

問題となる因果関係の抽出法の検討

Examination of extraction methods of problematic cause-and-effect relationship

松岡 功将

Kosuke Matsuoka

岡山大学 太田研究室

Ohta Laboratory, Okayama University

概要 Web 検索結果から因果関係を自動抽出し、抽出した要因と結果を結んで因果関係を可視化する因果関係ネットワークが提案されている。しかし従来の因果関係ネットワークを見ても、可視化された因果関係からユーザが今後どのように行動すべきかという改善策を見つけることは難しい。本研究はそのような改善策を合わせて可視化することを目的とする。本稿では、そのために問題となる因果関係を含む文書の抽出法を検討する。

1. はじめに

新聞やテレビといったメディアを通じて報じられるニュースの中には、多くの事象が複雑に絡み合っ引き起こされる出来事がある。それらを即座に深く理解することは容易でない。そこで青野ら[1]は Web 検索結果から因果関係を自動抽出し、抽出した要因を始点ノード、結果を終点ノードとし、それらを結ぶエッジにより因果関係を可視化する因果関係ネットワークを提案した。本稿では、問題点を含む文書から因果関係を得るには、どのように問題となる因果関係を含む文書を抽出すべきかを検討する。本稿でいう問題となる因果関係とは、現在おこっている CO2 排出が原因で地球温暖化が起こるといようなものことで、解決されていないが改善方法はあるような事柄である。

2. 因果関係抽出のための手がかり表現による武田らの因果関係抽出

本稿では、武田ら[2]の導入した手がかり表現を用いて因果関係を文章から取得する。以下ではこれらの手がかり表現を用いて要因事象を取得することを「要因検索」と呼び、結果事象を取得することを「結果検索」と呼ぶ。表 1, 2 は武田ら[2]の定めた 8 種類の手がかり表現である。因果関係の検索には、グループ毎に優先度を変えて手がかり表現を使用する。表の初めに出てくるものほど優先度が高い。これらの手がかり表現で検索した後に、「により」のみで抽出された場合については因果関係に妥当性がないと判断し、表 3 に示す 25 種類の手がかり表現を用いて再検索する。これは「により」が因果だけでなく手段なども表すからである。また表 3 の 25 種類の手がかり表現には優先度の違いはなく、すべての手がかり表現を平等に用いて因果関係かどうかを判定する。次にその文書を抽出する手順について述べる。

表 1 要因検索で使用する手がかり表現

優先度	手がかり表現
1	に伴う、に伴い、を理由に、が理由で、の影響で
2	により
3	の原因として、の原因は

表 2 結果検索で使用する手がかり表現

優先度	手がかり表現
1	の原因として、の原因は
2	に伴う、に伴い、を理由に、が理由で、の影響で
3	により

表 3 再検索で使用する手がかり表現

を背景に	の背景に	を受け
を受けて	を反映し	を反映して
に、が響く	が、に響く	は、が響く
に、が響いた	が、に響いた	は、が響いた
に、が響き	が、に響き	は、が響き
に、が響いている	が、に響いている	は、が響いている
によって	から	をきっかけに
が影響した	に支えられて	のため
の結果		

2.1. 要因検索と結果検索

因果関係を取得するために表 1,2 の手がかり表現に、因果関係を調べたい事象を組み合わせ検索式を生成し、Bing search API[3]を利用して検索結果を取得する。その後検索結果から要因事象や結果事象を取得する。

以下では、要因事象を X、結果事象を Y とする。表 1 を利用する場合は X と Y の関係を調べる検索式は以下の Q₁, …, Q₈ のようになる。表 2 の場合は手がかり表現の優先度が表 1 と異なるため、これに応じた Q₁, …, Q₈ となる。

Q₁ = “X に伴う Y”

Q₂ = “X に伴い Y”

Q₃ = “X を理由に Y”

Q₄ = “X が理由で Y”

- Q₅ = “X の影響で Y”
- Q₆ = “X により Y”
- Q₇ = “Y の原因として X”
- Q₈ = “Y の原因は X”

これらを用いて要因または結果を検索する。要因検索では、因果関係を知りたい事象の要因を検索するために、Y を調べる事象名に置き換え X を削除した検索を用いる。これにより Y の要因となる事象を検索することができる。結果検索では、因果関係を知りたい事象の結果を検索するために、X を調べる事象名に置き換え Y を削除した検索式を用いる。これにより X の結果となる事象を検索することができる。

なお、8 つの式 Q₁, …, Q₈ 全てを必ず検索するわけではない。表 1 または表 2 の優先度に従って検索し 50 件を取得した時点で検索処理を終了する。

2.2. 再検索

2.1 節の検索式を用いて抽出した因果関係のうち、手がかり表現「により」のみによって抽出された中には因果関係とはみなせない不適切な表現も含まれるため、再検索を行い因果関係の妥当性を判定する。再検索で用いる手がかり表現は表 3 に示した手がかり表現を用いる。生成する検索式は、因果関係を調べる二事象について要因事象を X、結果事象を Y とすると“X を背景に Y”、“X を反映して Y”、“Y に X が響き”、“X に支えられて Y”などようになる。これらの検索式を用いて検索し、複数の検索式で検索結果を得ることができた場合、その因果関係を妥当であると判定する。

3. 問題となる因果関係を含む文書集合の抽出

武田ら[2]によって導入された 2.1 節, 2.2 節の方法を使うことにより、因果関係となる要因と結果が抽出される。しかし単純な因果関係だけでは、それに対してどのような対策や改善策をとればよいか不明瞭である。そこで問題点を含む文書集合だけを獲得して、その文書集合から改善策を見つけだすようなシステムを本研究では検討している。

問題となる因果関係を含む文書集合だけを獲得する方法として、まず手がかり表現を利用してそのような文書集合を集める方法が考えられる。しかしそうしたとしても、検索した文書が因果関係を含むかどうか分からないという問題がある。そこで「の問題として」「の問題は」「の問題となる」の 3 つの手がかり表現を使って問題となる因果関係を含む文書を検索し、その文書から武田ら[2]の方法で因果関係が抽出できるか実験により評価した。

表 4 はこの 3 つの手がかり表現のみで文書を検索して因果関係を抽出した結果である。なおこの際の検索フレーズは“水質汚染の問題は”などである。抽出した因果関係の

表 4 「水質汚染」による因果関係の抽出

	獲得文書の数	因果関係の数
の問題は	50	1
の問題となる	50	0
の問題として	50	0

表 5 「水質汚染」の AND 検索による因果関係の抽出

	獲得文書の数	因果関係の数
の問題は	50	20
の問題となる	50	6
の問題として	50	7

数は極端に少なすぎるため、武田ら[2]による 8 種類の手がかり表現のうち「に伴う」と、AND 検索を設定した検索結果を利用することで抽出できる因果関係を増やしたのが表 5 である。

具体的には“水質汚染に伴う”と“水質汚染の問題は”などとの AND 検索を行った。結果を表 5 に示す。因果関係の数は表 4 に比べて増えている。

4. おわりに

本稿では、問題となる因果関係を含む文書を検索するための手がかり表現に注目し、それによってどの程度の因果関係が抽出されるかを実験により確かめた。今後の課題として、問題となる因果関係の定義の明確化がある。またその問題に対する改善策の獲得とその因果関係ネットワークへの組み入れについても検討したい。

参考文献

- [1] 青野荘志, 太田学, “要因検索による因果関係ネットワークの構築と因果知識の獲得”, DEIM Forum 2010, B9-1, 2010.
- [2] 武田真輝, 太田学, “因果関係ネットワーク構築のための要因検索法の改良”, WebDB Forum 2011, G2-1, 2011.
- [3] Microsoft が提供する API(Web Only), <http://datamarket.azure.com/dataset/bing/search/>