Vol.2009-DBS-149 No.9
2009/11/20

要因検索による因果関係ネットワークの構築

青 野 壮 志[†1]　太 田 学[†1]

本研究では，ユーザが興味のある事象の要因を検索し，さらにその因果関係を可視化できる要因検索システムを提案する．この要因検索システムは，膨大な Web 文書からの関連文書の検索，構文解析による因果関係の抽出，抽出した因果関係のクラスタリング，因果関係ネットワークの構築の 4 つの機能で構成されている．提案手法は再帰的な要因検索を行うことにより，ユーザが興味を示している事象の間接的な要因を含めて検索できる．さらにそれらのつながりを表現した因果関係ネットワークが生成されるので，ユーザは要因間の関係を容易に把握できるようになる．

Construction of a Causal Network by Searching Factors

Hiroshi Aono[†1] and Manabu Ohta[†1]

We propose a factor search system by which users can search for factors of interesting events and browse their visualized causal relations. This factor search system is composed of the following four functions: i) retrieval of related documents from myriad of Web documents, ii) extraction of causal relations by parsing the retrieved documents, iii) clustering of the extracted causal relations, and iv) construction of a causal network. The proposed method can retrieve indirect factors of interesting events by searching for factors recursively. In addition, it can help users to understand causal relations by showing the causal networks visualizing such relations.

1. はじめに

我々は新聞やテレビなどを通じて，様々な社会現象について知ることができる．しかし，多くの事象が絡み合って引き起こされている社会現象について深く理解することは容易ではない．社会現象などについて理解を深めるには，それに関連する事象間の因果関係を把握することが有効であり，意志決定やリスク回避などにも役立つと考えられる．そこで本稿では，ユーザが因果関係を把握したいと考える事象を検索語として与えると，その要因を検索・抽出してさらに因果関係ネットワークを構築する手法を提案する．

本稿では 2 節で関連研究，3 節で提案する要因検索，4 節で提案システムの実装，5 節で提案システムの評価実験，6 節でまとめと今後の課題について述べる．

2. 関連研究

文書から因果関係を自動抽出する研究では，文の接続関係を用いる方法[1)–4)] と手がかり表現を用いる方法[3)–6)] が提案されている．接続関係を用いる方法では，重文・複文を解析対象としており，それらを単文に分割したときの各単文の接続関係から因果関係を抽出する．佐藤ら[1)] や佐藤・堀田[4)] の研究では，取り出した因果関係の表現形式を格フレームを用いて整理している．一方，手がかり表現とは,「に伴い」や「を理由に」のように要因とその結果を結びつける表現であり，因果関係を含むかどうか判断する際の手がかりとなる．なおこれを乾ら[3)] は「接続標識」，佐藤・堀田[4)] は「手がかり標識」と呼んでいる．坂地ら[5)] や石井ら[6)] の研究では，因果関係のもつ構文パターンを用いることによって，重文・複文に限らず，手がかり表現を含む全ての文を対象にして因果関係を抽出している．

興味の中心となっている事象とその周辺の事象の関わり方によっては，事象間に成立する因果関係の種類も異なると考えられる．このことから，因果関係を分類する方法[2),3)] が提案されている．Khoo ら[2)] の研究では，医療関係の文書データから因果関係を抽出することを目的に，因果関係の構文パターンや分類は医療分野に特化したものを人手で作成している．乾ら[3)] の研究では，分野に依存しない分類手法を提案しており，事象を「事態」と「行為」に分け，その組み合わせにより，因果関係を cause 関係，effect 関係，precond 関係，means 関係の 4 つに分類している．

抽出した因果関係を可視化する手法[4),6)] も提案されている．佐藤・堀田[4)] の研究では，因果関係を把握したいとユーザが考える主題を表す単語（入力キーワード）に関連する Web 文書を獲得し，その事象全般の因果関係ネットワークを構築している．因果関係を含む文節から得られる重要単語を，事象ノードを表すキーワードとしている．エッジはノード間，すなわち事象間の因果関係と共起関係を表現しており，それぞれ片方向のエッジ，双方向のエッジで表現される．共起関係の強さは事象ノードのキーワードの類似度の大きさであり，これをノード間の距離の近さに置き換えて表現している．石井ら[6)] の研究では，オンライ

†1 岡山大学大学院自然科学研究科
Graduate School of Natural Science and Technology, Okayama University

ンニュース記事を 1 日単位で取得して，新たに抽出された因果関係からネットワークを更新できるようにしている．彼らは因果関係を含む文節から得られる重要単語だけでは情報が少なすぎて，出来事のつながりをうまく表現できないと考えており，ニュース記事のタイトル中の重要語や文中の共起語を用いて，キーワードを拡張している．このキーワードはノードの類似度の算出に利用され，類似度の高いノードはマージされる．他のノードとマージされなくなった場合，そのノードの表している因果関係は重要度（話題性）が低いと考え，ネットワークから除外している．

佐藤・堀田[4] の研究において，入力キーワードに関連する因果関係とはそのキーワードの検索結果に含まれる因果関係であった．本研究では，Web 検索エンジンを用いて得られた Web 文書から因果関係を抽出し，入力キーワードの要因となっている事象をさらに要因検索することで，階層的に因果関係を獲得していく．これにより，一見しただけでは関連の不明な間接的要因を網羅的に抽出し，さらにそれらのつながりをネットワークにより可視化する．

3. 要因検索システム

本研究で提案する要因検索システムの処理を**図 1** で説明する．まず，ユーザは要因検索システムにキーワードを入力する．例えば,「景気悪化」の要因について知りたい場合には,“ 景気悪化 ”と入力する．要因検索システムは，この入力キーワードに手がかり表現を組み合わせた検索式，例えば“ に伴う景気悪化 ”を生成する．その検索式を用いて，Yahoo!の検索結果を取得する．つづいて CaboCha[8] を用いて検索結果のタイトルおよびスニペットを係り受け解析し，因果関係を抽出する．さらに抽出された要因のうち重要度の高い要因については，その要因の要因を繰り返し検索する．最後に，JUNG[9] を用いて，抽出された因果関係を可視化する．以下，各処理の詳細について述べる．

3.1 文書データの取得

興味のある事象をキーワードとして単純に Web 検索すると，検索結果の中には因果関係の記述が存在するものとそうでないものが混在する．一方提案システムは，手がかり表現を用いてユーザの指定した特定の事象に関連する因果関係ネットワークを構築する．そのため，因果関係と手がかり表現の両方を含んでいる文書データが必要であり，検索エンジンに与える検索式が重要となる．例えば,「景気悪化」の因果関係を抽出したい場合,“ に伴う景気悪化 ”のように手がかり表現と入力キーワードを組み合わせたフレーズを検索することで文書データを取得する．

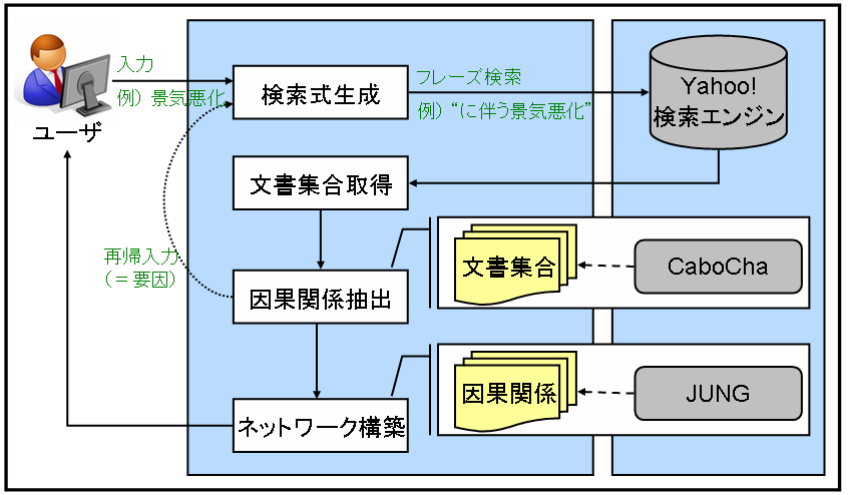

図 1 要因検索システムの処理の流れ
Fig. 1 A processing flow of the factor search system.

特に入力キーワードを構成する形態素が多い場合には，十分な検索結果を取得できないことがあるため，その場合は以下のようにキーワードを拡張する．

- **助詞の除去**
  入力キーワードに含まれる助詞を除去する．
- **キーワードの分割**
  助詞の位置でキーワードを分割する．また，入力キーワードの最後尾のサ変接続の名詞の手前で分割する．

例えば“ 米国の経済悪化 ”というキーワードが入力された場合には，**図 2** のように 4 つの検索式が生成される．まずは検索式 1 を用いて検索結果を取得する．ここで十分な検索結果が取得できない場合には，検索式 2，3，4 をこの順番で用いて検索結果を取得し，十分な検索結果を取得した時点で検索を終了する．以降，入力キーワードとそれを拡張したもの，すなわち図 2 の例では“ 米国の経済悪化 ”と“ 米国経済悪化 ”,“ 経済悪化 ”,“ 悪化 ”を，まとめて手がかりキーワードと呼ぶ．

3.2 因果関係の抽出

3.1 節の方法で取得した文書データには，手がかり表現と手がかりキーワードを含む文が

Vol.2009-DBS-149 No.9
2009/11/20

入力キーワード＝“米国の経済悪化”，手がかり表現＝{を理由に}の場合

【基本】
検索式1 ＝ “を理由に米国の経済悪化”

【助詞の除去】
検索式2 ＝ “を理由に米国経済悪化”

【キーワードの分割】
検索式3 ＝ 米国 AND “を理由に経済悪化”
検索式4 ＝ 米国 AND 経済 AND “を理由に悪化”

図 2 検索式拡張の例
Fig. 2 An example of query expansion.

必ず存在する．本研究では，検索した事象の要因表現は，その文の手がかり表現の直前の位置に出現するものとして抽出する．また，手がかり表現の直後，つまり手がかりキーワードを含む箇所が結果表現，すなわち興味対象の事象である．要因表現と結果表現は，係り受け解析器 CaboCha[8] を用いて係り受け解析を行うことによって取り出す．係り受け解析の出力例を**表 1** に示し，要因表現と結果表現および因果関係の定義について以下で説明する．

(1) **要因表現**

手がかり表現よりも前に存在する文節を，手がかり表現に近い文節から順に前方に辿っていく．このとき係り受け解析で係り先が手がかり表現を含む文節よりも後ろになるまで，文節を連結した文字列を取り出す．ただし，接続助詞や終助詞が現れた時点で連結処理は終了する．そして，取り出した文字列の末尾の助詞を除去した文字列を要因表現とする．表 1 では,「企業は」の係り先が手がかり表現「伴う」よりも後ろになっているので,「米国で起きた金融危機に」という文字列が取り出される．末尾処理により，要因表現は「米国で起きた金融危機」となる．

(2) **結果表現**

手がかりキーワードを含む文節を取り出す．ただし，その文節が係助詞や連体助詞，並立助詞で終わっている場合，その文節の係り先の文節まで連結した文字列を取り出す．取り出した文字列の末尾の助詞を除去した文字列を結果表現とする．表 1 では,「景気悪化に」から末尾の助詞を除去した「景気悪化」が結果表現となる．

(3) **因果関係**

抽出した要因表現と結果表現のペアを 1 つの因果関係とする．因果関係ネットワークを構築する際に，各因果関係の強さを表すために，因果関係 c_x の重み $weight(c_x)$ を以下のように定義する．

$$weight(c_x) = \sum_d \frac{cf_d(c_x)}{cf_d(C)} \quad (1)$$

ここで，d は因果関係 c_x が抽出された文書，$cf_d(c_x)$ は d から抽出された因果関係 c_x の数，$cf_d(C)$ は d から抽出された因果関係の総数である．

表 1 係り受け解析の出力例
Table 1 An example of the outputs of dependency structure analysis.

ID	文節	係り先
1	企業は	8
2	米国で	3
3	起きた	4
4	金融危機に	5
5	伴う	6
6	景気悪化に	7
7	対応する	8
8	ために	−

3.3 因果関係のクラスタリング

細かな表現の差異によりほぼ等価な因果関係を別の因果関係と判断していては，類似した因果関係を大量に抽出してしまう．そこで，因果関係のクラスタリングを以下の手順で行う．

a. 各要因表現から最も重要な語（以下，代表語）を 1 語抽出する．
b. 代表語間の類似性から因果関係の類似度を算出する．
c. すべての因果関係間の類似度が閾値未満であった場合，クラスタリングを終了する．
d. 最も類似度の高い因果関係を統合する．**b.** に戻る．

本研究では結果表現の最も重要な語は総じて手がかりキーワードであるため，要因表現から抽出する代表語の類似度を因果関係の類似度とする．因果関係の統合とは，類似する 2 つの因果関係において，重みの小さい方の因果関係を消去して，重みの大きい方の因果関係を更新することである．このとき更新後の因果関係は，統合前の 2 つの因果関係の重みの和と両方の代表語を保持する．

3.3.1 特徴語の抽出

要因表現内から次の形態素を抽出し，連続して現れる場合は連結する．このような語を特徴語として抽出し，この中から後述する 3.3.2 項の方法で代表語を選択する．

- 名詞（非自立と代名詞は除く）
- 連体助詞の「の」

3.3.2 代表語の選び方

代表語は要因表現内の要因を端的に表す語が望ましい．この代表語は類似度の算出以外にも，3.4 節で述べる因果関係ネットワークのノード名や，要因検索を繰り返し行う際の入力キーワードにも利用するため，非常に重要である．要因にはまた「〜した」といった表現が接続することが多い．そのため，代表語はサ変接続の名詞となる特徴語とした．そのような特徴語が要因表現内に複数出現する場合や 1 つも存在しない場合には，特徴語を構成する名詞の数が多いものを優先するようにした．これは，単に「悪化」というよりも「景気の悪化」といった方が事象が具体的になるからである．また，それでも複数の候補がある場合には，文中で手がかり表現に近い語を代表語とした．表 1 では,「米国で起きた金融危機」という要因表現から「金融危機」という代表語が選ばれる．これを因果関係の代表語とし，代表語のスコアは因果関係の重みをそのまま用いる．このスコアはノード名を決定する際に用いる．

3.3.3 因果関係の類似度の算出

因果関係の類似度の算出には 3.3.2 項の方法で選ばれた代表語を用いる．代表語の意味を考慮した類似度の算出は困難なため，代表語を形態素に分割し，それらの形態素の一致度を因果関係の類似度とする．すなわち，代表語 w_i と w_j の類似度 $Sim(w_i, w_j)$ を以下のように定める．

$$Sim(w_i, w_j) = (\frac{equals}{const_i} + \frac{equals}{const_j}) \quad (2)$$

ここで，$const_i$ は代表語 w_i を構成する名詞の数，$equals$ は代表語 w_i，w_j を構成する名詞のうち共通する名詞の数である．因果関係の代表語が 1 つである場合，$Sim(w_i, w_j)$ がそれぞれの代表語をもつ因果関係の類似度となる．この類似度が閾値 0.5 よりも高かった場合，因果関係を統合する．この統合によって，1 つの因果関係が複数の代表語をもつようになる．そのため，一般に代表語を複数もつ因果関係 c_x と c_y の類似度 $Sim(c_x, c_y)$ は以下のように，その因果関係の各代表語間の類似度の平均値となるように定義する．

$$Sim(c_x, c_y) = \frac{1}{|R(c_x)||R(c_y)|} \sum_{w_i \in R(c_x)} \sum_{w_j \in R(c_y)} Sim(w_i, w_j) \quad (3)$$

ここで，$R(c_x)$ は因果関係 c_x のもつ代表語の集合，$|R(c_x)|$ はその代表語数である．

3.4 因果関係ネットワークの構築

因果関係はその重みによってランク付けされており，因果関係ネットワークの構築には上位の因果関係のみを用いる．本システムでは，因果関係ネットワークを JUNG (Java Universal Network/Graph)[9] を用いて可視化した．要因とその結果を，それぞれ始点ノードと終点ノードに配置することで因果関係ネットワークを表現している．すなわち，2 つのノードとそれを結ぶエッジが 1 つの因果関係を表している．終点ノードの名前は手がかりキーワードであり，因果関係の代表語がそのノードに接続している始点ノードの名前となっている．ここで，因果関係が複数の代表語をもつ場合，スコアが高い代表語を始点ノードに表示する．

我々は因果関係をより深く分析できるようにするため，始点ノード名となっている事象の要因をさらに検索する．つまり，始点ノード名をそのまま入力キーワードとして要因検索を行う．これを繰り返し行うことで，詳細な因果関係ネットワークが構築できる．ユーザはこの因果関係ネットワークから，事象の要因を間接要因も含めてすべて閲覧でき，また要因として取り出された事象間の関連性を把握できる．

要因検索を繰り返し行う過程で，すでに抽出された要因と類似する要因が抽出されることがあるが，類似ノードを多数生成することはネットワークの可読性を下げる．そのため，類似度が閾値より高いノードはマージする．ノード間の類似度は，それぞれのノード名を決める際に用いた代表語の集合を用いて，3.3.3 項で述べた類似度算出方法により求める．ノードをマージした場合，先に生成されていたノードが残り，ノードに張られるエッジは両方を統合したものとなる．また，因果関係のもつ代表語はそれぞれの代表語集合を合わせたものとなる．

4. 実　　装

提案する要因検索システムの実装について説明する．

4.1 設 定 項 目

本研究では要因検索を実行するために以下のように設定した．

(1) **検索文書数**

1 回の要因検索につき，Yahoo!の検索エンジンが取得する検索結果を最大 50 件用いる．

(2) **手がかり表現**

“ に伴い ”,“ に伴う ”,“ が理由で ”,“ を理由に ”の 4 つの言語表現を用いる．我々は既に，ある事象の要因と結果の両方を含む文の中から，要因と結果を結びつける手がかり表現を自動抽出する方法[7] を提案している．その方法を用いて抽出した手がかり表現の中

Vol.2009-DBS-149 No.9
2009/11/20

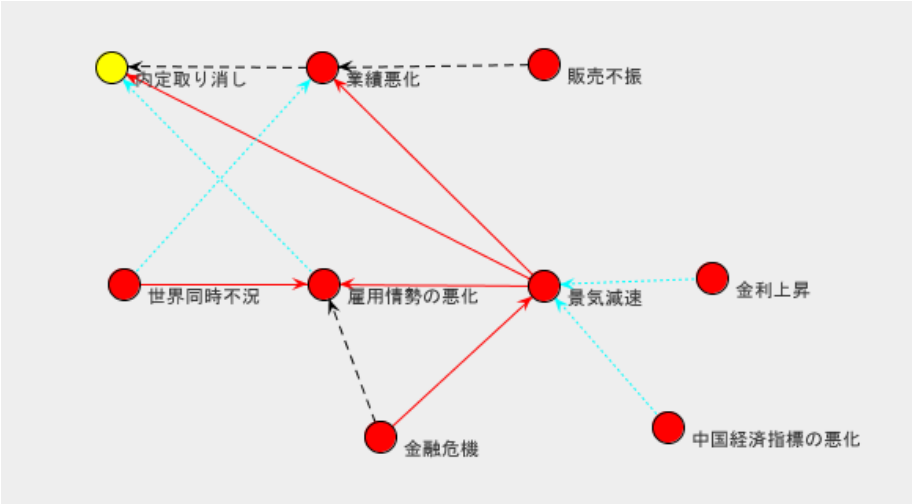

図 3 2 段階要因検索による因果関係ネットワーク
Fig. 3 The causal network of factor search results of phase 2.

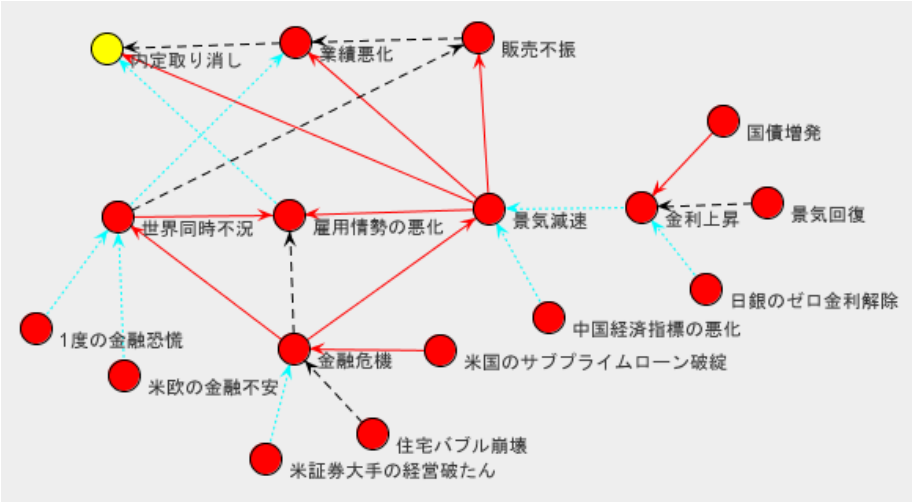

図 4 3 段階要因検索による因果関係ネットワーク
Fig. 4 The causal network of factor search results of phase 3.

から，因果関係の手がかりとして適切と考えるものを選択した．

(3) **因果関係の抽出**

1 回の要因検索において重みが大きい因果関係を最大 3 個抽出する．

4.2 要因検索の例

2008 年の景気悪化に伴い問題となっている“ 内定取り消し ”について要因検索を行った．ここで，“ 内定取り消し ”を入力キーワードに要因を取り出すことを 1 段階要因検索，1 段階要因検索によって取り出された最大 3 つの要因それぞれに関して，さらに要因検索を行うことを 2 段階要因検索と定義する．同様にして，n 段階要因検索も定義する．“ 内定取り消し ”の 2 段階要因検索により構築された因果関係ネットワークを**図 3** に示す．さらにもう 1 段階要因検索を行った場合の因果関係ネットワークを**図 4** に示す．

各事象の因果関係はこれらの図にあるように有向グラフで表現され，始点ノードが要因，終点ノードがその結果を表している．ユーザの入力したキーワードは黄色のノードで表現される．また，エッジの種類は赤色の実線，黒色の破線，青色の点線の 3 種類であり，赤が強い関連性をもつ因果関係を表現しており，黒，青となるに従って関連性が弱くなる．つまり，図 3 から“ 内定取り消し ”の主な要因は「景気減速」であり，“ 景気減速 ”の主な原因は「金融危機」であることがわかる．検索過程で同じ要因や類似する要因が取り出されることがあるため，図 4 の「世界同時不況」のように，図 3 の段階ですでに生成されていた「金融危機」から新たにエッジが張られることもある．

5. 実　　験

5.1 節では類似ノードをマージする場合とマージしない場合において，検索段階別のノード数の変化を測定して，マージ処理と検索回数がノード生成にどのように影響するか検証する．また 5.2 節では検索式を拡張することの効果について考察し，5.3 節では抽出した因果関係の精度を評価し，5.4 節では構築した因果関係ネットワークを評価する．なお本実験は 4.1 節で述べた設定で行った．

5.1 ノード数の検証

“ 内定取り消し ”の要因検索を 1 から 7 段階へと変化させたときのノード数の変化を調べた．類似ノードをマージする場合とマージしない場合におけるノード数の変化を**表 2** に示す．

マージありでは，7 段階目の検索で新たなノードは得られず，ノード数は収束している．それに対して，マージなしでは，ノード数はまだ増加している．このことから，類似ノード

Vol.2009-DBS-149 No.9
2009/11/20

表 2 ノード数の変化
Table 2 Changes in number of nodes.

	要因検索	1段階	2段階	3段階	4段階	5段階	6段階	7段階
ノード数	マージあり	4	8	14	18	21	22	22
	マージなし	4	9	20	35	60	94	142

のマージはネットワークの可読性を維持するために有効であることがわかる．4.2 節の図 4 では,「景気後退」が「景気減速」にマージされ,「米国発の金融危機」は「金融危機」にマージされた．しかし，3.3.3 項で定義した類似度では「景気回復」と「景気減速」を類似ノードとして検出してしまう可能性がある．また，マージなしでは「減速」や「改善」といった，単独では意味の具体性が乏しいノードが生成されていた．このようなノードの名前が入力キーワードに利用されると，関連性の不明確な因果関係を抽出することが多かった．

両方の実験において，3 段階以降の要因検索では，入力キーワードである“ 内定取り消し ”との関係性の乏しい因果関係や，形態素の類似性からは判断できない類似した因果関係が抽出され，あまり有用な因果関係は発見できなかった．そのため，要因検索は 3 段階程度に留めるのが妥当と考えられる．

5.2 検索式拡張の効果

4.2 節の図 4 が生成される過程では，ほとんどの入力キーワードに対して，1000 件以上の検索結果が存在し，この上位 50 件を取り出して因果関係を抽出することができた．そのため，図 2 の方法で検索式を拡張する必要がなかった．入力キーワードを“ 民主党支持 ”に変えてみたところ，検索式を拡張しなかった場合，拡張した場合の検索過程で，取得した検索結果数は**表 3** に示すように異なっていた．このとき構築された因果関係ネットワークを，それぞれ**図 5**，**図 6** に示す．

表 3 から“ 民主党支持 ”と“ 高速道路無料化 ”,“ 30 日の総選挙 ”,“ 道路公団廃止 ”の 4 つのキーワードに関して，検索式拡張により検索結果数が大きく増加しており，本実験で必要と考える 50 件の検索結果を取得することができた．その結果，図 5 に比べて検索式を拡張した図 6 の方が，期待通り多くの因果関係を抽出できている．

5.3 抽出した因果関係の精度

4.2 節で説明した図 4 を生成する過程において，400 件の検索結果を取得し，それらのタイトルおよびスニペットの中から，検索したフレーズを含む 431 文を取得した．検索結果のタイトルやスニペットは，途中で文が省略されていることが多々あり，431 文のうち 68 文は因果関係の取得において必要な箇所が省略されていた．本システムではこの 68 文のうち

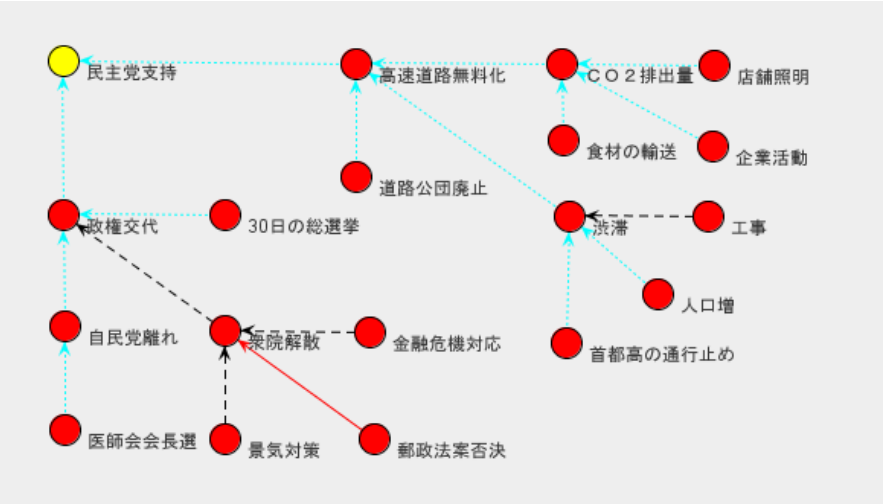

図 5 “ 民主党支持 ”の因果関係ネットワーク（検索式拡張なし）
Fig. 5 The causal network of “Support for the Democratic Party of Japan” w/o query expansion.

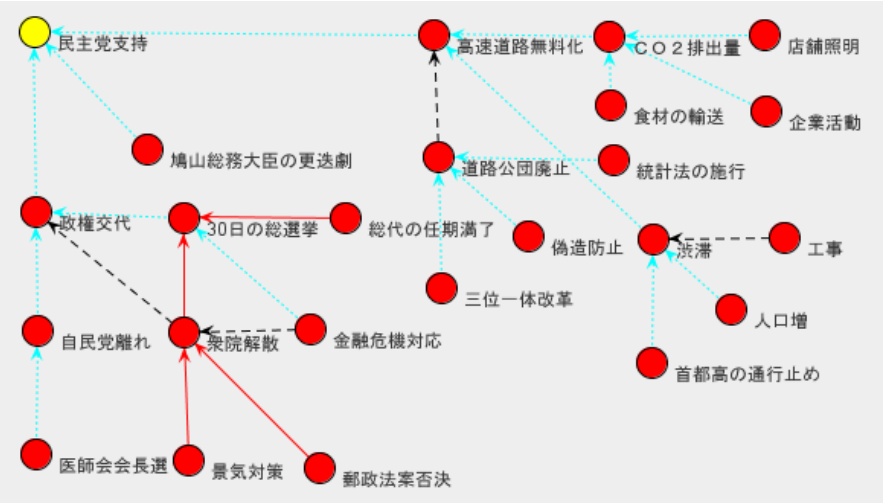

図 6 “ 民主党支持 ”の因果関係ネットワーク（検索式拡張あり）
Fig. 6 The causal network of “Support for the Democratic Party of Japan” w/ query expansion.

表 3 検索結果数
Table 3 The numbers of retrieval results.

入力キーワード	検索結果数	
	拡張なし	拡張あり
民主党支持	11	370
鳩山総務大臣の更迭劇	−	0
政権交代	335	335
高速道路無料化	28	143
衆院解散	804	804
30 日の総選挙	0	703
自民党離れ	3	7
道路公団廃止	0	333
ＣＯ２排出量	20700	20700
渋滞	82300	82300

表 4 因果関係の抽出精度
Table 4 Precision of causal relation extraction.

	適合率		
	要因表現	結果表現	両方
坂地らの手法	0.901	0.761	0.761
提案手法	0.953	0.969	0.923

45 文を自動的に因果関係の抽出対象から除外することができた．また，手がかり表現の前後が因果関係を表していない文が 431 文のうち 5 文あった．因果関係を含む文からの因果関係の抽出精度を調査するために，ノイズである 73 文を除外した 358 文を調査対象とする．

坂地ら[5] は景気の動向を示す記事 200 記事の中から，手がかり表現の直後に結果表現が存在する文書を 113 文書取得し，そこから抽出した因果関係の適合率を算出している．我々も同様に，要因表現と結果表現，その両方が正しかった場合の適合率を算出した（**表 4**）．坂地らの研究では，例えば「農産物の国際価格は，生産の増加を背景に総じて下落気味になろう」という文から,「を背景に」を手がかりにして,「農産物の国際価格は，総じて下落気味になろう」を結果表現として抽出している．それに対して，本研究では因果関係の結果と手がかり表現からなるフレーズで検索するため，結果表現が手がかり表現の前後に分かれることはなく，特に結果表現の精度が高かった．本研究での抽出失敗の主な原因の一つは,「リーマン・ブラザーズ」といった未知語への対応不足である．また,「金融危機に伴う景気悪化や少子化対策など克服すべき問題が」という文から，結果表現として「景気悪化や少子化対策」を抽出したが，本来は「金融危機に伴う景気悪化」と「少子化対策」が並立しているため不適切であった．

本研究では因果関係ネットワークのノード名を 1 単語としているため，要因表現と結果表現の中から，事象の要因と結果を端的に表す語をそれぞれ選ぶ必要がある．前者は 3.3.2 項で説明した代表語であり，本実験において正しく抽出できた 341 個の要因表現の中から，代表語を適合率 321/341，すなわち 0.941 で抽出することができた．後者は手がかりキーワードとしているが,“ 金利上昇 ”の要因を検索した際,「金利上昇を抑制する」要因である「景気低迷」と「金利上昇を促進する」要因である「景気拡大」が，同じ“ 金利上昇 ”の要因として扱われていた．このように,「抑制」や「促進」といった重要な情報が欠落している事例が 5 件見られたため，適合率は 342/347，すなわち 0.986 であった．

本研究では 1 文中から複数の因果関係を抽出することには未対応である．本実験では，358 文から 358 個の因果関係を抽出したが，実際には 425 個の因果関係が記述されていた．抽出できなかった因果関係の抽出が今後の課題の一つである．

5.4 構築した因果関係ネットワークの評価

図 4 には 22 個の因果関係が表示されている．そのうち,「1 度の金融恐慌」というノードは，特徴語の抽出ルールの不備による不適切なノード名で,「100 年に 1 度の金融恐慌」として表示されるべきものであった．

図 6 には 25 個の因果関係が表示されている．この実験では“ 高速道路無料化 ”の要因が「渋滞」や「ＣＯ２排出量」となっている．これは「渋滞やＣＯ２排出量を理由に高速道路無料化に反対する」といった文から抽出されたものであり,「反対」という重要な情報がなければ因果関係を正しく把握できなかった．また，キーワードを分割したことにより,「道路公団が発行しているハイウェイカードを偽造防止を理由に廃止する」といった文から,「偽造防止」が「道路公団廃止」の原因という誤った因果関係を抽出していた．

本研究では，5.3 節で説明したように因果関係の抽出精度は高く，さらに重みの大きい因果関係のみを可視化していることから，因果関係ネットワークの可読性はよいと予想した．確かに図 4 について因果関係の精度は 21/22，すなわち 0.955 と高かったが，図 6 については 19/25，すなわち 0.760 と低かった．この理由として，検索式を拡張したことにより，手がかり表現の前後が必ずしも興味対象の事象の因果関係ではなくなったことが挙げられる．

6. ま と め

本稿では，ユーザが興味のある事象の要因を検索し，さらにその因果関係を可視化することのできる要因検索システムを提案した．実験により，構築した因果関係ネットワークの

Vol.2009-DBS-149 No.9
2009/11/20

ノード数や検索結果数，抽出した因果関係の精度と構築した因果関係ネットワークを評価した．本研究では，手がかり表現を含むフレーズ検索の検索結果文書を用いることによって，高精度の因果関係抽出と，一定の可読性をもつ因果関係ネットワークの構築が実現できた．また，再帰的な要因検索を行うことにより，事象の間接的な要因を含めて検索することができた．

今後の課題として，因果関係ネットワークの可読性を向上させるために必要なキーワードをノードに付与することが挙げられる．また，生成した因果関係ネットワークにグラフマイニングの技術を応用して，因果知識を獲得できるようにしたい．

参 考 文 献

1) 佐藤浩史，笠原　要，松澤和光：テキスト上の表層的因果知識の獲得とその応用，電子情報通信学会技術研究報告， Vol.98, No.640, pp.27-32 (1999).
2) Khoo, C.S.G., Chan, S. and Niu, Y.: Extracting Causal Knowledge from a Medical Database Using Graphical Patterns, In: Proceedings of 38th Annual Meeting of the ACL, Hong Kong, pp.336-343 (2000).
3) 乾　孝司，乾健太郎，松本裕治：接続標識「ため」に基づく文書集合からの因果関係知識の自動獲得，情報処理学会論文誌， Vol.45, No.3, pp.919-933 (2004).
4) 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， Vol.4, pp.66-74 (2006).
5) 坂地泰紀，竹内康介，増山　繁，関根　聡：構文パターンを用いた因果関係の抽出，言語処理学会第 14 回年次大会論文集， pp.1144-1147 (2008).
6) 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 (DEIM Forum) 2009 論文集， C5-6 (2009).
7) 青野壮志，太田　学：ニュース記事に含まれる出来事の要因検索，電子情報通信学会，ISS 特別企画「学生ポスターセッション」，情報・システムソサイエティ誌 2009 年総合大会特別号， p.26 (2009).
8) 日本語係り受け解析器 CaboCha
http://chasen.org/〜taku/software/cabocha/
9) ネットワーク可視化・分析ツール JUNG
http://jung.sourceforge.net/