)、(由于, 因而) (because, thus), (because, for this), (because of, therefore), (because, for this), (only | unless, only), (because, even [in]), (already, but), (if, then), (because, thus), (already, just), (already, therefore), (if, just), (if, so), (because, so), (because, therefore), (because, so [to]), (because, then), (because, cause), (because, even) **Effect–cause matching conjunction causal prompt words:** ([之]所以, 因为)、([之]所以, 由于)、([之]所以, 缘于) (Therefore, because), (Therefore, due to)
Prepositions	为了、依据、为、按照、因[为]、按、依赖、照、凭借、由于 for, base on, in accordance, rely on, take, because
Adverbs	以免、以便、为此、才 lest, so, for this, only

Due to differences in Chinese and English, duplicates have been deleted.

Example 3: “如果加息预期变成现实，势必形成货币、信贷下降共振，其结果是同时抑制投资和消费需求。”(“ If the expectation of rising interest rates

Table 3 Verbs causal prompt words

Part of Speech	Causal Prompt Words
Verb	牵动, 导向, 使动, 导致, 勾起, 指引, 使, 予以, 产生, 促成, 造成, 造就, 促使, 酿成, 引发, 渗透, 促进, 引起, 诱导, 引来, 促发, 引致, 诱发, 推进, 诱致, 推动, 招致, 影响, 致使, 滋生, 归于, 作用, 使得, 决定, 攸关, 令人, 引出, 浸染, 带来, 挟带, 触发, 关系, 渗入, 诱惑, 波及, 诱使, 满足involve, guide, cause, evoke, give, produce, bring up, infiltrate, promote, induce, advancing, inducing, pushing, influencing, causing, breeding, attributing, acting, making, determining, critical, eliciting, infiltrating, bringing, entraining, triggering, relationship, infiltration, temptation, spreading, satisfy

Due to differences in Chinese and English, duplicates have been deleted becomes a reality, it will inevitably form a resonance of the decline in currency and credit, and the result will be to suppress investment and consumer demand at the same time.”). Example 3 is a candidate causal sentence guided by the idiom “its result.” The pair of causal events: (expectation of rising interest rates, becomes, reality) (, suppress, investment and consumer demand).

Some are super words that cross grammatical units, such as “由此可见/由此可知/由此看来/由此得出结论” “为什么……原因” “为什么……是因为……”。(“it can be seen/known from this/seen from this/concluded from this”, “why…reason”, “why…because…”.)

3.2 Verb causal prompt word set expansion

The core of causal sentence recognition is to find more causal prompt words and use causal prompt words to find more candidate causal sentences. This paper is based on the seed verb causal prompt words and uses the Gensim toolkit to train a word2vec model based on the Wikipedia Chinese corpus. In this model, the first 3 verbs (verb phrases) with the most similar meanings of the seed verb causal prompt words are found as new verb causal prompt words are expanded to the verb causal prompt word set. After multiple rounds of expansion operations, the verb causal prompt word set does not change much, and a relatively complete verb causal prompt word set is finally formed. The appendix contains an expanded set of verb causal prompt words, with 170 verbs (verb phrases) in total.

3.3 Finding rules for candidate causal sentences

Candidate causal sentences in financial texts often appear in the form of complex sentences, with multi-cause-one-effect, one-cause-multi-effect, multi-cause-multi-effect, and one-cause-one-effect multi-branch structures. According to the characteristics and patterns of causal sentence description in financial texts, the causal prompt words, syntactic features, cause parts, and effect parts are considered when designing candidate causal sentence discovery rules, and two types of discovery rules are designed.

(1) Candidate causal sentences with branch structure.

Rule1 multi-cause-one-effect. < conjunction cause prompt words ><cause part 1>< conjunction cause prompt words ><cause part 2>[,|<conjunction effect prompt word>]<effect part>.

Note: Generally, two conjunction cause prompt words are used together. After they are used together, they can be combined with the conjunction effect prompt words, and the conjunction effect prompt words are sometimes omitted. The sentence expression is that two cause events lead to a result event.

Example 4: 因为经济潜在增长速度可能会要下降，因为人口年龄结构的变化、劳动供给增长率下降,所以我们对未来投资报酬有所担心。(Because the potential economic growth rate may decline, because the age structure of the population changes and the labor supply growth rate declines, so we are worried about future returns on investment.)

Analysis: Example 4 is a candidate causal sentence with two causes and one effect guided by “因为-因为-所以”(“because-because-so”). What the author expresses is: ((经济潜在增长速度，下降，)∪(劳动供给增长率，下降，))(我们，担心，未来投资报酬).((potential economic growth rate, decline,)∪(labor supply growth rate, decline,)) (we, worry, future return on investment).

Rule2 one-cause-multi-effect. < conjunction cause prompt words ><cause part> [, |<verb causal prompt word>|<conjunction effect prompt word >] <effect part 1> [, |<conjunction effect prompt word>] < effect part 2>.

Note: Usually two effect prompt words are used together or the conjunction effect prompt word is used in conjunction with the verb causal prompt word. After the combined use, it can be used with the conjunction cause prompt word, and the conjunction effect prompt word is sometimes omitted. The sentence expression is that one cause event leads to the occurrence of two effect events.

Example 5: “由于原材料价格上升导致企业生产成本上升, 从而影响了企业毛利。”(The increase in the price of raw materials has led to an increase in the production cost of the company, which has affected the company’s gross profit.)

Analysis: Example 5 is a candidate causal sentence with one cause and two effects guided by “由于-导致-从而”(“because-cause-thus”). What the author expresses is: (原材料价格，上升，)((企业生产成本，上升，)∪(，影响，企业毛利)). (the price of raw materials, increase,) ((the production cost of the company, increase,)∪(, affect, the company’s gross profit)).

Rule3 multi-cause-multi-effect. < conjunction cause prompt words ><cause part 1>,<conjunction effect prompt word><effect part 1>[,|;]<conjunction cause prompt word><cause part 2>,<conjunction effect prompt words><effect part 2>.

Note: Generally, several candidate causal sentences are used together to describe the same topic, and there is a parallel relationship between several candidate causal sentences.

Example 6: 因为政策扶持部门所占的资源越来越多,而市场部门能获得的资源就少了, 因为大量资本用在了政策扶持部门,所以市场部门的实际利率就上升了。(Because the policy support department accounts for more and more resources, the market department can obtain fewer resources. Because a large amount of capital is used in the policy support department, the real interest rate of the market department rises.)

Analysis: Example 6 is a candidate causal sentence of two causes and two effects guided by the two pairs of causal prompt words “因为-而”“因为-所以”(“because-and” and “because-so”). What the author expresses is: ((policy support department, accounted for, more and more resources) (market department, obtain, fewer resources))∪((a large amount of capital, used in, a policy support department) (market department, real interest rate, rise)).

Rule4 one-cause-one-effect. [<preposition causal prompt word>|<conjunction cause prompt word>|NULL] <cause part> [, |<conjunction effect prompt word>|<verb causal prompt word>|<adverb causal prompt word >] <effect part>.

Note: Generally, a preposition causal prompt word or a conjunction cause prompt word comes in front, or a verb causal prompt word, an adverb causal prompt word is centered, or a pair of matching conjunction causal prompt words connect a candidate causal sentence. The sentence expression is that a cause event leads to an effect event and emphasizes the importance of the effect.

Example 7: 新技术降低了运输和通讯成本, 促进了国际贸易。(New technologies have reduced transportation and communication costs and promoted international trade.)

Analysis: Example 7 is a candidate causal sentence guided by“促进”(“promote”). What the author expresses: (new technology, reduce, transportation and communication costs) (, promote, international trade).

Rule5 one-effect-one-cause. <conjunction effect prompt word><effect part>, <conjunction cause prompt word><cause part>

Note: The effect part comes before the cause part. This sentence expresses that an effect event is caused by the occurrence of a cause event, which emphasizes the importance of the cause.

(2) Candidate causal sentences with idiomatic phrases or super word structures

Rule6 [<cause part>|<effect part>] idiomatic phrases or super word [<effect part>|<cause part>]

Note: Due to differences in habitual phrases or super words, there are two different structures: antecedents and effects, or antecedents and effects.

Example 8: 为什么我们的互助基金发展了13年, 被认为是成功的, 其中一个很重要的原因就是所有权归私人所有。(One of the most important reasons why our mutual fund has been developed for 13 years and is considered to be successful is that the ownership is privately owned.)

Analysis: Example 8 is a candidate causal sentence guided by the super word “为什么-原因”(“why-cause”). What the author expresses is: (所有权, 归, 私人所有)(互助基金, 是, 成功)。((ownership, is, privately owned) (mutual fund, is, successful)).

4 Candidate causality classification

This paper proposes a combined classification model that adds a CNN model to the basic BERT model: BERT-CNN model, as shown in Fig. 1. The BERT-CNN model has two characteristics: one is to use CNN to transform the specific task layer of BERT to obtain the local feature representation of the text; the other is to input the local features and output category C into the transformer after the CNN layer in the encoder. This facilitates the use of self-attention mechanisms to capture important segments (phrases) of the text.

1) Input Encoder

The input of the BERT-CNN model is N tokens, $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_N\}$. The first token is [CLS] as the start character, and the last token is [SEP] as the end character of the sentence. Like BERT [55], the initial input text is encoded as the sum of the three vectors of token, segment, and position.

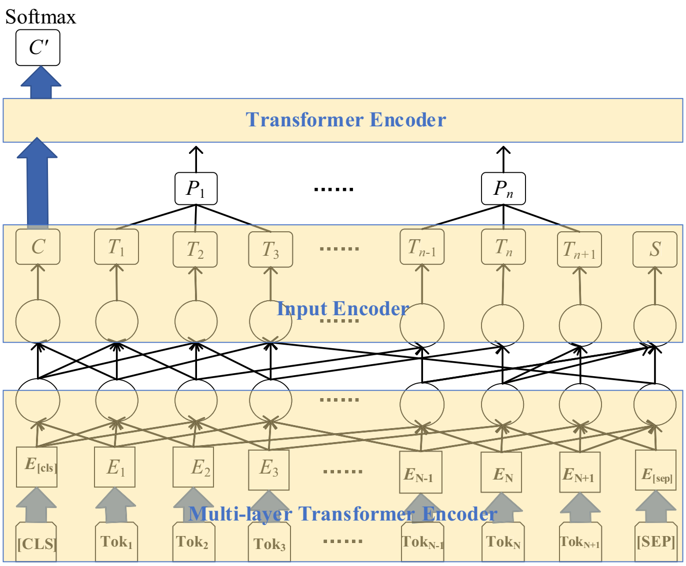

Fig. 1 BERT-CNN

2) Multi-layer Transformer Encoder

We use a multi-layer two-way transformer mechanism to map the input representation into a sequence of context embedding vectors $\mathbf{C} = \{\mathbf{c}, \mathbf{T}, \mathbf{s}\}, \mathbf{C} \in \mathbb{R}^{d*(N+2)}$. Where $\boldsymbol{c}$ stands for [CLS] and $\boldsymbol{s}$ stands for [SEP]. $\mathbf{T} = \left\{\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_N\right\}$ Represents the contextual representation of the text.

3) Local CNN Encoder

To make the text performance focus on the local information in the text, such as short sentences or phrases. We set the window of the convolution kernel to $k*1$, then the local CNN encoding output vector is expressed as:

$$\mathbf{P} = CNN_k(\mathbf{T}), \mathbf{P} \in \mathbb{R}^{d*(N-k+1)} \tag{1}$$

In: $\mathbf{P}=\{\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_n\}, n = N - k + 1$.

Through the convolution operation, a phrase expression with a window size of k is obtained, like $\mathbf{P}_i = (\mathbf{T}_i, \mathbf{T}_{i+1}, \ldots, \mathbf{T}_{i+k-1})$.

4) Transformer Encoder

Similar to multi-layer Transformer encoding, in order to integrate the text information in the vector ***P***, we use the Transformer encoder to map the local text representation vector ***P*** to the entire text representation. The Transformer encoding input is $\{\mathbf{C}, \mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_n\}$, and C' corresponds to the entire text representation. Obviously, through the Transformer encoder, ***C***' obtains independent token information from ***C*** and partial information of the text from ***P***.

5) Output Layer

The *softmax* function of the output layer is:

$$\hat{\mathbf{y}} = soft\max\left(\tanh(\mathbf{W}\mathbf{C}' + \mathbf{b})\right) \tag{2}$$

In: ***W*** is the weight matrix, and ***b*** is the bias value, $\mathbf{y}_i, \hat{\mathbf{y}}_i$ which is the true value and the predicted value, respectively.

Using the cross-entropy loss function:

$$L = \sum_i \mathbf{y}_i \log \hat{\mathbf{y}}_i \tag{3}$$

5 Experimental results and analysis

5.1 Data set

In order to verify the actual effect of the BERT-CNN model on text classification tasks, a binary classification experiment was carried out on the candidate causal sentence corpus and a multi-topic classification experiment on 10 topics was carried out on the THUCNews[1] corpus.

1) Candidate causal sentence corpus.[2] In this study, a total of 2,132 news reviews from 2001 to 2019 were crawled from the "China Economic Forum" website to construct a corpus of Chinese financial reviews. The corpus contains 366,383 sentences with a total of 1,465,540 words. Most of the sentences are between 60 and 80 words in length. Among them, there are 6,998 sentences with a true causal relationship, and 22,415 candidate causal sentences that need further classification.
2) THUCNews corpus. THUCNews Corpus is a Chinese corpus launched by the Natural Language Processing Laboratory of Tsinghua University. The corpus is generated by filtering and filtering historical data from the Sina News RSS subscription channel from 2005 to 2011. It contains 740,000 news documents (2.19 GB), all in UTF-8 plain text format. Based on the original Sina News classification system, 14 candidate classification categories were re-integrated:

[1] http://thuctc.thunlp.org/

[2] https://pan.baidu.com/s/1oLUFAIJeeiF9HLBC97HHZA password: 1111.

finance, lottery, real estate, stocks, home furnishing, education, technology, society, fashion, current affairs, sports, constellations, games, and entertainment. In this experiment, 200,000 news headlines were extracted from the THUCNews corpus, and the text length was between 20 and 30 words. There are 10 categories in total, including finance, real estate, stocks, education, technology, society, current affairs, sports, games, and entertainment. 20,000 news in each category.

5.2 Data preprocessing

Use the LTP-Cloud[3] to preprocess the experimental data such as word segmentation, part-of-speech tagging, and dependency syntax analysis. The experiment uses the Pytorch architecture, using Nivida 1070i, GPU with 8G video memory to classify candidate causal sentences.

The input of the BERT-CNN model is based on sentence-level, and the text content is segmented in the preprocessing part. In the experiment, the punctuation marks such as "?", "!", and "." appearing in the text are used as segmentation identifiers to segment the text content. The distribution of sentence length (including the number of sentences) in the training set and the test set after each article is segmented in the corpus is roughly the same. The length of most sentences in the candidate causal sentence corpus is concentrated between 80 and 100 (THUCNews corpus sentence length set (Between 20 and 30), so the sentence length of the candidate causal sentence corpus is fixed to 100 during preprocessing (the sentence length of the THUCNews corpus is fixed to 30), and sentences with a length less than a fixed value are marked with "[PAD]" Fill in, and the sentences longer than the fixed value will be truncated.

5.3 Evaluating indicator

In order to select the standard data set, this paper selects 3 undergraduates with such experience in accounting and software engineering basic courses as the annotators. A voting mechanism in which the minority obeys the majority is used to determine the final result. The main evaluation indicators of the classifier are accuracy rate *ACC* (Accuracy), precision rate P (Precision), recall rate R (Recall), and $F1$ value (F-score). Accuracy refers to the proportion of correctly classified samples to the total number of samples. Precision refers to the proportion of correctly classified positive samples to the number of positive samples predicted by the classifier. Recall rate refers to the proportion of correctly classified positive samples to true positive samples. The ratio of the number of samples. In order to balance the relationship between the precision rate and the recall rate, the $F1$ value (that is, the harmonic average) is usually introduced as the evaluation index of the classification in the classification task, and the calculation formula is shown in formula (4–7).

$$\mathrm{ACC} = \frac{\mathrm{t}rue\text{ - }positive + \mathrm{t}ure\text{ - }\mathrm{ne}gtive}{positive + negtive} \tag{4}$$

[3] https://www.ltp-cloud.com/.

Table 4 BERT-CNN model parameters

Parameters	Value
Vector Dimensions	768
Text length	100
Size of convolution kernel	(2,3,4)
Number of convolution kernels	256
learning rate	2.5e-5
Dropout	0.1
Epoch	5
Warm-up	0.05

$$\mathrm{p} = \frac{true\text{ - }positive}{ture\text{ - }positive + false\text{ - }positive} \tag{5}$$

$$R = \frac{true\text{ - positive}}{true\text{ - positive} + false\text{ - n}egative} \tag{6}$$

$$F1 = \frac{2P \times R}{P + R} \tag{7}$$

5.4 Parameter setting

Our implementation of BERT-CNN is based on the Pytorch implementation of BERT (https://github.com/huggingface/transformers). Due to the limited computational performance, BERT-base model contains an encoder with 12 Transformer blocks, 12 self-attention heads, and the hidden size of 768. BERT takes an input of a sequence of no more than 512 tokens and outputs the representation of the sequence. The sequence has one or two segments that the first token of the sequence is always [CLS] which contains the special classification embedding and another special token [SEP] is used for separating segments.

We further pre-train with BERT on 1 NVIDIA GeForce GTX 1070 Ti GPU, with a batch size of 32, max sequence length of 100, learning rate of 2.5e-5, Size of convolution kernel is (2,3,4), Number of convolution kernels 256, train steps of 100,000 and warm-up steps of 10,000. The parameters of the BERT-CNN network model involved in the experiment are shown in Table 4. Use tanh as the activation function of the convolutional layer. The experiment uses the data set introduced in Sect. 5.1. In the two-classification task of financial text candidate causal sentences, there are 22,415 sentences in the training set, 3000 sentences in the test set, and 3000 sentences in the verification set.

5.5 Experimental design and results analysis

5.5.1 Finding candidate causal sentences

In this paper, the frequency of each candidate causal sentence discovery rule appears through statistics of the financial review corpus, as shown in Table 5. To avoid

Table 5 Frequency and coverage of candidate causality sentence recognition rules

Rule	Type	Frequency
Rule1	Multi-cause-one-effect	186
Rule2	One-cause-multi-effect	238
Rule3	Multi-cause-multi-effect	18
Rule4	One-cause-one-effect	28,876
Rule5	One-effect-one-cause	26
Rule6	idiomatic phrases or super word structures	69

Table 6 Findings of candidate causal sentences

Ruleset	P	R	$F1$
Jiang [21]	82.36%	91.87%	86.86%
Ittoo and Bouma [23]	68.49%	85.46%	76.04%
Ours	85.76%	93.64%	89.53%

conflicts between rules and improve the matching efficiency of candidate causal sentences, each sentence is matched in sequence from low to high frequency of the rules, and a total of 29,413 candidate causal sentences are identified.

Jiang [21] uses Chinese conjunctions causal prompt words, verb causal prompt words, preposition causal prompt words, and adverb causal prompt words to summarize 8 candidate causal sentence recognition rules, but only discusses the recognition of simple causal sentence candidates for one cause and one effect. Ittoo and Bouma [23] summarized and sorted out the candidate causal sentence discovery rules composed of 80 verbs, prepositions, or conjunctions, such as cause (by, of), lead to, result in, and affect (with, by). In the experiment, the rules of Ittoo and Bouma [23] are translated into corresponding Chinese causal rules for comparative experiments.

In the experiment, we compare the rules of Jiang [21] and Ittoo and Bouma [23] with the rules of this article on the task of candidate causal sentence recognition. The experiment selects 3000 sentences, and the results of candidate causal sentence discovery are shown in Table 6.

It can be seen from Table 6 that in the design of candidate causal sentence discovery rules, this paper fully considers the characteristics of multi-cause-one-effect, one-cause-multi-effect, multi-cause-multi-effect, and one-cause-one-effect multi-branch structure of Chinese financial texts, making the rules the recall rate is greatly improved compared with the existing research. In the experiment, we compared the number of candidate causal sentences found before and after the expansion of the verb causal prompt word set on Rule 2 and Rule 4. In Rule 2, there are 237 candidate causal sentences found in verb causal prompt words, and in Rule 4, there are 18,950 candidate causal sentences found in verb causal prompt words. The experimental results are shown in Table 7.

Table 7 Comparison of the results of extended verb causal prompt word set

Rule	Type	Number of candidate causal sentences found in the seed verb causal suggestion word set	Number of candidate causal sentences found by expanding the set of causal verbs	Growth rate
Rule2	One-cause-multi-effect	115	237	106.08%
Rule4	One-cause-one-effect	8350	18,950	126.95%

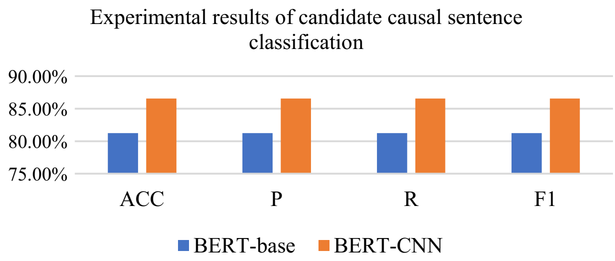

Fig. 2 Experimental results of candidate causal sentence classification

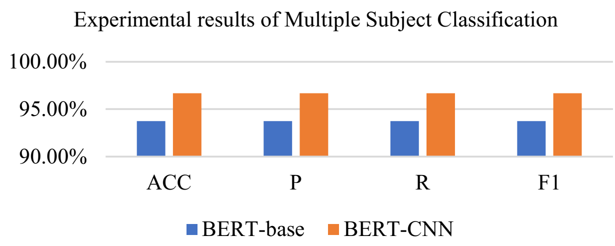

Fig. 3 Experimental Results of Multiple Subject Classification

The experimental results show that the expanded verb causal suggestion word set has doubled the number of candidate causal sentences found before expansion, indicating that the effect of expanding the verb causal suggestion is more obvious.

5.5.2 Candidate causal sentence two classification

Based on the financial text corpus, the experiment conducted a comparative analysis of the two-classification effects of the BERT-base and BERT-CNN models. In order to maintain the fairness of the comparison, the experiment sets the epoch of each model to 5, the learning_rate to 2.5e-5, and the warm-up to 0.05. The experimental results are shown in Fig. 2.

From the experimental results, because the BERT-CNN model first uses CNN to transform the specific task layer of BERT to obtain the local feature representation of the text; then, after the CNN layer, the local feature and the output category C are input into the transformer encoder. This is conducive to using the self-attention mechanism to capture important segments (phrases) of the text, so the accuracy, recall, and F1 value are greatly improved compared with the BERT-base model.

5.5.3 Multi-topic classification

Based on the THUCNews corpus, we did a topic classification experiment on 10 topics. The experiment compares and analyzes the multi-topic classification effects of BERT-base and BERT-CNN models. In order to maintain the fairness of the comparison, the experiment sets the epoch of each model to 5, learn rate to 2.5e-5, and warm-up to 0.05. The experimental results are shown in Fig. 3.

From the experimental results, compared with the BERT-base model, the $F1$ value of the BERT-CNN model in the multi-topic classification task increased by 2.93 percentage points. This proves the effectiveness of the BERT-CNN model on classification tasks.

Since the average length of sentences in the THUCNews corpus is shorter than the average length of sentences in the candidate causal sentence corpus, the experimental results show that the classification accuracy of short texts is higher than that of long texts, indicating that the semantic features of short texts are easier to obtain and are more beneficial for accurate classification.

6 Conclusion

Because the sentence structure of the causal sentence in financial texts is relatively complex, there are characteristics of multi-cause-one-effect, multi-effect-one-cause, and multi-cause-multi-effect. When designing candidate causal sentence recognition rules, programs should consider causal prompt words, syntactic features and punctuation, cause part and effect part. In this paper, 47 seed verb causal prompt words are selected in the task of candidate causal sentence recognition, and the semantic similarity is used to expand the set of verb causal prompt words to provide more identification clues for discovering new candidate causal sentences.

Aiming at solving the problem that the mask prediction mechanism of the BERT model causes inconsistent text representation during pre-training and downstream task fine-tuning, which affects the classification effect, this paper proposes a BERT-CNN combined model. The combined model has two advantages: One is that the convolutional neural network CNN is applied to text classification tasks, and multiple kernels of different sizes are used to extract key information in sentences (similar to n-grams with multiple windows sizes), to better capture local correlations. The second is to input partial segment (phrase) information and BERT pre-training results into the transformer structure, and use the self-attention mechanism to make the final text representation focus on the important part of the text. Experiments on two text classification data sets show that these two features enable our model to obtain competitive results.

The rule-based causal sentence recognition needs to constantly revise the rules, which makes the method proposed in this paper difficult to meet the requirements of all causal sentence recognition. It is necessary to adjust the recognition rules for different corpora.

When using LTP-cloud for dependency parsing, it is found that the parsing effect of multi-part of speech words is poor. The error of part of speech tagging is transferred to the result of syntactic analysis, resulting in inaccurate recognition of causal trigger words, which affects the classification effect.

In future research, it is also necessary to continuously maintain the candidate causal sentence recognition rules, and at the same time improve the effect of the candidate causal sentence classification algorithm.

Appendix

See Table 8

Table 8 Verbs causal prompt words

Seed verb causal prompt word set (47 words)	牵动, 导向, 使动, 导致, 勾起, 指引, 使, 予以, 产生, 促成, 造成, 造就, 促使, 酿成, 引发, 渗透, 促进, 引起, 诱导, 引来, 促发, 引致, 诱发, 推进, 诱致, 推动, 招致, 影响, 致使, 滋生, 归于, 作用, 使得, 决定, 攸关, 令人, 引出, 浸染, 带来, 挟带, 触发, 关系, 渗入, 诱惑, 波及, 诱使, 满足 Involve, guide, cause, evoke, guide, give, produce, bring up, cause, infiltrate, promote, induce, advancing, inducing, pushing, influencing, causing, breeding, attributing, acting, making, determining, critical, eliciting, infiltrating, bringing, entraining, triggering, relationship, infiltration, temptation, spreading, Satisfy
Extended verb causal cue vocabulary (170 words)	导致, 保证, 汹涌, 惹起, 未加, 促发, 唤起, 增添, 着实, 诱导, 酿, 夹杂, 形成, 扎实推进, 关系, 作用, 渗透, 催生出, 孕育, 刺痛, 夹带, 确保, 指引, 更让人, 蛊惑, 加进, 感同身受, 激起, 积极作用, 可能引发, 促成, 对立, 增进, 孕育出, 使得, 事关, 培育出, 促进, 取向, 并使, 偶发, 适配, 促进作用, 一致同意, 加深, 所带, 唤醒, 教唆, 诱致, 大力推进, 导至, 令人, 不予, 巨大作用, 会发生, 点燃, 溶入, 指南, 驱使, 矛盾激化, 下决心, 使, 引起, 唆使, 点出, 关乎, 以致, 推进, 介导, 以致于, 示范作用, 挟带, 落脚点, 引出, 出发点, 以至于, 满足, 会带来, 招致, 酿成, 消极影响, 影响, 催生, 助推, 抑止, 促使, 负面效应, 溶化, 诱发, 诱骗, 萌发, 诱使, 帮凶, 致使, 渗入, 重压, 扩散, 滋长, 决定, 导向, 推动, 重创, 提示, 突发, 加以, 浸染, 惹来, 波及, 逼使, 予以, 引诱, 颤抖, 矛盾, 决意, 压制, 大力发展, 带来, 重挫, 未予, 保障, 祸及, 攸关, 加之, 造成, 重击, 造就, 构成, 牵涉到, 以使, 加添, 归到, 产生, 归入, 迫使, 腐化, 滋生, 引发, 遭致, 平添, 触发, 助力, 突发性, 引导出, 诱惑, 带出, 引爆, 浸润, 滋养, 牵动, 引致, 归为, 熔化, 蔓延, 殃及, 提议, 冲击, 凝固, 下定决心, 引来, 深化, 勾起, 满足用户, 激酶, 怂恿, 阻遏, 归于, 招来, 触动, 表示同意, 包含 cause, guarantee, turbulent, arouse, not add, induce, add, actually, brew, mix, form, solid advance, relationship, effect, infiltrate, spawn, breed, tingling, entrainment, ensure, guide, more people, confuse, empathize, positively affect, may cause, promote, oppose, enhance, nurture, make, matter, orient, and make, occasionally, adapt, agree, deepen, bring, awaken, instigate, vigorously advance, lead, deny, great effect, will happen, ignite, dissolve, drive, intensify contradictions, make up your mind, to cause, point out, relate to, so as to advance, mediate, so as to, demonstration effect, foothold, elicitation, starting point, so that satisfaction, will bring, negative influence, influence, boost, restrain, urge, negative effect, melt, deceive, sprout, accomplice, press, spread, grow, determine, push, hurt, prompt, sudden send, infect, force, give, tremble, contradict, resolve, suppress, vigorously develop, frustrate, fail, protect, harm, affect, strike, bring up, constitute, involve, so that, return, produce, fall into, corrupt, trigger, assist, sudden, lead out, tempt, bring out, detonate, nourish, be attributed, propose, impact, solidify, attract, satisfy the user, kinase, deter, attribute, Invite, touch, include

Due to differences in Chinese and English, duplicates have been deleted.

Acknowledgements This work was sponsored by the National Natural Science Foundation of China [grant number 61972184, 61562032, 61662027, 61762042], Postgraduate Innovation Special Foundation of Jiangxi [grant number YC2017-B065], Modern Agricultural Research Collaborative Innovation Project of Jiangxi [grant number JXXTCXQN201906], and Key R & D Project of Jiangxi Science and Technology Department [grant number 2019ACB60016].

References

1. Tan PN, Steinbach M, Kumar V (2005) Introduction to Data Mining, (First Edition). Addison-Wesley Longman Publishing Co. Inc.
2. Hashimoto C, Torisawa K, De Saeger S, et al (2012). Excitatory or inhibitory: A new semantic orientation extracts contradiction and causality from the web. In: Proc of the Conf on Empirical Methods in Natural Language Processing and Natural Language Learning (ACL), pp. 619–630.
3. Radinsky K, Davidovich S, Markovitch S (2012). Learning causality for news events prediction. In: Proc of the 21st International Conference on World Wide Web (ACM), pp. 909–918.
4. Hashimoto C, Torisawa K, Kloetzer J, et al (2014). Toward future scenario generation: Extracting event causality exploiting semantic relation, context, and association features. In: Proc of the Annual Meeting of the Association for Computational Linguistics (ACL), pp. 987–997.
5. Hashimoto C (2019). Weakly supervised multilingual causality extraction from Wikipedia. In: Proc of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 2988–2999
6. Mehwish R, Roxana G (2010) Another look at causality: Discovering scenario-specific contingency relationships with no supervision. In: Proc of the 4th IEEE International Conference on Semantic Computing (ICSC), pp. 361–68.
7. Peters ME, Neumann M, Lyyer M, et al (2018) Deep contextualized word representations. arXiv preprint http://arxiv.org/abs/:1802.05365
8. Radford A, Narasimhan K, Salimans T, et al (2018) Improving language understanding by generative pre-training. https://s3-us-west-2.amazonaws.com.
9. Howard J, Ruder S (2018) Universal language model fine-tuning for text classification. arXiv preprint http://arxiv.org/abs/:1801.06146.
10. Devlin J, Wei Chang M, Lee K, Toutanova K (2018) Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint http://arxiv.org/abs/:1810.04805.
11. Pota M, Ventura M, Catelli, R et al. (2020). An effective bert-based pipeline for twitter sentiment analysis: a case study in Italian. Sensors, pp. 21(1), 133.
12. Conneau A, Schwenk H, Barrault L, et al (2016) Very deep convolutional networks for text classification. arXiv preprint http://arxiv.org/abs/1606.01781.
13. Li W, Gao SB, Zhou H, et al (2019) The automatic text classification method based on bert and feature union. In: Proc of the IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS), pp. 774–777.
14. Chung J, Gulcehre C, Cho K H, et al (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint http://arxiv.org/abs: 1412.3555.
15. Garcia D (1997) COATIS: An NLP system to locate expressions of actions connected by causality links. Proc of the 10th European workshop on knowledge acquisition. Springer, Modeling and Management, pp 347–352
16. Khoo CSG, Kornfilt J, Oddy RN et al (1998) Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. Literary Linguist Comput 13(4):177–186
17. Khoo CS, Chan S, Niu Y (2000). Extracting causal knowledge from a medical database using graphical patterns In: Proc of the Annual Meeting of the Association for Computational Linguistics (ACL), pp. 336–343.
18. Shen T, Zhou TY, Long G D (2018). DiSAN: directional self-attention network for RNN/CNN-free language understanding. In: Proc of the Thirty-Second AAAI Conference on Artificial Intelligence, pp. 1015–1025.
19. Shen T, Zhou T Y, Long G D, et al (2018) Bi-directional block self-attention for fast and memory-efficient sequence modeling. arXiv preprint http://arxiv.org/abs/1804.00857

20. Yang Z, Yang D, Dyer C, et al. (2017). Hierarchical attention networks for document classification. In: Proc of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp.1069–1079.
21. Jiang NQ (2012). Research on causality extraction method of emergencies in Chinese text (Dissertation). Dalian University of Technology.
22. Hashimoto C, Torisawa K, Kloetzer J, et al (2015). Generating event causality hypotheses through semantic relations In: Proc of the AAAI Conf on Artificial Intelligence (AAAI), pp. 2396–2403.
23. Ittoo A, Bouma G (2011) Extracting explicit and implicit causal relations from sparse. Lect Notes Comput Sci, Nat Lang Proc Inform Sys 6716:52–63
24. Hidey C, McKeown K (2016) Identifying causal relations using parallel Wikipedia articles. In: Proc of the Annual Meeting of the Association for Computational Linguistics (ACL), pp.1424–1433.
25. Kayesh H, Islam MS, Wang J, et al (2019). Event causality detection in tweets by context word extension and neural networks. In: Proc of the 20th Parallel and Distributed Computing: Applications and Technologies (IEEE): pp. 355–360.
26. Huang G, Xu M, Lin X et al (2017) Shuffle dog: characterizing and adapting user-perceived latency of android apps. IEEE Trans Mob Comput 16(10):2913–2926
27. Zhang Y, Huang G, Liu X, et al (2012). Refactoring android Java code for on-demand computation offloading. ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages, and Applications.
28. Song H, Huang G, Chauvel F et al (2011) Supporting runtime software architecture: a bidirectional-transformation-based approach. J Syst Softw 84(5):711–723
29. Chen CM, Chen L, Gan W et al (2021) Discovering high utility-occupancy patterns from uncertain data. Inf Sci 546:1208–1229
30. Chen CM, Huang Y, Wang KH, et al (2020) A secure authenticated and key exchange scheme for fog computing. Enterprise Information Systems, 1–16.
31. Quamer W et al (2021) SACNN: self-attentive convolutional neural network model for natural language inference. Trans Asian Low-Resour Lang Inform Proc 20(3):1–16
32. Huang G, Liu T, Mei H, et al (2004) Towards Autonomic Computing Middleware via Reflection. International Computer Software and Applications Conference.
33. Huang G, Luo C, Wu K, et al (2019) Software-Defined Infrastructure for Decentralized Data Lifecycle Governance: Principled Design and Open Challenges. IEEE International Conference on Distributed Computing Systems.
34. Chen X, Li M, Zhong H et al (2021) DNNOff: offloading DNN-based intelligent IoT applications in mobile edge computing. IEEE Trans Ind Inform, Publish Online,. https://doi.org/10.1109/TII.2021.3075464
35. Chen X, Chen S, Ma Y et al (2019) An adaptive offloading framework for android applications in mobile edge computing. SCI CHINA Inf Sci 62(8):82102
36. Lin B, Huang Y, Zhang J et al (2020) Cost-driven offloading for dnn-based applications over cloud, edge and end devices. IEEE Trans Industr Inf 16(8):5456–5466
37. Chen X, Zhu F, Chen Z et al (2021) Resource allocation for cloud-based software services using prediction-enabled feedback control with reinforcement learning. IEEE Trans Cloud Compu, Publish Online,. https://doi.org/10.1109/TCC.2020.2992537
38. Chen X, Lin J, Ma Y et al (2019) Self-adaptive resource allocation for cloud-based software services based on progressive QoS prediction model. SCIENCE CHINA Inform Sci 62(11):219101
39. Chen X, Wang H, Ma Y et al (2020) Self-adaptive resource allocation for cloud-based software services based on iterative qos prediction model. Futur Gener Comput Syst 105:287–296
40. Huang G, Chen X, Zhang Y et al (2012) Towards Architecture-based management of platforms in the cloud. Front Comp Sci 6(4):388–397
41. Chen X, Li A, Zeng X et al (2015) Runtime model based approach to IoT application development. Front Comp Sci 9(4):540–553
42. Wang, B (2018). Disconnected recurrent neural networks for text categorization. In: Proc of the 56th Annual Meeting of the Association for Computational Linguistics, Long Papers, (1):1024–1034.
43. Wang S, Huang M, Deng Z (2018). Densely connected CNN with multi-scale feature attention for text classification. In: Proc of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI), pp. 4468–4474.
44. Kim Y (2014) Convolutional neural networks for sentence classification. arXiv preprint http://arxiv.org/abs/1408.5882 .
45. Tai KS, Socher R, Manning CD (2015) Improved semantic representations from tree-structured long short-term memory networks. arXiv preprint http://arxiv.org/abs/1503.00075.

46. Priyadarshini, I., Cotton, C (2021) A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis. The Journal of Supercomput, pp.1–22.
47. Huang G, Ma Y, Liu X et al (2015) Model-based automated navigation and composition of complex service mashups. IEEE Trans Serv Comput 8(3):494–506
48. Liu X, Huang G, Zhao Q et al (2014) iMashup: a mashup-based framework for service composition. SCIENCE CHINA Inf Sci 54(1):1–20
49. Huang G, Liu X, Ma Y et al (2019) Programming situational mobile web applications with cloud-mobile convergence: an internetware-oriented approach. IEEE Trans Serv Comput 12(1):6–19
50. Huang G, Mei H (2006) Yang F (2006) Runtime recovery and manipulation of software architecture of component-based systems. Autom Softw Eng 13(2):257–281
51. Zhang X, Zhao J, Lecun Y (2015) Character-level convolutional networks for text classification. In: Proc of the Neural Information Processing Systems. MIT Press, pp. 1–9.
52. Lei Z, Yang Y, Yang M, Liu Y (2018) A multi-sentiment-resource enhanced attention network for sentiment classification. arXiv preprint http://arxiv.org/abs/1807.04990.
53. Kalchbrenner N, Grefenstette E, Blunsom P (2014) A convolutional neural network for modelling sentences. arXiv preprint http://arxiv.org/abs/1404.2188.
54. Johnson R, Zhang T (2017). Deep pyramid convolutional neural networks for text categorization. In: Proc of the 55th Annual Meeting of the Association for Computational Linguistics, 1:1024–1034.
55. Kruengkrai C, Torisawa K, Hashimoto C, et al (2017) Improving event causality recognition with multiple background knowledge sources using multi-column convolutional neural networks. In: Proc of the national conference on artificial intelligence, pp. 3466–3473.
56. Tang D, Bing Q, Liu T (2015) Document modeling with gated recurrent neural network for sentiment classification. In: Proc of the 2015 Conference on Empirical Methods in Natural Language Processing. pp.1422–1432.
57. Huang M, Qian Q, Zhu XY (2017) Encoding syntactic knowledge in neural networks for sentiment classification[J]. ACM Trans Inform Sys 35(3):1–27
58. Zhou C, Sun C, Liu Z, et al (2015) A C-LSTM neural network for text classification. arXiv preprint http://arxiv.org/abs/1511.08630.
59. Xiao Y, Cho K (2016) Efficient character-level document classification by combining convolution and recurrent layers. arXiv preprint http://arxiv.org/abs: 1602.00367.
60. Li M, Hsu W, Xie X et al (2020) SACNN: self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network. IEEE Trans Med Imag 99:1–1
61. Xiao LQ, Zhang HL, Chen WQ, et al (2018). Learning what to share: leaky multi-task network for text classification. In: Proc of the 27th International Conference on Computational Linguistics, pp.2245–2255.
62. Rajpurkar P, Zhang J, Lopyrev K, et al (2016) SQuAD: 100,000+ questions for machine comprehension of text. arXiv preprint http://arxiv.org/abs/1606.05250.
63. Maas AL, Daly RE, Pham PT, et al (2011) Learning word vectors for sentiment analysis. [C]// Proc of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 1. Association for Computational Linguistics (ACL), pp.3564–3574.
64. Sang E.F, De Meulder F (2003) Introduction to the CoNLL-2003 shared task: language independent named entity recognition. arXiv preprint cs/0306050 (2003).
65. Jain, Praphula Kumar, et al. (2021). "SpSAN: Sparse self-attentive network-based aspect-aware model for sentiment analysis." Journal of Ambient Intelligence and Humanized Computing 1–18.
66. Joulin A, Grave E, Bojanowski P, et al (2017). Bag of tricks for efficient text classification. In: Proc of the 15th Conference of the European Chapter of the Association for Computational Linguistics (ACL), pp. 2017(2): 427–431.
67. Peng P (2004) A review of causal connective components in modern Chinese. Chinese learning, pp. 2004(2): 45–49.
68. Dong ZD. http://www.keenage.com/zhiwang/c_zhiwang_r.html.
69. Xing FY (2001) A study of Chinese complex sentences. commercial press, China.

Authors and Affiliations

Chang-Xuan Wan[1,3] **· Bo Li**[1,2,3]

Bo Li
lbsinuo2010@163.com

1 School of Information and Technology, Jiangxi University of Finance and Economics, Nanchang 330013, China

2 Jiangxi University of Science and Technology, Ganzhou 340000, China

3 Jiangxi Key Laboratory of Data and Knowledge Engineering, Jiangxi University of Finance and Economics, Nanchang 330013, China