Proceedings of the 4th Financial Narrative Processing Workshop @ LREC 2022, pages 128–130
Marseille, 24 June 2022

ExpertNeurons at FinCausal 2022 Task 2: Causality Extraction for Financial Documents

Joydeb Mondal, Nagaraj Bhat, Pramir Sarkar, Shahid Reza

Oracle AI Services
Bengaluru, India
{ joydeb.mondal, nagaraj.b.bhat, pramir.sarkar, shahid.reza }@ oracle.com

Abstract

This paper provides a novel approach based on transformer models and POS (part of speech) features with an ensemble approach for causality extraction of financial documents for FinCausal 2022 task 2. We provide a solution with intelligent pre-processing and post-processing to detect the number of cause and effect in a financial document and extract the same. Our given approach achieved 90% as F1 score(weighted-average) for the official blind evaluation dataset.

Keywords: financial information extraction, BERT, Causal Inference, Part Of Speech tagging.

1. Introduction

Causality extraction is the extraction of relationship between events in financial documents like finance reports/news etc. Generally the financial causality contains the set of cause and effect span. Extracting such relationship could help gather valuable insights from the documents. The dataset we considered here has both single as well as multiple cause effect relationships. Our approach is based on the sequence labelling of cause effect relationships with Part Of Speech feature support in a BIO scheme. The sequence labelling approach should help deal with extraction of causal text with variable length. We also explore the ensemble method with various transformer models. Our proposed solution outperforms the results on evaluation dataset provided in the task.

2. Dataset

The purpose of the task is to extract cause and effect. The trial & practice set are provided as a csv file with the headers of Index; Text; Cause; Effect. All are separate by semicolon (;). Below are the details of the header field:

- Index : Id of the sample
- Text : Sample text
- Cause: Sequence of text referring to as the cause of the event.
- Effect: Sequence of text referring to as the effect of the event.

Blind/evaluation dataset have only Index and Text.

We noticed that the dataset had samples with multiple cause effect relationships where in a single cause in the text can be mapped to multiple effects or vice versa.

Below table provides the details of Training and evaluation set. Along with the current task samples, we also used the samples from 2020 task.

Data Type	Sample Count
Train	1541
Dev	343
Test	343

Table 1: Data Stats

Index	Text	Cause	effect
1	The increase in net interest income was due primarily to a $152.9 billion increase in average outstanding loans, a $32.6 billion increase in average securities, partially offset by a 78 basis point decrease in earning asset yields.NIM was 3.22% for 2020, down 20 basis points compared to the prior year.	a $152.9 billion increase in average outstanding loans, a $32.6 billion increase in average securities, partially offset by a 78 basis point decrease in earning asset yields.	The increase in net interest income
2	Additional increases in noninterest income were primarily due to higher insurance income driven by improved production levels and acquisitions.	higher insurance income driven by improved production levels and acquisitions.	Additional increases in noninterest income

Table 2: Two Dataset Samples

3. Proposed Approach

3.1 Pre-processing

We have used Stanford CoreNLP Stanza (Manning et al., 2014; Qi et al., 2020) model to tokenize each sample text and created the POS tag and corresponding token.

For sample of multiple cause-effect events, we added an index as special number token and the part of speech tag as

'CD' before each sample to represent it separately with respect to inputs for the model. For extracting causal relations we have used BIO(Begin, Inside, and Outside) tagging scheme with 'C for cause and 'E' for Effect as labels to represent the positional information of the tokens and the semantic roles of the causal events.

Cause			Effect		
Token	POS Tag	BIO Tag	Token	POS Tag	BIO Tag
The	DT	B-E	It	PRP	B-C
Sunshine	NNP	I-E	is	VBZ	I-C
State	NNP	I-E	consistently	RB	I-C
drew	VBD	I-E	one	CD	I-C
in	IN	I-E	of	IN	I-C
. . .	. . .	. . .	. . .	. . .	. . .
17.7	CD	I-E	low	JJ	I-C
billion	CD	I-E	taxes	NNS	I-C

Table 3: Pre-processed Dataset Samples

3.2 Applied Method

We have used pretrained text encoder BERT which generally performs very well in many NLP tasks (Devlin et al., 2018). We use the BERT-base cased model as a pretrained model which consists of 12 transformer layers with hidden dimension of 768. We have used huggingface (Wolf et al., 2019). library which is the most commonly used for this kind of pretrained models. We also experimented with uncased versions and noticed that the cased version performed much better. Hence all of our base models adopted the cased version.

We started with a baseline structure, where we finetuned the BERT-base cased model into simple token classifier where we have added a linear layer given the tokens as inputs and corresponding sequence labels as target.

We have taken the max length as 350 (based on the max text size in the given sample set), batch size as 32, and initial learning rate is set to 5e-05, and we used cross entropy as the loss function. We use cross entropy loss along with Adam Optimizer.

We have extended the model architecture with POS embedding features. We have used POS tags as an embedding and concatenated it with the last hidden state output of BERT's embedding and pass it through the final linear layer. We have used Tesla V100-SXM2 with 16 core to train our system.

3.3 Post-Processing

The predictions from the models are in the form of BIO tags. After concatenating B & I tags we are infer the cause and effect. We also added a set of heuristics to find out the cause-affect pair.

- For prediction, we send the index value to detect multiple events
- If any event has a length less than 4, then we merge it.

We select the longest cause-effect pair if multiple causal chains are present in a given data instance.

4. Evaluation

We have trained multiple pretrained models with the typical loss functions on the train dataset and evaluated the results on the provided blind dataset as well as the test data. We extracted F1 score, Recall, and Precision from codalab evaluation and added our computed F1 score on model. Transformer models including RoBERTa (Robustly Optimized BERT Pre-training Approach) (Liu et al., 2019), BERT Base (Devlin et al., 2018), BERT Large-Cased Whole Word Masking (Devlin et al., 2018) (BWM) were experimented with different hyper-parameter settings. The mentioned results on Table3 indicates the effectiveness of our approach.

Models	F1	Recall	Precision	Exact Match	Test Data Eval Score
Bert base	0.90	0.90	0.90	0.70	0.87
Bert large	0.90	0.90	0.90	0.70	0.89
Roberta	0.90	0.90	0.90	0.70	0.88
Bert base+ Bert large+ Roberta	0.90	0.90	0.91	0.71	0.88

Table 4: Evaluated Model Result on test data

Error analysis show that the cases that were missed were mostly due to wrong linking of cause and effect in multiple cause/inference scenario (Cases that had one cause mapped to two/more effect and vice versa.).

5. Conclusion

In this paper, we explore the causal inference for Fincausal Task 2. Our approach involved experimenting with various transformer models viz, BERT, Roberta, Bert Large with part of speech feature support. We observed that the best results were achieved with an ensemble model of Bert base, Roberta and Bert large with max voting strategy. In future, we would like to explore the pretrained Finance BERT with cause effort link modeling. This should improve the errors due to multiple cause effect linking.

6. Bibliographical References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. CoRR, abs/1810.04805.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In Association for Computational Linguistics (ACL) System Demonstrations, pages 55–60.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew. 2019. Huggingface's transformers: State-of-the-art natural language processing. CoRR, abs/1910.03771.