

Economic Causal-Chain Search and Economic Indicator Prediction using Textual Data

Kiyoshi Izumi and Hitomi Sano and Hiroki Sakaji

School of Engineering, the University of Tokyo

Hongo 3-8-1, Bunkyo, Tokyo 113-8656, JAPAN

izumi@sys.t.u-tokyo.ac.jp

Abstract

This paper proposes a method that uses causal information extracted from textual data to predict economic indicators. The method automatically extracts causal information included in each sentence using machine learning and natural language processing methods. The extracted cause-effect expressions are stored in an economic causality database. Then, the method can generate causal chains from the given text using the word similarity between a result expression and a cause expression in the database. The causal chains are used to predict how the numerical values of economic indicators will change in the future due to spillover effects from the given text.

1 Introduction

The causality drives economic phenomena. The people involved in the phenomenon predict the future and decide their actions based on their perception of causality. As a result of the accumulation of these actions, the behavior of the entire economic system is determined. For example, consider a causal series (causal chain) that starts with "Aging society. " The aging society has negative effects on the economy in terms of causing a decline in the labor force. It also has positive effects, causing the increasing demand for products for the elderly. Thus, to predict the future of economic phenomena, it is essential to analyze the perception of cause and effect that people have.

It is, however, difficult to statistically analyze the economic causality from numerical data alone because the economic causality involves human behavior. The key to causality is how humans perceive the causal event and their actions in response to it. The perception of economic causality can change over time. It is almost impossible to extract an objective and universal causal series by statistical analysis of numerical data, like natural science.

Therefore, in this study, we analyze textual data in an economic area that contain human-perceived causal relationships and construct a database of economic causality. We propose an algorithm that constructs causal sequences derived from phrases representing specific events and presents economic indicators related to spillover effects or potential causes. Using this method, we can search for spillover effects and potential causes based on causal information expressed in text and use the relationships between events and economic indicators to predict changes in numerical values.

The main contributions of this research are as follows:

- We developed a new method for integrating causality search using textual data (unstructured data) and economic indicators (structured data) prediction.
- This method enables us to predict the change of economic indicators by tracing a causal sequence from a text representing an event of interest.
- This method can give prediction results in an explainable form that is intuitively understandable by humans.

2 Related Works

In recent years, much research concerning causal information extraction from natural language is based on neural networks.

For example, Dasgupta et al.(Dasgupta et al., 2018) proposed a method for extracting causal information using Long short-term memory (LSTM) architecture.

Furthermore, concerning English causal information extraction, various methods were proposed at the Financial Narrative Processing Workshops (FNP), which is a workshop of Colling 2020 because it is included in the Shared Task FinCausal 2020 of the workshop.

Two types of tasks were set in the shared task, extraction of causal information sentences and extraction of causal-effect expressions from causal information sentences.

Most proposed methods for extracting causal information sentences are based on BERT consisting of Transformer and achieved high performance (Ionescu et al., 2020; Gordeev et al., 2020; Gupta, 2020).

Additionally, BERT based method has also been proposed in the extraction causal expressions task (Imoto and Ito, 2020).

Researches concerning the construction of causal chains, such as Ishii et al. (Ishii et al., 2012), Alashri et al. (Alashri et al., 2018), and Zhao et al. (Zhao et al., 2017) exist.

Ishii et al. proposed a method of constructing causal networks by extracting causal expressions from newspaper articles and combining them with SVO based on the hypernym-hyponym relation dictionary.

Alshri et al. proposed a concept-based causal chain construction method.

Zhao et al. proposed a method for constructing causal chains by a method that considers the cause-to-effect and effect-to-cause paths.

In addition, various applications of causality other than constructing causal chains are expected.

For example, in the world of robotics, Causal World (Ahmed et al., 2020) a new benchmark that considers causality has been proposed.

In the existing research mentioned above, it is limited to causal extraction and causal chain construction, and it is not clear what kind of event or concrete numerical value it is related to.

Therefore, several methods are required to use it for actual economic analysis.

On the other hand, this research can be linked to a numerical index, and the change of the numerical value can be predicted.

It is the novel point of this research.

3 Economic causality detection

First, we analyze textual data containing causal economic information recognized by humans and extract cause-effect expressions from them. In this system, we extracted causal relationships from the text of financial statements, which listed companies regularly publish to disclose their business performance and financial status, using a method that uses cue expressions (Sakaji et al., 2017).

| |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Input: A list of cause-effect expressions CI $CI_i = (\text{Cause Expression } c_i, \text{Effect Expression } e_i, \text{Company } cp_i, \text{Date } d_i)$</p> <p>Output: A list of causal chain LCC</p> <pre> 1: $LCC \leftarrow \emptyset$ 2: for each $(c_i, e_i, cp_i, d_i) \in CI$ do 3: for each $(c_j, e_j, cp_j, d_j) \in CI$ do 4: $similarity \leftarrow getSimilarity(e_i, c_j)$ 5: if $similarity \geq threshold$ then 6: $LCC \leftarrow LCC + (c_i, e_i, cp_i, d_i, c_j, e_j, cp_j, d_j)$ 7: end if 8: end for 9: end for 10: return LCC </pre> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Figure 1: Economic causal-chain construction

- Textual data: approximately 20,000 financial summary texts issued by approximately 2,300 companies between October 2012 and May 2018.
- Extracted causal-effect expressions: 1,078,542 pairs.

The extracted causal-effect expressions are stored in the database along with the publishing date and the company name of the financial summary containing the causality information.

4 Economic causal-chain construction

To construct causal chains from the economic causality database, we connect an effect expression of a cause-effect expression and a cause expression of another cause-effect expression. We show an algorithm (Izumi and Sakaji, 2020) for constructing causal chains in Figure 1.

In Figure 2, “Company” indicates the company that issues the financial statement summary from which the cause-effect expression has been extracted. Additionally, “Date” is the date the financial statement summary was published. In Figure 1 $getSimilarity(e_i; c_j)$ is a function that calculates the similarity between the effect expression e_i and the cause expression c_j . Our method estimates the similarities based on vectors of word embedding. First, our method obtains the word embedding average of the words included in the expressions. Here, we define the average obtained from the effect expression e_i as \bar{W}_{e_i} and the average obtained from the cause expression c_j as \bar{W}_{c_j} . $\bar{W}_{e_i}, \bar{W}_{c_j} \in R^m$ and m is the dimension size of word embedding. Then, our method calculates a cosine similarity between \bar{W}_{e_i} and \bar{W}_{c_j} and employs the similarity as the similarity between the effect expression e_i and

the cause expression c_j . Finally, our approach acquires pairs of cause-effect expressions as a causal chain when the similarities are larger than a threshold.

5 Numerical economic forecasting using causal information

A causal chain is constructed using the algorithm in the previous section from the user’s input phrase. Then the following method is used to estimate the numerical values that are expected to change as ripple effects and potential factors (Figure 2).

5.1 Forward search

In the case of forward causal-chain (ripple effect) search (Figure 2a), the procedure is as follows.

1. Construct a causal chain that represents the spillover effects from a text representing a particular phenomenon.
2. The text contained in the cause and effect expressions that appear in the causal chain becomes the expression related to the ripple effect from the first specific phenomenon.
3. When the text related to the spillover effect is given, the system presents the related numerical index from the text using the learning results using the combination of the numerical index and the related text set.
4. For the text data given by the user, the numerical indexes related to the result representation are presented as prediction results as economic indicators that are likely to change as a ripple effect.

5.2 Backward search

In the case of backward causal-chain (latent cause) search (Figure 2b), the procedure is as follows.

1. Construct a causal chain representing the potential causes from the text representing a particular phenomenon.
2. The text contained in the cause and effect expressions that appear in the causal chain becomes the expression related to the potential cause of the first specific phenomenon.
3. When the text related to the potential factor is given, the system presents the related numerical index based on the results of learning

using the combination of the numerical index and the set of related texts done in advance.

4. For the text data given by the user, the numerical indexes related to the causal expressions are presented as prediction results as economic indicators that are likely to change as potential factors.

6 Experiments to predict economic indicators

In this section, we show the experimental results using proposed methods. In these experiments, we used the alternative data provided by the company “Deep Data Research” to predict economic indicators. This data consists of monthly corporate reports published on their official homepages. Therefore, this alternative data includes the economic indicators, the numeric values, and some economic texts describing the company’s situation. We used almost 150,000 data with non-blank texts in these experiments out of 470,000 data (from Jan.2015 to Dec.2020).

6.1 Experimental Methods

1. Extracting causal information:

Given arbitrary texts and periods extract causal information to causal-chain search.

2. Extracting the economic indicators:

Economic indicators that have the causal information in the economic texts of alternative data are extracted, and each frequency is counted.

3. Calculating the relevance:

The odds ratio between causal information and economic indicators is calculated. Furthermore, the economic indicators which have the highest odds ratio are detected as the results of prediction. The odds ratio (represented as “R”) is shown as equation(1)-(3).

$$R = \frac{(P + 0.5) * (1 - Q + 0.5)}{(1 - P + 0.5) * (Q + 0.5)} \quad (1)$$

$$P = \frac{Pa}{Pb} \quad (2)$$

$$Q = \frac{Qa}{Qb} \quad (3)$$

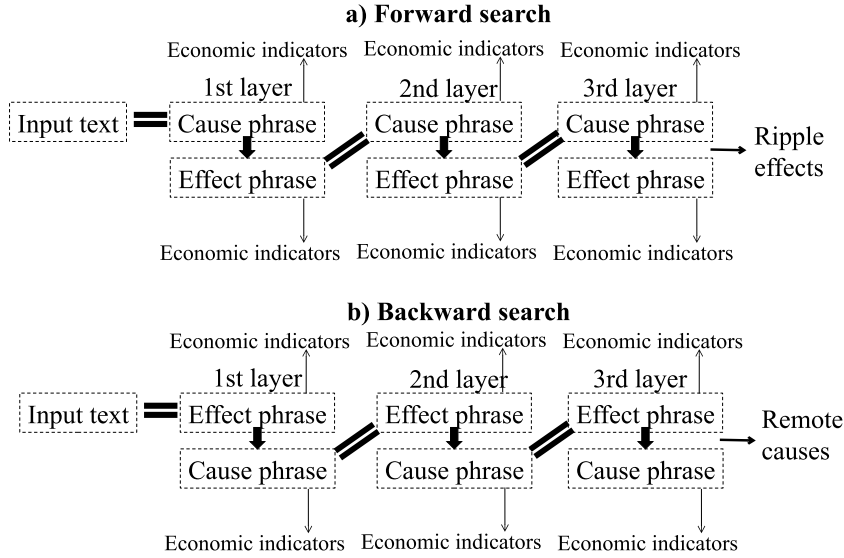


Figure 2: Economic causal-chain and related economic indicators.

- Pa: The frequency of target indicator
- Pb: The summary of the target indicator
- Qa: The frequency of all indicators
- Qb: The summary of all indicators

6.2 Experimental Results

We show the experimental results to predict economic indicators related to causal information in the following 4 cases. The First 3 cases give the texts (“Infectious disease,” “US presidential election,” and “Global warming”) to causal-chain search. The following case gives a text “Olympics” and two divided periods to causal-chain search.

Prediction of spillover in the case of “Infectious diseases”

The text “Infectious disease” was given to the causal-chain search, and the causal information was extracted, repeating the spillover effects of multiple layers. Then, the economic indicators related to the causal information were predicted using our proposed methods. The causal information of the second layer in the causal-chain search and the top 5 economic indicators that are predicted to be strongly related to the causal information is shown in Figure 3. Furthermore, among these economic indicators, the numerical change of “room occupancy rate” is shown in Figure 4.

Regarding Figure 4, the transition of “Guest room occupancy rate” decreased sharply around May 2020, when the lockdown was announced in Japan due to the spread of COVID-19. Since the numerical values fluctuate more than usual, it

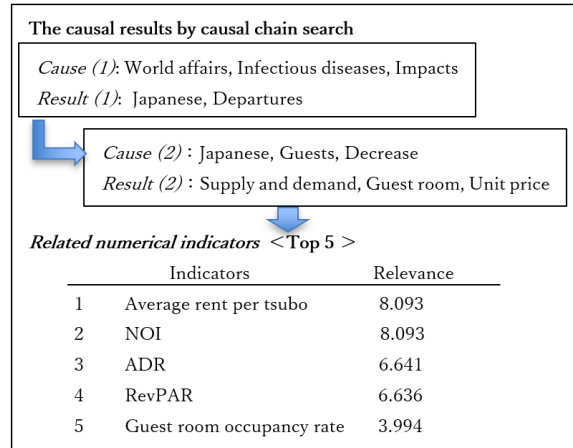


Figure 3: Prediction of spillover in the case of “Infectious diseases”

is considered that the predicted related economic indicators are strongly related to the causal information that spillover from “infectious diseases.”

Prediction of spillover in the case of “US presidential election”

The text “US presidential election” was given to the causal-chain search, and the extracted causal information of the second layer in the causal-chain search was “Yen depreciation,” “High stock prices,” “Economy,” and “Boom”. The top 3 economic indicators that are predicted to be strongly related to the causal information are shown in Figure 5. Furthermore, the numerical change of “Amount of foreign exchange” among these economic indicators is shown in Figure 6.

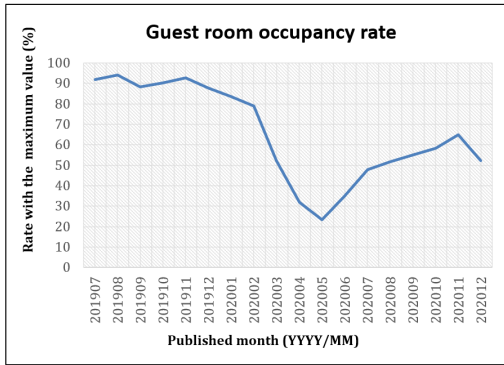


Figure 4: Changes in “Room occupancy rate”

| Indicators | Relevance |
|------------------------------|-----------|
| 1 Deposited assets | 1.200 |
| 2 Amount of foreign exchange | 1.069 |
| 3 Number of accounts | 1.052 |

Figure 5: Prediction of spillover in the case of “US presidential election”

Regarding Figure 6, the amount of foreign exchange rose sharply during both presidential elections (November.2016 and November.2020). COVID-19 caused the rise in February.2020, but the related indicators that spilled over from the “US presidential election” changed at the time of the election.

Prediction of spillover in the case of “Global warming”

The text “Global warming” was given to the causal-chain search, and the extracted causal information of the first layer in the causal-chain search was “Greenhouse gas,” “Emission,” “Reduction,” “Target.” The top 3 economic indicators that are predicted to be strongly related to the causal information are shown in Figure 7. Furthermore, among these economic indicators, the numerical change of “CO2 reduction” is shown in Figure 8.

“Reduction of greenhouse gases,” which is one of the causal information that spilled over from “Global warming,” is in progress as a global goal. And the related numerical index “CO2 reduction” (Figure 8) is on an upward trend over the long term. Therefore, it is considered that the text “Global warming,” causal information, and the economic indicators are strongly related.

Prediction of spillover in the case of

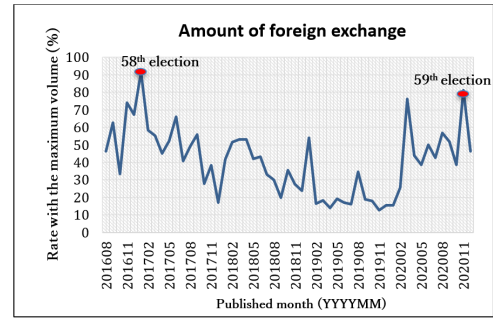


Figure 6: Changes in “Amount of foreign exchange”

| Indicators | Relevance |
|--------------------|-----------|
| 1 Lumber purchase | 1.200 |
| 2 CO2 reduction | 1.069 |
| 3 Power generation | 1.052 |

Figure 7: Prediction of spillover in the case of “Global warming”

“Olympics”

The text “Olympics” and two target periods were given to the causal-chain search, and economic indicators were predicted using the proposed method.

In the case where “January.2016-December.2016” held at the Rio de Janeiro Olympics was given as the target period, the extracted causal information was “Golf,” “Industry,” “Revitalization,” and “Expectation.” These results show the spillover effect of the new addition of golf as a new Olympic sport. The economic indicators predicted from the causal information were “Number of stores,” “Number of customers,” “Unit price per customer,” and “Number of stores opened.”

On the other hand, in the case where “January.2020-December.2020” held at the Tokyo Olympics was given as the target period, the extracted causal information was “Development project,” “Construction period,” “Review,” and “Postponement.” These results show the spillover effect that the new COVID-19 expanded worldwide during this period, and the Olympic Games were postponed for one year. The economic indicators predicted from the causal information were “Number of stores closed” and “Occupancy rate.”

In this way, the causal-chain search visualizes the chain of spillover with given texts and periods. Therefore, applying our proposed method makes

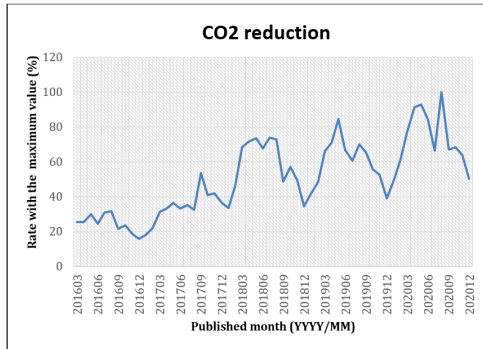


Figure 8: Changes in “CO2 reduction”

it possible to obtain related economic indicators from events that are usually difficult to see the connection and predict trends through causal-chain search.

7 Conclusion

This paper proposes a method that uses causal information extracted from economic texts to predict numerical indicators related to economic and financial fields, such as macroeconomic indicators and stock prices.

Using the proposed method, we identified numerical indicators that are expected to change due to spillover effects from three keywords: “infectious diseases,” “U.S. presidential election,” and “global warming.” The hotel occupancy rate, a numerical indicator related to “infectious diseases,” dropped sharply around April 2020, when a state of emergency was declared nationwide in response to the spread of the new coronavirus infection in Japan. The trading volume of foreign exchange markets, which is a numerical indicator related to the “U.S. presidential election,” has risen substantially during the presidential election periods of November 2016 and November 2020. CO2 reduction, a numerical indicator related to “global warming,” continues to rise. This numerical change indicates that measures to reduce greenhouse gases are continuing.

Furthermore, by giving a target period to the causal chains, more relevant indicators can be extracted. In the case of extracting the relevant indicators of the “Olympic Games,” the number of customers and the number of stores opened were extracted related to golf in 2016, and occupancy rate and the number of stores closed were extracted related to COVID-19 in 2020. In this way, the causal chain can extract numerical indicators that are highly relevant to given text data and periods.

In future work, we will add a method for polarity analysis of texts that appears in the middle of a causal series. This method allows us to predict whether the spillover effects and potential factors estimated from the causal series will impact the relevant numerical indicators in the direction of increasing or decreasing changes.

References

- Ossama Ahmed, Frederik Träuble, Anirudh Goyal, Alexander Neitz, Yoshua Bengio, Bernhard Schölkopf, Manuel Wüthrich, and Stefan Bauer. 2020. [Causalworld: A robotic manipulation benchmark for causal structure and transfer learning](#).
- Saud Alashri, Jiun-Yi Tsai, Anvesh Reddy Koppela, and Hasan Davulcu. 2018. Snowball: Extracting causal chains from climate change text corpora. In *2018 1st International Conference on Data Intelligence and Security (ICDIS)*, pages 234–241.
- Tirthankar Dasgupta, Rupsa Saha, Lipika Dey, and Abir Naskar. 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.
- Denis Gordeev, Adis Davletov, Alexey Rey, and Nikolay Arefiev. 2020. Liori at the fincausal 2020 shared task. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 45–49.
- Sarthak Gupta. 2020. Finlp at fincausal 2020 task 1: Mixture of berts for causal sentence identification in financial texts. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 74–79.
- Toshiya Imoto and Tomoki Ito. 2020. Jdd fincausal 2020, task 2: Financial document causality detection. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 50–54.
- Marius Ionescu, Andrei-Marius Avram, George-Andrei Dima, Dumitru-Clementin Cercel, and Mihai Dascalu. 2020. Upb at fincausal-2020, tasks 1 & 2: Causality analysis in financial documents using pre-trained language models. In *Proceedings of the 1st Joint Workshop on Financial Narrative Processing and MultiLing Financial Summarisation*, pages 55–59.
- Hiroshi Ishii, Qiang Ma, and Masatoshi Yoshikawa. 2012. Incremental construction of causal network from news articles. *Journal of Information Processing*, 20(1):207–215.
- Kiyoshi Izumi and Hiroki Sakaji. 2020. Economic causal-chain search using text mining technology.

In *Artificial Intelligence. IJCAI 2019 International Workshops*, pages 23–35. Springer International Publishing.

H. Sakaji, R. Muroso, H. Sakai, J. Bennett, and K. Izumi. 2017. Discovery of rare causal knowledge from financial statement summaries. In *The 2017 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFER)*, pages 602–608.

Sendong Zhao, Quan Wang, Sean Massung, Bing Qin, Ting Liu, Bin Wang, and ChengXiang Zhai. 2017. Constructing and embedding abstract event causality networks from text snippets. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, page 335–344. Association for Computing Machinery.