

The Financial Causality Extraction Shared Task (FinCausal 2022)

Dominique Mariko¹, Kim Trottier² and Mahmoud El-Haj³

¹Yseop, France ²HECMontreal, Canada ³Lancaster University, UK

¹lab@yseop.com, ²kim.trottier@hec.ca, ¹m.el-haj@lancaster.ac.uk

Abstract

We present the FinCausal 2020 Shared Task on Causality Detection in Financial Documents and the associated FinCausal dataset, and discuss the participating systems and results. The task focuses on detecting if an object, an event or a chain of events is considered a cause for a prior event. This shared task focuses on determining causality associated with a quantified fact. An event is defined as the arising or emergence of a new object or context in regard to a previous situation. Therefore, the task will emphasise the detection of causality associated with transformation of financial objects embedded in quantified facts. A total number of 7 teams submitted system runs to the FinCausal task and contributed with a system description paper. FinCausal shared task is associated with the 4th Financial Narrative Processing Workshop (FNP 2022) (El-Haj et al., 2022) which is held at the The 13th Language Resources and Evaluation Conference (LREC 2022) in Marseille, France, on June 24, 2022.

1. Introduction

Financial analysis needs factual data, but also explanation on the variability of these data. Data state facts, but provide little to no knowledge regarding how these facts materialised. The Financial Document Causality Detection Task aims to develop an ability to explain, from external sources, the reasons why a transformation occurs in the financial landscape, as a preamble to generating accurate and meaningful financial narrative summaries. Its goal is to evaluate which events or which chain of events can cause a financial object to be modified or an event to occur, regarding a given external context. This context is available in the financial news, but due to the high volatility of such information, mapping an external cause to a given consequence is not trivial.

FinCausal 2022 shared task follows the successful FinCausal shared tasks on 2020 (Mariko et al., 2020) and 2021 (Mariko et al., 2021). In this edition we chose to propose only the data and task details of the Causality Task, formerly named Task 2, which is a causality detection task. The training and evaluation sets have been augmented with extracts of the Management Discussion and Analysis (MD&A) sections from 10k filings found on the EDGAR Company Filings database of the U.S. Securities and Exchange Commission (SEC).

2. Data

The data are extracted from a corpus of 2019 financial news provided by Qwam¹, collected from 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. This edition contains the training data from 2021 (2020 data slightly augmented with 643

examples added in the Practice data set), in addition to the newly created SEC data presented below. For a detailed overview of the corpus creation and 2020 edition systems, see (Mariko et al., 2020). Data are released under the CC0 License.

2.1. 2022 Augmentation

2022 data have been augmented from the 2021 data samples with the following

- 537 data points have been added to the training data
- 934 data points have been added to the blind test data

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

A data sample for the task is provided in Table 1. Interesting results (up to 95.50 F1 score) had been achieved during the 2020 and 2021 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.

¹<https://www.qwamci.com/>

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 1: Three examples from FinCausal 2021 Corpus - Practice dataset

4. Participants and Systems

In total, 7 different teams have participated and submitted their system to FinCausal 2022, the teams are presented in Table 2.

SPOCK team addressed the information extraction problem with span-based and sequence tagging neural network models. Specifically, they fine tuned pre-trained language models to perform text span classification and sequence labeling tasks. They trained a span-based causality extraction system by fine tuning the BERT-Base model. This model resulted in an F1 score of 89.36 and Exact Match score of 81.67. Their best performing model was an ensemble of sequence tagging models based on the BIO scheme using the RoBERTa-Large model, which achieved an F1 score of 94.70 to win the FinCausal 2022 challenge.

DCU-Lorcan employed advanced pre-trained language models (PLMs) to facilitate the causality extraction task as PLMs have been proven to be effective in many NLP tasks including text classification, text generation especially on span extraction/sequence labeling task such as Named-entity Recognition and Question Answering. Building on PLMs, they also propose a heuristically-induced post-processing strategy to refine the system predictions. Their best system (BERT-large + post-process) achieved F-1, Recall, Precision and Exact Match scores of 92.76, 92.77, 92.76 and 68.60 respectively.

LIPI reused the implementation of two the state of the art approaches (Nayak et al., 2022) and (Kao et al., 2020). They trained the CEPN architecture proposed by (Nayak et al., 2022) separately on FinCausal-2020 and FinCausal2021 datasets and evaluated them on the FinCausal2022 data set. Subsequently, they combined the entire labelled dataset available up to 2022 and re-trained the same architecture.

iLab introduced graph construction techniques to inject cause-effect graph knowledge for graph embedding. The graph features combining with BERT embedding, then are used to predict the cause effect spans.

Their results show that their proposed graph builder method outperforms the other methods with and without external knowledge.

MNLP focused their approach on employing Nested NER using the Text-to-Text Transformer (T5) generative transformer models while applying different combinations of datasets and tagging methods. Their system reports accuracy of 79% in Exact Match comparison and F-measure score of 92% token level measurement.

ExpertNeurons proposed a solution with intelligent pre-processing and post-processing to detect the number of cause and effect in a financial document and extract them. This approach achieved 90% as F1 score (weighted-average) for the official blind evaluation dataset.

ATL presented two independent transformer based deep neural network architectures for the causal sentence classification and cause-effect relation extraction task. They have used the fine-tuned Bidirectional Encoder Representations from Transformers (BERT) language model cascaded with a sequence-labeling architecture.

5. Evaluation

We used CodaLab to allow participants to train and test their systems. Table 3 shows the FinCausal 2022 results run on our blind test set. A baseline was provided on the trial samples for the Causality Task Tasks 2)². Participating systems were ranked on blind Evaluation datasets based on a weighted F1 score, recall, precision for Task 1, plus an additional Exact Match for Task 2. Regarding official ranking, weighted metrics from the scikit-learn package³ were used for both Tasks, and the

²https://github.com/yseop/YseopLab/tree/develop/FNP_2020_FinCausal/baseline

³https://scikit-learn.org/stable/modules/model_evaluation.html#multiclass-and-multilabel-classification

Team	Affiliation
SPOCK	Rensselaer Polytechnic Institute and IBM
DCU-Lorcan	Dublin City University
LIPI	Fidelity Investments, Jadavpur University
iLab	National Institute of Advanced Science and Technology, Japan Advanced Institute of Science and Technology, and Tokyo Institute of Technology
MNLP	George Mason University
ExpertNeurons	Oracle
ATL	TCS Research

Table 2: FinCausal 2022 participating teams and their affiliations

official evaluation script is available on Github⁴. Participating teams were allowed to submit up to 100 runs, while only their highest score was withheld to represent them during the evaluation phase⁵. Only the scores validated during the evaluation phase of the competition are displayed below.

Team	F1	R	P	EM
SPOCK	0.95 (1)	0.95 (1)	0.95 (1)	0.86 (1)
ilab	0.94 (2)	0.94 (2)	0.94 (2)	0.83 (2)
DCU Lorcan	0.93 (3)	0.93 (3)	0.93 (3)	0.69 (5)
LIPI	0.92 (4)	0.92 (4)	0.93 (4)	0.79 (3)
MNLP	0.92 (5)	0.92 (5)	0.92 (5)	0.79 (3)
Expert Neurons	0.90 (6)	0.90 (6)	0.91 (6)	0.71 (4)
ATL	0.64 (7)	0.65 (7)	0.62 (7)	0.21 (7)

Table 3: FinCausal 2022 Results. R: Recall, P: Precision, F1, F1 Measure, EM: Exact Match.

6. Conclusion

In this paper, we present the framework and the results for the FinCausal Shared Task. In addition, we present the new FinCausal dataset built specifically for this shared task. We plan to run similar shared tasks in the near future, possibly with some augmented data, in association with the FNP workshop.

Acknowledgements

We would like to thank our dedicated annotators who contributed to the building the FinCausal Corpus: Yagmur Ozturk, Estelle Labidurie, Minh Anh Nguyen, Aurélie Nomblot, Lilia Ait Ouarab, Laura Sasso and Léonie Muller, as well as the FNP Committee for their gracious support.

7. Bibliographical References

Mahmoud El-Haj, et al., editors. (2022). *Proceedings of the 4th Financial Narrative Processing Workshop*,

⁴https://github.com/yseop/YseopLab/tree/develop/FNP_2020_FinCausal/scoring

⁵<https://codalab.lisn.upsaclay.fr/competitions/3802>

Marseille, France, 24 June. The 13th Language Resources and Evaluation Conference, LREC 2022.

Kao, P.-W., Chen, C.-C., Huang, H.-H., and Chen, H.-H. (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.

Mariko, D., Abi-Akl, H., Labidurie, E., Durfort, S., De Mazancourt, H., and El-Haj, M. (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), December. COLING.

Mariko, D., Akl, H. A., Labidurie, E., Durfort, S., de Mazancourt, H., and El-Haj, M. (2021). The financial document causality detection shared task (FinCausal 2021). In *Proceedings of the 3rd Financial Narrative Processing Workshop*, pages 58–60, Lancaster, United Kingdom, 15-16 September. Association for Computational Linguistics.

Nayak, T., Sharma, S., Butala, Y., Dasgupta, K., Goyal, P., and Ganguly, N. (2022). A generative approach for financial causality extraction. *arXiv preprint arXiv:2204.05674*.