

ブログのコミュニティ分析による因果関係事象の抽出

高見真也[†] 田中克己[†]

特定のコミュニティ内において、ある話題が及ぼす影響が話題になることがある。それらは、マスメディアから提供されるニュースとしては大きく扱われないが、近年ブログの普及に伴い、一般消費者から提供される情報によりウェブ上で話題になることが多い。本研究では、ブログにおいて同じ話題に関するバースト特性がコミュニティごとに異なる点に着目し、ある話題と因果関係にある事象を発見するための手法を提案する。

Finding causal relations by analyzing blog communities

SHINYA TAKAMI[†] and KATSUMI TANAKA[†]

The influence that a certain topic exerts may get into the news in a specific community. General consumers often discuss them actively on the Web as Blog spreads in recent years though they are not treated on a large scale as news offered from the mass media. In our research, we pay attention to the difference of the bursty characteristics in each Blog community, and we propose the technique to find causal relations from the blogosphere.

1. はじめに

我々は、ブログ記事に代表される一般消費者が提供する情報をマイニングすることにより、社会的知識が獲得できると考えている。

社会的知識とは、主観性を残した情報を統計学的に処理することによって得られる知識のことで、断片的な情報を統合することによって生成される。また、時間が経てば、その内容が変化する場合もあり、世論の一部と捉えることができるものもある。

このような知識の獲得手法は、専門家により整理された信頼性の高い情報があらかじめ存在すると仮定した、従来の情報検索のアプローチとは大きく異なる。そこで、我々はそれを新たな知識発見プロセスと位置づけ、「Social Knowledge Mining」と呼んでいる。

社会的知識の獲得に関する研究としては、検索語に接続助詞を付加する事による同位語の獲得や、文書構造を考慮した出現頻度計算による詳細語の獲得などがある。

本論文では、社会的知識のうち、因果関係に関する知識をブログ記事から抽出する手法を紹介する。

2. 因果関係とは

因果関係に関する知識についての研究は、人工知能をはじめ、多くの分野において古くから関心が向けられてきたテーマである。ここではまず、因果関係とはどのような特徴をもった関係であるかを整理する。

2.1 相関関係と先後関係

因果関係と比較される関係として、相関関係と先後関係がある。これらの関係も因果関係と同様に2つの事象間の関係を示すものである。

相関関係とは、2つの事象の変化が連動している関係であり、事象 P がある基準に対して、例えば時系列的に増加すれば、事象 Q もそれに呼応しているかの様に、増加または減少しているような場合をいう。

ただし、相関関係の場合、事象 Q の変化は、事象 P によってもたらされたものとは限らない。なぜなら、第三の事象が共通の原因である可能性もあるからである。また、それぞれの事象が複数の原因によるものである場合、一部が重複している可能性はあっても、その原因集合がまったく同じであるとも限らない。

相関関係の例としては、以下のようなものがある。「ビールがよく売れる (P) 夏は電気代が高い (Q)」

この場合、ビールがよく売れることと電気代が高いことには直接の因果関係はない。おそらく、気温が高いという第三の事象が原因として考えられるが、他に

[†] 京都大学大学院 情報学研究所 社会情報学専攻
Department of Social Informatics,
Graduate School of Informatics, Kyoto University

も原因が存在する可能性もある。そのため、相関関係は、因果関係の必要条件であっても十分条件ではない。

先後関係とは、2つの事象が発生した時期に前後関係が存在している関係であり、事象 P が発生した後に、事象 Q が発生したような場合をいう。

因果関係が、原因と結果を表す関係だと考えると、それらには当然時系列的な前後関係が存在するため、先後関係は因果関係にとって必要条件ではあるが、先後関係の場合、事象 P がたまたま先に（事象 Q が後に）起こっただけという可能性もある。

先後関係の例としては、以下のようなものがある。「おみくじで凶をひいた (P) 後に怪我をした (Q)」

この場合、おみくじで凶をひいたからといって、怪我をしたわけではない。また、怪我をするためにはおみくじで凶をひく必要があるわけでもない。そのため、これらの事象間に因果関係は存在しない。

以上のように、相関関係と先後関係は、因果関係にとって必要条件にはなり得るとしても、十分条件ではないため、「因果関係 \neq 相関関係 \cap 先後関係」であることが分かる。

2.2 因果関係の特徴

因果関係とは、2つの事象が原因（事象 P ）と結果（事象 Q ）の関係にあるものをいう (1)。ただし、因果関係は2つの事象間の関係を表すものであるが、原因と結果という観点から考えると、それぞれが複数存在する可能性もある。つまり、事象 P_i は複数の原因のうちの一つであるかもしれない (2)、 Q_j は複数の結果のうちの一つであるかもしれない (3)。

$$P \rightarrow Q \quad (1)$$

$$P_1, P_2, P_3, \dots, P_n \rightarrow Q \quad (2)$$

$$P \rightarrow Q_1, Q_2, Q_3, \dots, Q_n \quad (3)$$

このように、因果関係は、論理学でいう「 $P \supset Q$ 」とは異なる関係であることが分かる。事象 P は、事象 Q よりも時系列的に先に発生し、事象 Q の原因であることは間違いないが、事象 P が発生する度に、必ず事象 Q が発生することを確約しているわけではない。

そこで、事象 P は、複数の原因集合 ($\sum_k P_k$) に含まれる原因の一つであり、事象 Q は複数の結果集合 ($\sum_k Q_k$) に含まれる結果の一つであると考え、本研究では因果関係を以下のように定義する。

$$P_i \in \sum_k P_k \rightarrow Q_j \in \sum_k Q_k \quad (4)$$

つまり、因果関係とは、過去に発生した現象の観察結果であり、その関係は時間によって変化する可能性がある。また、対象となる2つの事象以外にも、原因または結果に関係する事象が存在する可能性があるこ

とから、因果関係は主観的な側面を持つと考えられ、その関係を因果関係と判定するためには暗黙の前提を考慮する必要がある。そのため、相関関係や先後関係と違い、因果関係を定量的に評価するのは難しいといえる。

2.3 因果関係の種類

因果関係を扱う多くの談話理解研究においては、2つの事象が有意志的な「行為」か、非有意志的な「状態」であるかに着目した分類がなされている¹⁾。また、2つの事象が共に有意志的な行為である場合、その主体が同一か、異なるかにより区別する必要があるが、因果関係には表1に示す、5種類の関係種別が存在するといわれている。

対象となる2つの事象の有意志性による分類は、先に述べた因果関係の主観性と、その判別には暗黙の前提が必要であるという特徴とを併せて、2つの事象をつなぐ接続語がそれぞれ異なる点に着目した研究に利用されている。

3. 因果関係知識の獲得

このように、因果関係は原因となる事象 P と、その結果となる事象 Q から成り、それらが因果関係であるという情報が加わる事で、因果関係知識となる。

3.1 因果関係表現の自動獲得

因果関係はその関係を定量的に判断することが難しいため、因果関係に関する研究の多くは、事象 P と事象 Q の関係が因果関係であるか否かに着目し、その判定を目指すのではなく、因果関係知識であると思われる表現情報を獲得することを主な目的としている。

このような因果関係知識の獲得に関係するものとして、接続語による識別や、確率モデルを利用した機械学習的なアプローチにより、因果関係に関する知識、あるいは因果関係を表す言語表現を文書集合から自動獲得しようとする研究がいくつか行われている。

まず、2つの事象を接続する言語表現として、接続語など手がかりとなる語句を利用する事で因果関係知識の獲得を目指し²⁾³⁾、名詞句間の因果性を判定するベイジアン・ネットワークを構築⁴⁾しようとする研究が行われている。

また、動詞並列句に注目し、接続語などに依存しない手法⁵⁾ やナーブバイズ分類器を用いた自動獲得手法⁶⁾ なども提案されている。また、既知の因果関係知識を教師データとして用いた確率因果モデルを構築し、因果関係の判定に利用しようとする研究⁷⁾ もある。

これらの研究で対象となる因果関係知識は、新聞記事など比較的自然言語処理に適した文書からの獲得が

表 1 因果関係の種類
Table 1 type of causal relations

P → Q	関係種別	説明
状態 → 状態	原因 (cause)	P が起こった結果, Q が起こった
行為 → 状態	効果 (effect)	P した結果, Q が起こった
状態 → 行為	前提条件 (precondition)	P が起こったので, Q した
行為 → 行為	手段-主体同一 (means) 応答-主体相違 (enable)	A が Q するために, P した A が P したために, B が Q した

中心となっている。また、因果関係にある 2 つの事象において、文書内共起率が高いという前提が多くの研究で利用されている。

3.2 社会的知識としての因果関係

我々もまた、因果関係知識の獲得を目的としているが、「太陽が昇ると、気温が高くなる」といった、因果関係を知識と考えた場合、自然科学的な根拠が客観的に示されるような知識ではなく、因果関係の特徴である主観的な側面を持つ知識に注目した。

主観的な側面を持つ因果関係知識とは、事象 P と事象 Q が因果関係にあるという情報そのものが人間の意図や意思に強く影響される場合をいう。そのような場合、因果関係が否かにおいて、特定の時区間においてのみ真であったり、少なくともどちらかの事象が有意志的行為である可能性が高い。このような特徴をもった知識は、社会的知識の一種であると考えられる。

一方、ブログの爆発的な普及により、ウェブ上には一般消費者が提供した情報が大量に溢れている。ブログ記事の多くは、日記の延長線上にあるかのように口語的、また主観的である。そこで、我々はこのような一般消費者から提供される主観的な情報を対象にする事で、社会的知識としての因果関係知識をブログ記事から獲得することを目指している。

4. ブログ記事における因果関係知識

4.1 時系列的な相関とバースト特性

一般消費者が提供する情報として注目されている CGM (Consumer Generated Media) の一種であるブログは、社会的知識を獲得するための対象として適している。

ブログには他の記事を容易に引用できる機能があり、ブログ記事から時間情報を取得する事は難しくないため、時系列的な話題の盛り上がり度 (バースト特性) を測定する事ができる⁸⁾。そのため、ある瞬間に爆発的に利用されはじめた話題を抽出する技術として、Bursty Mining という技術も存在する。

このように、ブログ記事からは時間情報をもとにした先後関係やバースト特性の類似度をもとにした話題

の相関関係を、定量的に評価する事ができる。

4.2 コミュニティ評価型モデル

ブログ記事の多くは専門家ではなく一般消費者が書いたものであるため、主観的な内容のものが多い。ブログ記事を書く一般消費者のことをブロガーと呼ぶが、彼らを書くブログ記事の特徴から、以下の 2 つの仮定を置いた。

仮定 1 自分に関係することを話題にしやすい

仮定 2 相関関係よりも因果関係を話題にしやすい

ブロガーは自発的な情報提供者であり、自身が興味をもった内容を引用したり、その感想として記事を書くことが多い (**仮定 1**)。また、他人に見られていることを意識しながら記事を書く傾向があり、取り上げる話題については定量的な評価が可能で、議論に発展する可能性が低い相関関係よりも、むしろ主観的な側面をもつ因果関係を好むのではないかと考えられる (**仮定 2**)。

これらの仮定が正しいとすれば、因果関係にある事象 P と事象 Q のうち、原因となる事象 P が与える影響 (結果となる事象 Q) が特定のコミュニティに深く関係する場合、該当コミュニティ内において、事象 P と事象 Q は高い共起性をもった話題として取り上げられる可能性が高い。

なお、本論文では、ブログにおけるコミュニティを「特定のテーマに属する記事を頻繁に書くブロガーが書いたブログ記事集合」と定義する。複数のテーマに属する記事を書くブロガーの場合、カテゴリ毎に分類された記事集合が、それぞれのコミュニティに属すると考える。そこで、このようなコミュニティ毎に異なる話題のバースト特性と共起性を利用した因果関係知識の獲得手法をコミュニティ評価型モデルとして提案する。

まず、ある話題 T_i が、コミュニティ C_x において話題にされた件数を $Tf_{C_x}(T_i)$ と定義する。コミュニティ C_x の規模を S_{C_x} として、評価値 TfC (Topic frequency in Community) を次のように定義する。

$$TfC = \frac{Tf_{C_x}(T_i)}{\sum_k Tf_{C_x}(T_k)} \cdot \log \frac{\sum_k S_{C_k}}{S_{C_x}} \quad (5)$$

ある話題において、特定の時区間におけるバースト特性が高いコミュニティを発見するためには、この TfC 値を各コミュニティ毎に測定し、評価値の高いものを選択する。

次に、求めたコミュニティ内において、話題 T_i と因果関係にある可能性が高い話題 T_j を発見するために、以下の2種類の評価法を考案した。

$$Case1: \frac{TfC_x(T_i \cap T_j)}{TfC_x(T_i)} \quad (6)$$

$$Case2: \frac{\sum_k C_k}{\sum_k (C_k \supset T_j)} \cdot \log \frac{\sum_k TfC_x(T_k)}{TfC_x(T_j)} \quad (7)$$

$Case1$ は、コミュニティ内での共起性を示しており、 $Case2$ ではさらに、話題 T_j のコミュニティ依存度とコミュニティ内での話題特性を測定している。

5. 予備実験

社会的知識としての因果関係知識をコミュニティの特性を利用して獲得する提案手法の妥当性を評価するため、Google の検索結果を中心とした予備実験を行った。

5.1 実験方法

今回の実験では、因果関係知識の獲得を目的とするのではなく、因果関係知識が提案手法により獲得し得ることを示す事を主な目的としている。そのため、既知の因果関係知識を用いて、事象 P と事象 Q が、提案手法により抽出される得る特徴を備えているかどうかを検証する形で行った。

実験方法としては、事象 P を検索エンジンの検索語として使用し、コミュニティ毎に限定した検索結果から、それぞれ共起率の高い名詞句を抽出し、提案手法における $Case1$ または $Case2$ の特徴を示しているものが、事象 Q に関するものかどうかを調査した。

今回実験に使用した既知の因果関係知識は、 P 「紀子様のご懐妊された」を原因として、 Q_1 「皇室典範の改正が延期された」、 Q_2 「ベビー（子供）用品を扱う会社の株価が上昇した」という2種類の結果を持つものである。

まず、検索エンジンに適した検索語にするため、事象 P を主体 (S) と行為/状態 (V) に分解し (8)、 S_p として「紀子様」、 V_p として「ご懐妊」という2つの名詞句に単純化し、併せて検索語として使用した。

$$S_p + V_p \rightarrow S_q + V_q \quad (8)$$

ブログ記事を対象にした検索エンジンでは、ブログ記事本体のみが検索対象になっているため、コミュニティを限定した検索を行うのが難しい。そこで、ブログ記事だけではなくその周辺の情報も検索対象として

扱う Google を検索エンジンとして利用した。また、検索結果から共起率の高い名詞句を抽出する対象として、今回は Google の検索結果上位 100 件のスニペットを使用した。

なお、今回の実験では、ブログのコミュニティを対象とした検索を、コミュニティを限定するための検索語を追加することで擬似的に行った。さらに、精度を高めるため、「blog」という検索語を追加した。コミュニティを限定するために使用した検索語は、「ニュース」「政治」「経済」「株」「映画」「料理」「スポーツ」「音楽」「車」「ゲーム」「生活」の11種類である。

また、今回は既知の因果関係知識を用いたこともあり、厳密には時系列的なバースト特性は観測しておらず、検索結果件数をもとに各種の特性評価を行った。

5.2 実験結果

まず最初に、 TfC 値の高いコミュニティを見つけるため、検索語として、各コミュニティを限定するためのキーワードを入力し、その後、各コミュニティ毎に事象 P を表すキーワードを検索語に追加して検索件数を測定した。その結果を表 2 に示す。「総数」は各コミュニティを表す検索語のみの件数、「絞り込み」はさらに事象 P を表す検索語を追加した場合の件数である。表 3 は、検索語として「blog」を追加した場合の結果である。

これらの結果から、「blog」という検索語を追加することで、コミュニティ「ニュース」の順位が下がり、コミュニティ「株」の順位が上昇しているのが分かる。また、コミュニティ毎の差が大きくなっている事から、コミュニティを限定する方法として、「blog」を検索語に追加することが有効であると考えられるため、以降の結果は基本的に「blog」を追加したものとする。

次に、 TfC 値の高いと考えられるコミュニティのうち、「政治」「経済」「株」を対象に共起率の高い名詞

表 2 Google の検索結果「紀子様 ご懐妊」
Table 2 Search Result by Google
with "princess kiko pregnancy"

カテゴリ	総数	絞り込み	割合 (%)
ニュース	288,000,000	48,800	0.0169
政治	197,000,000	32,700	0.0166
経済	154,000,000	25,200	0.0164
株	120,000,000	14,900	0.0124
映画	241,000,000	29,800	0.0124
料理	130,000,000	14,900	0.0115
スポーツ	342,000,000	31,300	0.0092
音楽	292,000,000	23,300	0.0080
車	268,000,000	19,600	0.0073
ゲーム	339,000,000	17,500	0.0052
生活	738,000,000	35,600	0.0048

句上位 25 件を測定した結果を表 4 に示す。また、比較対象として、コミュニティを限定しない場合（基準）と、WebQA 的な手法として「影響」を検索語に追加した場合の結果も併せて表 4 に含めた。

この結果から、事象 Q_1 に関係する 3 種類の名詞句、「皇室」「典範」「改正」については、すべての場合において、上位 25 件に含まれている事が分かる。ただし、上記 3 種類の名詞句の TF 値を相加平均または相乗平均にて評価した場合、コミュニティや手法毎の差が明確になるが、コミュニティを限定しない「基準」の評価は依然高く測定された。

そのため、事象 Q_1 については、提案手法のうち $Case1$ の事象に該当するが、 TfC 値の高いコミュニティに限定することによる効果はあまり期待出来ないことが分かった。また、「秋篠宮」という名詞句がすべてのコミュニティで上位に位置しているが、これは事象 P の主体 S_p と密接に関係するものであり、事象 Q に関係するものとの識別は今後の課題である。

ところで、事象 Q_2 に関係する名詞句は表 4 には存在していないが、事象 Q_2 に関係すると思われる「ベビー」という名詞句はコミュニティ「株」においてのみ測定された。

そこで、「ベビー」という名詞句が提案手法における $Case2$ の事象に該当すると考え、「ベビー」よりも TF 値が高い名詞句で、かつコミュニティ「株」以外には上位にランクインしていない名詞句を調査した結果、 TF 値が高い順に、「ライブドア（ライブ+ドア）」「取引」「投資」「堀江」「会社」が存在していることが分かった。

「ベビー」を含むこれらの名詞句について、 $Case2$ の評価値を計算するため、事象 P の代わりにこれらの名詞句を検索語として使用した場合の検索結果件数を測定した（表 5）。この結果から、「ベビー」は他の

名詞句と比較して、 $Case2$ の評価値が高いことが分かる。

今回の予備実験では、提案手法のうち、 $Case1$ の場合については、他の手法と比較して、優位性を証明することは出来なかったが、 $Case2$ の場合については、提案手法を用いる事が事象 P と因果関係にある事象 Q を獲得するために有効な手段である可能性が高いことが示された。

6. おわりに

因果関係には論理学の「 $P \supset Q$ 」とは異なり、暗黙の前提が存在するため、因果関係の知識が存在しなければ、因果関係の正否を判断することはできない。そのため、因果関係知識を獲得する対象は、新聞記事など比較的自然言語処理に適した文書が中心であった。

我々は社会的知識としての因果関係知識を獲得する対象としてブログ記事に着目し、バースト特性に時系列的な相関関係が見られる 2 つの話題において、事後に発生した話題が、特定のコミュニティ内でバーストした場合、それらが因果関係にある可能性が高いことを利用したコミュニティ評価型モデルを提案した。

そして、そのような特徴を有し因果関係を持つ話題の存在が Google の検索結果を用いた予備実験により確認できた。

謝辞

本研究の一部は、文部科学省 21 世紀 COE 拠点形成プログラム「知識社会基盤構築のための情報学拠点形成」（リーダー：田中克己、平成 14~18 年度）ならびに、文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」、計画研究「情報爆発時代に対応するコンテンツ融合と操作環境融合に関する研究」（研究代表者：田中克己、A01-00-02、課題番号 18049041）によるものです。ここに記して謝意を表すものとします。

表 3 Google の検索結果「blog 紀子様 ご懐妊」
Table 3 Search Result by Google
with "blog princess kiko pregnancy"

カテゴリ	総数	絞り込み	割合 (%)
政治	30,000,000	22,900	0.0763
経済	24,900,000	17,700	0.0711
株	21,500,000	11,100	0.0516
ニュース	67,900,000	31,700	0.0467
スポーツ	51,900,000	22,000	0.0424
映画	77,200,000	21,200	0.0275
料理	36,800,000	9,520	0.0259
車	51,000,000	11,900	0.0233
音楽	71,200,000	16,300	0.0229
ゲーム	64,800,000	11,200	0.0173
生活	139,000,000	22,500	0.0162

表 5 Google の検索結果「blog 株」
Table 5 Search Result by Google
with "blog stock"

検索語	検索件数	TF 値 (表 4)
ライブドア	7,440,000	0.05230
取引	5,280,000	0.04404
投資	7,450,000	0.02753
堀江	1,480,000	0.01927
会社	10,400,000	0.01927
ベビー	387,000	0.01376

表 4 Google の検索結果「blog 紀子様 ご懐妊 [政治 | 経済 | 株 | 影響]」
Table 4 Search Result by Google
with "blog princess kiko pregnancy [politics|economy|stock|influence]"

No.	基準 / TF 値	+政治 / TF 値	+経済 / TF 値	+株 / TF 値	+影響 / TF 値					
1	紀子	0.044300	紀子	0.038309	紀子	0.037985	紀子	0.034132	紀子	0.045419
2	懐妊	0.043220	懐妊	0.036988	ご	0.036099	ご	0.032755	懐妊	0.041481
3	ご	0.041869	様	0.036988	懐妊	0.036099	懐妊	0.032480	ご	0.034917
4	様	0.039708	ご	0.035931	様	0.033405	様	0.031104	様	0.032029
5	秋篠宮	0.013506	政治	0.022457	経済	0.017511	株	0.027801	影響	0.020478
6	皇室	0.010265	皇室	0.012153	秋篠宮	0.014009	秋篠宮	0.009359	秋篠宮	0.019953
7	典範	0.007834	秋篠宮	0.011096	皇室	0.011045	ニュース	0.007432	皇室	0.014439
8	さま	0.007563	典範	0.008190	ニュース	0.009698	皇室	0.006606	妃	0.014439
9	妃	0.006753	問題	0.007926	妃	0.008082	日記	0.005780	さま	0.013914
10	ニュース	0.005673	改正	0.005548	典範	0.007004	の	0.005505	典範	0.012602
11	改正	0.005673	的	0.005548	さま	0.005927	ドア	0.005230	ニュース	0.008926
12	子	0.005402	妃	0.005548	改正	0.005927	ライブ	0.005230	子	0.007876
13	日記	0.004862	家	0.005284	政治	0.005388	さま	0.004955	改正	0.007614
14	トラック	0.004322	こと	0.004756	日本	0.004849	こと	0.004679	論議	0.006563
15	バック	0.004322	年	0.004756	子	0.004580	月	0.004679	こと	0.004988
16	こと	0.003782	さま	0.004227	問題	0.004041	年	0.004679	そう	0.003938
17	問題	0.003512	小泉	0.004227	国会	0.003502	取引	0.004404	日	0.003938
18	年	0.003241	ニュース	0.003963	日記	0.003502	典範	0.004404	日記	0.003938
19	の	0.002971	月	0.003699	の	0.003233	妃	0.004404	の	0.002888
20	御	0.002971	御	0.003435	小泉	0.003233	よう	0.003578	さ	0.002625
21	記事	0.002701	子	0.003435	こと	0.002963	改正	0.003578	トラック	0.002625
22	日	0.002701	よう	0.003170	人	0.002963	さん	0.002753	バック	0.002625
23	話	0.002701	天皇	0.003170	年	0.002963	投資	0.002753	月	0.002625
24	お	0.002431	日	0.003170	ランキング	0.002694	話	0.002753	今	0.002625
25	人	0.002431	日本	0.002906	社会	0.002694	ん	0.002477	問題	0.002625

参 考 文 献

- 1) Allen, J.: *Natural language understanding (2nd ed.)*, Benjamin-Cummings Publishing Co., Inc., Redwood City, CA, USA (1995).
- 2) Girju, R.: Automatic detection of causal relations for Question Answering, *Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering*, Morristown, NJ, USA, Association for Computational Linguistics, pp. 76–83 (2003).
- 3) Inui, T., Inui, K. and Matsumoto, Y.: Acquiring causal knowledge from text using the connective marker tame, *ACM Transactions on Asian Language Information Processing (TALIP)*, Vol. 4, No. 4, pp. 435–474 (2005).
- 4) Sanchez-Graillet, O. and Poesio, M.: Acquiring bayesian networks from text, *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC-2004)*, pp. 955–958 (2004).
- 5) Torisawa, K.: An unsupervised learning method for commonsensical inference rules on events, *Proceedings of the Second CoLogNet-ElsNET Symposium*, pp. 146–153 (2003).
- 6) Chang, D.-S. and Choi, K.-S.: Causal Relation Extraction Using Cue Phrase and Lexical Pair Probabilities, *Proceedings of the First International Joint Conference on Natural Language Processing 2004 (IJCNLP-2004)*, pp. 61–70 (2004).
- 7) Takamura, H., Inui, T. and Okumura, M.: Latent Variable Models for Semantic Orientations of Phrases, *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, pp. 201–208 (2006).
- 8) Kleinberg, J.: Bursty and hierarchical structure in streams, *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (SIGKDD-2002)*, New York, NY, USA, ACM Press, pp. 91–101 (2002).