

人物関係の可視化におけるニュースストリームの理解支援 Supporting News stream understanding by presenting relation diagram

鈴木 聖乃[†] 岩崎 有基[‡] 松下光範[†]
Akino Suzuki Yuki Iwasaki Mitsunori Matsushita

1. はじめに

Web ニュースサイトは、いつでもニュース記事を読覧できるといったアクセスの容易性や、情報の更新の速さを利点として持つ反面、頻繁に更新されるため記事数が膨大になる傾向がある。そのため、ユーザはニュースの情報を整理したりニュース全体を俯瞰的に把握したりすることが困難になる。特に経済や政治などのジャンルのニュースは長期的にわたり報道される傾向があるうえ、ニュース記事中で言及される人物や組織が多いため、ニュースの内容を正確に理解するには継続的にニュースを読覧し、人物や組織間の関係性を把握しつつそのニュースの全貌を俯瞰する必要がある。そこで本研究では、「人物関係図を生成することは文書の内容理解に有効である」という懸念の指摘 [1] に基づき、一連のニュース記事中から人物名と組織名を人物関係図として可視化することにより、ニュースの内容理解を支援する可視化インタフェースを提案する。

2. デザイン指針

本章では提案インタフェースを作成するにあたり考慮すべきことについて記述する。

本研究では、経済と政治の分野を対象ジャンルに設定し、普段そのジャンルのニュース記事を読覧しない人を対象ユーザと定めた。継続的にニュースを読覧しないユーザにとって、ニュース全体の流れを把握することは困難である。そのため、ニュースの内容を正確に理解するためには一連のニュース記事の日付をさかのぼって読み、記事中に登場する人物名や組織名との関係を整理しつつニュースの全容を把握することとなる。

また、菊池らによると時系列情報を用いることで、ニュース内容の話題が変化していく様子を視覚的に捉えるようにする手法が有効であるとされている[2]。このことからニュース記事を読む際に人物関係図を表示することでユーザのニュース記事理解のための支援ができると考える。

以上よりユーザの行動を考慮すると、そのニュースの前後に起きた事象の理解やニュース記事中の人物や組織の関係性の把握がシステムに求められる支援の要諦となる。

そこで本稿では、

- (1) ニュース全体の把握と記事ごとの把握がシームレスにできること
- (2) ニュース記事中の重要な人物名と組織名が明確であること
- (3) ニュース記事内で関連している人物名や組織名の繋がりがわかること

の 3 点をシステムが満たすべきデザイン指針として定めた。

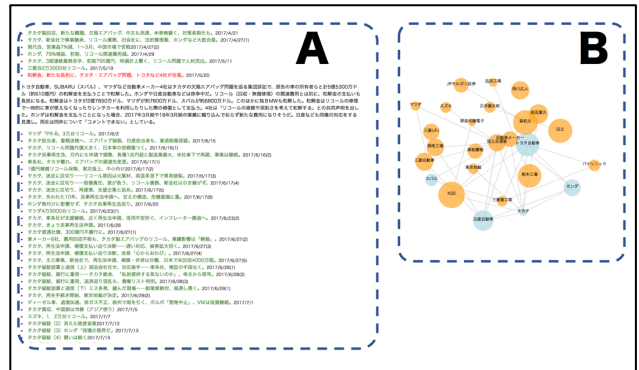


図 1 提案システムの概要図

3. 人物名の抽出

本稿では、経済と政治に関するニュースから人物名と組織名を抽出し、共起関係を元に人物関係図の作成を行った。ニュース記事は人物が複雑に関係している「森友事件」と「リコール問題」を日経テレコンより収集し、データとして用いた。

ニュース記事から該当する人物、組織名を抽出するために、あらかじめニュース記事群を形態素解析にかけ、辞書の作成を行った。ニュース記事には表記揺れが存在し、例えば「安倍」、「安倍晋三」、「安倍総理」などがある。そのため、辞書作成において表記揺れを考慮し、人物の名前の表記が苗字のみの場合、フルネームである場合、首相や総理など特定の名前ではないが、どの人物か特定できる場合も同一人物として登録している。

4. 共起関係による相関図作成

本研究により作成した辞書を用い、人物名と組織名との関係を抽出し、抽出した情報から相関図を作成した。

馬場らの研究では、小説テキストからセリフ情報と入退場情報を用いることで人物の係数度を計算可能としている[3]。しかし、ニュース記事ではセリフ情報や入退場情報が必ずしも入手できるとは限らない。

そこで、ニュース記事の本文の 1 文中に「人物名」もしくは「組織名」が 2 回以上登場した場合、その単語同士は共起関係にあるとして処理をした。その例として「田中さんは花子さんに本を貸した」といった 1 文が存在した場合、「田中」と「花子」の人物名が計 2 回登場しているため、関係があるとして抽出する。この共起関係を、辞書を元に抽出したところ「リコール問題」では 88 個、「森友事件」では 289 個の関係が抽出できた。しかし、重要でない人物名や組織名まで抽出されているため、すべてを人物相関図として可視化してしまうと視認性が低下する。そこで、生成された相関図に表示させる単語の条件を以下のように設けた。

[†] 関西大学総合情報学部, Faculty of Informatics, Kansai University

[‡] 関西大学大学院総合情報学研究科, Graduate School of Informatics, Kansai University

- リコール問題
 - (1) テーマに関する全記事中でその単語の出現回数が 5 回以上であるもの
 - (2) 条件 1 を満たした単語が含まれた文中から共起関係の組を生成できるもの
 - (3) 条件 2 を満たし生成された関係の組の中の単語の出現回数が全体で 3 回以上であるもの
- 森友事件
 - (1) テーマに関する全記事中でその単語の出現回数が 30 回以上であるもの
 - (2) 条件 1 に当てはまる単語を含む文章から狂気関係を用いて関係の組を生成できるもの
 - (3) 条件 2 より生成された関係の組の中の単語の出現回数が全体で 10 回以上であるもの

森友事件の記事数はリコール問題の記事数に比べて多く、出現回数を統一してしまうと単語の出現回数が少ないリコール問題の関係が抽出できなくなる。そのため、「リコール問題」と「森友事件」のテーマには条件 1 と 3 に指定している出現回数は上位 30 単語前後になるようにそれぞれ基準を定めている。これらの処理により抜き出された単語と関係を用いて人物相関図の作成を行う。

5. 提案システム

提案システムの概観を図 1 に示す。このシステムはニュース記事のタイトルと本文を表示する記事ペイン (図 1-A) と、記事から生成された人物関係図を表示する関係図ペイン (図 1-B) から構成される。記事ペインには、ニュース記事のタイトルのみが表示されており、ニュース記事のタイトルをクリックした際、そのタイトルの色が赤色に変化し、その下にニュース記事の本文が表示される (図 1-A)。これにより、ユーザはどのタイトルのニュース記事を閲覧しているのかを把握することが可能になる。関係図ペインには人物関係図が表示されており、ニュース記事から抽出した人物名と組織名がオレンジ色のノードで表示されている。関係図ペインに表示するノードの数には制限を設けた。ニュース記事中に登場した回数が少ない場合、ニュース記事の内容自体に関係していない可能性があるため、提案システムでは経済のテーマは 5 回以上、政治のテーマは 10 回以上出現している単語をノードとして表示した。1 文中で共起関係にある人物名と組織名はニュース記事中で関係がある単語と捉え、それらのノードをつなぐことによって関係の有無を示すこととした。また、記事ペイン中のタイトルをクリックした際、その記事に関係するノードの色がオレンジ色から水色に変化する (図 1-B)。これにより、ユーザは記事ごとの人物や組織の関係性を把握することが可能になる。

6. 実験

人物関係図の提示がニュース記事の理解に及ぼす効果について調べるために、図 1 の提案システムを用いてニュースを閲覧する実験協力者群 (提案システム条件) と紙媒体を用いて時系列順に並べられたニュースを閲覧する実験協力者群 (紙媒体条件) のニュースの理解度を測る比較実験を行った。実験協力者は計 24 名で、いずれも普段ニュースを閲覧しない学生を対象にした。実験では「リコール問題」のテーマに 12 名 (提案システム条件 6 名、紙媒体条件 6 名)、「森友学園」のテーマに 12 名 (提案システム

条件 6 名、紙媒体条件 6 名) の協力者を割り当てた。なお、ニュースの閲覧形態 (提案システム条件、紙媒体条件)、テーマの種類 (リコール問題、森友学園) の割り当ては全てランダムとした。実験では、ニュース内容の理解度を検証するために 300 字程度の要約文生成とニュース内容の穴埋めの 2 種類の課題を実施した。要約文生成は、ニュース記事の閲覧を許し、ニュース内容の穴埋めは要約文生成の後にニュース記事の閲覧を許さずに解答させた。ニュース内容の穴埋めは全 5 問で、選択語群は提示しなかった。また、どちらの課題にも時間制限は設けなかった。

要約文生成は、記述された固有名詞 (人物名、組織名、地名) の数を比較することで検証した。Wilcoxon の符号順位付検定の結果、提案システム条件と紙媒体条件で有意差が見られなかった (リコール問題: $W(6,6) = 19.5$, n.s., 森友事件: $W(6,6) = 18.0$, n.s.)。実験終了後に行ったアンケートによると、提案システムの画面 (図 1) では記事中の文字の色や大きさに変化がなく、重要人物や重要な組織がわかりにくいという意見があった。そのため、実施した要約では人物名や組織名が記述されず、固有名詞の数に差が生じなかったことが推察される。一方、ニュース内容の穴埋めは、U 検定の結果、提案システム条件と紙媒体条件で有意差がみられた (リコール問題: $U(6,6) = 2.34$, $p < 0.05$, 森友事件: $U(6,6) = 2.76$, $p < 0.05$)。提案システム条件の方が課題の得点が有意に高かったことから、人物関係図の提示がニュース内容の理解を支援することが示唆された。以上の結果から、経済や政治など長期にわたるニュース記事の内容を理解させるために人物関係図を提示することの優位性が確認された。

7. おわりに

本稿では、普段ニュースを閲覧しない人を対象に、経済や政治の分野の一連のニュース記事を人物関係図と共に提示することで、そのニュース内容の理解を支援するインタフェースを提案した。提案インタフェースと紙媒体を用いて実験協力者に対して比較実験を行った。このことから、経済や政治などの長期にわたるニュース記事の内容を理解させるために人物関係図を提示することの優位性が確認された。今後は、ニュース記事の表示方法の改善や関係図ペインから記事ペインへの逆引きを可能にしたインタフェースの改善を図る。

謝辞

本研究の遂行にあたり、文部科学省科学研究費 (課題番号: 15H02780) の助成を受けた。記して謝意を表す。

参考文献

- [1] 縣啓治, 伊藤雄一, 高嶋智毅, 北村喜文, 岸野文郎: 物語テキストから進行状況に応じて登場人物の存在状態と関係を推定する手法, 第 18 回インタラクティブシステムとソフトウェアに関するワークショップ(2010).
- [2] 菊池匡晃, 岡本昌之, 山崎智弘: 階層型クラスタリングを用いた時系列テキスト集合からの話題推移抽出, 日本データベース学会論文誌, Vol.7, No.1, pp.85-90(2008)
- [3] 馬場こづえ, 藤井敦: 小説テキストを対象とした人物情報の抽出と体系化”, 言語処理学会第 13 回年次大会発表論文集, pp.574-577(2007)