Economic Causal-Chain Search using Text Mining Technology

Kiyoshi Izumi and Hiroki Sakaji
School of Engineering, the University of Tokyo, Japan
izumi@sys.t.u-tokyo.ac.jp and sakaji@sys.t.u-tokyo.ac.jp

Abstract

In this research, we extract causal information from textual data and construct a causality database in the economic field. Furthermore, we develop a method to produce causal chains starting from phrases representing specific events and offer possible ripple effects and factors of specific events or situations. Using our method to Japanese textual data, we have implemented a prototype system that can display causal chains for user-entered words. A user can interactively edit the causal chains by selecting appropriate causalities and deleting inappropriate causalities. In this project, we will apply our method to English textual data such as financial news articles and financial reports. The economic causal-chain search algorithm can be applying to various financial information services.

1 Introduction

Economic news articles and financial reports contain various descriptions of cause and effect between economic factors such as price movements, product sales, employment, and trades. For example, "Hospital operator reconsiders London IPO because of Brexit uncertainty" and "the higher prices are likely to take a toll on manufacturers as well as consumers because the economy has decelerated greatly this quarter" appeared in Bloomberg Market News on March 21 2019.

It is beneficial to construct a database of economic causality and analyze the relationship between causality for both financial professionals and non-specialists. Such technology can support professionals' report writing and businesses. For non-specialists, the technology can help them understand the implicit information about causal relationship behind the specialized texts.

It is, however, difficult to analyze the causality between economic phenomena only by the statistical analysis of numerical data. That is because human activities produce a causal relationship between economic phenomena. Human activities are determined by mental processes such as cognition, thinking, and emotion. Thus, economic causality is influenced by social and cultural situations. It is almost impossible to extract objective and universal causality by statistical analysis of numerical data like natural scientific phenomena.

2 Technical ideas

In this program, we analyze economic text data that seems to contain causality recognized by humans and construct a database of causality related to the economic field. Furthermore, we develop a method to search for causal chains derived from phrases representing specific events. Using this method, we implement a system that can display causal chains for user's input words and select appropriate sequences or delete inappropriate sequences. Our method consists of the following steps.

1. Step 1 extracts sentences that include cause-effect expressions (causal sentences) from Japanese financial statement summaries using a support vector machine.
2. Step 2 obtains cause-effect expressions from the extracted sentences using syntactic patterns.
3. Step 3 constructs economic causal chains by connecting each cause-effect expression.

Step 1 and Step 2 are applied a method of [Sakaji *et al.*, 2017].

2.1 Step 1: Extraction of Causal Sentences

We developed a method for extracting causal sentences from economic texts. Since this method uses a support vector machine (SVM) for extraction, we will now explain how to acquire features from financial statement summaries. To extract causal sentences, our method uses the features shown in Fig. 1. We employ both syntactic and semantic features.

Syntactic features
- Pairs of particles

Semantic features
- Extended language ontology

Other features
- Part of speech of morphemes just before clue expressions
- Clue expressions
- Morpheme unigrams
- Morpheme bigrams

Figure 1: Table example

Proceedings of the First Workshop on Financial Technology and Natural Language Processing (FinNLP@IJCAI 2019), pages 61-65, Macao, China, August 12, 2019.

We aim to use expressions that are frequently used in cause and effect expressions in sentences as syntactic features

2.2 Step 2: Extracting Cause-Effect Expressions

We employ a method by [Sakaji *et al.*, 2008] to extract cause-effect expressions using four syntactic patterns. We analyzed sentence structures and used a pattern matching method with syntactic patterns is shown in Fig. 2. In Fig. 2, “Cause” indicates a cause expression, “Effect” indicates an expression of effect and “Clue” indicates a clue expression.

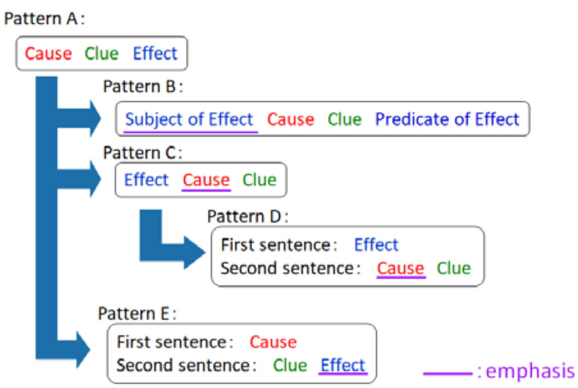

Figure 2: A syntactic patterns list

2.3 Step 3: Constructing Causal Chains

To construct causal chains, our method [Nishumura *et al.*, 2018] connects an effect expression of a causal expression and a cause expression of another causal expression. We show an algorithm of causal chains construction in Fig. 3.

In Fig. 3, “Company” indicates the company that issues the

Input: A list of cause–effect expressions CI
CI_i = (Cause Expression c_i, Effect Expression e_i, Company cp_i, Date d_i)
Output: A list of causal chain LCC
$LCC \leftarrow \emptyset$
for each $(c_i, e_i, cp_i, d_i) \in CI$ **do**
  for each $(c_j, e_j, cp_j, d_j) \in CI$ **do**
    $similarity \leftarrow getSimilarity(e_i, c_j)$
    if $similarity \geq threshold$ **then**
      $LCC \leftarrow LCC + (c_i, e_i, cp_i, d_i, c_j, e_j, cp_j, d_j)$
    end if
  end for
end for
return LCC

Figure 3: Construction of causal chains

financial statement summary from which the cause–effect expression has been extracted. Additionally, “Date” is the date the financial statement summary was issued. In Fig. 3, $\mathrm{getSimilarity}(e_i; c_i)$ is a function that calculates similarity between the effect expression e_i and the cause expression c_j. Our method calculates the similarities based on vectors of word embedding. First, our method obtains word embedding average of the words included in the expressions. Here, we define the average obtained from the effect expression e_i as $\widetilde{W}_{e_i}$ and the average obtained from the cause expression c_j as $\widetilde{W}_{c_i}$. Where $\widetilde{W}_{e_i}, \widetilde{W}_{c_i} \in R^m$ and m is the dimension size of word embedding. Then, our method calculates a cosine similarity between $\widetilde{W}_{e_i}$ and $\widetilde{W}_{c_i}$ and employs the similarity as similarity between the effect expression e_i and the cause expression c_j. Finally, our method acquires pairs of cause-effect expressions as causal chain when the similarities are larger than a threshold.

3 Evaluation

In this section, we evaluate our method until Step 2. For evaluation, we use 30 pdf files of Japanese financial statement summaries and 30 documents of newspaper articles concerning business performance as test data. As results of human tagging, the 30 pdf files include 478 cause-effect expressions, the 30 documents include 51 cause-effect expressions. For classification of causal sentences, we use tagged 3,360 sentences that include 1,454 causal sentences. The tagger is an individual investor with 15 years of investment experience. We use MeCab (http://taku910.github.io/mecab/) for Japanese language morphological analyzer and CaboCha for Japanese dependency parser [Kudo *et al.*, 2002]. Moreover, we employ linear kernel as SVM kernel, and *SVM*^{Light} as SVM.

3.1 Evaluation Results

Table 1 shows experiment results. From Table 1, the method presents good performance for Japanese financial statement summaries and newspaper articles concerning business performance. Results of newspaper outperforms results of financial summaries. Because the method was developed for extracting cause-effect expressions from newspaper articles. However, results of financial summaries satisfy sufficient performance to construct causal chains. Therefore, we think that the method performance is enough to construct causal chains from financial texts.

Table 1: Evaluation results

	Precision	Recall	F1	Number of extracted expressions
Financial summaries	0.82	0.62	0.71	360
Newspaper	0.93	0.75	0.83	34

4 Prototype system

Based on the above-mentioned causal chain construction algorithm, the program of the basic framework of the economic causal chain search system for Japanese texts was implemented. You can try this system at `http://socsim.t.u-tokyo.ac.jp/ccs/`. Based on this system, we will develop English version of the system. The behavior of this system is as follows. First, the user enters the start text (Fig. 4). The user can select the search direction, from cause to result or from result to cause. It is

Figure 4: Text box for entering start text

also possible to limit the search period of textual data.
Click the search button to the right of the text box to display the causality chain from the input text (Fig. 5). By default, three causal relationships are displayed in descending order of similarities. If you want to see more causal relationships, you can click the "More" button to increase the display of causality nodes. If for each node of the causal relationship, the user determines that it is not appropriate, you can delete the node by pressing the delete button at the upper right of each node.

Figure 5: Display of causal relationships

If you want to further extend the causal chain from each node, click the ">" button on the right of each node, and the related causality is added with the clicked node as the terminal node (Fig. 6).
You can build the above-mentioned causal chain repeatedly and construct the causal chain required by the user, you can save the constructed causal chain in a file.

5 Application images

The current prototype system uses only small sized Japanese text, earnings summaries of Japanese firms. In order to improve the precision and recall of acquired causal chains, it is necessary to expand the text data. In this research program, we study the following two things to improve the causality database.

1. Expansion of a causal database using new text data such as news articles.


Figure 6: Extended causal chains

2. Extraction of causal information from English documents such as Form 10-k and press releases, and English database construction.

Our economic causal-chain search system and algorithm can be applied to various financial information services for both individual investors and financial professionals. Within the program period, we will implement the application service prototype program for any of the following application services. After the end of the program, we would like to launch some of the following services in collaboration with financial institutions or financial information vendors.

5.1 Services for Individual Investors

For non-specialists, the technology can help them understand the implicit information about causal relationship behind the specialized texts. One of the causes of this difficulty is the large gap between the knowledge of everyday life and that of finance. From everyday events to financial market trends, there is a causal-chain with financial specific knowledge. The proposed method can implement a service that provides knowledge to fill this gap.

(a) Presentation of background information in financial documents.

Using our algorithm, a user can search related stocks and possible factors derived from keywords and phrases in news articles and economic document-level (Fig. 7). By the influence search, a user can know which stocks' price may be affected by specific economic events and situations denoted in the documents. By the factor search, a user can know

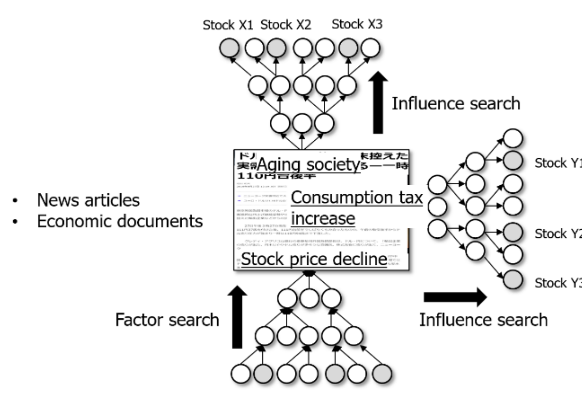

Figure 7: Influence and factor search from financial texts

possible causes of specific economic events and situations.

(b) Question Answering System.

Our algorithm can be applied to an interaction agent service provided by financial institutions for individual investors. Individual investors often want to ask basic questions to financial specialists and advisors. Because face-to-face advice from financial professionals is expensive, automated question answering leads to service penetration (Fig. 8).

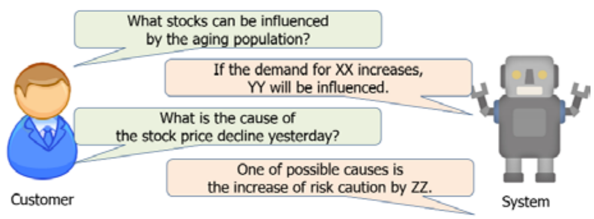

Figure 8: Question Answering System for Individual Investors

5.2 Supports for Financial Professionals Services

Our algorithm can be applied to a business support system for financial professionals in various departments of financial institutions such as market analysts and financial sales.

(a) Support for market report writing.

The proposed method can help market analysts decide the content when writing a report. For example, they search whether there is any influence from a certain event to the market to be explained, and decides whether this event should be written in the report. In addition, for a certain price movement, it is possible to search for potential factors and check whether there are any factors that should be written in the report.

(b) Sales support

Similar to the question answering system for individual investors mentioned above, when salespersons of a financial institution sell their financial products to a customer, they can search for stocks related to personal interests of the customer. If related stocks are searched in advance in relation to the interests of the customers, they can support sales activities. Also, for questions from customers, the above-mentioned question answering system can provide candidates for the contents to be answered.

6 Related work

Much work has been done on the extraction of causal information from texts. Inui et al. proposed a method for extracting causal relations (*cause*, *effect*, *precond* and *means*) from complex sentences containing the Japanese resultative connective “ため(*tame*: because)” [Inui *et al*., 2004], as this is a strong indicator of causal information. Khoo et al. proposed a method to extract cause–effect information from newspaper articles by applying manually created patterns [Khoo *et al*., 1998], as well as a method to extract causal knowledge from medical databases by applying their graphical patterns [Khoo *et al*., 2000]. Chang et al. proposed a method to extract causal relationships between noun phrases using cue expressions and word pair probabilities [Chang *et al*., 2006], defined as the probability that the pair forms a causal noun phrase. Girju proposed a method for automatic detection and extraction of causal relations based on cue phrases [Girju, 2003] where causal relations are expressed by pairs of noun phrases. Girju used WordNet [Fellbaum, 1998] to create semantic constraints for selecting candidate pairs, so her method cannot extract unknown phrases that are not in WordNet. Bethard et al. proposed a method for classifying verb pairs that have causal relationships [Bthard *et al*., 2008] using an SVM for classification. Sadek et al. proposed a method for extracting Arabic causal relations using linguistic patterns [Sadek *et al*., 2016] represented using regular expressions. In contrast, our method not only extract cause-effect expressions but also construct causal chains.

Ishii et al. proposed a method for constructing causal chains using WordNet and SVO tuples [Ishii *et al*., 2012]. They employ method of [Sakaji *et al*., 2008] for extracting cause-effect expressions. Alashri et al. proposed a method to extract causal relations and construct causal chains from large text corpora related to climate change [Alashri *et al*., 2018]. However, their method can not construct causal chains when expressions consist of noun phrases. Because their method targets expressions that include Subjects, Verbs and Objects (SVO). On the other hand, our method is able to construct causal chains from expressions that consist noun phrases only.

7 Conclusions

We develop a method to produce causal chains starting from phrases representing specific events and offer possible ripple effects and factors of specific events or situations. Using our method to Japanese textual data, we have implemented a prototype system that can display causal chains for user-entered words. A user can interactively edit the causal chains by selecting appropriate causalities and deleting inappropriate causalities. In this project, we will apply our method to English textual data such as financial news articles and financial reports. The economic causal-chain search algorithm can be applying to various financial information services.

References

[Alashri *et al*., 2018] Saud Alashri, Jiun-Yi Tsai, Anvesh Reddy Koppela, Hasan Davulcu: Snowball: Extracting Causal Chains from Climate Change Text Corpora, In the 2018 1st International Conference on Data Intelligence and Security (ICDIS), pages 234-241, 2018.

[Bthard *et al*., 2008] Steven Bthard, James H.Martin: Learning Semantic Links from a Corpus of Parallel Temporal and Causal Relations, In Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies, pages 177-180, 2008.

[Chang *et al*., 2006] Du-Seong Chang, Key-Sun Choi: Incremental Cue Phrase Learning and Bootstrapping Method for Causality Extraction Using Cue Phrase and Word Pair Probabilities, Information Processing and Management, pages 662-678, 2006.

[Fellbaum, 1998] Christiane Fellbaum: WordNet: An Electronic Lexical Database, The MIT Press, 1998.

[Girju, 2003] Roxana Girju: Automatic detection of causal relations for Question Answering, In ACL Workshop on Multilingual Summarization and Question Answering, pages 76-83, 2003.

[Inui *et al*., 2004] Takashi Inui, Kentaro Inui, Yuji Matsumoto: Acquiring Causal Knowledge from Text Using the Connective Marker tame, Journal of Information Processing Society of Japan, pages 919-933, 2004.

[Ishii *et al*., 2012] Hiroshi Ishii, Qiang Ma, Masatoshi Yoshikawa, Incremental Construction of Causal Network from News Articles, Journal of information processing, pages 207-215, 2012.

[Khoo *et al*., 1998] Christopher S.G. Khoo, Jaklin Kornfilt, Robert N. Oddy, Sung Hyon Myaeng: Automatic Extraction of Cause-Effect Information from Newspaper Text Without Knowledge-based Inferencing, Literary and Linguistic Computing, pages 177-186, 1998.

[Khoo *et al*., 2000] Christopher S.G. Khoo, Syin Chan, Yun Niu: Extracting Causal Knowledge from a Medical Database Using Graphical Patterns, In Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, pages 336-343, 2000.

[Kudo *et al.*, 2002] Taku Kudo, Yuji Matsumoto: Japanese Dependency Analysis using Cascaded Chunking, In CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002 (COLING 2002 Post-Conference Workshops), pages 63-69, 2002.

[Nishumura *et al*., 2018] Kohei Nishumura, Hiroki Sakaji, Kiyoshi Izumi: Creation of Causal Relation Network using Semantic Similarity, the 32nd Annual Conference of the Japanese Society for Artificial Intelligence, p. 1P104, 2018.

[Sadek *et al*., 2016] Jawad Sadek, Farid Meziane: Extracting Arabic Causal Relations Using Linguistic Patterns, ACM Transactions on Asian and Low-Resource Language Information Processing, pages 1-20, 2016.

[Sakaji *et al*., 2008] Hiroki Sakaji, Satoshi Sekine, and Shigeru Masuyama: Extracting causal knowledge using clue phrases and syntactic patterns. In 7th International Conference on Practical Aspects of Knowledge Management (PAKM), pages 111–122, 2008.

[Sakaji *et al*., 2017] Hiroki Sakaji, Risa Murono, Hiroyuki Sakai, Jason Bennett, and Kiyoshi Izumi: 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, 2017.