

因果関係ネットワークの増分的な構築について

石井 裕志* 馬 強† 吉川 正俊‡

京都大学 情報学研究科 社会情報学専攻

〒 606-8501 京都市左京区吉田本町

1. はじめに

現在、新聞やニュース番組、Web ニュースで報道されているように、多くの事象が毎日起こっている。それらの事象の中には、展開の速いものや、他の事象と複雑に絡んでいるものも多い。そのような事象を正しく理解するためには、利用者が事象の起こった背景知識を理解していることが必要である。そのためにはたくさんのニュースを読み、多くの労力が必要である。よって、利用者に対して、事象の背景知識を提示するシステムが必要である。システムが事象の背景を示すために、事象の起こった原因と結果を示す因果関係を提示することが有用であるが、一つ一つの因果関係そのものは事象の背景を十分に示すことはできない。事象の起こった大元の理由や事象によって波及した結果といった因果関係の連鎖を表現する、因果関係ネットワークを構築することが有効な手段である。

事象とその相互関係は、時間の経過とともに変化する。しかしながら、従来の因果関係ネットワークの構築手法では、因果関係の時系列特徴が十分に考慮されておらず、事象間の関係を正確に整理できない場合がある。そこで、我々は、因果関係ネットワークを増分的に構築する手法を提案する。ニュース記事から事象間の因果関係を抽出して因果関係ネットワークを構築するが、結合や簡略といった操作を用いて因果関係ネットワークを増分的に更新する。しかし、従来手法である事象ノードに含まれるキーワード集合を用いた手法には二つの課題があり、一つは事象ノード間の同一性判定に必要な計算量が大きいという点、もう一つは同一性判定を行う事象ノードの順序によりネットワークの結果が異なるという点である。そこで我々は、キーワードの抽出手法に述語項構造を導入し、事象ノードをトピックと内容を表すキーワードの二つ組によって表現する、因果関係ネットワークの表現モデル TEC (Topic Event Casual network Model) とそれに基づく構築手法を提案し、この二つの課題の改善を行う。

2. 関連研究

因果関係ネットワークを構築する手法として、佐藤らの手法 [2] がある。佐藤らは、一般の文書や Web の文書から因果関係を抽出し、因果関係ネットワークを構築した。この手法では、因果関係の原因事象を表す文節、結果事象を表す文節から重要キーワードを抽出し、その

事象SVO構造(S:オバマ,
V:目指す,
O:雇用状況の回復)
トピックキーワード(オバマ, 政策)

▼ 1

事象SVO構造(S:オバマ,
V:発表する,
O:一般家庭の支援策)
トピックキーワード(オバマ, 政策)

例)「オバマが雇用状況の回復を目指す」ことが原因となり、結果として「オバマが一般家庭の支援策を発表」したことを示すグラフ

図 1: TEC モデルにおける因果関係の表現例

キーワードで元の事象を表現した。しかし、因果関係を含む文節から得られるキーワードだけでは、元の事象を十分に表現できなかった。これを解決するために、本研究では事象を表すキーワードを拡張する。また、佐藤・堀田の手法は、キーワードの類似したノード同士を近くに配置し、因果関係のつながりを表現しようとした。しかし、類似事象を示すノードが多い場合には因果関係のつながりを理解しづらい。本研究では、類似事象を示すノード同士をマージすることで因果関係のつながりを表現する。また、佐藤・堀田の手法は、一度生成したネットワークが整理されることはないため、ネットワークが複雑であると利用者が理解できない。本研究では、ネットワークの整理を逐次行いながら、ネットワークの構築を行う。

3. TEC モデル

我々は、因果関係ネットワークを構築するために、TEC モデル (Topic-Event Causal relation model) について研究を行っている。TEC モデルでは、因果関係の原因事象、結果事象それぞれに事象ノードを作成して、原因の事象ノードから結果の事象ノードに有向枝を張る。各事象ノードは記事の主題を表す部分 (topic) と事象の詳細を表す部分 (event) からなる。また、枝は因果関係の重要度を表すラベル (枝の重要度と呼ぶ) を保持する。

TEC モデルにおける因果関係の表現例を図 1 に示す。事象の詳細を、「動作の主体」、「動作」、「動作の目的」を表す 3 つのフレーズによって構成される事象 SVO 構造で表現する。各フレーズは単語と単語の係り受け構造を保存しており、図 1 の例では、「雇用状況」が「回復」に係っていることが保存されている。事象の主題をトピックキーワードによって表現する。枝の重要度は因果関係の重要度を表し、因果関係の頻度によって決定する。

従来の因果関係ネットワークの表現手法では、事象ノードは事象を表すキーワード集合とキーワードの重要度を保持し、ノード間のキーワードの類似度の高いノード対を、類似事象を表すノードとしてマージしていた。

Incremental Construction of Causal Network

*Hiroshi ISHII(ishii@db.soc.i.kyoto-u.ac.jp)

†Qiang MA(qiang@i.kyoto-u.ac.jp)

‡Masatoshi YOSHIKAWA(yoshikawa@i.kyoto-u.ac.jp)

Graduate School of Informatics,

Kyoto University Yoshida Honmachi, Sakyo, Kyoto 606-8501

この手法では、マージ後のノードのキーワード集合を、マージ前のノード対のキーワード集合の和集合としていた。マージ前とマージ後でノードの持つキーワード集合が異なるため、マージを行うノードの順序の違いにより、ノードがマージされる場合とされない場合があるという問題が存在した。

我々は、日本語文法 SVO 構造で事象を表現する方法を提案する。類似事象のマージでは、シソーラスを用いて概念語レベルで SVO 構造のマッチングによる事象の同一性判定を行い、構造がマッチしたノードをマージする。この方法では、マージの前後でノードの持つ構造が同じであり、マージを行うノードの順序に関わらず同じネットワークを得ることができる。

4. TEC モデルによる因果関係ネットワークの構築

TEC モデルにおける因果関係ネットワークの構築手法について述べる。まず、「ため、」や「を受けて」のような因果関係を含む文を示す手がかり表現を用いて、文書から因果関係を含む文節を抽出する [1]。抽出した因果関係に対して事象ノードと因果関係の枝を作成し、次に事象ノードの事象 SVO 構造とトピックキーワードを抽出する。事象 SVO 構造は、因果関係を含む文節を述語項構造解析し、動作の主体、動作、動作の目的を表すフレーズを抽出する。トピックキーワードは、記事の抽出時に記事のタイトルで用いられた頻度の高い単語を関連記事集合ごとに抽出する。次に、事象 SVO 構造を用いて事象の同一性判定に基づいてノードのマージを行い、枝の重要度を利用してネットワークを簡略化する。時間の経過による事象と因果関係の変化に対応するため、この工程を繰り返し、ネットワークを増分的に構築する。

事象の同一判定手法とネットワークを増分的に構築する手法について述べる。

4.1 事象の同一性判定

事象の同一性判定を行うことで、複数の因果関係を結びつけ、因果関係の連鎖を表現することができる。事象の同一性判定は 2 段階の計算によって行われる。1 段階目の計算で、トピックキーワードが同じノードごとに事象ノードをグループ化する。2 段階目の計算は、グループ化されたノード内で事象ノード間のマッチングを行い、マッチしたノード対を同一事象を表現していると判定してをマージする。トピックキーワードは関連記事集合ごとに同じなので、1 段階目の計算は記事集合間同士でキーワードを比較すればよく、計算量は少ない。そして、2 段階目の計算は、全ての事象ノード間でマッチングする時に比べて、グループ内でマッチングすることで計算量が少ない。この手法により同一性判定にかかる計算量を削減する。

図 2 に同一性判定計算の例を示す。ノード間の事象 SVO 構造を比較し、SVO の各フレーズが一致する時にノードをマージする。各フレーズは、係り受け構造を比較し、図 2 では、「一般家庭」と「一般世帯」、「支援策」と「保護策」を比較する。各語は概念辞書を用いることによって単語の表記揺れや上位下位概念を考慮して比較する。図 2 では「発表する」と「掲げる」、「家庭」と「世

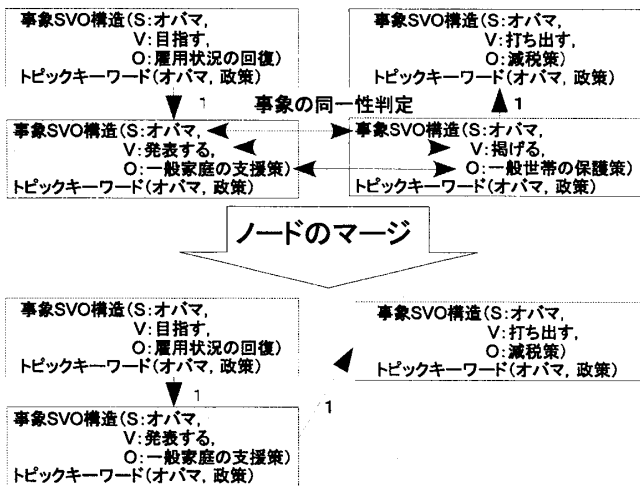


図 2: 事象の同一性判定

帯」、「支援」と「保護」が同一概念であり、SVO 構造がマッチするのでノードをマージする。

4.2 ネットワークの増分的な構築

時間の経過に伴う事象と因果関係の変化に対応するため、因果関係ネットワークを増分的に構築する。記事は日々追加されるので、我々はまず、一定期間に追加された記事から因果関係を抽出し、追加された因果関係の中で事象ノードのマージとネットワークの整理を行い因果関係ネットワークを構築する。次に、追加された記事から作成した因果関係ネットワークとそれまでに構築された因果関係ネットワーク間で事象ノードのマージとネットワークの整理を行う。これを繰り返すことで、因果関係ネットワークを増分的に構築する。

5. おわりに

本稿では、SVO 構造を利用した因果関係のネットワーク上での表現モデルと、それを用いた因果関係ネットワークの構築方法を提案した。類似事象のマージを SVO 構造のマッチによって行うことで、増分的にネットワークが構築できるようになると考えられる。今後、提案手法を使ったシステムの実装・実験を行う。また、時間軸をモデルに取り入れ、手法の改善を行っていく。謝辞 本研究の一部は、科研費(20700084 と 20300042)の助成を受けたものである。

参考文献

- [1] Sakaji, H., Sekine, S. and Masuyama, S.: Extracting Causal Knowledge Using Clue Phrases and Syntactic Patterns, *Proceedings of the 7th International Conference on Practical Aspects of Knowledge Management (PAKM2008)*, pp. 111-122 (2008).
- [2] 佐藤岳文, 堀田昌英: Web マイニングを用いた因果ネットワークの自動構築手法の開発, *社会技術研究論文集*, pp. 66-74 (2006).