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

青野　壮志† 　太田　学†

† 岡山大学大学院自然科学研究科　〒 700–8530 岡山県岡山市北区津島中 3–1–1
E-mail: †{aono,ohta}@de.cs.okayama-u.ac.jp

あらまし　ある事象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る．本研究では，事象間の因果関係を見つける手がかりとなる表現を基に，Web 文書から要因を検索，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．提案する要因検索システムでは，因果関係の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構築する．因果関係ネットワークに関して評価実験を行い，提案手法の有効性を確認した．また，提案手法を用いて構築した因果関係ネットワーク内の特徴的な部分構造を基に，事象間の関係分析を行った．

キーワード　因果関係，可視化，Web マイニング，部分構造，グラフマイニング

Construction of a Causal Network and Acquisition of Causal Knowledges by Searching Factors

Hiroshi AONO† and Manabu OHTA†

† Graduate School of Natural Science and Technology, Okayama University
Tsushimanaka 3–1–1, Kita-ku, Okayama-shi, Okayama, 700–8530 Japan
E-mail: †{aono,ohta}@de.cs.okayama-u.ac.jp

Abstract　For deeply understanding an event, it is effective to grasp causal relations between related events. In this paper, we propose a method of extracting factors from retrieved Web documents and constructing a causal network by cue phrases used for finding causal relations between events. The proposed factor search system constructs a causal network in which factors and their results are respectively represented as source and sink nodes and each pair of them constitutes a causal relation. We also confirmed the effectiveness of the proposed system by experiments. Moreover, we analyzed relations between events by using characteristic substructures in the causal networks constructed by the proposed system.

Key words　causal relation, visualization, web mining, substructure, graph mining

1. ま え が き

新聞やテレビなどを通じて報じられる内容の中には，多くの事象が絡み合って引き起こされているものがあり，それを深く理解することは容易ではない．このとき，関連する事象間の因果関係を把握することは，その内容理解を助け，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，因果関係を把握したい事象を検索語として与えると，その事象の要因を詳細に抽出する手法を提案する．本研究では要因として抽出された事象をさらに要因検索するので，階層的に獲得した因果関係をグラフを用いて可視化する．また，因果知識としてそのグラフの部分構造から読み取れる事象間の関係について報告する．

本稿では 2 節で関連研究，3 節で手がかり表現，4 節で提案する要因検索，5 節で提案システムの評価実験，6 節で事象間の関係分析，7 節でまとめと今後の課題について述べる．

2. 関 連 研 究

本研究に関連する先行研究について述べる．

2.1 因果関係抽出

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,「に伴う」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・複文のみならず，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．本研究でもこの方法を利用する．

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，医療分野に特化した因果関係の構文パターンの作成や分類を人手で行っている．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 関係の四つに分類している．

2.2 因果関係ネットワーク構築

抽出した因果関係を可視化する研究もある．佐藤・堀田 [4] の研究では，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，オンラインニュース記事を一日単位で取得して，新たに抽出された因果関係から因果関係ネットワークを更新できるようにしている．事象は因果関係を含む文節から得られる重要単語だけでなく，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表現される．また，因果関係を含む文節から取り出した主語，動詞，目的語の三つのキーワードの一致判定を行い，ノードをマージしている．一致判定には概念辞書を用いることで，表記のゆれを吸収し，同義の単語を同じものとして扱うようにしている．抽出した因果関係は日々蓄積していくため，他のノードとほとんどマージされない因果関係を削除することで，因果関係ネットワークの可読性を高めている．

佐藤・堀田 [4] の研究では，入力キーワードの検索結果に含まれる因果関係全体を抽出することで，因果関係ネットワークを構築している．それに対して本研究では，入力キーワードを結果にもつ因果関係を抽出し，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．また，本研究では石井ら [6] のようにキーワードの一致判定により因果関係を整理する．本研究では事象を一つのキーワードで表現し，そのキーワードを構成する形態素の文字列としての類似度と類語辞典を併用して一致判定を行う．

2.3 部分構造解析

階層的に因果関係を獲得することにより，ある事象の起こった根本的な要因や事象の変遷の様子を獲得できる可能性がある．しかし，それにはある程度の規模をもつ因果関係ネットワークが必要となり，複雑になりすぎると可読性が下がることがある．そのため，グラフ構造を解析して，事象間の関係を明らかにできれば有用である．部分構造パターンを発見することはグラフマイニングの主要な技術の一つであり，様々なアルゴリズムが提案されてきた [7]．部分構造パターンは，残りの部分の構造や将来の構造を予測する手がかりにもなる [8]．構造予測に利用できる情報は大きく分けて,「ノードの情報」と「構造の情報」である．前者はノード自身がもっている情報であり，例えば，SNS におけるノードである参加者の個人情報（住所や年齢，趣味など）が挙げられる．後者は対象とするノードの周辺のリンク構造を表す情報であり，例えば,「友達の友達が友達である確率が高い」という観察から，A と B，B と C がそれぞれリンクでつながっている場合，A と C を直接つなぐリンクが作られやすいということが考えられる．本研究では，特徴的な部分構造が含むノード（事象）とそれらを結び付ける有向エッジを利用して，事象間の関係を分析する．

3. 手がかり表現

本研究では手がかり表現を用いて因果関係を抽出する．本節では人手で一つの手がかり表現を与えることにより，その他の手がかり表現を Web 文書から自動抽出する方法 [9] について述べる．手がかり表現の抽出には，係り受け解析器 CaboCha(注1)で取得できる品詞情報と文節の係り先情報を利用する．本研究では以下の条件にあてはまる表現を手がかり表現とする．

条件 1 一文中の要因とその結果の間の単語列.

条件 2 格助詞または連体助詞「の」で始まる.

条件 3 二つ以上の形態素で構成される(注2).

条件 4 助詞，助動詞または動詞－自立で終わる.

3.1 要因と結果のペアの取得

4 節で述べる提案する要因検索システムでは，検索語として与えた事象の要因を検索することができる．しかし，要因を検索するためには少なくとも一つ以上の手がかり表現が必要である．そこで，ある事象 R と手がかり表現を一つ人手で与えることにより，事象 R の要因となる事象の集合 $F = \{f_1, f_2, ..., f_k\}$ を検索し，要因である事象 f_i とその結果である事象 R のペアを取得する．

3.2 手がかり表現の抽出

3.1 節で取得した要因と結果を用いて，以下の検索式を生成する．

$$(f_1 \text{ OR } f_2 \text{ OR } ... \text{ OR } f_k) \text{ AND } R$$

この検索結果から要因と結果の両方を含み，結果よりも前の位置に要因が出現する文を取得する．そして，このような文から手がかり表現を抽出する．まず手がかり表現の条件 1 より，この文の要因と結果の間にある単語列が手がかり表現の候補となる．このとき，要因の後に続く形態素は手がかり表現の先

(注1)：CaboCha - http://chasen.org/〜taku/software/cabocha/
(注2)：「が」,「から」といった表現は除外した.

表 1　係り受け解析の出力結果（1）
Table 1　The output of dependency structure analysis (1).

ID	文節	形態素	品詞	係り先
1	景気悪化が	景気	名詞－一般	3
		悪化	名詞－サ変接続	
		が	助詞－格助詞	
2	ますます	ますます	副詞－助詞類接続	3
3	深刻化して	深刻	名詞－形容動詞語幹	5
		化	名詞－接尾	
		し	動詞－自立	
		て	助詞－接続助詞	
4	企業は	企業	名詞－一般	5
		は	助詞－係助詞	
5	内定取り消し	内定	名詞－サ変接続	－
		取り消し	名詞－一般	

頭の形態素となるため，条件 2 より格助詞(注3)または連体助詞「の」である必要がある．条件 2 を満たす場合，次のように手がかり表現を抽出する．

(1)　要因を含む文節内の要因の後に続く形態素を取り出す．
(2)　結果を含む文節の直前の文節までの文字列を (1) の形態素に連結する．ただし，その途中の文節について文節と係り先文節が連続していない場合，その文節までの文字列とする．
(3)　取得した単語列から副詞を除去する．
(4)　取得した単語列が条件 3, 4 を満たす場合，手がかり表現とする．

例えば，要因を「景気悪化」，結果を「内定取り消し」として検索すると,「景気悪化がますます深刻化して企業は内定取り消し」といった単語列が取得される．この単語列は要因と結果の両方を含み，結果よりも前の位置に要因が出現している．この単語列の係り受け解析結果は表 1 のようになる．まず，要因を含む文節内の要因の後に続く格助詞「が」を取り出す．次に,「深刻化して」の係り先がその次にくる文節 4 ではないことから，この文節までの文字列を格助詞「が」に連結して「がますます深刻化して」を取得する．この単語列には「ますます」という副詞が含まれているため，これを除去して「が深刻化して」となる．最後に，この単語列は二つ以上の形態素から構成され，接続助詞「て」で終わるため手がかり表現となる．

3.3　手がかり表現の抽出実験

3.1 節で述べた要因と結果のペア取得のために手がかり表現として“ に伴う ”を用い,“ 内定取り消し ”の要因と考えられる事象を検索した．その結果，検索結果 200 件の中から“ 内定取り消し ”の要因集合 $F = \{$景気悪化，金融不況，経済悪化，リーマンショック，業績悪化$\}$ を抽出した．この要因と結果のペアを用いて 3.2 節の方法で検索し，検索結果 500 件の中から手がかり表現を 29 種類抽出した．その一部を表 2 に示す．この中から抽出頻度の高い手がかり表現“ を理由に ”,“ に伴い ”,“ が理由で ”を選択し,“ に伴う ”と併せて四つの手がかり表現を提案する要因検索システムでは用いる．

表 2　抽出した手がかり表現
Table 2　Extracted cue phrases.

を理由に，に伴い，が理由で，を受けて，に端を発する，によるリストラで，による影響で，の影響で，が深刻化して，の影響を受け，を理由とする，を原因とする，を機に

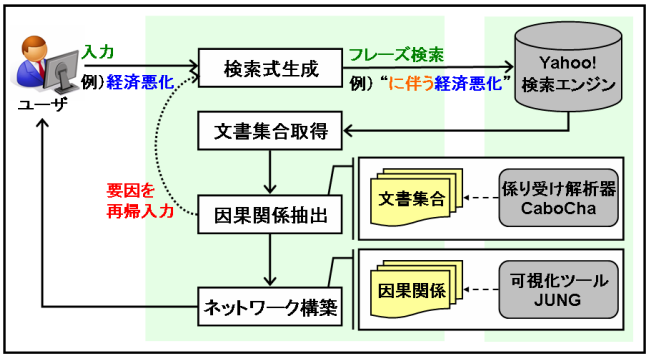

図 1　要因検索システムの処理の流れ
Fig. 1　A processing flow of the factor search system.

4.　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,「経済悪化」の要因について知りたい場合には,“ 経済悪化 ”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“ に伴う経済悪化 ”を生成する．この検索式を用いて，Yahoo!の検索結果を取得する．つづいて CaboCha 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(注4)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．

4.1　文書データの取得

本研究では，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．手がかり表現には“ に伴う ”,“ に伴い ”,“ を理由に ”,“ が理由で ”の四つの表現を用いる．例えば,「米国の経済悪化」の因果関係を抽出する場合,“ に伴う米国の経済悪化 ”といった検索フレーズ p_1 を生成する．他の手がかり表現も同様に用いて，検索フレーズ集合 $P = \{p_1, p_2, p_3, p_4\}$ を生成する．これらの検索フレーズを用いて，次の検索式を生成する．

$$p_1 \text{ OR } p_2 \text{ OR } p_3 \text{ OR } p_4$$

この検索式で十分な検索結果を取得できない場合には，それぞれの検索フレーズを図 2 のように拡張する．検索式 2，3，4 をこの順番に用いて検索結果を取得していき，十分な検索結果を取得した時点で検索を終了する．

(注3)：ただし,「で」は後ろに文の主語となる語が続くことが多いため除外した．

(注4)：JUNG - http://jung.sourceforge.net/

入力キーワード＝“米国の経済悪化”，手がかり表現＝“に伴う”の場合

検索式1＝“に伴う米国の経済悪化”

【助詞の除去】

検索式2＝“に伴う米国経済悪化”

【キーワードの分割】

検索式3＝米国 AND “に伴う経済悪化”

検索式4＝米国 AND 経済 AND “に伴う悪化”

図 2　検索式拡張の例

Fig. 2　An example of query expansion.

表 3　係り受け解析の出力結果（2）

Table 3　The output of dependency structure analysis (2).

ID	文節	形態素	品詞	係り先
1	企業は	企業	名詞－一般	8
		は	助詞－係助詞	
2	米国で	米国	名詞－固有名詞	3
		で	助詞－格助詞	
3	起きた	起き	動詞－自立	4
		た	助動詞	
4	金融危機 を	金融	名詞－一般	5
		危機	名詞－一般	
		を	助詞－格助詞	
5	理由に、	理由	名詞－一般	7
		に	助詞－格助詞	
		、	記号－読点	
6	経済悪化 が	経済	名詞－一般	7
		悪化	名詞－サ変接続	
		が	助詞－格助詞	
7	深刻化した	深刻	名詞－形容動詞語幹	8
		化	名詞－接尾	
		し	動詞－自立	
		た	助動詞	
8	ことに対して	こと	名詞－非自立	－
		に対して	助詞－格助詞	

4.2　因果関係の抽出

入力キーワードを“ 経済悪化 ”とし，検索した文書データに次のような検索フレーズを含む文が存在したとする．

「企業は米国で起きた金融危機 を理由に、
経済悪化 が深刻化したことに対して...」

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 を用いて係り受け解析を行うことによって取り出す．上記の検索フレーズを含む文の係り受け解析の出力結果を表 3 に示し，要因表現と結果表現の抽出方法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 3 では,「企業は」の係り先が手がかり表現“ を理由に ”の文節よりも後ろになっているので,「米国で起きた金融危機を」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**

手がかり表現よりも後ろに存在する文節を，手がかり表現に近い文節から順に後方に辿っていく．このとき係り受け解析で手がかり表現の係り先の文節になるまで，文節を連結した文字列を取り出す．ここで，文節に連体助詞の「の」や並立助詞の「や」が含まれている場合，その文節の係り先の文節までさらに連結する．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 3 では，手がかり表現の係り先は「深刻化した」であるため,「経済悪化が深刻化した」が結果表現となる．

次に，要因表現および結果表現を代表する語（以下，それぞれ要因および結果と呼ぶ）を抽出する．この要因と結果のペアを一つの因果関係と定義する．要因は要因表現内の末尾の特徴語とする．この特徴語 F は主語になり得る語であり，以下のように定義する．

```
<N> ::= <名詞(注5)> | <カタカナ>
<P> ::= <連体助詞「の」>
<F> ::= <N> | <N><F> | <F><P><F>
```

一方結果は入力キーワードとするため，先の例では,「金融危機」という要因と「経済悪化」という結果が抽出される．本研究ではこのようにして抽出した要因と結果，すなわち因果関係を可視化する．また，要因や結果の事象の詳細を知りたい場合には要因表現と結果表現が役立つ．

因果関係 c_x の重み $weight(c_x)$ は次のように定義する．

$$weight(c_x) = \sum_d \frac{cf_d(c_x)}{cf_d(C)} / (keysplit + 1) \qquad (1)$$

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，C は d から抽出された因果関係の集合，$cf_d(C)$ は d から抽出された因果関係の総数であり，$keysplit$ は図 2 の検索式 3，4 のようにキーワードの分割を行った場合の分割回数(注6)である．入力キーワードをそのまま用いる方が信頼性の高い因果関係が得られると考えることから，分割されたキーワードで抽出した因果関係の重みは小さくする．

4.3　因果関係のクラスタリング

類似する因果関係を大量に抽出してそのまま可視化すると，因果関係ネットワークの可読性が下がるため，因果関係のクラスタリングを行う．因果関係のもつ結果は総じて入力キーワードであるため，要因の類似性を因果関係の類似性とする．まず，要因を形態素に分割し，それらの形態素の類似度を算出する．要因 f_i から見た要因 f_j との類似度 $Sim(f_i, f_j)$ と，平均類似度 $Ave(f_i, f_j)$ を次のように定める．

(注5)：代名詞と非自立の名詞は除く．

(注6)：キーワードを分割していない場合は 0.

$$Sim(f_i, f_j) = \frac{equals_{i,j}}{const_i} \quad (2)$$

$$Ave(f_i, f_j) = \frac{Sim(f_i, f_j) + Sim(f_j, f_i)}{2} \quad (3)$$

ここで，$const_i$ は要因 f_i を構成する名詞の数，$equals_{i,j}$ は要因 f_i，f_j を構成する名詞のうち共通する名詞の数である．まず，$Sim(f_i, f_j)$ または $Sim(f_j, f_i)$ が 1.0 のとき，この二つの因果関係を統合する．また，$Ave(f_i, f_j)$ が 0.5 以上の場合には，Weblio の類語辞典(注7)を用いて，一致していない形態素の類語を調べる．このとき，一致しない形態素の一方がもう一方の類語であった場合には因果関係を統合する．類語辞典を用いることによって,「景気悪化」と「景気回復」のように「景気」という一つの形態素が一致するが，残りの形態素の意味が全くことなる因果関係を統合するという問題が発生しない．

類似する二つの因果関係のうち，重みの小さい方の因果関係は重みの大きい方の因果関係に統合する．すなわち，統合後の因果関係の重みは統合前の二つの因果関係の重みの和とし，統合後の因果関係のもつ要因は因果関係の重みの大きい方の要因とする．統合時には因果関係のもつ要因を統合前の二つの要因の和集合とする方法も考えられるが，その方法では統合の順番によってクラスタリング結果が異なるという問題が生じた [10]．そのため，因果関係の重みの大きい方の要因のみ残す方法を提案する．

4.4 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できるので，その上位の因果関係のみを用いて因果関係ネットワークを構築する．提案システムでは要因とその結果を，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構築する．すなわち，因果関係ネットワークは有向グラフで，その中の二つのノードとそれを結ぶエッジが一つの因果関係を表す．

我々は因果関係をより詳細に分析するために，始点ノード名となっている要因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，階層的に因果関係を取得し，それらのつながりを可視化する．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を把握できる．

この再帰的な要因検索の過程において，すでにノード名となっている事象と類似する事象が要因として抽出されることがある．ここで類似ノードをそのまま可視化すると，因果関係ネットワークの可読性が下がるため，類似ノードを統合する．新たに抽出した要因事象 f_i とノード名となっている事象 f_j について，4.3 節で述べた方法を用いて類似判定を行い，因果関係の統合条件に合う場合に統合する．ノードを統合すると，因果関係ネットワークは例えば図 3 に示すように変化する．

4.5 実　行　例

“ 内定取り消し ”の要因検索を例に，提案システムにおける因果関係ネットワークの構築について説明する．手がかり表現には“ に伴う ”,“ に伴い ”,“ を理由に ”,“ が理由で ”の四つを用い，さらに各回の要因検索は以下の設定で行う．

(1) **取得する検索文書**

Yahoo!の検索結果を最大 50 件

(2) **可視化する因果関係**

重みの大きい因果関係を最大三つ

つまり，1 回の要因検索で入力キーワードと手がかり表現を組み合わせた検索フレーズを用いて検索結果を 50 件取得し，その中から抽出した因果関係のうち重みの大きいものを最大三つ可視化する．例えばユーザが要因検索システムに入力した“ 内定取り消し ”を入力キーワードとして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三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 4 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 5 に示す．

各事象の因果関係はこれらの図のような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示している．ユーザが入力したキーワードは黄色のノードで表示される．また，エッジの色は因果関係の重みを示しており，赤が最も大きく，黒，青となるに従って小さくなる．つまり，図 4 から“ 内定取り消し ”の主な要因は「景気の悪化」であり,“ 景気の悪化 ”の主な要因は「100 年に一度の金融危機」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,「住宅バブル崩壊」のように複数の事象の要因となる事象がある．

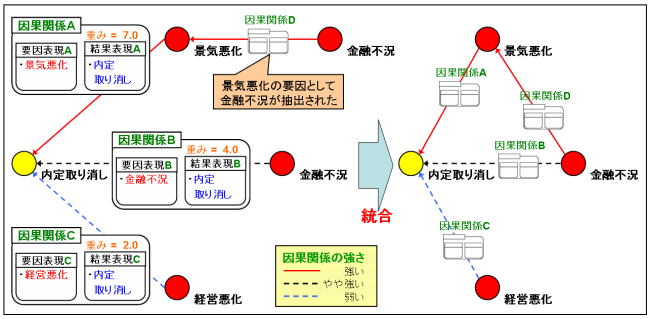

図 3　ノードの統合

Fig. 3　Merging of nodes.

5. 実　　験

“ 内定取り消し ”の 3 段階要因検索結果を基に，提案システムを評価した．5.1 節で提案システムが取得した文書データから抽出した因果関係の抽出精度，5.2 節では因果関係ネットワークとして可視化した因果関係の精度について述べる．

5.1 因果関係の抽出精度

図 5 を生成する過程で取得した因果関係を含む文，358 文を評価対象とした因果関係の抽出精度について述べる．単純に比較することはできないが，本実験結果を坂地ら [5] の実験結果と比較して示す [10]．両手法における要因表現と結果表現の適

(注7)：Weblio - http://thesaurus.weblio.jp/

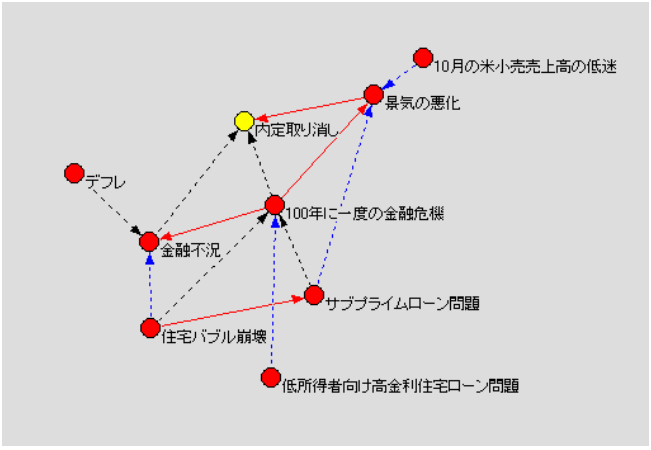

図 4　2 段階要因検索による因果関係ネットワーク
Fig. 4　The causal network of factor search results of phase 2.

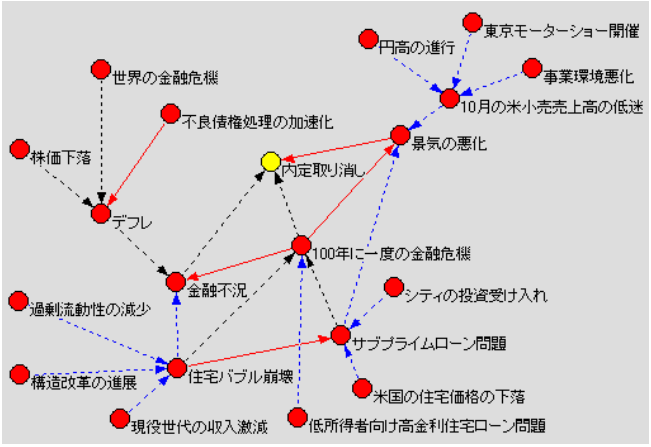

図 5　3 段階要因検索による因果関係ネットワーク
Fig. 5　The causal network of factor search results of phase 3.

表 4　要因表現と結果表現の抽出精度
Table 4　Extraction accuracy of factor and result expressions.

	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

合率と，要因表現と結果表現が両方正しい場合を正解とする適合率を表 4 に示す．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，結果表現の精度は高かった．

また，提案手法は要因と結果を抽出する．それぞれの適合率は 0.9 以上であり，高い精度で因果関係を抽出することもできた [10]．しかし，“ 金利上昇 ”の要因を検索した際，「金利上昇を促進する」要因と「金利上昇を抑制する」要因の両方が抽出される事例があった．要因は特徴語一語，結果は入力キーワードであるが，事象の表現には不十分である場合がある．そのため，要因表現および結果表現の中からどの語を抽出し，どのように利用するかについてはさらに検討する必要がある．

5.2　可視化した因果関係の精度

図 5 には 20 個のノードとそれらを結ぶエッジにより，24 の因果関係が可視化されている．そのうち，不適切な因果関係を示したものは四つだったため，可視化に利用した因果関係の精度は 20/24，すなわち 0.833 である．不適切な因果関係は，主に入力キーワードを分割した検索式を用いた場合に抽出されていた．図 2 に示す検索式 4 を用いた場合には，例えば「投資成果は投資対象市場の下落に伴う悪化に加え、」や「高齢化の進展に伴う悪化が予測される長期財政見通し」といった，直接“ 米国の経済悪化 ”について述べられていない文が抽出されることがあった．この場合，“ 米国の経済悪化 ”の要因として「投資対象市場の下落」と「高齢化の進展」を抽出してしまう．このような不適切な因果関係を抽出しないようにするには，検索式に含まれる語をすべて含む文のみを用いるようにしたり，検索式の生成の仕方を工夫したりすることが考えられる．また，本実験において不適切な因果関係の重みは総じて小さく，青のエッジで表現されていた．このことから，可視化する因果関係を決定する重みの閾値を調整することで，不適切な因果関係を可視化しないなどの改善策も考えられる．

6.　因果関係ネットワークの分析

構築した因果関係ネットワーク内の部分構造を用いた，事象間の関係分析結果について報告する．なお本実験で用いた入力キーワードは“ 内定取り消し ”，“ 円高 ”，“ 地球温暖化 ”の 3 語であり，2009 年 12 月に要因検索を行った．“ 円高 ”，“ 地球温暖化 ”の要因検索結果をそれぞれ図 6，図 7 に示す．

本研究では階層的に因果関係を抽出するため，図 8 に示すように，エッジを 1 本しかもたないノードを含む部分構造は頻出する．そこで，このように提案システムの特性上頻出する部分構造は除き，すべてのノードについて入次数と出次数の和が 2 以上となる部分構造を分析した．

6.1　部分構造の分類

因果関係はある事象とその結果発生した事象を表しているため，事象の時系列に対応しているとも考えられる．また，複数の因果関係からなる部分構造では，1 番最初に起こる事象と最終的に起こる結果が重要となる．この最初に発生したと考えられる事象は入次数 0 のノードで表され，最終的に起こったと考えられる事象は出次数 0 のノードで表される．本研究では，これらのノードに注目して，部分構造に含まれる入次数 0 のノードの数と出次数 0 のノードの数に基づいて部分構造を分類する．例えば，図 9 のようにノードを追加したり，エッジを追加したりしたとしても，入次数 0 のノード（赤のノード）と出次数 0 のノード（青のノード）の数は変化しないため，これらは類似した特徴をもつ部分構造とみなす．以降，例えば入次数が 0 のノードが一つ，出次数が 0 のノードが二つの部分構造を 1-2 構造のように表記する．

6.2　考　　察

実際に抽出した部分構造とその具体例を図 10 に示す．

1-1 構造は出次数 0 のノードが一つだけあり，その事象に至るまでの複数の経緯が可視化されている．図 10 の (A) において 1 ⇒ 2 ⇒ 3 はより詳しい経緯，一方 1 ⇒ 3 は途中を短絡した因果関係を示している．1-1 構造において根本的な要因は事象 1

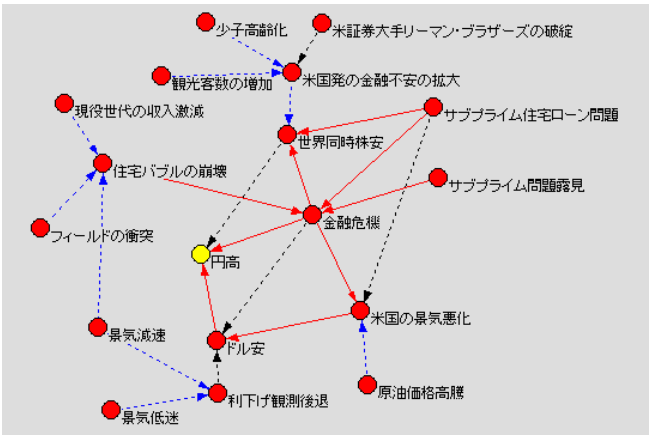

図 6 “ 円高 ”の因果関係ネットワーク

Fig. 6 The causal network of “Strong yen”.

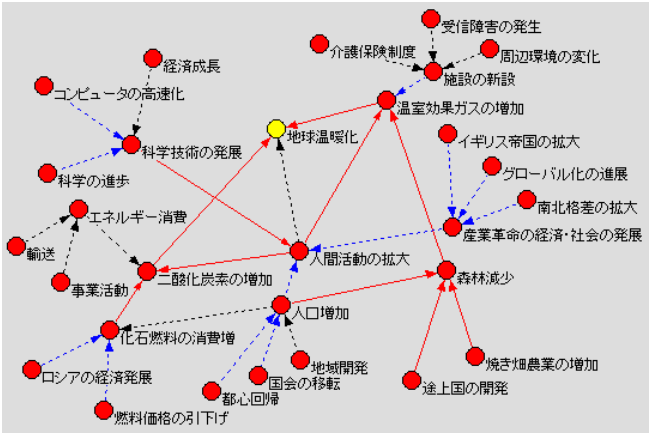

図 7 “ 地球温暖化 ”の因果関係ネットワーク

Fig. 7 The causal network of “Global warming”.

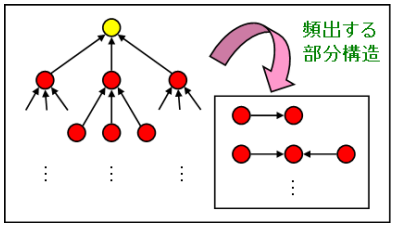

図 8 提案システムの特性上頻出する部分構造

Fig. 8 Frequent substructures characteristic of the proposed system.

であり，この要因ノードと出次数 0 の結果ノードだけ残し，この要因ノードとその結果ノードをエッジで結べば因果関係の一種の要約が得られる．また，事象 1 から最終結果に至る別々の経路上の事象，例えば，(B) の事象 2 と事象 3 はそれぞれ同じ事象に起因して，同じ事象を引き起こすという共通点をもつことから，字面では判断できない類似事象であったり，これらの事象がほぼ同時期に発生したりしていると考えられる．(D) は (C) に比べてエッジが多く，(C) では判別できない事象 3 と事象 4 の関係，例えばそれらの発生順を推測することができる．

2-2 構造は入次数 0 と出次数 0 のノードが共に二つずつある．図 10 の (F),(G) では，間接的な因果関係も含めると事象 4 と

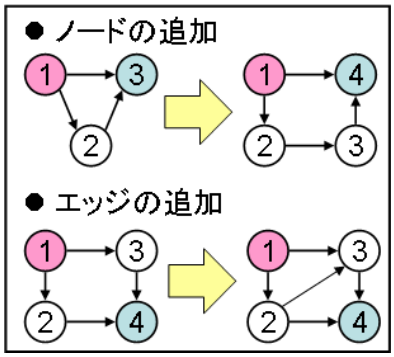

図 9 類似した特徴をもつ部分構造

Fig. 9 Substructures with similar features.

事象 5 は事象 1 と事象 2 という共通の異なる事象を要因にもつ．このように共通の異なる事象から引き起こされる事象間，また共通の異なる事象を引き起こす事象間にはそれぞれ何らかの関連があると予測される．具体例をみると，「世界同時株安」と「ドル安」，「住宅バブル崩壊」と「サブプライムローン問題」など，実際関連のある事象が抽出されている．3-3 構造など入次数 0 と出次数 0 のノードがさらに増えた場合でも，複数の共通要因をもつ事象や複数の同じ結果をもたらすような事象を発見することで，同様に関連のある事象を抽出できると考えられる．

1-2 構造のように，入次数 0 のノードと出次数 0 のノード数が異なる部分構造には，入次数 0 のノード数と出次数 0 のノード数の等しい部分構造が含まれている．例えば図 10 の (I) には (A) の構造が二つ，(E) の構造が一つ含まれている．(A) の関係分析結果を用いると，根本的な要因は事象 1 であることから，事象 2 を省略して一種の要約を行うことができる．また，(E) の関係分析結果を用いると，事象 1 と事象 2，事象 3 と事象 4 にはそれぞれ関連があることがわかる．しかし，事象 1 と事象 2 の間にエッジがあることから，事象 1 の方が発生時期が早いと推測できる．

実験では循環を含む部分構造は抽出されなかったが，因果関係から発生順が読み取れることから，仮に抽出されるとすれば，循環に含まれるそれぞれの事象はほぼ同時発生する，またはそれらの事象がスパイラルのように発生しているなどの推測が可能となる．

部分構造を分析することにより，事象間の類似性や発生順を把握したり，ネットワークを縮約したりできることがわかった．本研究では誘導部分グラフだけではなく，すべての部分グラフを取り出し，エッジ数の少ない比較的単純な構造について分析した．図 10 の (J) は (H) の部分構造にエッジを追加した構造である．(H) では「人口増加」と「人間活動の拡大」の内容や発生時期は類似していると推測されるが，(J) ではこの二つのノード間の発生順も読み取れる．基本的にエッジ数の多い部分構造を取り出す方がより詳細に分析することができるが，ユーザがどの程度詳細な分析結果を要求するかに応じて，抽出する部分構造の複雑さを制御する必要がある．

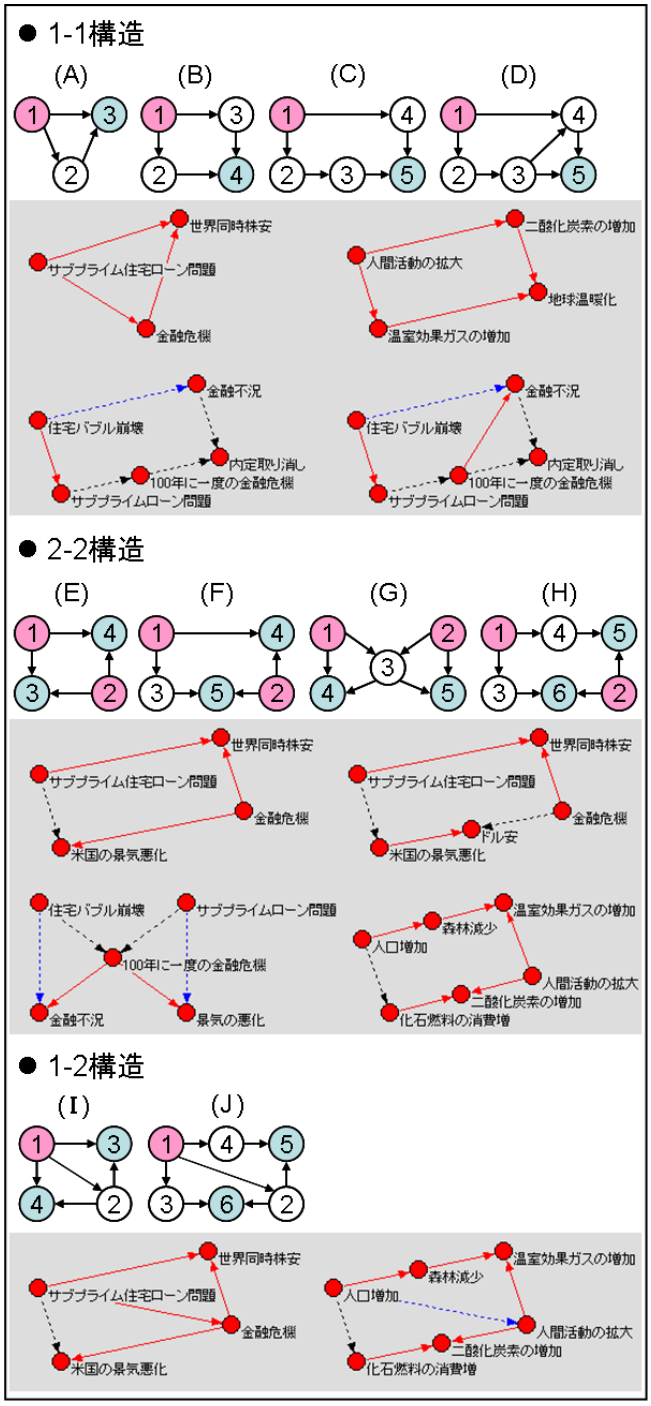

図 10　抽出した特徴的な部分構造

Fig. 10　Extracted characteristic substructures.

7. 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．提案システムでは，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，一定の可読性をもつ因果関係ネットワークを構築することができる．また，因果関係ネットワークの部分構造に基づいて事象間の関係を分析し，事象の類似性や発生順を推測したり，ネットワークを縮約したりできる可能性があることがわかった．

事象間の関係分析を自動化することにより，因果関係ネットワークから有用な因果知識を獲得できるようにすることが今後の課題である．また，複雑なネットワークをうまく縮約して可読性を高める方法についても検討していきたい．

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：SVO 構造を用いた因果関係ネットワーク構築手法について，情報処理学会研究報告， Vol.2009-DBS-149, No.10, pp.1-8, November (2009).

[7] 浅井達哉，有村博紀：半構造データマイニングにおけるパターン発見技法，電子情報通信学会論文誌， Vol.J87-D1, No.2, pp.79-96 (2004).

[8] 鹿島久嗣：ネットワーク構造予測，人工知能学会誌， Vol.22, No.3, 344-351 (2007).

[9] 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).

[10] 青野壮志，太田　学：要因検索による因果関係ネットワークの構築，情報処理学会研究報告， Vol.2009-DBS-149, No.9, pp.1-8, November (2009).