

The Financial Document Causality Detection Shared Task (FinCausal 2021)

Dominique Mariko

dmariko
@yseop.com

Hanna Aki Akl

habi-akl
@yseop.com

Estelle Labidurie

elabidurie
@yseop.com

Stephane Durfort

sdurfort
@yseop.com

Hugues de Mazancourt

hdemazancourt
@yseop.com

Mahmoud El-Haj

m.el-haj
@lancaster.ac.uk

Abstract

We present the FinCausal 2021 Shared Task on Causality Detection in Financial Documents and discuss the participating systems and results. A total of 6 teams submitted runs across the task and 4 of them contributed with a system description paper. This task is associated with the 3rd Financial Narrative Processing Workshop (FNP 2021), held at Lancaster University, UK, on September 15-16, 2021.

1 Introduction

This shared task is a follow up session of FinCausal 2020 presented at COLING 2020. In this edition we chose to propose only the data and task formerly named Task 2, which is a causality detection task.

2 Data

The data are extracted from a corpus of 2019 financial news provided by Qwam, collected on 14.000 economics and finance websites. 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 topics considered relevant for finance information.

For this edition, the training data have been slightly augmented with 643 examples added in the Practice data set, evaluation data remaining the same as in the 2020 edition. For a detailed overview of the corpus creation and 2020 edition systems, see [Mariko et al. \(2020\)](#). Data are released under the CC0 License.

3 Task

The purpose of this task is to extract, from provided text sections, the chunks identifying the causal sequences and the chunks describing the effects.

The trial and practice samples were provided to participants as csv files with headers: *Index*; *Text*; *Cause*; *Effect*

- **Index:** ID of the text section. Is a concatenation of [file increment . text section index]
- **Text:** Text section extracted from a 2019 news article
- **Cause:** Chunk referencing the cause of an event (event or related object included)
- **Effect:** Chunk referencing the effect of the event

Average statistics on the causes and effects chunks detected in the causal text sections are provided in Table 1. A data sample for the task is provided in Table 2. Interesting results (up to 94.72 F1 score) had been achieved during the 2020 edition, one of the remaining difficulty being the prediction of complex causal chains considered during the annotation process, leading to one text section possibly containing multiple causes or effects.

4 Evaluation

A baseline was provided on the trial samples. Participating systems were ranked on blind Evaluation datasets based on a weighted F1 score, recall, precision, plus an additional Exact Match. Regarding official ranking, weighted metrics from the scikit-learn package were used, and the official evaluation

Metric	Trial	Practice	Evaluation	Total
Total number of text sections	641	1752	638	3031
Total number of uncausal text sections	500	1404	452	2090
Total number of multicausal text sections	141	348	186	941
Average character length of causal chunks	113.73	109.30	112.48	-
Average character length of effect chunks	107.79	102.56	99.66	-

Table 1: Task Data Distribution

Index	Text	Cause	Effect
0009.00052.1	Things got worse when the Wall came down. GDP fell 20% between 1988 and 1993. There were suddenly hundreds of thousands of unemployed in a country that, under Communism, had had full employment.	Things got worse when the Wall came down.	GDP fell 20% between 1988 and 1993.
0009.00052.2	Things got worse when the Wall came down. GDP fell 20% between 1988 and 1993. There were suddenly hundreds of thousands of unemployed in a country that, under Communism, had had full employment.	Things got worse when the Wall came down.	There were suddenly hundreds of thousands of unemployed in a country that, under Communism, had had full employment.
23.00006	In case where SGST refund is not applicable, the state is offering a 15% capital subsidy on investments made in Tamil Nadu till end of 2025.	In case where SGST refund is not applicable	the state is offering a 15% capital subsidy on investments made in Tamil Nadu till end of 2025

Table 2: Three examples from FinCausal 2021 Corpus - Practice dataset

script is available on Github ¹. Participating teams were proposed to submit at most 20 runs, and to choose themselves which run they wished to display on the leaderboard.

Results for the task are provided in Table 3. Last line displays the best 2020 result for the task by Kao et al. (2020) (please note the 2020 edition training set was smaller than the one proposed in 2021). One of the challenge of this task was to rebuilt the correct span of causal chunks, according to the annotation scheme.

5 Participating systems

Most participants took advantage of the systems proposed in 2020 and worked on augmenting them with different strategies.

The winning system by NUS-IDS adapts Kao et al. (2020)’s BERT-CRF with Viterbi decoder, augmenting the initial BIO-scheme with dependency tree relations, mapping the text dependencies into a directed graph and concatenating them with the BERT and POS embeddings, framing their solution as a Graph Neural Network performing a token classification task.

The system ranked second by DSC-IITISM choose to frame the task as a sequence labelling

problem, using additional BIO-scheme with transformers embeddings, the best model averaging ensemble for XLNet + GPT-2 + BWM models. A post processing optimization method is proposed, selecting the longest cause-effect pair when multiple causal chains are present in a given data instance to tackle the Exact Match problem.

The third system by NVJPSFI builds on Becquin (2020), using grid-based ensemble learning among multiple n-best outputs for BERT, RoBERTa and ALBERT models, optimizing the Exact Match metric.

The fourth system by LIORI experiments on sequence to sequence as well as token sequence classification methods, eventually using an ensemble method averaging ALBERT and XLNet.

6 Conclusion

Participants have built interesting augmentations from 2020 edition systems. This and the additional training data proposed for the 2021 shared task allowed to level up the results. Tackling the complex causality problem remains a task in itself though, and would need a dedicated session and specific data to be fully addressed.

¹https://github.com/yseop/YseopLab/tree/develop/FNP_2020_FinCausal

Team	F1	Recall	Precision	EM	BIO	LM	TF	Ens	HS
NUS-IDS (1)	95.56	95.56	95.57	86.05	x	x			
DSC-IITISM (2)	95.51	95.50	95.54	87.77	x		x	x	x
NVJPFISI (3)	94.77	94.78	94.78	87.62			x	x	x
LIORI (4)	94.65	94.67	94.65	78.37			x	x	
NTUNLPL (2020 edition)	94.72	94.70	94.79	82.45	x	x			

Table 3: Task results and approaches adopted by the participating teams. BIO refers to any adaption of the IOB scheme. LM refers to any language model embedding features, excluding the use of Transformers models. TF refers to Transformers architecture and associated embeddings. Ens corresponds to Ensemble Learning method. HS implies some heuristics has been used in the final computation, mostly to adapt the span.

Acknowledgements

We thank the FNP 2021 Committee for the opportunity to propose this shared task.

7 References

References

- Guillaume Becquin. 2020. [GBe at FinCausal 2020, task 2: Span-based causality extraction for financial documents](#). In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 40–44, Barcelona, Spain (Online). COLING.
- Pei-Wei Kao, Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020. [NTUNLPL at FinCausal 2020, task 2:improving causality detection using Viterbi decoder](#). In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 69–73, Barcelona, Spain (Online). COLING.
- Dominique Mariko, Hanna Abi-Akl, Estelle Labidurie, Stephane Durfort, Hugues De Mazancourt, and Mahmoud El-Haj. 2020. [The financial document causality detection shared task \(FinCausal 2020\)](#). In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 23–32, Barcelona, Spain (Online). COLING.