

# 話題の因果関係の視覚化によるニュース記事群の経時的変化の理解支援

杉原 昂紀<sup>†</sup> 畑 玲音<sup>†</sup> 松下 光範<sup>†</sup>

<sup>†</sup> 関西大学大学院総合情報学研究科 〒569-1095 大阪府高槻市霊山寺町 2-1-1

E-mail: †{k029061,k223167,m\_mat}@kansai-u.ac.jp

**あらまし** 長期間報道されているニュースは時間と共に複数の話題が組み合わさりながら変化するため、ニュースを途中から閲覧するユーザはそのニュースについて理解することが困難である。ニュースの全貌を理解するには話題の変遷を把握することが望ましいが、大量の記事から必要な情報を選ぶのはコストがかかるため、話題の変遷を直観的に理解する支援が求められる。この問題を解決するため、本研究では話題間の因果関係に着目した可視化手法を提案する。提案手法では、ニュースをクラスタリングし、因果関係を基にした話題間の関係性を可視化することでニュースの理解支援を行う。提案手法を使ったユーザ実験の結果、因果関係と時系列の可視化はユーザのニュース理解に一定の効果があることが示唆された。

**キーワード** 因果関係, ニュース, 情報可視化

## A method for visualizing causal relationships between topics in news articles to grasp their changes over time

Koki SUGIHARA<sup>†</sup>, Reon HATA<sup>†</sup>, and Mitsunori MATSUSHITA<sup>†</sup>

<sup>†</sup> Faculty of Informatics, Kansai University 2-1-1 Ryozenji-cho, Takatuki-shi, Osaka 569-1095 Japan

E-mail: †{k029061,k223167,m\_mat}@kansai-u.ac.jp

**Abstract** News reported over a long period undergoes various changes, making it difficult for users who begin reading it later to understand it. When the user wants to grasp a complete picture of the news, he or she must understand how the topic has changed. However, selecting necessary information from many articles is costly, which calls for support in intuitively understanding the development of topics. To address the problem, we propose a visualization method that highlights causal relationships between topics. The proposed method first clustered a set of news articles, and then the relationships between topics based on causality are visualized to aid in understanding the news. The results of user experiments using this proposed method suggest that visualizing causality and chronological sequences assists users' news comprehension to some extent.

**Key words** causal relationship, news, visualization

### 1. まえがき

かつて、ニュースを閲覧する方法は新聞や雑誌などの紙媒体が主流であったが、PC やスマートフォンが普及した現在では電子化されたニュース記事をブラウザやアプリケーションなど様々な方法で閲覧することが一般的になっている。ニュースアプリケーションの普及も顕著で、その利用者数は2017年時点で4,683万人、2020年時点では5,671万人と年々増加している<sup>(注1)</sup>。これらのニュースアプリケーションやサイトで配信されるニュースの記事数は非常に多く、例えばYahoo!ニュース<sup>(注2)</sup>

では1日で約6,000本のニュース記事が配信されている<sup>(注3)</sup>。一度インターネット上で公開されたニュース記事は、掲載から時間が経過してもアクセス可能であり続けるため、ユーザが閲覧可能なニュースの記事数は増加していく。大量のニュース記事に手軽にアクセスできることや過去のニュース記事にもアクセスできることは電子化のメリットであるものの、ユーザが取得可能なニュース記事の増加は、自らの関心に沿ったニュースを収集したり、一連のニュース記事を俯瞰しながら整理したりする際のコストの増大を招く。

石井ら[5]は因果関係の提示がニュースの理解支援に有効で

(注1) : <https://ictr.co.jp/report/20211220.html> (2023/10/20 確認)

(注2) : <https://news.yahoo.co.jp/> (2023/10/20 確認)

(注3) : <https://news.yahoo.co.jp/newshack/pr/infographics/index.html> (2023/10/20 確認)

あるとし、複数のニュース記事から因果関係を抽出し、それらを結合して因果関係ネットワークを構築することでこの問題の解決を試みている。この研究では、「～を受けて」や「～を背景に」などの因果関係の存在を示す手掛りとなる表現（以下、手掛り表現と記す）に着目して文章を原因部分と結果部分に分けて各々をノードで表現し、それらのノードを結合・削減することによってネットワーク図を作成している。しかし、長時間報道されるニュースでは、時間経過と共に関連する記事が増え続け、その記事間の因果関係も膨大となる。そのため、長期間報道されるニュースを把握するためにこのシステムを使用した場合、ユーザが情報過多に陥ることが懸念され、問題の解決には繋がらない。ユーザが必要とする情報を適切に提示するには、ニュース記事を意味的に集約し、それらを話題単位でユーザに提示する事が望ましい。

こうした観点の下、本研究では報道されたニュースを話題ごとに集約し、話題の因果関係をユーザに提示することでニュースの理解を支援することを目指す。その実現の端緒として、本稿ではニュース記事をクラスタリングし、話題間の因果関係を可視化するインタフェースを提案する。

## 2. 関連研究

ニュース記事の理解を目的とした研究は、自然言語処理や情報可視化などの研究分野で様々な視点から行われている。

齋藤ら [11] は、ニュースが報道された時期に着目し、報道されたニュースを話題ごとにクラスタリングし、話題ごとの記事数を時系列に沿って可視化することにより、ニュースの推移の理解支援を行った。この手法では、報道された話題の時系列的变化を読み取ることが可能であるものの、話題間の関係性は読み取れない。また、政治や経済、国際など長期間報道されやすい特定の記事紙面のみを対象としているため、一般家庭に与えた影響など、周辺情報への派生を知ることは出来ない。

河田ら [10] は、情報変化グラフと重要情報グラフの2つを作成し、時系列情報に関する情報変化や、出現する単語の重要性の理解支援を目指した。どちらもニュース記事から抽出された形態素をノードとし、情報変化グラフでは、ノードの単語が前区間から存在したかどうかを色で表現することで、時間経過による文書中に存在する単語の出現や消失が確認できる。重要情報グラフでは、報道された日時別に重要なノードを計算し、ノードの色で表現することにより、高頻度で出現する単語と、それに付随する低頻度な情報が確認できる。しかし、1つの話題に対するニュース記事のみを使用し、記事数も14記事と少ない。長期間報道されるニュースを包括的に理解するには不十分であると考えられる。

文書中の時間経過に伴う話題変遷の抽出や、文書中に存在する話題の検出には Latent Dirichlet Allocation (以下、LDA と記す) [2] や Dynamic Topic Models (以下、DTM と記す) [1] をはじめとしたトピックモデルが用いられることが多い。

武田ら [7] は、ブログ記事からライフイベントに関する有用な情報を抽出するために、時間経過に伴って内容も変化するブログ記事から DTM を用いてトピックを抽出した。DTM を使

用することにより、時間経過によるトピックに含まれる単語の重要度の変化を確認できる。しかし、DTM では期間ごとのトピック数は同数にしか設定できないという制約がある。長期間にわたって報道された一連のニュースの場合、その話題数は常に一定ではなく、話題数の増加や減少が起こりえる。例えば、ウクライナとロシアの話題では、ウクライナとロシアの紛争に関する第一報から小麦の価格の高騰までに、ロシアとウクライナの戦争、ロシアが小麦の輸出を制限、アメリカの利上げ、円安の発生などが話題として報道された。これらは全て別の時期に報道されたわけではなく、複数の話題が重なって報道された時期や、どの話題も報道されなかった時期も存在する。このように、報道される話題は流動的であるため、DTM では正確にトピックの検出ができないと考えられる。

松本ら [8] はこの問題を解決するために、LDA を用いてトピックに分類した後、トピックの類似性からトピックの統合を行い、その後時間軸で分けられたトピックの推移関係の発見を行う手法を提案した。DTM を用いて時間軸で分割した後トピックの分類を行うのではなく、LDA を用いて全ての記事を分類した後に、時間軸で分割することにより、流動的にトピックの数を変化可能であったと報告している。一方でトピックの関係性は単語のコサイン類似度で求めている。文章は単語の組み合わせによって成り立っているため、単語の類似度と文章の類似度は同じとは限らない。例えば、前述したウクライナとロシアの話題では、小麦とロシアの単語分散表現における類似度は近くないと考えられるが、話題の関係性では比較的近い場所に存在する。そのため単語のコサイン類似度を用いて話題の関係性を求めるのは適切でない。

西川らの研究では、専門知識を持たない者を対象とし、特定のテーマに対して言及された記事同士の関連性を把握しやすくすることにより、テーマに対する理解の支援を行う可視化インターフェースを提案した [9]。文書内に含まれるトピックに階層関係があるという仮定に基づき、トピックの推定を行うモデルである hierarchical Latent Dirichlet Allocation (以下、hLDA と記す) を用いてトピックの階層構造を可視化した、その結果、階層構造の有効性と提案システムの有用性が示唆された。一方で、hLDA 法は文書内のトピック数や階層数を予め決定する必要があるため、本来の文書に含まれるトピック数や階層数との誤差が生じかねない。

以上より、話題の数が流動的に変化する長期間報道されるニュースにおいて、時系列と因果関係を可視化するシステムを作成する。

## 3. デザイン指針

本章では、システムを作成するにあたり、システムが満たすべきデザイン指針を定義し、そのデザイン指針を満たすシステムの作成手法について述べる。

本研究で対象とするユーザは、ニュースを普段から閲覧していなかった人を対象としている。対象ユーザが興味のあるニュースを閲覧し、ニュースについて理解を深めたいと感じた時、関連する他の記事を閲覧することが考えられる。その際、

対象ユーザは知識不足によりニュースの内容を時系列で追うことや、内容の因果関係を理解することにコストがかかってしまう。本研究では、ユーザのニュース理解促進のため、ニュースの内容の時系列や因果関係を可視化することで支援を行う。ニュース記事は1つの話題に対して様々な視点で書かれることが多い。対象ユーザは複数記事に対して類似性の判断ができず、読む記事が増加してしまう。そのため、ニュース事を意味的に集約することが求められる。因果関係を把握するためには、様々な日時の記事を閲覧する必要がある。しかし、キーワード検索では比較的新しい記事が検索の上位に入り、古い情報を探しにくい問題や、キーワードから発展した記事を見つけにくい問題が挙げられる。そのため、ユーザが求める話題の前後一定期間のニュース記事のみが閲覧可能なシステムが求められる。

以上より、システムが満たすべきデザイン指針を以下の3点に集約した。

- ニュース記事を話題単位で閲覧できること
- 話題の報道時期が把握できること
- トピック間の因果関係が把握できること

図1に、これらのデザイン指針を満たす可視化表現を示す。この可視化表現では、クラスタリングされたクラスタが時系列に並んでいる。クラスタをクリックすることで、クラスタにあてはまる記事群を閲覧可能になり、因果関係にあるクラスタにエッジが引かれる。因果関係を辿ってニュース記事を見ることによって、ユーザの理解が促進されることが考えられる。

この可視化表現を作成するための手法を、4.章に示す。

## 4. 実 装

本章では提案手法を実装するために行った処理について述べる。特定期間内に報じられたニュース記事をクラスタリングし、全ての記事から因果関係にある単語セットを抽出する。クラスタリングによって得られる特徴語と因果関係にある単語セットを照合することで、クラスタ間の因果関係を定義した。その後、クラスタとクラスタ間の因果関係を可視化することで提案システムを実装した。

### 4.1 クラスタリング

報道された話題の変化をユーザが把握するために、システムは与えられたニュース記事を話題ごとにクラスタリングする必要がある。2.章で述べたように、LDAやDTMによるクラスタリングではクラスタ数を人手で決める必要があり、話題推移の途中で出現したり消滅したりする話題を正確に追うことができない。そのため、本研究ではHierarchical Dirichlet Process Latent Dirichlet Allocation (以下、HDP-LDAと記す)[4]を用いてクラスタリングを行う。この手法は、LDAをノンパラメトリックベイズに拡張した手法であり、LDAではトピック数を人手で決定する必要があるが、HDP-LDAでは適切なトピック数が自動で推定されるという特徴を持つため、話題推移を正確に追うことが可能になる。

新聞記事には時事問題に関する記事の他にも、川柳やイベントに関する広告、コラムなど様々な記事が収録されているため、全ての文章を対象とすると不必要な情報が増えてしまう可能性

がある。そのため新聞記事のうち、長期間報道されるニュースに関連のある特定のキーワードを持つ記事を対象とした。ニュース一般において、内容の容易な理解・処理に重要な役割を果たすのはニュースの最初の発話(リード)である[3]。よって、本研究では形態素解析を用いてリード文から記事の特性が強く表出されると考えられる名詞のみを抽出した。Slothlib<sup>(注4)</sup>に記載されている名詞に加えて、mecab-ipadic-NEologdで未知語と設定されている名詞をストップワードに設定した。また、新聞記事に高頻度で出現する「○月」「○日」「写真」「発表」もストップワードに追加した。

形態素解析によって抽出された名詞を用いて、HDP-LDAによるクラスタリングを行った。HDP-LDAでは、1つのデータは1つのクラスタに属するのではなく複数のクラスタに属するものと考え、データがクラスタに属する確率を出力する。複数のクラスタに同じデータが入っている場合、因果関係抽出の際、同じデータが入ったクラスタが関係性を有すると判断されやすくなる可能性がある。そのため、出力された各記事におけるクラスタに属する確率が最も高いと判断されたクラスタの記事のクラスタとした。クラスタに属する確率が一定の閾値を超えたものを使用した。また、本研究では複数の記事が同じ話題でまとまったクラスタ間の関係性を表示するため、記事数が少ないクラスタは対象にすべきではない。従って、クラスタ内の記事数が5個以下のクラスタは対象から除外した。同時に、クラスタリングによって得られたクラスタ間を結びつけるため、HDP-LDAによって推定された各クラスタの特徴語の、上位3件を抽出した。

### 4.2 話題の報道時期の推定

3.節で述べたとおり、システムには話題の報道時期の提示が求められる。そのため、話題と報道時期の紐付けが必要となる。しかし、前節のクラスタリングではHDP-LDAを使用している。HDP-LDAでは時間軸での分類はなされないため、クラスタに存在する記事の日付から、クラスタの報道時期を推定する必要がある。話題が最も注目されたタイミングで報道された記事には、その話題を象徴する事象が書かれていると考えられる。そのため、各クラスタ内の記事が最も多く報道された月をそのクラスタが報道された月とした。新しい事実の発覚により、同じ話題が複数回ピックアップされて報道されることもあるため、最も多く報道された月が複数ある場合、それら全てを報道された月とした。また、長期間報道されたニュースを対象としているため、その話題が1ヶ月のみならず、月を跨いで数ヶ月間報道が続く場合がある。よって、最も多く報道された月の前後の月が、2つ以上の記事を持つ場合も、報道された月とした。

### 4.3 クラスタの因果関係の抽出

クラスタの因果関係を抽出する方法として、まず、テキストから因果関係を持つ文章を抽出可能である坂地らの手法[6]を用いる。ニュース記事から因果関係のある文章を抽出し、原因部分に存在する単語と、結果部分に存在する単語をそれぞれ抽出する。次に、HDP-LDAによって抽出されたトピックを特徴

(注4) : <https://ja.osdn.net/projects/slothlib/> (2023/10/20 確認)

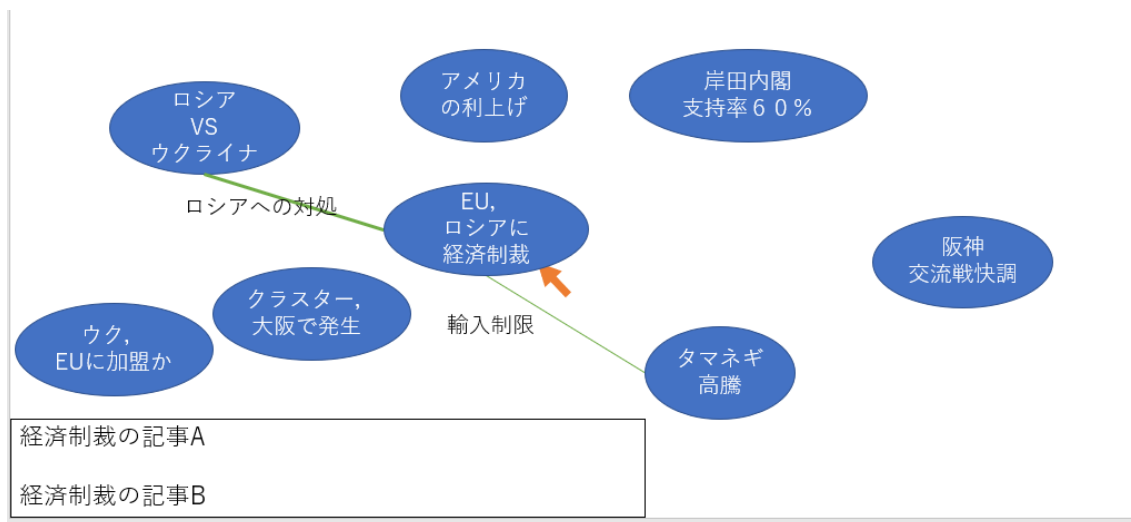


図1 想定する可視化表現。報道された話題のクラスタと、クラスタ間の因果関係が確認できる。

づける単語（以下、特徴語と記す）と、原因部分の単語が一致した場合、原因を表すクラスタとする。同様に結果を表すクラスタを推定し、因果関係にあるクラスタを推定した。

キーワードが含まれる記事から、手掛かり表現を定めて、これらの表現がある文章を抽出した。本研究で定めた手掛かり表現を表1に示す。

表1 手掛かり表現

“よる”, “ため”, “背景”, “受ける”, “伴う”, “反映し”, “きっかけ”, “響く”, “から”, “その結果”
---

このうち、「よる」は品詞が動詞であった場合、「から」は接続助詞であった場合、「反映し」は「反映」という単語の後にレンマ化したものが「する」であった単語が続いていた場合、因果関係があると判断した。それ以外の手掛かり表現には、形態素解析を行った後に形態素をレンマ化し、得られた見出し語が設定した手掛かり表現に一致している場合に因果関係があると判断する手法を採用している。抽出した因果関係を持つ文章から、坂地らの手法で原因表現と結果表現を抽出した。例えば、「中国経済の成長鈍化を背景に原油などの資源価格が大きく下落したことが影響した」という文章では、「背景に」が手掛かり表現になり、「中国経済の成長鈍化」が原因表現、「原油などの資源価格が大きく下落」が結果表現となる。クラスタリングで抽出されたクラスタの特徴語に、原因部分の単語がある場合、そのクラスタを原因クラスタとした。上記の例では、特徴語の上位3件が「構築、成長、簡単」であるクラスタが存在し、原因表現に「成長」という単語があるので、これを原因のクラスタとした。同様に結果クラスタを定め、各文章の手掛かり表現に対する原因のクラスタと結果のクラスタを推定した。原因のクラスタと結果のクラスタの両方が一致しているものがあった場合、それらのクラスタを原因と結果で結ばれているクラスタとした。また、通常因果関係がある文章では、原因部分の内容は結果部分の内容よりも時間的に前に存在する。そのため、原因部分の内容が結果部分の内容よりも時間的に後に存在するのは不自然で

ある。よって、原因クラスタが結果クラスタよりも時間的に新しかった場合、その因果関係は適切でないと考え、可視化の際には使用しなかった。

#### 4.4 トピックと因果関係の可視化

クラスタをノード、因果関係を有向エッジとして因果関係ネットワーク図を作成した。システムのインタフェースを図2に示す。このシステムは、それぞれの話題の因果関係を表示する領域（以下、可視化領域と記す）（図2-A）と、話題に含まれているニュース記事の日付と本文を表示する領域（以下、記事領域と記す）（図2-B）から構成される。可視化領域では、一連のニュース記事を話題ごとに分類し、ノードとして表現している。横軸で時間経過を表すことで、月ごとのトピック変遷が確認できる。また、ノードの特徴語がノード内に表示されている。各有向エッジは因果関係を表している。エッジを辿ることによって、話題の推移理由を確認できる。事象の各ノードをクリックすることで、該当する期間と話題を持つ記事の日付と本文が記事領域に表示される。これらの機能を用いることで、ニュースの詳細な理解につながることを期待される。本研究では長期間報道される報道される話題の変化を理解することを目的としており、ニュース記事内部の詳細な情報を提供するものではない。HDPで得られる特徴語では必要以上に詳細な情報を提示してしまうと考え、可視化の際にはTF-IDFによって得られた単語を使用した。

## 5. 実験

作成したシステムがユーザのニュース理解にどのような影響を及ぼすか検証するため、ユーザ実験を行った。本研究では、毎日新聞社が提供する2016年版の新聞記事コーパスを使用する。この新聞記事コーパスには2016年に毎日新聞に掲載された記事83,668本が収録されている。実験データとして、2016年度毎日新聞記事コーパスから「増税」「消費税」いずれかのキーワードを持つ678本を使用した。

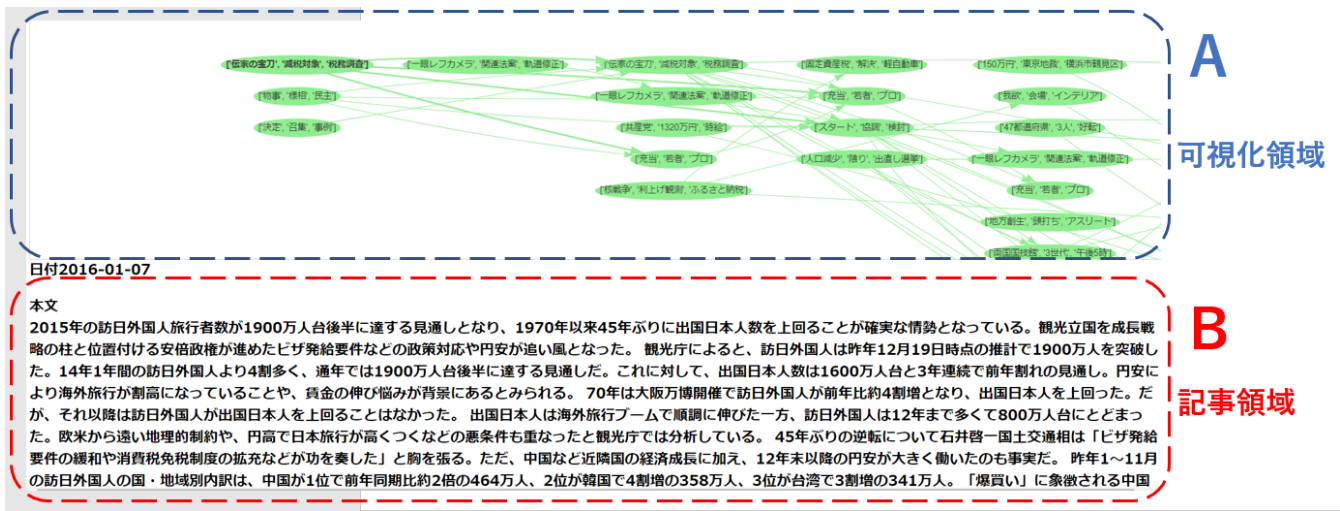


図2 提案システムの外観。因果関係を表示する領域と話題に含まれているニュース記事の日付と本文を表示する領域で構成される。可視化領域にはクラスターと因果関係が表示されている。クラスターには特徴語が表示されている。

表2 クラスターの精度検証に使用したクラスターの特徴語

	特徴語
クラスター1	倒産, けむ, 勝利者
クラスター2	宮沢, 足枷, 参院選
クラスター3	充当, 若者, プロ
クラスター4	独自, 大阪県警, 1.5%
クラスター5	一眼レフカメラ, 関連法案, 軌道修正
クラスター6	47 都道府県, 三人, 好転
クラスター7	支持率, 清田暁, 学級

### 5.1 クラスターリングの有効性に関する実験

提案手法のクラスターリングの有効性を検証するため、ユーザ実験を行った。本実験では、4.1節で得られたクラスターにある記事のリード文を提示し、実験対象者がクラスターに名前をつけ、その名前に何本が対応しているか回答してもらうことでクラスターリングの精度を検証した。この際、実験参加者へ与える情報が増加しタスクが難化することを防ぐため、リード文のみを提示した。実験参加者146名に対し、7つのクラスターを持つ各6本のリード文を提示した。全ての設問に対し一様な回答を行ったものや記述欄に適していない回答などを行った参加者の回答を無効な回答とし、有効な回答128件を取得した。使用したクラスターの特徴語と記事数を表2に示す。

実験の結果、提示した全ての記事が設定した名前に対して対応していると回答した実験参加者は18.64%、半数以上が対応していると回答した実験参加者は49.2%となり、全実験対象者が対応していると回答した記事数の平均は全体で4.2本であった(図3)。

### 5.2 クラスター間の関係性に関する実験

次に、2つのクラスターにあるリード文を提示し、それらのクラスターはどのような関係にあるかを選択肢から回答してもらうことで、クラスター間の関係性が適切に判断されたかを検証した。ユーザに提示した選択肢を以下に示す。

- 全く関係がない

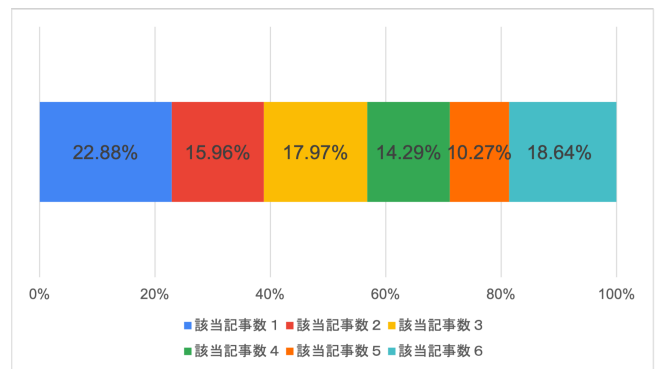


図3 クラスターリングの精度評価

- 関係はないが共通項が存在する
- クラスターAを原因としてクラスターBが起こった
- クラスターBを原因としてクラスターAが起こった
- クラスターAがクラスターBの続報である
- クラスターBがクラスターAの続報である
- クラスターAがクラスターBを包含する
- クラスターBがクラスターAを包含する
- 同じ話題から派生したクラスターである
- その他

この際、日付の情報は因果関係や続報関係、派生関係に対して実験参加者の回答に影響を及ぼす可能性があるため、システムでは提示していた日時を提示せず新聞記事の内容のみを提示した。さらに、実験参加者へ与える情報が増加しタスクが難化することを防ぐため、新聞記事のうち、リード文のみを提示した。実験参加者には、システムが因果関係を持つと判断した7組と、因果関係が持たないと判断した2組の計9組を提示した。提示したリード文があるクラスターの特徴語を表3に、提示したクラスターの組み合わせを表4に示す。因果関係でない組み合わせは、報道時期がより早いと推定されたクラスターを原因クラスター、遅いと判断されたクラスターを原因クラスターとしている。

表 3 関係性検証に使用したクラスタの特徴語

	特徴語
クラスタ 8	伝家の宝刀, 減税対象, 税務調査
クラスタ 9	スタート, 協調, 検討
クラスタ 10	共産党, 1320 万円, 時給
クラスタ 11	物事, 様相, 民主
クラスタ 12	両国国技館, 3 世代, 午後 5 時
クラスタ 13	総裁任期, 衆議院本会議, 日々

表 4 使用したクラスタの組み合わせと因果関係

原因クラスタ番号	結果クラスタ番号	因果関係の有無
8	9	あり
8	10	なし
10	9	あり
9	12	あり
8	12	あり
9	13	あり
9	11	なし
12	11	あり
12	13	あり

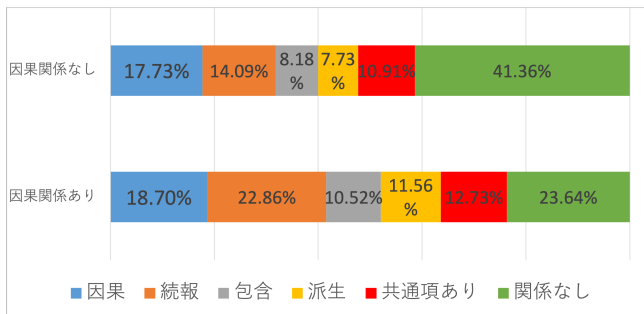


図 4 関係性の回答率

116 人に対してこの実験を行い、全ての設問に対し一様な回答を行ったものや記述欄に適していない回答などを行った参加者の回答を無効な回答とし、有効な回答 110 件を取得した。

実験の結果を図 4 に示す。実験の結果、システムが因果関係を持つと判断したクラスタ間に、「因果関係」「続報関係」「包含関係」「同じ話題からの派生」のいずれかを回答した参加者は 63.6%、システムが因果関係に持たないと判断したクラスタに、「関係はないが共通項あり」「関係なし」と回答した参加者は 52.2%であった。

しかし、因果関係にあるとシステムが判断したクラスタ間に、「因果関係」と回答したのは 18.7%、原因と結果の順番も正しく回答したのは 8.8%に留まった。

### 5.3 システムの有効性に関する実験

作成したシステムがユーザの理解を支援するか検証するために、実験を行なった。本実験では、タスクの特性上、レポートを作成する機会の多い大学生 5 名を実験の対象とした。実験参加者にはシステムを使用させ、「増税再延期に関して、その内容とニュースの流れについて要約した文章を作成してください。」という指示を行なった。ユーザが初見の状況下で可視化表現から得た情報を調査するために、システムに関する説明を行わず、システム使用開始の 5 分後に説明を行った。システムに関する

表 5 実験参加者のクリック行動の分析

参加者	因果関係	より新しいクラスタ	同じ月のクラスタ	その他
1	8	17	50	6
2	2	8	21	7
3	3	2	2	1
4	8	8	7	5
5	4	5	18	1

説明では、横軸が時間軸であること、特徴語がそのノードを代表する語であること、話題ごとにノードが分類されていること、エッジが因果関係を基に作られたものであること、について説明を行った。その後、時間制限を設けず、要約が終了までシステムを使用してもらった。その後、

- 普段のニュース視聴の多寡
- 可視化表現の探索への影響
- インタフェースの使用感

について聞き取り調査を行った。レポートの作成にかかった時間は平均 39.5 分 (max64.4 分, min22.8 分) であった。

次に、探索時にノードをクリックする行為に着目し、実験参加者の探索の傾向を調べた。実験参加者のクリックのうち、平均 32.8%がクリックしたノードよりも新しいニュースを、53.6%が同じ月のクラスタをクリックしていた。また、平均 13.7%がエッジで繋がっているクラスタをクリックしていた。

次に、実験参加者が作成した要約から、どのクラスタから引用を行ったかを特定し、クラスタ間の関係性を用いて要約内容の傾向を調べた。その結果を表 5 に示す。

要約内容の 96.4%が、1 つ前のクラスタより時系列が後のクラスタ、あるいは同じ月のクラスタについて述べられており、システムが報道された話題の変化の把握を容易にしていることが示唆された。また、クラスタ間の関係性のうち、14.3%が因果関係であった。

聞き取り調査では、ニュース視聴の多寡については、5 名全員が「ほとんど見ない」と回答した。このことから、実験参加者のニュースに関する事前知識が乏しく、今回のニュースに関しても知識がほとんどない状態で実験参加していたことがわかる。可視化表現の探索への影響については、時間軸に関しては 5 名全員が時間軸に沿って可視化表現を確認していると回答した。特徴語に関しては、5 名中 4 名から特徴語を参考にしたという回答が得られた。一方で、「因果関係を重視していて、特徴語はあまり使用しなかった」「詳細な説明を受ける前は使用していたが、説明を受けた後は使用しなかった」といった回答も得られた。因果関係に関しては、5 名中 3 名が「詳細な説明を受けた後は因果関係を参考にして要約を作成していた」という回答が得られた。また、「類似したクラスタを探すために、因果関係は意識して確認していた。実際に似たようなクラスタが多かったように思う」という意見が得られた。

## 6. 考察と展望

### 6.1 クラスタリングに関する考察と展望

5.1 節で示したクラスタリング精度検証の結果、クラスタリ

ングの妥当性が示唆された。しかし、クラスタの代表語（表2参照）と、ユーザが設定したクラスタの名前はほとんど一致せず、提示した全てのクラスタにおいて、「消費税」「増税」「安倍」「アベノミクス」等が多く見られた。よってクラスタの代表語と実際のクラスタの内容が合致しているとは言い難い結果となった。これは、ユーザに提示した記事集合が「消費税」「増税」という単語が存在している記事であると伝えていなかったため、実験参加者は計算機が分類する話題よりも大きな話題で括ってしまったと考えられる。この問題は、ユーザにリード文を提示する際に、事前に大きな話題についてユーザに提示することで解決出来ると考えられる。今回対象とした記事数は678本であり、ほとんどのクラスタにある記事数が10本を超えなかった。クラスタの記事数が少ない場合、クラスタが必要以上に分割されていることが考えられる。必要に応じてクラスタの結合を行うことで改善を図る必要がある。

## 6.2 因果関係に関する考察と展望

4.3節で得られた文章を確認すると、手掛かり表現が因果関係以外の意味で使われている文章が見られた。例えば、「関係者によると、24日に初会合を開く党内の「戦略本部会議」で、元財務官僚の高橋洋一氏を招いて延期に伴う財源確保の方法などについて話し合うという。」という文章では、「関係者によると」の「よる」は情報源を表す使い方であり、因果関係を表すものではない。これらの文章を使用すると、間違った原因表現や結果表現が抽出され、結果として間違った話題間の因果関係が得られたと考えられる。因果関係を検証する実験では、実験参加者にリード文を提示したが、リード文のみで実験参加者が話題の関係性を把握できるかは疑問が残る。例えば、ロシアとウクライナの例では、ロシアとウクライナの紛争の問題と、小麦が高騰したことのみを伝えたことのみを伝えても、そこにどのような関係があるかはわからない。ここにどのような関係があるかは、ロシアへの経済制裁や、それに伴う小麦の輸出制限を知らなければ把握出来ない。そのため、話題に沿ったリード文とともに、話題間に因果関係を持つと判断される原因となった文章を提示する必要があると考えられる。

## 6.3 システムに関する考察と展望

「因果関係は意識して確認していたが、あまり似ているクラスタが見つからず、逆に似ていると感じたクラスタに因果関係がなかったことも多かった」、「クラスタ間の繋がりが不明瞭なので使わなかった」という回答が得られた。この回答をした2名は、総クリックのうち、因果関係にあるクラスタをクリックしたのは9.1%のみであり、要約文にも因果関係にあるクラスタを使用していなかった。そのため、因果関係により信頼性を持たせる必要があると考えられる。「因果関係を重視して要約を行った」と回答した参加者は、増税を再延期したことについて詳しく要約しているが、その後不信任決議案が棄却され、解散総選挙になったことまで書かれていなかった。また、因果関係にあるクラスタをクリックする確率が、他の実験参加者が13.6%であるのに対し、因果関係を重視して要約を行った参加者は19.4%と、大きく変わらなかった。これらのことから、因果関係を重視した結果、それ以外のクラスタを軽視してしまい、

重要な情報を見落としてしまうことが考えられる。因果関係のみならず、様々な関係性も示すことにより、より多くの情報に目を向けさせることができると考えられる。また、「因果関係を示す矢印が交差する部分が多く、どのクラスタが繋がっているのかわかりにくかった」という意見が2人から得られた。クラスタの内容や因果関係の有無によって、クラスタの配置を変更し見やすくすることが求められる。

## 7. おわりに

本稿では長期間報道されたニュースの話題の変化を把握するために、クラスタ間の因果関係を提示するデバイスを提案した。HDP-LDAを用いてニュース記事を話題にクラスタリングし、因果関係を含む文から原因・結果表現を抽出し、その結果に基づき話題間の因果関係を推定し可視化した。実験の結果、リード文によるクラスタリングの妥当性と、因果関係を持つ話題に関係性があることが示唆された。また、因果関係と時系列の可視化は、ユーザのニュース理解に一定の効果があることが示唆された。今後は、因果関係にある文章の抽出手法の改良や、話題間における因果関係推定手法の改良に取り組み、提案した可視化表現の妥当性向上を図る。

## 文 献

- [1] Blei, D. M. and Lafferty, J. D.: Dynamic Topic Models, *Proceedings of the 23rd international conference on Machine learning*, pp. 113–120 (2006).
- [2] Blei, D. M., Ng, A. Y. and Jordan, M. I.: Latent dirichlet allocation, *the Journal of machine Learning research*, Vol. 3, pp. 993–1022 (2003).
- [3] Sion, A.: テレビ・ニュースのリードにおける主題の特徴について, 筑波応用言語学研究, Vol. 11, pp. 85–96 (2004).
- [4] Teh, Y., Jordan, M., Beal, M. and Blei, D.: Sharing Clusters among Related Groups: Hierarchical Dirichlet Processes, *Advances in Neural Information Processing Systems*, Vol. 17 (2004).
- [5] 石井裕志, 馬強, 吉川正俊: 因果関係ネットワークの構築によるニュースの理解支援, 第1回データ工学と情報マネジメントに関するフォーラム論文集, C5-6 (2009).
- [6] 坂地泰紀, 増山繁: 新聞記事からの因果関係を含む文の抽出手法, 電子情報通信学会論文誌, No. 8, pp. 1496–1506 (2011).
- [7] 武田直人, 関洋平, 森下民平, 稲垣陽一: ライフイベントの経験に有用なトピックの抽出と評価, 情報処理学会論文誌データベース, Vol. 11, No. 2, pp. 16–29 (2018).
- [8] 松本直彰, 湯本高行, 山本岳洋, 大島裕明: 時間的推移を考慮した報道トピックの抽出と再構成, 情報処理学会研究報告, Vol. 2020-IFAT-139, No. 1, pp. 1–6 (2020).
- [9] 西川奈都月, 盛山将広, 内藤峻, 松下光範: 初学者を対象としたニュース記事中のトピックの関係性に基づく可視化インタフェースの提案, 第15回インタラクティブ情報アクセスと可視化マイニング研究会予稿集, pp. 62–67 (2017).
- [10] 河田裕成, 赤石美奈, 細部博史: 事象情報グラフの操作による時系列情報の可視化, 第32回人工知能学会全国大会, 20103 (2018).
- [11] 齋藤藍, 松下光範: 長期間報道されるニュース記事を対象とした話題推移の可視化による内容理解の支援, 情報処理学会研究報告, Vol. 2022-HCI-196, No. 8, pp. 1–8 (2022).