Inter-sentence and Implicit Causality Extraction from Chinese Corpus

Xianxian Jin[1], Xinzhi Wang[1(✉)], Xiangfeng Luo[1(✉)], Subin Huang[1,2], and Shengwei Gu[1,3]

[1] School of Computer Engineering and Science, Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China
{xianxianjin,wxz2017,luoxf,huangsubin,gushengwei}@shu.edu.cn
[2] School of Computer and Information, Anhui Polytechnic University, Wuhu 241000, China
[3] School of Computer and Information Engineering, Chuzhou University, Chuzhou 213000, China

Abstract. Automatically extracting causal relations from texts is a challenging task in Natural Language Processing (NLP). Most existing methods focus on extracting intra-sentence or explicit causality, while neglecting the causal relations that expressed implicitly or hidden in inter-sentences. In this paper, we propose Cascaded multi-Structure Neural Network (CSNN), a novel and unified model that extract inter-sentence or implicit causal relations from Chinese Corpus, without relying on external knowledge. The model employs Convolutional Neural Network (CNN) to capture important features as well as causal structural pattern. Self-attention mechanism is designed to mine semantic and relevant characteristics between different features. The output of CNN and self-attention structure are concatenated as higher-level phrase representations. Then Conditional Random Field (CRF) layer is employed to calculate the label of each word in inter-sentence or implicit causal relation sentences, which improves the performance of inter-sentence or implicit causality extraction. Experimental results show that the proposed model achieves state-of-the-art results, improved on three datasets, when compared with other methods.

Keywords: Causality extraction · Causality · Event extraction

1 Introduction

In recent years, the automatic extraction of causal relation in the field of NLP has attracted increasing attention of researchers. Most existing methods focus on extracting intra-sentence or explicit causality, while neglecting the causality that expressed implicitly or hidden in inter-sentences. In this paper we focus on mining causality in text, and make effect to improve the result of extracting implicitly or inter-sentence causality by building deep learning model.

H. W. Lauw et al. (Eds.): PAKDD 2020, LNAI 12084, pp. 739–751, 2020.
https://doi.org/10.1007/978-3-030-47426-3_57

Causality of text is defined as the relationship between cause and effect, which may be linear or undirected. Causal relation of text exists in many scenario, such as product or social event reviews. The causality can be formalized as the relationship between event e_1 and event e_2, where the event e_2 is considered as a result of the event e_1 [17]. The representation of causal relations can be concluded into four forms: explicit causal relation whose sentence contains explicit causal connective such as "lead to", implicit causal relation whose sentence does not contain causal connective, intra-sentence causal relation whose cause-effect are distributed in short text between two adjacent punctuations, and inter-sentence causal relation whose cause-effect is broken by punctuations. Examples of the four forms are shown in Table 1. Automatic extraction of causality from texts plays an important role in natural language processing application, such as providing event basis for question answer system [19].

Table 1. The forms of causal relation.

Forms	Sentence	Causal-effect
Intra-sentence or explicit	Rising food prices will led to CPI continue to rise	"Rising food prices" → "CPI continue to rise"
Inter-sentence or explicit	RMB appreciation, result in surrounding house prices rise	"RMB appreciation" → "house prices rise"
Intra-sentence or implicit	Reduced food production, falling prices of gold and silver	"Reduced food production" → "falling prices of gold and silver"
Inter-sentence or implicit	According to analysis, loss of controlling interest, many companies are affected, appear passive overweight	"Loss of controlling interest" → "appear passive overweight"

Rule-based [9] methods or traditional machine learning methods [1] contribute a lot on causality mining. However, there are still many drawbacks. More specifically, rule-based methods heavily depend on manually designed language patterns, such as lexical-syntactic patterns [11]. Recently, deep learning methods are widely used in NLP tasks including causality extraction. Some researchers [3] focused on finding language expression of causality and extracting causal triplets from explicit causality sentences. In Table 1, for sentence like the first line, phrase "lead to" indicate the causality blocked by adjacent punctuation explicitly. For sentence like the second line, explicit cause and effect distribute in two adjacent blocks. For sentence like the third line, the cause "Reduced food production" and the effect "falling prices of gold and silver" are hidden implicitly. As being distributed in nonadjacent sentient blocks or lacking of causality indicators, mining inter-sentence or implicit causality is harder when compared with that of intra-sentence and explicit causality. Regardless of implicit causality in nonadjacent sentence blocks like the last line.

This paper proposes model CSNN a novel and unified inter-sentence or implicit causality extraction model. The CSNN model can be described in three steps. Firstly, CNN [2] captures local important features from different blocks of sentences. Secondly, fully considering the extracted advanced features correlation, we cluster the extracted advanced features, and establish semantic relevant characteristics between different features using self-attention. Thirdly and finally, BiLSTM [5] is employed to capture long dependencies between cause and effect using the extracted feature before extracting implicit causal relations or inter-sentence causal relations.

The main contributions of this paper are summarized as follows:

1. We propose a novel and unified model, which can automatic extract inter-sentence or implicit causal relations from Chinese corpus and does not rely on other external knowledge.
2. The model can not only capture local important features from different blocks of sentences, but also obtain long dependencies between cause and effect hidden in different sentence blocks.
3. Experiment results on three differenitemt datasets demonstrate that our model achieves state-of-the-art F1-score (F) in the task for extracting inter-sentence or implicit causal relations.

The paper is organized as follows: the related works on causal relation extraction is shown in Sect. 2. Details of the proposed causal relation extraction model is introduced in Sect. 3. Experimental and results are presented in Sect. 4. Finally, conclusion is given in Sect. 5.

2 Related Work

As complex structure and diverse forms, the extraction of causality in text is still a challenging task. Existing extraction of causality from nature language texts mainly considered two different methods: statistical methods and non-statistical methods.

Non-statistical Approaches: Numerous rule-based [9] methods have been dedicated on causality mining. Kontos and Sidiropoulou [13] used causal language patterns and hand-crafted causal relation templates to detect causal relations which hidden in contexts. Garcia [4] produced a system COATIS, which explorate the rules through contextual and linguistic features to extract causal relations from French texts.

Early work in this area heavily relied on hand-crafted rules and linguistic features. Which limit greatly on flexibility and hard to scale to others corpus. Due to the complexity of causal relation expressions in texts, the precision and recall are also dramatically low.

Statistical Approaches: In recent years, statistical methods have shown promising results in extracting causality from texts. Fu et al. [10], turned the causality extraction problems into sequence labeling problems for the first time.

They extracted causality from Chinese text with two-layer conditional random field. Dasgupta [3] focused on finding language expression of causal relation and proposed a method that combine word-level embedding with other linguistic features to extract causality from sentences. Li and Mao [16] extracted a method named Knowledge-oriented Convolutional Neural Network (KCNN) to extract causal relation. Li et al. [15] proposed a neural causality extractor named SCIFI, which can directly extract causal triplets from explicit causal relation sentences. The methods focused on intra-sentence or explicit causality. Such as "Attrition of associates will affect scheduled release of product causing high business impact." [3]. For sentence like the last line of Table 1, implicit causality in nonadjacent sentence blocks, only considered the dependencies between causal words is not enough.

Our work attempts to extract important features as well as causal structural pattern automatically, combine semantic and relevant characteristics between different features as higher-level phrase representations, to improve the result of inter-sentence or implicit causal relation extraction.

3 Cascaded Multi-structure Neural Network

In this section, our proposed inter-sentence or implicit causality extraction model is presented in detail. The overview model CSNN is shown in Fig. 1.

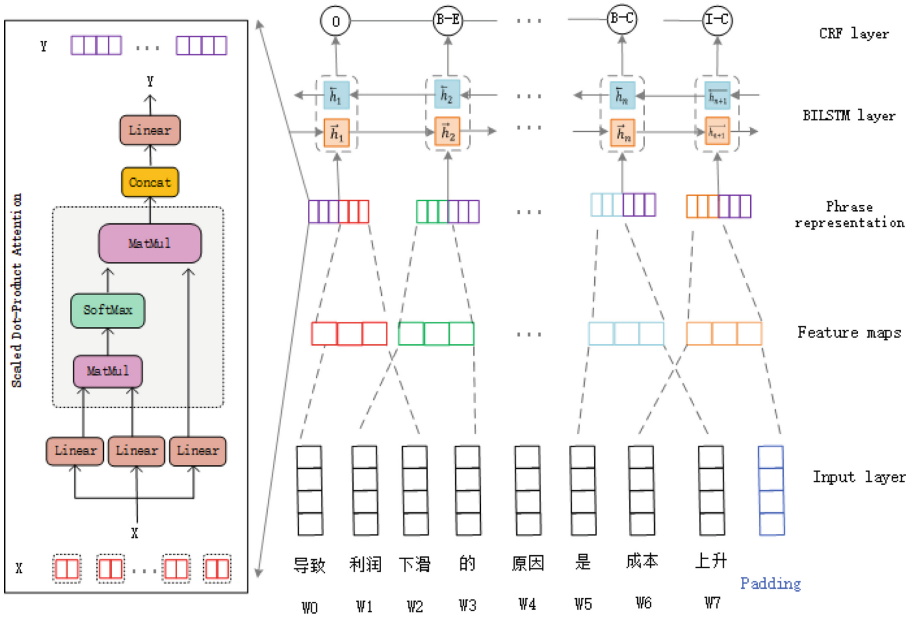

Fig. 1. Overview of the CSNN model for Cause-Effect relation extraction. Left: Semantic and relevant characteristics extraction process.

3.1 Model Architecture

In this subsection, we mainly explain how CSNN is employed to extract inter-sentence or implicit causality from Chinese corpus. The model is composed of two modules: 1) CNN is designed to capture important features within the windows of causality. Self-attention mechanism is designed to mine semantic and relevant characteristics among different features. 2) Long dependencies between cause and effect using BiLSTM [5] are established, which obtain deeper contextual semantic information.

In addition, CSNN expresses sentence as vector embedding. In this regard, a sentence is expressed as a vector $S \in R^{l \times e}$, where l is the number of words in sentences and e is embedding size. Besides, x_i is a vector standing for an e-dimentional word embedding of the *i-th* word in sentences.

$$S = (x_1, x_2, x_3, ..., x_l) \tag{1}$$

Convolution Layer: CNN [2] utilizes max-pooling to extract data features. However, max-pooling often produces poor results in causality extraction due to loss of information. More specifically, max-pooling only maintains the feature with the highest activation, which discarding all other features even though they seem to be useful. Therefore, all the features after convolution are used as windows features in causality extraction task.

The convolution layer is aimed to capture important features within the windows of causality, and compress these important cues into feature map. In general, let $w \in R^{h \times e}$ be the filter for convolution operation, where h is the number of words to generate a new feature. For example:

$$C_i = f(w \cdot x_{i:i+h-1} + b) \tag{2}$$

Where $b \times R^n$ is a hyper-parameter and f is a nonlinear function such as sigmoid, ReLU etc. In addition, ReLU [6] is used as the nonlinear function. The CSNN model employ multiple filters in convolution to generate multiple feature maps. For example, generated features for each window *i-th* can be expressed as:

$$W = [C_1, C_2, C_3, ..., C_m] \tag{3}$$

Where C_i is the feature map generated with the *i-th* filter. W_i is the feature representation generated from m filters for the window vector at position i and stride adopted 1.

In learning process, although each eigenvalue represents the overall characteristics of a window, the internal tendencies of each part of the feature is different. The internal semantic correlations between different features play a crucial role in the extraction of causal relations. Then, these features are grouped and self-attention mechanism is used to mine semantic and relevant characteristics between different features. Number of group is and each of group features is $d = m/t$.

Self-attention Layer: Self-attention is a special attention mechanism according to a sequence to compute its representation and has been successfully applied in many tasks, such as machine translation [20] and language understanding [21].

In our model, the multi-head attention mechanism [23] is employed to mine internal related information between features. As depicted in the left part of Fig. 1, the scaled dot-product attention is used as attention function to compute the attention sources by following Eq. 4. The input consists of query matrix $Q \in R^{t \times d}$, keys $K \in R^{t \times d}$, values $V \in R^{t \times d}$ and d is the number of features in a group.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d}})V \tag{4}$$

The multi-head attention mechanism captures semantic relevant information from different features subspaces at different positions. First, we utilized h times different linear projections to get the queries, keys and values matrices, whose dimension is d/h. Then we perform the attention in parallel and concatenate the output values of h heads. The mathematical formulation is shown below:

$$H_i = Attention(QW_i^Q, KW_i^K, KW_i^V) \tag{5}$$

$$HHead = Concat(H_1, H_2, ..., H_h)W \tag{6}$$

Where the parameter matrices of *i-th* linear projections $W_i^Q \in R^{n \times (\frac{d}{h})}$, $W_i^K \in R^{n \times (\frac{d}{h})}$, $W_i^V \in R^{n \times (\frac{d}{h})}$. In addition, the outputs of CNN and self-attention structure are concatenated as higher-level phrase representations and fed into BiLSTM.

BiLSTM Layer: Causality sentence combines a positive relationship between cause-effect and a reverse relationship between effect-cause. Since the information loss is very serious in the long-distance transmission, LSTM [24] is not suitable for modeling causality. Given a sequence of input features $\{f_t\}$, the relation between the cause and the effect can be obtained through forward LSTM and can be enhanced through back LSTM:

$$\overleftarrow{h_t} = \overleftarrow{LSTM}(\overleftarrow{f_t}, \overleftarrow{h_{t+1}}) \tag{7}$$

$$\overrightarrow{h_t} = \overrightarrow{LSTM}(\overrightarrow{f_t}, \overrightarrow{h_{t-1}}) \tag{8}$$

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{9}$$

Then, the probabilistic matrix $P = \{p_1, p_2, p_3, \ldots, p_n\}$ is generated, whose size is $m*n$, n is the number of words while m is the number of tags.

Finally, the conditional random field [14] is used to adjust the previously predicted tag sequence, taking into account the interaction between adjacent tags. In the model, we obtained the last hidden layer to predict the score of the word with each possible label, and the weight matrix is used to learn the

conversion probability between different tags. Given a prediction sequence $y = \{y_1, y_2, y_3, ..., y_n\}$, the CRF score can be calculated as:

$$score(x,y) = \sum_{i=1}^{n+1} A_{y_{i-1},y_i} + \sum_{i=1}^{n} P_{i,y_i} \tag{10}$$

p_{i,y_i} is the prediction that the i-th word is the y_i label probability in a sentence. In addition, the likelihood of transitioning from label y_{i-1} to label y_i can be expressed as A_{y_{i-1},y_i}.

Viterbi algorithm is used to output the valid sequence of labels with the largest *score(x,y)*. The loss function minimizes the scores of other sequences, while maximizing the score of the correct tag sequence. The mathematical formulation is shown below:

$$E = log \sum_{y \in Y} exp^{s(y)} - score(x, y) \tag{11}$$

Where Y is the set of all possible label sequences for a sentence.

4 Experiments and Results

4.1 Datasets

Chinese encyclopedias in the web, such as Jinrongjie[1] and Hexun[2], which contain a large number of financial datasets with variety of causal relations, from which programmed datasets are extracted. Finally, 11568 web pages were crawled from Jinrongjie and 11742 web pages were crawled from Hexun. 1486 inter-sentence causality sentences and 500 implicit sentences was annotated as our datasets.

Chinese Emergency Corpus (CEC)[3] is an event ontology corpus developed by Semantic Intelligence Laboratory of Shanghai University. It has 332 articles including five categories: earthquake, fire, traffic accident, terrorist attack and intoxication of food, which are derived from Internet and processed by several steps. Finally, 966 inter-sentence causality sentences and 609 implicit causality sentences are extracted.

Financial Event Evolutionary Graph[4] is an event logical knowledge which was published by Harbin Institute of Social Computing and Information Retrieval Research Center (HIT-SCIR). It mainly contains two kinds of relationships, one is causal relation, and the other is a similar relation. Besides, it also contains relationship context sentence information. We obtain 556 intra-sentence causality sentences based on the context information of the datasets.

The above three datasets are used to develop and validate the effectiveness of proposed model generated from the web, the CEC and SCIR datasets. We are willing to share financial datasets with the community.

[1] http://www.jrj.com.cn.
[2] http://www.hexun.com.
[3] https://github.com/shijiebei2009/CEC-Corpus.
[4] http://eeg.8wss.com.

Besides, the advice of two experts is followed to annotate the data. First, determine whether a given sentence contains a causal event. Second, annotate which parts are the cause or an effect. The annotated dataset such as "China Wine Network has been <Cause> in a state of loss for a long time </Cause>, which has <Effect> great impact </Effect> on the performance of Qingqing Liquor.". Finally, BIO was used to mark the sentences ("B-X" represents the beginning of the Cause or Effect, "I-X" represents the middle and end of the Cause or Effect, "O" means not belonging to Cause or Effect).

4.2 Experiment Settings

During the experimental pretreatment process, Hanlp[5] is used as a tool for word segmentation. Word2vec is used to train these financial datasets. Skip-Gram algorithm [18,25] is used to get pre-trained 100-dimensional word embedding vectors on financial datasets instead of initialized randomly. Random vectors (All of the random vectors are sampled from a uniform distribution in the range of $[-0.5, 0.5]$) are used to express the words that did not occur in the embedding vocabulary.

Dropout [7] is applied to the last layer to avoid overfitting, which can reduce the coadaptation of hidden units by randomly dropping out a proportion of the hidden units during the training process. Moreover, different filters are adopted within CSNN model in different datasets for causal relation extraction from Chinese. The filter size is 5 in CEC, 3 in SCIR and financial datasets. The number of feature map is 100. The model is fit over 200 epochs, where the batch size is 16. The unit number of LSTM is chosen to be 100 hidden layers, the dropout is 0.2. Adam [12] is used with the learning rate of 0.006 for optimization.

4.3 Results and Analyses

To prevent the impact of uneven data distribution on the experimental results, the averaged F1 score is used to calculate from 10-fold cross validation for evaluation. In addition, the datasets are divided into 66% training set, 18% test set and 16% verification set. We perform a number of different classical experiments to verify the effectiveness of the proposed model.

IDCNN-Softmax: By increasing the width of the filter, this method obtains data on a wider input matrix, which makes up for the shortcomings of the last layer of neurons in the original CNN convolution to obtain only a small piece of information in the original input data. Then each tag is predicted using the Softmax classifier.

IDCNN-CRF [22]: Taking into account the interaction between adjacent tags, the model uses a CRF classifier to maximize the score of the correct tag sequence to obtain the best output sequence.

CNN: Peng Fei [16] firstly bases on this model, by adding human prior knowledge to capture the linguistic clues of causality extraction.

[5] https://github.com/hankcs/HanLP.

Table 2. Average F-scores (%) of the cause/effect extraction by six extraction models namely, IDCNN-Softmax, IDCNN-CRF, BiLSTM-Softmax, BiLSTM-CRF, CNN-BiLSTM-CRF and CSNN on CEC, SCIR and Financial Chinese datasets.

Model	CEC		Financial		SCIR	
	Cause	Effect.	Cause	Effect.	Cause	Effect.
IDCNN-Softmax	62.96	54.74	50.97	57.65	66.15	60.02
IDCNN-CRF	66.01	59.56	53.65	60.18	76.05	68.24
BiLSTM-Softmax	69.29	66.81	54.51	61.07	79.87	67.53
BiLSTM-CRF	73.87	70.98	58.68	64.61	81.11	70.09
CNN-BiLSTM-CRF	74.55	70.79	59.56	65.85	80.86	72.55
CSNN	**75.40**	**72.72**	**62.24**	**66.70**	**82.27**	**75.34**

BiLSTM-Softmax: Dasgupta [3] uses the bi-directional LSTM model to extract explicit causal relationships automatically, using additional language features and Softmax classifiers to independently predict causal labels.

The above two methods [3,16] are based on knowledge. Due to the lack of Chinese knowledge and the inability to add additional language features, model-based and adding CRF are used as our baseline, such as BiLSTM-CRF [8].

CNN-BiLSTM-CRF: The model uses CNN to recode the input vector bi-directional LSTM to capture longer dependencies and learn the semantic representation of causality.

The experimental results are shown in Table 2, we can easily find that the BiLSTM-based approach is superior to the CNN-based approach. The reason may be that the BiLSTM layer can capture semantic features and establish long dependencies between causal relationships more efficiently. Besides, BiLSTM layer can understand the relationship between causal features and effect characteristics through forward LSTM, and enhance the relationship between effects and causes through reverse LSTM. In addition, we find that the combination of CNN and BiLSTM approach captures more local feature information using CNN, and BiLSTM enhances connections to longer dependencies, yielding better results for causal relation extraction (Fig. 2).

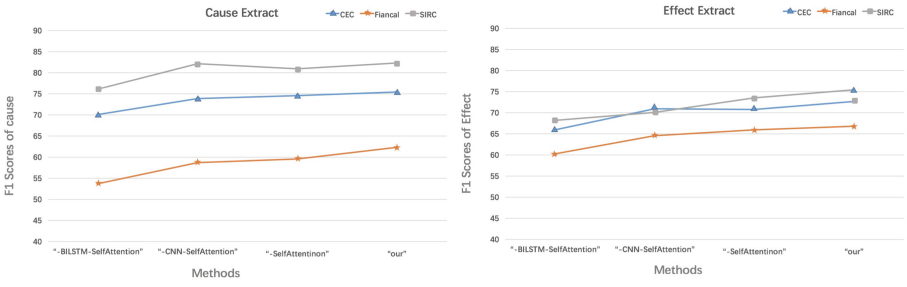

Fig. 2. Ablation analysis of our proposed framework CSNN. "Our" denotes the complete CSNN framework, while "-" denotes removing the component from the CSNN.

For best verifying whether our self-attention mechanism can learn semantics and relevant characteristics between different features, we demonstrate the effectiveness of the causal extraction task. Specifically, through ablation experiment can find that the self-attention mechanism provides significant improvements in the causal extraction task. These results are consistent with our hypothesis that the CSNN model can automatically learn semantics and related features between different features, which is useful for inter-sentence or implicit causality extraction. Besides, we also find that BiLSTM plays an important role in capturing dependencies between causal relationships, thereby improving the accuracy and recall of causality extraction.

Moreover, the result shows the role of self-attention in causality extraction task. The convolutional layer is aimed to capture important features within the windows of causality, but lacks internal correlation learning of important features. For example, in the sentence "the important reason for the repurchase is that Chengdian Medical Star's performance is not up to standard", the reason is "Performance is not up to standard" and the result is "for the repurchase". When the sliding window is 3, the model can learn the overall characteristics of the words "performance", "not" and "up to standard", but the relationship between "not "and" up to standard" can not be learned. Experiments show that in the causality extraction task, self-attention helps to define the cause or result boundary by learning the characteristics of internal words, and makes the

methods	a-Causal-Effect	b-Causal-Effect
Sentence	母公司产品**受市场影响**，毛利率下降。"The parent company's products are **affected by the market** and *the gross profit margin is declined.*"	**原材料价格下跌**，这是产品价格下跌的主要原因。**The price of raw materials fell**, which is the main reason for *the decline in product prices.*
CSNN	**受市场影响** 毛利率下降 **affected by the market** *the gross profit margin is declined*	**原材料价格下跌** 产品价格下跌 **the price of raw materials fell** *the decline in product prices*
CNN-BILSTM-CRF	**市场** 毛利率下降 **market** *the gross profit margin is declined*	**原材料价格下跌** 价格下跌 **the price of raw materials fell** *decline prices*
BILSTM-CRF	None 毛利率下降 None *the gross profit margin is declined*	**原材料价格下跌** 价格下跌 **the price of raw materials fell** *decline prices*
IDCNN-CRF	**市场** 下降 **market** *declined*	**价格下跌** 价格下跌 **decline prices** *decline prices*

Fig. 3. Extraction results by different models, where words are bold represent "Cause", italicize represent "Effect".

cause or result more closely connected in a window. Thus, the effect of causality extraction is improved.

Figure 3 shows some typical cases of inter-sentence or implicit causal extraction to show the advantages and disadvantages of our proposed model compared with other methods. Figure 3a is an example of an implicit causal relationship. Only CSNN captures the relationship between them internally and accurately extracts the causal-effect description. For the causality extraction model, learning the intrinsic relationship of causality is very helpful in defining the boundary.

Due to ambiguity of the boundary in Chinese causal relation, an event may be a cause or effect in different contexts. When the cause and effect are similar, it will lead to mistakes. As shown in Fig. 3b, in this regard, precision and recall can be promoted based on context information and background knowledge.

5 Conclusions

In recent years, automatic extraction of causal relation from texts has attracted increasing attention of researchers. Existing methods mainly focus on mining intra-sentence or explicit causality, while neglecting the causal relations that expressed implicitly or hidden in inter-sentences. In this paper, we proposed a new method named Cascaded multi-Structure Neural Network (CSNN) for the task of inter-sentence or implicit causality extraction. In the proposed model, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) with self-attention mechanism, and CRF layer are employed to capture semantic relevant characteristics. The output of CNN and self-attention are concatenated as high-level phrase representations, which improves the performance of inter-sentence or implicit causality extraction. The output of CRF layer is the label of each word in inter-sentence or implicit causal relation sentences. In the experiment, we created three annotation datasets on a wide range of public data, which will be released to facilitate ongoing research. The results shown that our model achieved stat-of-the-art performs by improving the F1-scores of causal extraction up to 82.27% and F1-scores of effect extraction up to 75.34% on the created datasets.

Acknowledgments. The research reported in this paper is supported in part by the National Natural Science Foundation of China under the grant No. 91746203, 61991415, 61625304 and the Ant Financial Services Group.

References

1. Blanco, E., Castell, N., Moldovan, D.: Causal relation extraction (2008)
2. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. J. Mach. Learn. Res. **12**(1), 2493–2537 (2011)
3. Dasgupta, T., Saha, R., Dey, L., Naskar, A.: Automatic extraction of causal relations from text using linguistically informed deep neural networks, pp. 306–316 (2018). https://doi.org/10.18653/v1/W18-5035

4. Garcia, D.: COATIS, an NLP system to locate expressions of actions connected by causality links. In: Plaza, E., Benjamins, R. (eds.) EKAW 1997. LNCS, vol. 1319, pp. 347–352. Springer, Heidelberg (1997). https://doi.org/10.1007/BFb0026799
5. Graves, A.: Supervised Sequence Labelling. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-24797-2
6. Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines Vinod Nair, pp. 807–814 (2010)
7. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 (2012)
8. Huang, Z., Xu, W., Yu, K.: Bidirectional LSTM-CRF models for sequence tagging (2015)
9. Ittoo, A., Bouma, G.: Extracting explicit and implicit causal relations from sparse, domain-specific texts. In: Muñoz, R., Montoyo, A., Métais, E. (eds.) NLDB 2011. LNCS, vol. 6716, pp. 52–63. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22327-3_6
10. Fu, J.-F., Liu, Z.-T., Liu, W., L., Zhou, W.: Event causal relation extraction based on cascaded conditional random fields. Pattern Recogn. Artif. Intell. **24**(4), 567–573 (2011)
11. Khoo, C., Chan, S., Niu, Y.: Extracting causal knowledge from a medical database using graphical patterns. In: Proceedings of 38th Annual Meeting of the ACL, Hong Kong (2002)
12. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. Kontos, J., Sidiropoulou, M.: On the acquisition of causal knowledge from scientific texts with attribute grammars. Literary Linguist. Comput. **4**(1), 31–48 (1991)
14. Lafferty, J., Mccallum, A., Pereira, F.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the ICML, January 2002
15. Li, Z., Li, Q., Zou, X., Ren, J.: Causality extraction based on self-attentive BiLSTM-CRF with transferred embeddings (2019)
16. Li, P., Mao, K.: Knowledge-oriented convolutional neural network for causal relation extraction from natural language texts. Pattern Recogn. Artif. Intell. **115**, 512–523 (2019)
17. Mackie, J.: The cement of the universe: a study of causation, vol. 42, no. 3, pp. 7930–7946 (1974)
18. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, vol. 6 (2013)
19. Oh, J.H., Torisawa, K., Hashimoto, C., Sano, M., De Saeger, S., Ohtake, K.: Why-question answering using intra- and inter-sentential causal relations. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Long Papers. vol. 1, pp. 1733–1743. Association for Computational Linguistics, Sofia, Bulgaria, August 2013
20. Paulus, R., Xiong, C., Socher, R.: A deep reinforced model for abstractive summarization (2017)
21. Shen, T., Zhou, T., Long, G., Jiang, J., Pan, S., Zhang, C.: Disan: directional self-attention network for RNN/CNN-free language understanding (2018)
22. Strubell, E., Verga, P., Belanger, D., McCallum, A.: Fast and accurate entity recognition with iterated dilated convolutions (2017)

23. Vaswani, A., et al.: Attention is all you need, pp. 5998–6008 (2017)
24. Wang, X., Yuan, S., Zhang, H., Liu, Y.: Estimation of inter-sentiment correlations employing deep neural network models (2018)
25. Xinzhi Wang, Hui Zhang, Y.L.: Sentence vector model based on implicit word vector expression. IEEE Access p. 17455–17463 (2018)