Proceedings of the 4th Financial Narrative Processing Workshop @ LREC 2022, pages 131–134
Marseille, 24 June 2022

ATL at FinCausal 2022: Transformer based Architecture for Automatic Causal Sentence Detection and Cause-Effect Extraction

Abir Naskar, Tirthankar Dasgupta , Sudeshna Jana, Lipika Dey
TCS Research
{abir.naskar, dasgupta.tirthankar, sudeshna.jana,lipika.dey}@tcs.com

Abstract

Automatic extraction of cause-effect relationships from natural language texts is a challenging open problem in Artificial Intelligence. Most of the early attempts at its solution used manually constructed linguistic and syntactic rules on restricted domain data sets. With the advent of big data, and the recent popularization of deep learning, the paradigm to tackle this problem has slowly shifted. In this work we proposed a transformer based architecture to automatically detect causal sentences from textual mentions and then identify the corresponding cause-effect relations. We describe our submission to the FinCausal 2022 shared task based on this method. Our model achieves a F1-score of 0.99 for the Task-1 and F1-score of 0.60 for Task-2 on the shared task data set on financial documents.

Keywords: Causality extraction, Explicit causality , Implicit causality, Inter-sentential causality, BERT Transformer

1. Introduction

The proliferation of advance Natural language Processing and Machine Learning techniques (Bui et al., 2010) has tremendously helped develop intelligent agents that can extract meaningful information from various sources like, web pages, blogs, news articles, tweets and social media posts. Assimilation of such information with proper reasoning strategies can help these agents in the quest for new knowledge. One of the key abilities of such an agent is to perceive an event and reason about its cause and the potential impacts through causal reasoning.

The concept of causality can be informally introduced as a relationship between two events e_1 and e_2 such that occurrence of e_1 results in the occurrence of e_2 (Girju and Moldovan, 2002; Chan et al., 2002). For example, in the sentence "*Aston Martin is recalling 7,256 vehicles because the seat heaters are getting too hot*", the event "*seat heaters are getting too hot*" is causing the event "*Aston Martin is recalling 7,256 vehicles.*". The extraction of causal relations from textual mentions is an important step for the improvement of many Natural Language Processing applications such as question answering (Sorgente et al., 2013; Blanco et al., 2008), information extraction, knowledge graphs and document summarization. In particular, it enables the possibility to reason about the detected events (Girju, 2003) beside creation of new insights and for the support of the predictive analysis. Natural language texts contain an abundance of such relations appearing in different forms. Even a single sentence expressing causal relations can be arbitrarily complex and varied in structure that makes the extraction task challenging. Indeed, there are few explicit lexico-syntactic patterns that are in exact correspondence with a causal relation while there is a huge number of cases that can evoke a causal relation not in a uniquely way.

Most of the traditional approaches of causality detection are are either based on pattern or rule engineering techniques or use statistical machine learning (ML) models (Khoo et al., 1998; Khoo et al., 2001). Rule based approaches are restricted to particular domains, and thus, cannot be generalized in a real-world scenario. On the other hand, ML models uses sparse features such as bag-of-words, part-of-speech tags and dependency relations, which can suffer from the drawbacks of time-consuming feature engineering problem. There is a recent surge of interest in deep neural network-based models that are based on continuous-space representation (Yih et al., 2015) of the input and non-linear functions. Thus, such models are capable of modeling complex patterns in data and since they do not depend on manual engineering of features, they can be applied to solve problems in an end-to-end fashion.

In this paper we present two independent transformer based deep neural network architectures for the causal sentence classification and cause-effect relation extraction task. We have used the fine-tuned Bidirectional Encoder Representations from Transformers (BERT) language model cascaded with a sequence-labeling architecture (Zhou and Xu, 2015). The proposed models solves the two tasks comprised of - (i). classifying sentences into two categories - causal and non-causal (ii). Labeling appropriate sub-sequences in a causal sentence as cause, effect and connective. The labeling of connectives is a unique proposition of the work, which along with its companion cause and effect pair, helps in detection of causal relations from complex sentences more effectively.

2. The Task Definition and Data sets

As part of the Financial Narrative workshop, the FinCausal-2022 [1] focused on detecting if an object, an event or a chain of events is considered a cause for a

[1] https://wp.lancs.ac.uk/cfie/shared-tasks/

	Task-1	Task-2
Avg. no. of sentences	1.3	1.6
Avg. no. of words	34.7	48.2
Max no. of word in document	298	176
Max no. of sentence in document	5	5
number of positive label	1281	N.A
number of negative label	12228	N.A

Table 1: Data statistics for task-1 and task-2.

prior event. This shared task focuses on determining causality associated with a quantified fact. Accordingly the shared task is composed of the following two sub-tasks:

- **Task 1:** is a binary classification task. The data set consists of a sample of text sections labeled with 1 if the text section is considered containing a causal relation, 0 otherwise. The data set is by nature unbalanced, as to reflect the proportion of causal sentences extracted from the original news and SEC corpus, with provisional distribution approximately 5% 1 and 95% 0.
- **Task-2:** is a relation extraction task. The text sections will correspond to the ones labeled as 1 in the Task 1 data set, though for the purpose of results evaluation, they will not be exactly the same in the blind test set. The purpose of this task is to extract, in a causal text section, the sub-string identifying the causal elements and the sub-string describing the effects.

The data are extracted from a corpus of 2019 financial news provided by QWAM. The original raw corpus is an ensemble of HTML pages corresponding to daily information retrieval from financial news feed. These news mostly inform on the 2019 financial landscape, but can also contain information related to politics, micro economics or other topic considered relevant for finance information. There are 13516 documents for task1. For task2, there are 2014 unique documents. For each document cause and effect parts are marked. For some documents there may be multiple cause and effect pair. Total 2290 pair are annotated for all documents. The details about the data statistics for both Task-1 and Task-2 [2]. is depicted in Table 1.

3. Overview of proposed causal entity extraction and classification framework

BERT (Bidirectional Encoder Representations from Transformers) (Vaswani et al., 2017) is widely used now a days in several NLP tasks and it actually works well in most of the cases. We implemented a sequence-to-sequence model for cause and effect term extraction and a binary classifier for classification of the causal documents. Initially several rule based systems (Mirza and Tonelli, 2016; Sorgente et al., 2013) are used to extract cause ans effect from sentences or to classify causal sentences. Then several deep learning models (Dasgupta et al., 2018) were came into fashion. We use transformer based architecture with pre-trained BERT to train our both models.

3.1. Architecture for causal document classification model

A BERT based classification model in figure:1 is used to classify a document is a causal sentence or not.

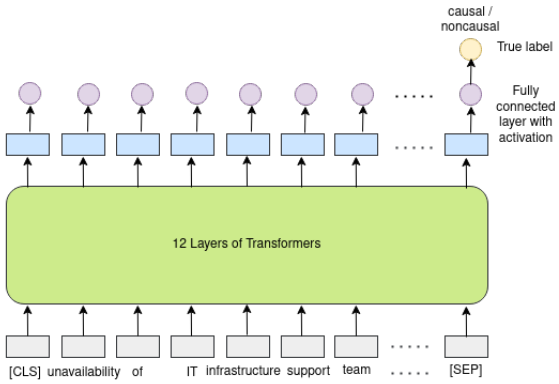

Figure 1: Proposed architecture for Causal sentence classification (Task-1).

In our proposed model we pass the document $D = \{w_1, w_2, \ldots, w_n\}$ which consists of n words and the goal of this sub task is to predict the binary label y of the document D, where $y \in \{0, 1\}$ where the label 0 stands for non causal document and 1 is for causal document. For example, take the document below,
D= *"If the energy sector in Canada continues down this steep decline that's been caused by legislation over the last three or four years, it will get so much worse for Canadians in terms of jobs and also in terms of revenue across all three levels of government which provide the social services and the public programs that Canadians deserve and expect."*
This example document is a causal document. And for another document,
D= *"While the Speaker's office disclaimed the leaked version, saying it is out of date, the draft reveals several noteworthy Democratic policy options likely being discussed including Medicare negotiation, capping drug prices at an International Price Index, capping out-of-pocket costs for Part D beneficiaries, and establishing an inflation rebate for drugs whose prices rise too fast."*
This is not a causal document.
For the purpose of many to one set up, we take the BERT output and send them into a fully connected layer for multiclass classification. Then the output of the fully connected layer is matched with the original label. To deal with over-fitting a Dropout mechanism in the fully connected layer is used.

[2] https://github.com/yseop/YseopLab

3.2. Architecture for cause effect term extraction model

An almost similar BERT based classification model in figure: 2 is used to identify the portion of the document as cause or effect or none of that. Here we pass

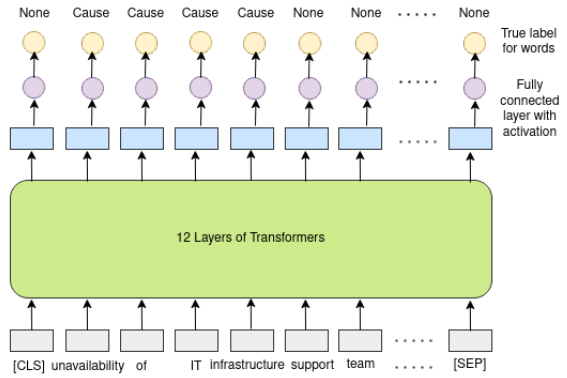

Figure 2: Proposed architecture for Cause-effect extraction (Task-2).

the full document $D = \{w_1, w_2, \ldots, w_n\}$ which consist n word tokens into the BERT based transformer module and the goal of this sub task is to predict the target t, where $t = \{t_1, t_2, \ldots t_n\}$ where $t_i \in \{cause, effect, none\}$. For example, for the document,

D= *"NPAs increased $703 million year over year, primarily due to PCI loans that would have been classified as nonperforming at December 31, 2019 and loans exiting certain accommodation programs related to the CARES Act.Noninterest income increased $3.6 billion for the year with nearly all categories of noninterest income being impacted by the Merger."*

In this example document, The cause portion is *"PCI loans that would have been classified as nonperforming at December 31, 2019 and loans exiting certain accommodation programs related to the CARES Act."*. And the effect part is *"NPAs increased $703 million year over year"*.

For that purpose we send the BERT output sequence into a fully connected sequence to sequence module for predicting the sequence tag $S = \{s_1, s_2, \ldots, s_n\}$. This is then matched with the original sequence label, $Y = \{y_1, y_2, \ldots, y_n\}$. The Dropout technique is used in the fully connected layer to cope up with the overfitting issue. The Cross Entropy Loss is used for back propagation.

4. Experiments and Results

4.1. Experimental settings

We use bert-base-uncased (Devlin et al., 2018) as default backbone network. This use 12 layers, 768-dimensional embeddings. Total 110 million parameters for 12 heads per layer. For both task we keep the hyperparameters same. We use Adam optimizer with learning rate 2×10^{-5}. The dropout rate is used was 0.1. We took batch size of 4 and run that for 10 epoch. We mostly set the same setup for both of our model. The entire data was broken three parts randomly. the 60% data is taken for training purpose, 20% is for evaluation and 20% for test purpose. We run the entire system and test our model in CPU only. It took around 50 minutes to complete 1 epoch.

4.2. Results

We had achieved F-measure 94.3 for Task1. For Task2 we have got exact match for 21.3% and when we proceed with token accuracy excluding the [CLS] and [SEP] tokens we have got the F-measure value as 63.6. Initially we had trained our system for 5 epochs, when we train it for 10 epochs we saw slight improve over accuracy. the precision, recall, F-measure calculated are given below.

	Task1	Task2
Precision	93.2	62.2
Recell	95.6	65.1
F-measure	94.3	63.6
Exact match	N.A	21.3

5. Conclusion

The key idea of the task and build the model is to automatically detecting the causal documents and extracting the cause and effect information. Initially several rules (Guo et al., 2020) and statistical models (Khoo et al., 1998; Khoo et al., 2001) were used for that purpose. in our end-to-end system the document is passed through our proposed model as input and output will be the extracted entities and the class where the document belongs to. Our proposed model focuses all the causes and effects in a document. But it fails to understand the relation between the cause and effect where multiple causal instances and their effect present in the document. For example if for one cause multiple effect happened, or may be there are multiple cause and effects present in the document, we are failing to identify which cause inspires which effect. In some cases the cause portion and the effect portion is so far away from one another that to identify their dependencies will be very difficult. And for our BERT based transformers model there is always a constraint about the number of tokens as input. And we need large corpus of annotated data for that. So we intended to work on those aspects to facilitate research.

6. Bibliographical References

Blanco, E., Castell, N., and Moldovan, D. (2008). Causal relation extraction. In *Lrec*.

Bui, Q.-C., Nualláin, B. Ó., Boucher, C. A., and Sloot, P. M. (2010). Extracting causal relations on hiv drug resistance from literature. *BMC bioinformatics*, 11(1).

Chan, K., Low, B.-T., Lam, W., and Lam, K.-P. (2002). Extracting causation knowledge from natural language texts. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 555–560. Springer.

Dasgupta, T., Saha, R., Dey, L., and Naskar, A. (2018). Automatic extraction of causal relations from text using linguistically informed deep neural networks. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 306–316.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Girju, R. and Moldovan, D. (2002). Mining answers for causation questions. In *AAAI symposium on mining answers from texts and knowledge bases*.

Girju, R. (2003). Automatic detection of causal relations for question answering. In *Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering-Volume 12*. Association for Computational Linguistics.

Guo, S., Jin, L., Yang, J., Jiang, M., Han, L., and An, N. (2020). Causal extraction from the literature of pressure injury and risk factors. In *2020 IEEE International Conference on Knowledge Graph (ICKG)*, pages 581–585. IEEE.

Khoo, C. S., Kornfilt, J., Oddy, R. N., and Myaeng, S. H. (1998). Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary and Linguistic Computing*, 13(4):177–186.

Khoo, C. S., Myaeng, S. H., and Oddy, R. N. (2001). Using cause-effect relations in text to improve information retrieval precision. *Information processing & management*, 37(1):119–145.

Mirza, P. and Tonelli, S. (2016). Catena: Causal and temporal relation extraction from natural language texts. In *The 26th international conference on computational linguistics*, pages 64–75. ACL.

Sorgente, A., Vettigli, G., and Mele, F. (2013). Automatic extraction of cause-effect relations in natural language text. *DART@ AI* IA*, 2013:37–48.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

Yih, W.-t., He, X., and Gao, J. (2015). Deep learning and continuous representations for natural language processing. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorial Abstracts*, pages 6–8.

Zhou, J. and Xu, W. (2015). End-to-end learning of semantic role labeling using recurrent neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1127–1137.