

令和 6 年 6 月 5 日現在

機関番号：12601  
研究種目：基盤研究(B)（一般）  
研究期間：2020～2022  
課題番号：20H02410  
研究課題名（和文）将来事象予見型の新しい被害予測技術「リアルタイム物理・社会現象予測AI」の開発

研究課題名（英文）Development of Real-time physical and social phenomenon forecasting AI

研究代表者  
廣井 悠（Hiroi, U）  
東京大学・先端科学技術研究センター・教授

研究者番号：50456141  
交付決定額（研究期間全体）：（直接経費） 15,350,000円

研究成果の概要（和文）：本研究は、これまでの災害現象から社会現象に至るまでのあらゆる因果や、この対応に関する膨大な経験をデータベース化し、これを根拠にデータサイエンスを用いて災害発生直後に次々と発生する事象を予見し、最適な対応を示唆する方法論の確立を最終的な目的とする。そのなかで、本研究においては膨大な因果データベースから災害の特徴を即時的に予測し、将来に発生する事象を定性的に予見しようとする方法論の確立を行った。結果として、新聞記事からほぼ自動的に精度高く因果ネットワークを作成する技術を確立した。

研究成果の学術的意義や社会的意義  
従来の地震被害予測手法は災害発生前の事前対策を評価・意思決定する予防的アプローチとして用いられることが多い。そのため、想定した災害に対する最適解の対応は事前に導き出せるものの、災害直後に「これから何が起きるのか」といった、実際の災害対応の参考となる情報を与えてくれるものでは必ずしもない現状がある。これに対して本研究では、これまでの防災研究の基本スキームを踏襲する従来型の被害想定手法とは全く異なる、将来事象予見型の新しい被害予測技術「リアルタイム物理・社会現象予測AI」の理論部分を開発した。この結果、任意のテキストデータからほぼ自動的に精度高く因果ネットワークを作成する技術が確立可能となった。

研究成果の概要（英文）：In this research, We want to make a methodology to predict events that will occur immediately after a disaster using data science based on a database of all causal factors from disaster phenomena to social phenomena and a vast amount of experience related to these responses, and to suggest optimal responses, finally. In this study, we made a methodology to immediately predict the characteristics of disasters from a huge causal database and to qualitatively predict events that will occur in the future. As a result, we can make a technique to create a causal network from newspaper articles almost automatically and with high accuracy.

研究分野：都市防災

キーワード：災害連鎖 機械学習 被害予測

## 1. 研究開始当初の背景

わが国では首都直下地震や南海トラフ巨大地震をはじめとした激甚な災害による影響が懸念されているが、従来の被害予測手法は災害発生前の事前対策を評価・意思決定する予防的アプローチとして用いられることが多かった。そのため、想定した災害に対する最適解の対応は事前に導き出せるものの、災害直後に「これから何が起きるのか」といった、実際の災害対応の参考となる情報を与えてくれるものでは必ずしもなかった。そこで本研究では、これまでの防災研究の基本スキームを踏襲する従来型の被害想定手法とは全く異なる、将来事象予見型の新しい被害予測技術「リアルタイム物理・社会現象予測 AI」の理論部分を構築することとした。

## 2. 研究の目的

上記の背景を踏まえ、本研究は、これまでの災害現象から社会現象に至るまでのあらゆる因果や、この対応に関する膨大な経験をデータベース化し、これを根拠にデータサイエンスを用いて災害発生直後に次々と発生する事象を予見し、最適な対応を示唆する方法論の確立を最終的な目的とする。そのなかで、本研究においては膨大な因果データベースから災害の特徴を即時的に予測し、将来に発生する事象を定性的に予見しようとする方法論の確立を行った。

## 3. 研究の方法

本研究の方法は、①自然言語処理を用いて新聞記事から災害事象とその因果関係を機械的に抽出し、**cascading disaster network** を作成する作業、そして上記の結果を踏まえ、②機械的な因果の抽出方法並びに **cascading disaster network** 作成の方法論を高度化する作業、に大別される。特に前者については、**ascading disaster** の代表例とも言える過去の大地震災害である阪神淡路大震災と東日本大震災のそれぞれ発生後 6 カ月間の災害事象を対象として、災害間の比較を行っている。詳しい研究手法等については、次章に示す。

## 4. 研究成果

### (1) 自然言語処理による cascading disaster network の作成

ここでは、阪神淡路大震災、東日本大震災の発生当日から半年後の新聞記事（阪神淡路大震災については 1995 年 1 月 17 日から 7 月 16 日まで、東日本大震災については 2011 年 3 月 11 日から 9 月 10 日まで）を災害因果抽出の対象とした。これらの期間に関して「朝日新聞記事データ（学術・研究用）」1995 年版及び 2011 年版のデータフォーマットにおいて、「本誌・地域誌：本誌、面名：1 面から 4 面（スポーツ面以外）及び特設ニュース、記事分類：災害または原子力事故」のすべてに該当する記事のなかから、記事本文に阪神淡路大震災の場合は「兵庫県南部地震」または「大震災」、東日本大震災の場合は「東北地方太平洋沖地震」または「大震災」が含まれるものを対象とした。なお、これらの条件を満たしていても、被害情報の数値の羅列のような明らかに因果関係文を含まない記事は除外している。方法論の詳細は冗長なため、Sakahira and Hiroi (2021) に譲るが、下記に、端的に説明する。

はじめに、手がかり表現と継起因果表現の 2 種類の表現方法を用いて抽出された因果関係文の候補に対して、機械学習による判別モデルを作成する。そのうえで、判別された因果関係文に対して、それら文の原因部分と結果文をノードとリンクとして災害因果ネットワークを作成している。手がかり表現とは、例えば、「を背景に」「のために」のような直接的に因果関係を表す表現のことである。日本語の新聞記事から 35 種類の手がかり表現 (Sakaji et. al 2011, 2014) を用いて因果文の候補を抽出した。しかしながら、日本語の新聞記事は複雑であるため、この段階で抽出された文の中には因果知識が含まれていないものがある可能性がある。そこで、ここでは構文的特徴と意味的特徴を説明変数とする機械学習モデルを作成した。各地震災害に関して、それぞれ発生後 1 カ月間の記事で抽出された因果文の候補のうち約 3 分の 1 にあたる量の文に対して因果関係の有無についてアノテーションを付けた。このうち 70% を学習用、残り 30% をテスト用とした。これら因果関係の有無を目的変数として、サポートベクターマシン (SVM) を用いた機械学習モデルを構築した。説明変数としては、1) 主節と基底節にぶら下がっている助詞の対、2) 主節と基底節にぶら下がっている節の名詞の上位概念、3) 手がかり文節の前の形態素解析における品詞、4) 35 種類の手がかり文節、5) 形態素ユニグラム、6) 形態素ビッググラムを採用した。他方で、継起因果表現とは、時間的な前後関係を表す継起関係に因果関係も含まれるものである。例えば、「高速道路やビルが倒壊して、道路をふさぎ消防車も入れなかった」という記事の文章のなかには、前述の手がかり表現は見られないが、「高速道路やビルが倒壊して」が原因、「道路をふさぎ消防車も入れなかった」が結果という関係が含まれている。日本語の新聞記事から 8 種類の継起因果表現 (Sakahira and Hiroi 2019) を用いて因果文の候補を抽出した。こちらも、この段階で抽出された文の中には因果知識が含まれていないものがある可能性がある。そこで、各地震災害に関して、それぞれ発生後 1 カ月間の記事で抽出された因果文の候補

のうち約 3 分の 1 にあたる量の文に対して 因果関係の有無についてアノテーションを付けた。このうち 70%を学習用、残り 30%をテスト用とした。これら因果関係の有無を目的変数として、SVM を用いた機械学習モデルを構築した。説明変数としては、手がかり表現の判別モデルから若干変更した。主な違いは、Sakahira and Hiroi (2019) に基づき、③継起因果表現の前後の主語の上位概念の一致・不一致を採用していることである。なお、これら構文的特徴と意味的特徴を得るために、構文解析器として CaboCha、形態素解析器として MeCab を用いた。その結果が Table. 1 である。

一方で、Cascading disaster network の作成については、各判別モデルで抽出された因果関係文について原因部と効果部に分けられた因果データベースを用いて因果ネットワークを作成した。因果ネットワークは、開始ノードを原因、終了ノードを結果とした有向グラフから構成されている。そして、この終点ノードを原因として因果データベースから結果を検索して、終点ノードに新しい終点ノードを追加することで因果ネットワークを更新した。具体的には、終点ノードの文と類似性が高い(コサイン類似度 0.7 以上)原因部を持つ因果関係文を因果データベースから検索した。そして、それら因果関係文の結果部を終点ノードの候補とした。終点ノード候補の中から内容を踏まえ人手で選択して、終点ノードを決定した。そして、今度はこの終点ノードを起点とし、その先の終点ノード候補を探索した。この処理を何度でも繰り返すことで因果ネットワークを構築した (Fig. 1)。なお、類似度の算出には、文に含まれる名詞と動詞の分散ベクトルの平均値を用いてコサイン類似度を算出した。分散ベクトルとコサイン類似度の算出は日本語の自然言語処理ライブラリである Ginza を用いた。

本研究ではこれら因果ネットワークから、cascading disaster の定義 (Pescaroli and Alexander 2015, Alexander 2018, Alexander and Pescaroli 2019) を踏まえ、新聞記事に含まれる個人的な被災体験などに関わる部を除外し、重要インフラや社会的事象を中心に再構成して cascading disaster network とした。本論文では以降、各地震災害について 1 カ月間の記事の 30%を用いた各判別をモデルのテスト結果と、これら各モデルを用いて 6 カ月間の記事に適用して作成した cascading disaster network を示す。なお、cascading disaster network を図示する際には、可読性と考慮して、新聞記事の日本語フルセンテンスそのままではなく、原文の意味を可能な限り残したうえで、筆者らが英語に翻訳しうえて要約した句を用いている。

表 2 は、各地震災害における手がかり表現と継起因果表現の各因果判別モデルのテスト結果を示す。平均値は、精度、リコール、F スコアについて、因果文と非因果文の数に応じて加重平均したものであり、合計はデータ数である。手がかり表現の因果判別モデルでの阪神淡路大震災の平均結果は、精度=0.75、リコール=0.74、F スコア=0.74 であった。一方、東日本大震災の平均結果は、精度=0.68、リコール=0.68、F スコア=0.68 であった。また、継起表現の因果判別モデルで阪神淡路の平均結果は、精度=0.75、リコール=0.76、F スコア=0.76 であった。一方、東日本大震災の平均結果は、精度=0.77、リコール=0.78、F スコア=0.77 であった。これらの値については、Sakaji et. al (2011, 2014) の実験結果と比較しても、同程度の精度であることがわかる。従って、本論文でも用いた両モデルは、新聞記事からの災害因果の抽出に適用できると言える。

最終的に、本研究で作成した阪神淡路大震災の cascading disaster network と東日本大震災の cascading disaster network をそれぞれ fig. 2 と fig. 3 に示す。阪神淡路大震災の cascading disaster network の主な特徴としては、地震からの一次被害において「建物倒壊」と「木造住宅の倒壊」が見られ、それらを原因した二次被害が多く抽出されていることである。「建物倒壊」を原因とした結果には「被災地の工場生産の停止」「アスベスト飛散」「病院カルテの喪失」「道路閉塞による消防車侵入の困難」「仮設住宅の不足と需要増」「圧死」「学校等への避難」がある。「木造住宅の倒壊」を原因とした結果には「仮設住宅の不足と需要増」「圧死」「学校等への避難」の他に、「同時多発火災・延焼」がある。また、「被災地の工場生産の停止」を原因とした結果が最後には「経済の先行き不安」や「倒産」まで伸びて長い連鎖となっていることも特徴である。これらの特徴は、阪神淡路大震災の震源が都市部を襲った直下型の地震であることを反映している。一方で、東日本大震災の cascading disaster network の主な特徴としては、地震による断層の破壊を原因とする「津波」が見られ、それを原因した「家屋流出」「塩害」「火力発電所への被害」「漁業施設への被害」「原子力発電への被害」「直接死・ショック死・関連死・行方不明」などが多くの被害が抽出されていることである。また、「原子力発電への被害」を原因とした結果が最後には「世界市場での株安」や「倒産」「国内企業の部品・素材調達の停滞」「海外企業の部品・素材調達の停滞」「避難生活によるストレス」「避難元の税収減」「避難所での疫病流行」まで伸びて長い連鎖となっていることも特徴である。これらの特徴は、東日本大震災の震源が発電所の近海で発生した地震であることを反映している。なお結果の詳細は、本研究の主な成果の一つである Sakahira and Hiroi (2022) の通りとなる。

## (2) 機械的な因果の抽出方法並びに cascading disaster network 作成の高度化

上記の結果を踏まえ、実用レベルに本研究の成果を落とし込むためには、方法論の精度向上が

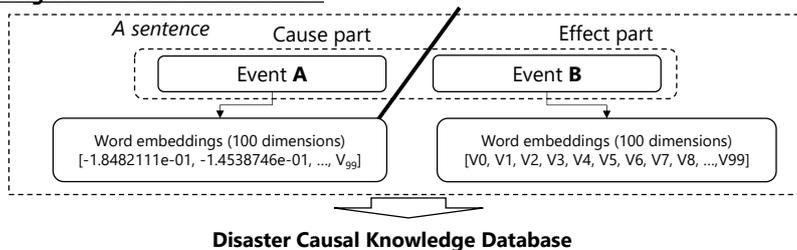
必要であるとの認識に至った。このため、本研究では上記の結果を踏まえて、より精度の高いプログラムを作成した。

はじめに、新聞記事からの因果判定及び因果抽出を、最新の自然言語処理・人工知能技術を用いた改善を行った。上記手法の主な課題は、cue phrase・succession expressionが含まれない文は強制的に除外される、上記を含む文に対して SVM による二値判定を行っているが精度が十分ではないという2点である。因果判定については、学習データ数、パラメータ数、計算速度のバランスが優れていると考えられる東北大学 BERT をファインチューニング (FT) することで、文構造・災害情報の知識にもとづいた因果有無の判定を行うこととする。まず、新聞記事が因果を含むかを人手で判定し、アノテーションデータを作成する。そして、アノテーションデータをもとに文書分類タスクとして BERT を学習し (日本語版 Wikipedia をコーパスとして学修)、学習パターンごとに BERT を作成した。その後、学習した 24 個の BERT を適用して因果有無の属性フラグを付与し、学習した 24 個の BERT を適用して因果有無の属性フラグを付与する方針を取った。因果抽出については、GPT を用いてプロンプトにより原因・結果を抽出する方針を取った。具体的には、新聞記事の中で原因結果がある文に対して原因・結果をアノテーションし、アノテーションデータをもとに GPT を FT (GPT-3.5-turbo-0125 をベースモデルに)、FT した GPT を因果判定で因果ありと判定した文に適用し、抽出した原因・結果それぞれに対して因果判定、そして原因・結果において因果ありと判定された場合はさらに原因・結果を抽出と言う流れである。またここでは、上記 (1) では行えていなかった、全ての文から文内容が記している事象の日時・場所を抽出することとした。類似度算出については、ネットワーク作成に時間が掛かるという課題があった。そのため、抽出した原因・結果が冗長になっており分散表現に不必要な情報が含まれるという点、副詞や新聞記事に使われやすい表現 (ex. 指摘、被害) による影響を受けやすいという点、因果の数が多いという点の課題解決を試みた。結果として、

- ① BERT による因果判定は、Precision を高めたい場合はモデルのパラメータ数やアノテーションデータ量を増やす、Recall を高めたい場合は因果なしの除外割合を高くすることが有効と考えられる
- ② GPT による因果抽出は FT によって一致度合いを Precision と Recall の両面で高めることができるが、パラメータ数が多い場合は FT による精度の向上度合いは小さいと考えられる
- ③ 日付や場所の抽出は文中から抽出するだけでなく、前の文からの継承を行うことによって精度を高めることができる。ただし時間抽出については継承の有効性は小さい

などが判明した。本研究の成果を用いて、熊本地震 (2016/04/14-20) を対象として作成した因果ネットワークが Fig4 である。今後の課題として、座標算出の精度改善への住所補完、類似度算出方法の精度改善、因果ネットワークの自動作成などが挙げられる。

#### Creating the disaster causal database



#### Creating the disaster causal network

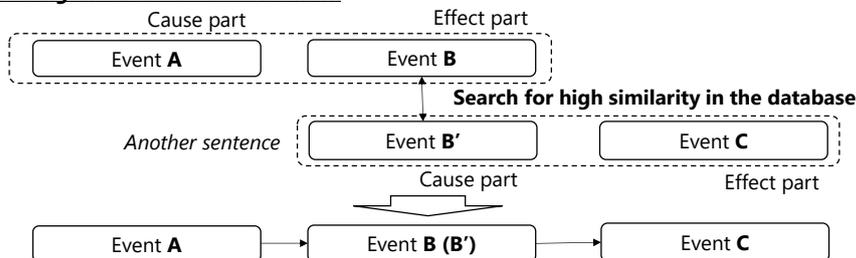


Fig. 1. Creating the disaster causal database and the causal network

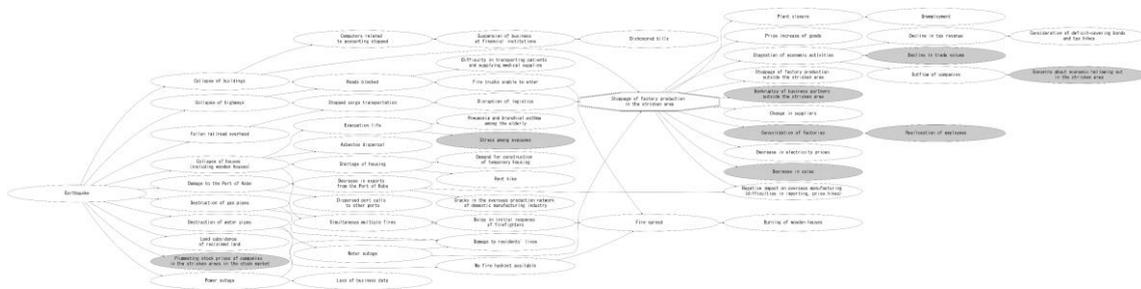


Fig. 2. Cascading disaster network for the Great Hanshin-Awaji Earthquake. Each node is an event, and the arrowed lines show the starting point, which is the cause, and the endpoint, which is the effect

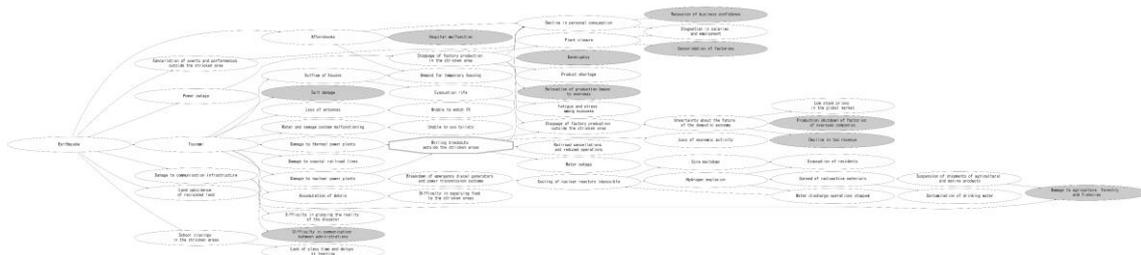


Fig. 3. Cascading disaster network for the Great East Japan Earthquake. Each node is an event, and the arrowed lines show the starting point, which is the cause, and the endpoint, which is the effect

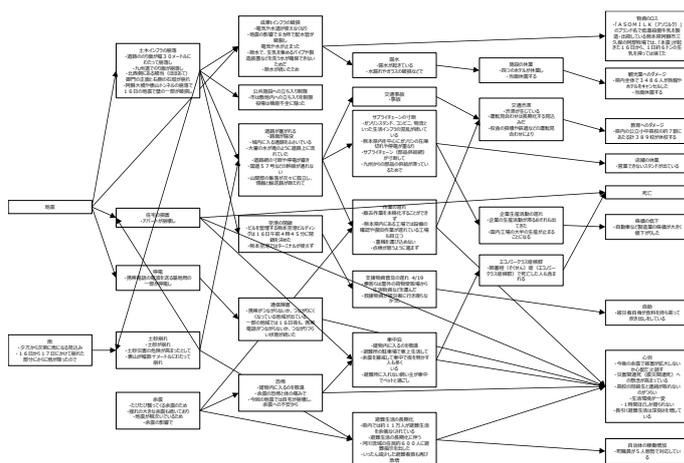


Fig. 4. Cascading disaster network for the 2016 Kumamoto Earthquake. Each node is an event, and the arrowed lines show the starting point, which is the cause, and the endpoint, which is the effect

Table 1. Test results of each causal discriminant model for cue phrases and succession expressions with causation in each earthquake disaster

Model	Earthquake	Sentence	Precision	Recall	F-score	Number
Cue phrases model	The Great Hanshin-Awaji Earthquake	Non-causal	0.83	0.76	0.79	206
		Causal	0.64	0.73	0.68	122
		Average/ Total	0.76	0.75	0.75	328
	The Great East Japan Earthquake	Non-causal	0.78	0.80	0.79	263
		Causal	0.59	0.55	0.57	136
		Average/ Total	0.71	0.72	0.71	399
Succession expressions with causation model	The Great Hanshin-Awaji Earthquake	Non-causal	0.80	0.82	0.81	174
		Causal	0.46	0.42	0.44	62
		Average/ Total	0.71	0.72	0.71	236
	The Great East Japan Earthquake	Non-causal	0.81	0.81	0.81	153
		Causal	0.71	0.70	0.71	101
		Average/ Total	0.77	0.77	0.77	254

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Fumhiro SAKAHIRA, U HIROI	4. 巻 Vol.25, No.3
2. 論文標題 Creating Disaster Chain Diagram from Japanese Newspaper Articles Using Mechanical Methods	5. 発行年 2021年
3. 雑誌名 Journal of Advanced Computational Intelligence and Intelligent Informatics	6. 最初と最後の頁 277-284
掲載論文のDOI（デジタルオブジェクト識別子） 10.20965/jaciii.2021.p0277	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Fumihiko Sakahira, U Hiroi	4. 巻 Volume 66
2. 論文標題 Designing cascading disaster networks by means of natural language processing	5. 発行年 2021年
3. 雑誌名 International Journal of Disaster Risk Reduction	6. 最初と最後の頁 102623
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.ijdrr.2021.102623	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計2件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 廣井悠
2. 発表標題 深層学習を利用した定性的リアルタイム被害予測手法の構築
3. 学会等名 第12回横幹連合コンファレンス
4. 発表年 2021年

1. 発表者名 廣井悠, 坂平文博
2. 発表標題 機械学習を用いた災害連関図の自動作成手法の検討
3. 学会等名 日本災害情報学会第22回大会予稿集
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	関谷 直也  (sekiya naoya)  (30422405)	東京大学・大学院情報学環・学際情報学府・准教授   (12601)	
研究 分担者	坂平 文博  (sakahira fumihiko)  (70578129)	大阪工業大学・情報科学部・准教授   (34406)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------