Pairwise Causality Structure: Towards Nested Causality Mining on Financial Statements

Dian Chen[†,‡] , Yixuan Cao[†,‡] , and Ping Luo[†]

[1] [†] *Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China*
[2] [‡] *University of Chinese Academy of Sciences, Beijing 100049, China*
{chendian, caoyixuan, luop} @ict.ac.cn

Abstract. Causality mining, which aims to find cause-effect relations in text, is an important yet challenging problem in natural language understanding. The extraction of causal relations is beneficial to practitioners in document-intensive industries. For instance, it enables investors and regulators in financial industries to quickly understand the correlation between events in financial statements. However, this problem is difficult since the expression of causality is diverse, and more importantly, nested. Specifically, causality often has a nested structure, where a pair of cause-effect can be the cause of another higher-level causality. Recent works deal with this problem by a bottom-up relation extraction solution, but it performs worse for relations on higher levels. In this study, we find that the nested causality structure can be transformed into a graph of pairwise causality between sentence segments. Then we propose a two-step solution: first, a segmenter disassembles a sentence into segments by detecting causality connectives; second, a relation classifier predicts whether a pair of segments has cause-effect relation or not. Two modules above are trained jointly in our proposed Causality Detection Network (CDNet). On a large dataset we collect, the precision of our model reaches 92.11% and the recall reaches 93.07% for this task. Compared with the existing state-of-the-art solution, the precision of our model is improved by 3.28% and 3.03% for recall. We also observe that the percentage of exactly correct sentences from prediction is 74.26% without post-processing, indicating the hardness of our problem and space for improvement.

Keywords: Causality Mining · Nested Relations · Pairwise Structure

1 Introduction

Causality occurs frequently in natural language, especially in narrative texts. It always carries useful but complicated semantic information. A powerful causality mining system can help improve a number of basic natural language understanding (NLU) tasks such as Query Answering [4] and Information Extraction [12]. Rather than causal inference, causal mining is to extract existing causal relations among segments in a descriptive sentence.

✉ Ping Luo is the corresponding author.

We call texts in financial documents as “financial statements”, which usually contain a large number of complex cause and effect relationships. In this paper, causality mining is used to empower financial practitioners in document-intensive industries. A complete and accurate causality mining system can support financial practitioners to make decisions, like the one shown in Fig. 1. In real-world applications, investors in financial industries usually analyze complex causal relationships of events in reports to support their decisions. Moreover, regulators need to check whether the explanations in financial documents are reasonable and acceptable to meet the requirements efficiently.

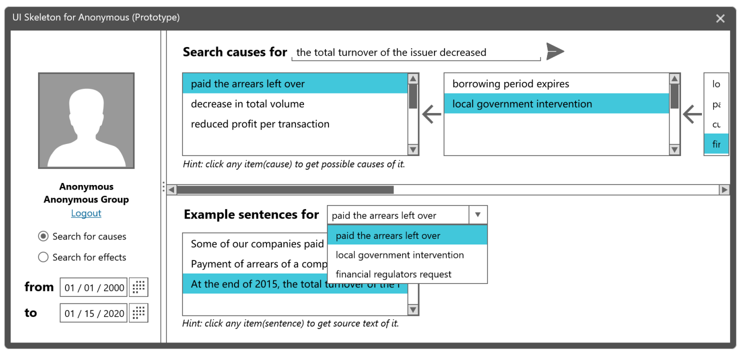


Fig. 1: An example application prototype for market movement and change analysis.

Fig. 2 shows what the mining task gets from a sentence in financial documents. In the example statement, the segment (s_2) is simultaneously *cause* and *effect* in the nested causal relations. There is a two-layer nested causal relation in this sentence: the cause s_3 leads to its effect s_2, The causal relation r_1 between them also plays the role of a cause in another causal relation r_3 with the effect s_1. As shown in Fig. 2(c), we call the structure with nested causal relations as a *Nested Causality Structure*. The goal of nested causality mining is to extract the nested causality structure from each sentence.

Recent advances in deep neural networks have yielded impressive applications in natural language processing, but the extraction for nested causality structures remains a challenge. The difficulty of causality detection lies in the fact that: causality structures have complex diversity in textual expression, especially in the news or financial documents. In detail, the causality relationship in a sentence is sometimes a hierarchy, which means a cause might be further decomposed into another cause-effect relation. We collect thousands of published financial documents as the corpus. Then we find that nearly 15% of sentences with causal relations contain nested causal relationships (Section. 5.1 for detail). It shows that mining for nested causality is an unavoidable problem.

There are many recent studies on relation extraction, but relatively few studies on detection for nested relations. A closely related solution is Iterative Neural Network (INN) [1], which extracts nested relations from texts. In INN, the state-of-the-art model for extracting nested binary relations among entities by DAG-LSTM, relations are predicted layer by layer. Cao [1] also mentioned a major problem for nested relations: The higher a relation is located, the harder it to be extracted. This problem is also encountered when extracting nested causality.

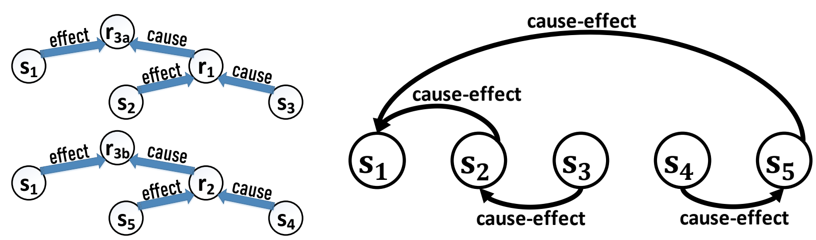
Text: **[At the end of 2015, the total turnover of the issuer decreased by 1.098 billion dollars compared with the end of 2014](s_1), mainly due to the following two reasons: [some subsidiaries failed to recover the last two months of costs](s_2) due to [insufficient company funds](s_3); and [paid the arrears left over from previous years](s_4), resulting in [higher expenditures in 2015 than in 2014](s_5)**

(s_1) mainly due to the following two reasons: (s_2) due to (s_3), and (s_4), resulting in (s_5).

(a) An example financial statement

ID	Trigger Words	Cause	Effect
r_1	due to	s_3	s_2
r_2	resulting in	s_4	s_5
r_{3a}	mainly due to … reasons	r_1	s_1
r_{3b}		r_2	s_1

(b) Common output form



(c) Nested Causality Structure (d) Pairwise Causality Structure

Fig. 2: The causality structure of an example statement from financial documents.

One of the difficulties of extracting nested relations lies in the unknown depth of the structure. For this, we propose a claim (Section. 3.2 for detail) that transform the output into a simpler representation called *Pairwise Causality Structure* (PC-Structure) on segment pairs, according to the nature of causality. The claim is, the cause-effect pair in a causal relation can be represented by its effect. In other words, when predict between one relation and another segment, we can use the representation of the effect-segment to stand for the relation. This conversion reduces the complexity of the structure. To verify the effectiveness of our proposal, we take experiments in Section. 5, the claim is supported by experimental results.

In our formulation, segments serve as the basic units of causality in sentences. A segment is a compact and independent text that describes an event. Most of the existing researches on causal relation extraction use the event denoters [3] or trigger words [14] as the basic elements for causality. For users with little NLP experience, the structure for presenting events shown in Fig. 2(d) is more straightforward, rather than relation tuples as shown in Fig. 2(b) or its graph representation as shown in Fig. 2(c). To separate different events, we use causal connectives or commas to divide the text into segments as events in our proposed method.

Our proposed representation “Pairwise Causality Structure” simplifies the representation for causality. Based on this, we propose a model called **C**ausality **D**etection **Net**works (CDNet) in Section. 4 to integrate the segmenter and classifier. However, whether the simplified representation will affect the performance of relation extraction remains a question. Experimental results shows our model is on par with INN on simple relations, but outperforms on nested ones. It proves the representation is an alternative new choice in other tasks about causality.

2 Related Works

The extraction for causal relations has long been a question of interest in a wide range of fields, such as Lexical Semantics [5] and Information Science [2]. A majority of effective works for this task are based on pattern mining, feature engineering and neural network. However, relatively little research carries out on nested relations.

Classic solutions for causality extraction are mainly based on rules and pattern mining. An existing causal-pair extraction method is proposed [6] based on part-of-speech tagging, syntactic analysis and causality templates. However, the composition of sentences is ever-changing, rules and templates do not have the ability to disambiguate and cope with complex sentences. Then, more kinds of features are collected to support causality extraction, such as causal connectives [13]. Solutions using pattern mining are able to extract any specific kind of causal relations, but there are strong limitations and difficulties when transfer to other fields.

Later, solutions based on deep learning are widely proposed, which usually performs well on robustness and scalability in detection or extraction tasks. They are mostly designed for single-layer relations [8,9], few works are able to apply on nested causality tasks. There exists a solution in Japanese [10] using cue phrases as trigger words. It firstly identifies cue phrases for causality by regular expressions or model, then recognize causal relations for answering why-questions by sequence labeling and feature engineering. It is suitable for extracting the causal relationship between sentences and sentences, but not able to deal with high-level nested relations. Therefore, we try another way, which focuses on the segmentation instead of entity identification.

In recent years, researchers start to define and learn nested relationships in the form of graphs. Iterative relation extraction (INN) model [1] is an effective method for nested relations, for each kind of relations with two participants, especially in the form of an ordered binary DAG. But the model keeps traversing until there is no more relations can be extracted among all levels. The process with no certain ending time brings a relatively large complexity of time and space. In their task, the extraction for nested causal structures is also one of the goals, we will take this as a baseline in Section. 5.

3 Problem Formulation

3.1 Nested Causality Mining

According to the definition from Cao [1], Nested Relation Extraction is the task that extracts semantic relations among elements (could be entities or other relations) from texts. Nested Causality Mining is a specific sub-task that extracts causal relations. In this paper, we focus on extracting nested causal relations from texts. But in practical applications, both sentence segmentation and relation extraction need to be completed.

Segment. The basic unit in this task is the "segment", which consists of a series of consecutive words in the sentence t, or a sub-sequence of the sentence, denotes as: $s := w_{i:j},\ 0 \leq i < j < |t|$. The example shown in Fig. 2(a) contains segments $s_{1:5}$.

Relation. In the task of Nested Causality Mining, a collection of segments is taken as input. The goal of this task is to extract a hierarchical structure as shown in

Fig. 2(c), that is for representing all "causal relations" among segments in a sentence. A causal relation is formed as a segment pair, one of which is used as a *cause* and the other performs as an *effect*. The example mentioned above contains basic relations: r_1 : $Relation(s_3, s_2)$, r_2 : $Relation(s_4, s_5)$, and high-level relations above them: r_{3a} : $Relation(r_1, s_1)$, r_{3b} : $Relation(r_2, s_1)$.

where r_{3a} and r_{3b} can form a ternary relation r_3 in post-processing. We call the expected output of this task as a *nested causality structure*, where r_i in the structure denotes a node for causal relation connecting with its effect and causes.

The extraction for flat cause and effect relation pairs is simple, but narrative statements in real-world usage usually contain complicated causal relations. A complex part of this task is the case when a causal relation contains another relation as its cause:

$$r \in \{Segment, Relation\} \times \{Segment\}$$

thus forming a nested causality structure as in Fig. 2(c). Similar to other kinds of nested relations, a causal relation between a causal relation and a segment is far more complex than one between segments. Mainly because the former one has a larger domain of the components, which means a more complicated connection at a higher level. Moreover, the structure for causality is a graph rather than a tree, take Fig. 2(c) as an example, the edge (s_3, r_2) is also allowed.

3.2 Pairwise Causality Structure

In the nested causality mining task, how to extract nested relations from events is the key phase. Complex structures and indistinguishable event boundaries have combined to stymie progress in the field. It would be better served by having a concise and well-defined structure. So that we propose the Pairwise Causality structure.

We notice that the causality is a special kind of nested structure: the effect is the key to a causal relation, the causal connectives are the keys to a sentence. We propose an assumption called "effect's representative". Here is a brief claim for it: while an existing causal relation $r_0 = (s_k, s_i)$ leads to another effect s_j, what actually cause the effect s_j directly is the effect s_i in the causal relation r_0.

According to basic symbolic annotations described in Section. 3.1, we formulate the nested relation as functions for concise, with segments or relations as arguments. As the transformation for nested causal relations in Section. 3.2, we also propose the reduce function $R(\cdot, s)$ to cover all kinds of causal relations in this task:

$$R(Relation(s_k, s_i), s_j) = \begin{cases} Relation(s_i, s_j), & \text{if } s_k \text{ leads to } s_i \text{ directly} \\ Relation(s_k, s_j), & \text{if } s_i \text{ leads to } s_k \text{ directly} \\ null, & \text{otherwise} \end{cases} \tag{1}$$

where $Relation(s_k, s_i)$ as the first argument should be a causal relation. Otherwise, the outer R function is meaningless for it has an invalid operand, which is neither a segment nor a causal relation.

Relation. Based on this, a nested causality structure is able to be transformed into a simpler structure. We can solve the nested causality mining task in another way, by solving the corresponding pairwise classification task. In this way, we only need to care about the relations of the form:

$$r \in \{Segment\} \times \{Segment\}$$

We call this kind of structure for causality as a *Pairwise Causality Structure*. Review the example in Fig. 2(d), it contains contains the same basic relations, but different high-level relations above them: $r_3 : Relation(s_2, s_1)$, $r_4 : Relation(s_5, s_1)$.

The causal relations r_1 ($r_1 : s_3 \rightarrow s_2$) over s_3 and s_2 also plays the role of cause for the segment s_1 ($r_1 \rightarrow s_1$). For easier understanding, here is another simple example in the same form as r_1 in Fig. 2(d):

``[I paid for dinner](s_1), **because** [he didn't have enough money](s_2) **due to** [his friend had just borrowed $200 from him]($s_3$).''

The cause s_3 in the causal relation $(s_3 \rightarrow s_2)$ contributes little to the effect s_1, while it contributes a lot for the effect s_2.

According to this, the original graph can be re-constructed into a pairwise causality structure, by associating the effect of a child-relation with its parent-relation recursively. Any causality structure with higher levels is also able to apply this by Algorithm. 1.

Algorithm 1: Transforming Algorithm:
constructing the PC-Structure from a nested causality structure.

Input: The Nested Causality Structure $G = (V(G), E(G))$ as aim of the task.
Output: The PC-Structure $G' = (V(G'), E(G'))$ for representing G in a simpler way.
initial $V(G') \leftarrow V(G)$; $E(G') \leftarrow E(G)$;
repeat
  for *each* $(u, v) \in E(G')$, u *is a segment node* **do**
    if $e_{u,v}$ *is an effect-relation edge* **then**
      Remove u from $V(G')$;
      $v \leftarrow u$ in $V(G')$; $e_{*,v} \leftarrow e_{*,u}$ in $E(G')$;
until $\forall d \in V(G')$, d *is a segment node*;
return G';

The transformation guarantees a limited number of inputs (only segments from segmentation), thus simplified the representation of the graph without extra nodes. A smaller variety of model parameters also contributes significantly to higher performance, both in learning efficiency and convergence speed. In addition, once we get the relations from prediction, it still needs to be converted back to the original causality structure. The transformation from prediction to the PC-Structure is feasible by Algorithm. 2.

After solving the simplified pairwise prediction problem, a nested structure can be reconstructed from a PC-Structure by the backtracking algorithm. Each edge (u, v) in the returned graph G stands for a causal relation, the whole tree with u as the root stands for the cause while the segment v stands for the effect. In practical applications, sometimes we choose not to generate the entire graph, but to select a specific segment as a root node then explore the causality sub-graph from it. The above algorithm is also valuable in re-construction, after arbitrarily deleting or adding edges in the graph.

Algorithm 2: Backtracking Algorithm:
constructing the causality structure from a PC-Structure.

Input: The collection S of segments in text and the PC-Structure $G' = (V(G'), E(G'))$.
Output: A directed acyclic graph G of the causality structure with all segments $s_i \in S$;
initial $V(G) \leftarrow S$; $E(G) \leftarrow \emptyset$; $D \leftarrow [0, |S| - 1]$;
Calculate in-degree $deg^+(*)$ and out-degree $deg^-(*)$;
while $D \neq \emptyset$ **do**
  for *each* d ***in*** D *such that* $deg^-(s_d) = 0$ **do**
    if $deg^+(s_d) > 0$ **then**
      Add node $(r_{c,d})$ into $V(G)$;
      Add edge $(r_{c,d}, s_d,$ 'effect'$)$ into $E(G)$;
      for *each* c ***in*** D *such that* $(s_c, s_d) \in E(G')$ **do**
        Add edge $(s_c, r_{c,d},$ 'cause'$)$ into $E(G)$;
        $deg^-(s_c) \leftarrow deg^-(s_c) - 1$; $deg^+(s_d) \leftarrow deg^+(s_d) - 1$;
    if $deg^+(s_d) = 0$ **then**
      Remove d from D;
return G;

4 Model Structure

In this section, we introduce the model structure of **C**ausality **D**etection **Net**works (CDNet). The model consists of a *Sentence Segmenter* and a *Relation Classifier*. As shown in Fig. 3, firstly a bi-directional LSTM layer encodes each token in the sentence for judging whether it is in a causal connective. It plays the role of a segmenter to split the sentence into segments and connectives. Secondly, the hidden state of each segment is generated with all tokens in it. Based on the representation of segments and causal connectives, another bi-directional LSTM layer encodes each segment for judging whether there exists a causal relation between segments.

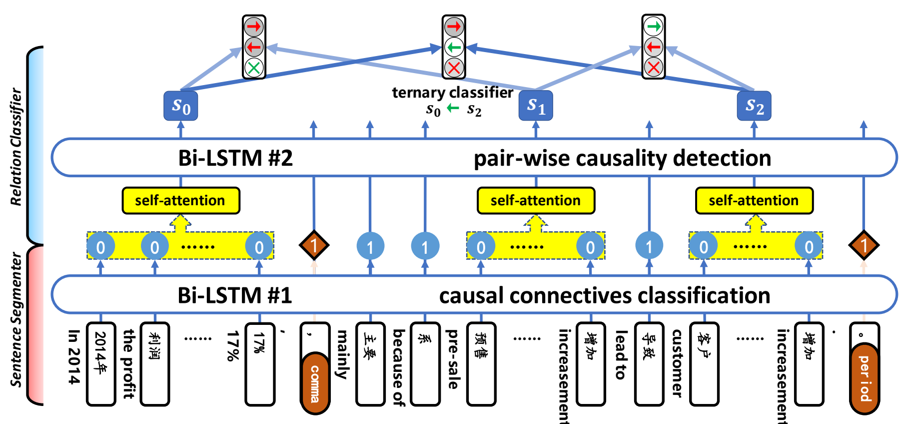


Fig. 3: An overview of the CDNet model structure.

4.1 Sentence Segmenter

As shown in Fig. 3, the sentence segmenter predict a label for each token, to indicate whether the token is used as a separator for segments.

According to Zhao's research [13], a causal pair can be evoked by different causal connectives, and the syntactic dependency structure of a sentence expressing causality varies for different connectives. Based on this, we expand the definition of "segments": they are separated by both commas[3] and causal connectives. Besides, the causal connectives not only help determine whether a sentence contains causality or not, but also plays an important role for sentence segmentation in the mining task.

To classify tokens by whether a token is in a causal connective or not, we apply a Bidirectional LSTM layer to achieve information about the context for each token. Then, a binary classifier handles the task of determining whether it is in a causal connective or not. After the segmentation step, we obtain the representation of each segment by mean-pooling or self-attention.

4.2 Relation Classifier

A segment pair is represented by the combination of two segments' hidden vectors, and the difference of the vectors $h(pair(s_i, s_j)) = [h_{s_i}; h_{s_j}; h_{s_j} - h_{s_i}]$. The third vector emphasizes the direction besides the difference between the two segments. which makes it easier for the model to learn that there cannot be causal relations for the same segment.

Our task focuses on explicit causations. It means, the explicit causal relationship is the main part of this study, both operands in a causal relation are necessary and not interchangeable. We use a classifier to predict causal relations in pairs. To avoid conflicts like "s_i leads to s_j but s_j also leads to s_i at the same time", a simple solution is to predict for each $pair(s_i, s_j), i < j$. There are three states for each pair, the collection of states is $C = \{\textit{effect-cause}, \textit{null}, \textit{cause-effect}\}$. In order to figure out the confidence for each pair, we model the loss function as:

$$\mathcal{L}_p = -\sum_{i,j=0}^{|S|} \sum_{k \in C} \left(y_{i,j}^{(k)} \log(p_{i,j}^{(k)}) + (1 - y_{i,j}^{(k)}) \log(1 - p_{i,j}^{(k)}) \right) \tag{2}$$

As shown in (2), the loss function of this task is defined as the cross entropy. $p_{i,j}^{(k)} = p(\hat{y}_{i,j}^{(k)} = 1 \mid s_i, s_j, t)$ denotes the probability from model's prediction $\hat{y}_{i,j} = k$, where $y_{i,j}^{(k)}$ is an one-hot vector which expresses the ground truth.

4.3 Joint Training

As shown in Fig. 3, we train the sentence segmenter and causal relation classifier jointly in the same network. The model calculates the hidden vectors of words, segments and pairs in order according to the steps of the task.

[3] punctuation marks with higher pause level than commas such as semicolon or period also work here, while lower ones do not.

Two sub-tasks share the same embeddings and hidden vectors of words. We share the parameters between modules in the joint training model, so that the joint training process is more efficient, especially for words' representation. The loss function consists of causal connectives ($\mathcal{L}_{cc}$) and the loss of pairwise causal relations ($\mathcal{L}_p$):

$$\begin{aligned}\mathcal{L} &= \alpha\mathcal{L}_{cc} + \mathcal{L}_p \\ &= -\alpha\sum_{i=0}^{|t|}\left(y_{cc(i)}\log(p_i) + (1 - y_{cc(i)})\log(1 - p_i)\right) \\ &\quad + \mathcal{L}_p, \;\; \mathcal{L}_p \text{ as mentioned in (2)}\end{aligned} \tag{3}$$

The sub-tasks for causality mining task are sequence labeling and relation extraction. The parameter α is designed for further adjustment.

During training for relation extracting, segments. There are two ways to obtain segments for training CDNet:

guided: The model obtains the causal connectives as known in the beginning. Then the model can jointly train the task of segmentation and relation detection.

predict: The segmentation used in the detection task is obtained from the sentence segmenter. If there exist faulty predictions in segmentation, the classification will also be affected by the involved words.

5 Experiment

5.1 Data Preparing

Financial documents are our target corpus for experiments, which often contain complex causal relations to explain changes in indicators. We collect paragraphs about indicator changes from 2,039 published financial documents as our corpus. Finally, 69,120 sentences are collected as our dataset for labeling. In detail, a total of 25 volunteers participated in the annotation task, each volunteer independently annotates at least 500 sentences to the final data set. For quality control, four more professional volunteers were selected to be responsible to deal with conflicts and extremely hard questions. and regularly explaining the complex problems to all the volunteers. The dataset is randomly divided into the training set and the validation set, by the ratio of 9:1. The distribution of our dataset is shown in Fig. 4. We notice that nested relations (i.e. the relation with a depth of more than 2) accounts for 14.87% of all (1028 in 6912).

For a fair comparison, we also use the same data set as INN [1], which has 10,000 labeled sentences with causal relations. 70% of the sentences are used for training.

5.2 Hyper-Parameter Settings

During the model training, there is no intentional adjustments to set the hyper-parameters for the model: we set embedding size to 128, LSTM hidden size to 256, dropout rate to 0.1, and Adam [7] optimizer's learning rate to 1e-03. Pre-trained word embedding is prepared by Glove [11] using all 2,039 financial documents. The most frequent 4,000

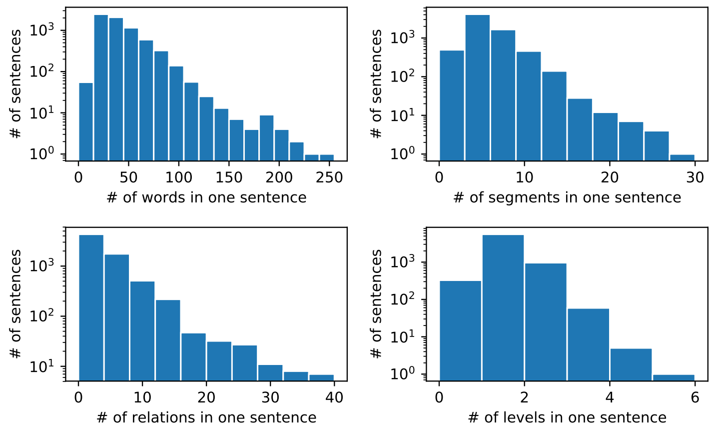


Fig. 4: The basic distribution of sentences in validating set.

common words are used as our vocabulary. We consider that words in causal connectives are more important than words in segments, so that all words in causal connectives are covered by this vocabulary.

5.3 Evaluation Metrics

The task of Nested Causality Mining is *pair-wise causality detection* on segments from *sentence segmentation*. Next, we introduce the metrics for evaluating both tasks.

Sentence Segmentation. We need to judge whether a token in the sentence is in the causal connectives, it is a task of sequence labeling. So we treat each token as a sample and calculate F1-Score for them.

Pairwise Causality Detection. We detect causal pairs from segment pairs like $pair(s_i, s_j)$. It is a ternary classification task, so-called "pairwise causality detection", used to determine the type of its relation $y_{i,j}$ in the state collection C. We treat it as "correct", only if a segment pair from prediction is exactly the same with one pair in the ground truth. The most common and strict method for evaluation is the *Complete Accuracy*. It means the order, the tokens in each segment, and the relation type between segments in the pair are all the same. In addition, we also use *F1-Score* on causal pairs as an evaluation indicator for classification. Note that if a sentence contains no causal relations, the precision and recall depend on the edge set from prediction: If it is an empty set, both precision and recall are set to 1, otherwise to 0.

5.4 Result Analysis

As causal relations also satisfy the requirements for an ordered binary DAG-LSTM, we select INN [1] as a baseline. The structure of INN is similar to the nested causality structure: At the begining, the first layer is used for predicting relations over basic segments. Next, INN predict for whether there is a relationship to be extracted, between an entity and a relation in the previous layer. The main difference between INN and our model is the representation of causal relations, when a prediction is over one causal relation and a segment. So that we take a comparison with INN to validate whether our simplified representation is feasible for nested causal extraction.

Table 1: Results (%) of evaluation for comparison.

Trained Models	Dataset Train	Valid	Pair Level P	R	F_1	Complete Accuracy
INN-Predict	INN[a] Trainset	INN Validset	78.51	79.41	78.96	-
CDNet-Predict			**88.29**	89.26	88.77	66.94
INN-Guided			81.63	80.99	81.31	-
CDNet-Guided			87.42	**90.73**	**89.05**	66.95
INN-Guided	Entire[b] Trainset	Entire Validset	88.83	90.10	89.46	73.96
CDNet-Guided			**92.11**	**93.07**	**92.59**	**74.26**
INN-Guided	Entire Trainset	Nested[c] Validset	86.47	84.76	85.61	61.19
CDNet-Guided			**92.48**	**91.39**	**91.93**	**70.29**

[a] the same dataset used in INN[1]. (10,000 sentences, 7:3)
[b] covers all sentences in our dataset. (69,120 sentences, 9:1)
[c] only contains sentences with nested causal relations. (1,028 sentences)

For a fair comparison, we use the same training mode of CDNet for each corresponding method in INN. With the same pre-trained word embeddings, corpus and dataset split ratio, an experiment for comparison between INN and CDNet is shown as results in Table. 1. Compared with INN model, our model performs better on classification for segment pairs. In Table. 1, F1-score is improved by 7.7% on the same dataset with INN's paper and 3.1% on our larger one. In addition, CDNet shows strengths in nested causality recognition on sentences with higher levels. Our model performs a 6.3% improvement on pair-level F1-Score and a 9.1% on whole-sentence accuracy. The accuracy from INN model is affected more than CDNet while dealing with nested relationships, mainly reflected in recall rate.

As mentioned in [1], when the number of layers increases, the accuracy of the prediction decreases. Review relations on the highest level in Fig. 2(c), the prediction for it depends on whether the operands are all correctly prepared. So that the error in previous layers propagates through layers, then affects extracting high-layer relations. Our proposed model is built in a simpler way: using the effect segment only to represent the relationship over it. In our method, the model drops the cause-operand in causal relations during prediction. The information loss should lead to a decline in performance, but the illustrative result from the comparison is able to support our claim: In causal relations, the effect is able to represent the causal relationship that contains it, to a large extent. The claim also provides an alternative way to simplify or optimize problems in other tasks involving nested causality.

Our model performs an obvious improvement on the nested dataset, but improves slightly on the entire dataset. It reveals that although we reduce the accuracy reduction caused by complex nesting, relations on flat ones has not been well solved. The efficacious representation for a segment or span is also an unsolved challenge, and will be another way to optimize the nested extraction tasks.

6 Conclusion

Overall, this paper aims to facilitate the resolution of problems in nested causality mining. We propose a nested causality mining model with acceptable performance. The model is based on a novel representation "pairwise causality structure" for nested causal relations, which is proved as a feasible solution to detect nested causal relations on segment levels. Benefit from this, we figure out the causality mining task can be conditionally simplified into a classification problem, between segment pairs. Besides, both the problem formulation and model structure are able to act as a starting ground for further extraction. However, it is still challenging to deal with texts that are more casually expressed (such as sentences in oral presentations), we leave it for further work.

Acknowledgments

The research work supported by the National Key Research and Development Program of China under Grant No. 2017YFB1002104, the National Natural Science Foundation of China under Grant No. U1811461.

References

1. Cao, Y., Chen, D., Li, H., Luo, P.: Nested Relation Extraction with Iterative Neural Network. In: CIKM. pp. 1001–1010. ACM (2019)
2. Duan, P., Yang, F., Chen, T., Shah, S.L.: Direct causality detection via the transfer entropy approach. TCST **21**, 2052–2066 (2013)
3. FU, J., LIU, Z., LIU, W., ZHOU, W.: Event Causal Relation Extraction Based on Cascaded Conditional Random Fields. PR & AI **24**(4), 567–573 (2011)
4. Girju, R.: Automatic detection of causal relations for Question Answering. In: MultiSumQA. pp. 76–83 (2003)
5. Hobbs, J.R.: Toward a useful concept of causality for lexical semantics. Journal of Semantics **22**(2), 181–209 (2005)
6. Ittoo, A., Bouma, G.: Extracting Explicit and Implicit Causal Relations from Sparse, Domain-Specific Texts. NLP & IS **6716 LNCS**, 52–63 (2011)
7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2015)
8. Li, Z., Li, Q., Zou, X., Ren, J.: Causality Extraction based on Self-Attentive BiLSTM-CRF with Transferred Embeddings. ArXiv (2019)
9. Miwa, M., Bansal, M.: End-to-End Relation Extraction using LSTMs on Sequences and Tree Structures. In: ACL. pp. 1105–1116 (2016)
10. Oh, J.h., Torisawa, K., Hashimoto, C., Sano, M., De Saeger, S., Ohtake, K.: Why-Question Answering using Intra- and Inter-Sentential Causal Relations. ACL pp. 1733–1743 (2013)
11. Pennington, J., Socher, R., Manning, C.: GloVe: Global vectors for word representation. In: EMNLP. pp. 1532–1543. Association for Computational Linguistics (2014)
12. Surdeanu, M., Harabagiu, S., Williams, J., Aarseth, P.: Using predicate-argument structures for information extraction. In: ACL (2003)
13. Zhao, S., Liu, T., Zhao, S., Chen, Y., Nie, J.Y.: Event causality extraction based on connectives analysis. Neurocomputing **173**, 1943–1950 (2016)
14. Zhao, S., Wang, Q., Massung, S., Qin, B., Liu, T., Wang, B., Zhai, C.: Constructing and Embedding Abstract Event Causality Networks from Text Snippets. In: WSDM. pp. 335–344 (2017)