

動向情報の抽出と要約  
—動向をまとめる—

Extraction and Summarization of Trend Information

難波英嗣

Hidetsugu Nanba

広島市立大学大学院 情報科学研究科  
〒731-3194 広島市安佐南区大塚東 3-4-1

Graduate School of Information Sciences, Hiroshima City University

## 1. はじめに

株価や内閣支持率の推移、世の中で注目を集める話題、ある分野での発明や研究の動向などを表す情報を、動向情報と呼ぶ。膨大な電子化情報の海から、ユーザが必要とする情報に効率的にアクセスするための技術が求められている今日、動向情報についても、さまざまなテキストから自動的に抽出したり、より短い文章や図表などで簡潔にまとめたりすることで、ユーザがより簡単に把握できるようにするための研究が活発に行われるようになってきている。本稿では、動向情報の抽出と要約に関する諸研究を概観する。

動向情報の抽出と要約は、一般に複数のテキストをシステムの入力とし、その内容を概観できる情報をユーザに提示することから、複数テキスト要約の一種と考えられている。複数テキスト要約は、一般には以下の手順で作成される[15]。

- (1) 関連する文書の自動収集(システム入力)
- (2) 重要文抽出
- (3) 文書間の共通点と相違点の抽出
- (4) 重要箇所抽出
- (5) 重要箇所の出力順序の決定
- (6) 書き換え
- (7) 要約結果の提示

動向情報の抽出と要約でも、大まかな処理手順は一般的な複数テキスト要約との共通点が多いが、以下の二点が異なる。

- 手順2において、テキストから抽出する単位が多くの場合、文ではなく語や句である。
- 手順3において、一般的な複数テキスト要約では、文書間の共通点の抽出に焦点が当てられることが多いのに対し、動向情報の抽出と要約では、むしろ相違点に焦点が当てられる。

特に二点目は、動向情報の抽出と要約の結果を利用する多くのユーザが、マーケティングや技術動向調査などを目的とした動向分析と結びついており、複数のテキストの概要を知るためというよりも、むしろそこから新たな知識を見出すことに主眼を置いていることにも関係する。

現在、動向情報の抽出と要約に関する研究には以下の3つの流れがある。

1. 株価や内閣支持率に関するテキスト集合中から、統計量の推移を示すグラフを生成する。
2. ある事項が一定期間にテキスト集合中で何回言及されたのかを調べ、その回数に基づいて動向分析を行う。
3. 特許や論文などの技術文献を対象に、ある分野でどのような研究や技術開発が行われているかを

分析する。

本稿では、各項目について、それぞれ2節、3節、4節で述べる。特に4節では、筆者らが中心となって実施した研究プロジェクト「特許マイニングタスク」を紹介する。最後に、5節で本稿をまとめる。

## 2. 特定の統計量に関するテキスト集合からの動向情報の抽出と要約

本節では、動向情報の抽出と要約に関する研究の第一の流れとして、株価や内閣支持率に関する新聞記事やBlogなどのテキスト集合から、統計量の推移を示すグラフを生成する諸研究について述べる。その代表的な処理手順は、統計量と時間表現の対をテキスト集合から抽出し、統計量を時系列に並べグラフとして提示する、というものである。具体的な手順を日経平均に関する以下の新聞記事を例に説明する。

5日の東京株式市場は続落し、日経平均株価は前日終値比218円33銭安の、14,042円91銭で取引を終えた。

この記事には、日経平均株価に関する2つの統計量「218円33銭」と「14,042円91銭」が記述されている。前者は前日終値との今日の終値の差分を、後者は今日の終値そのものを示しており、それぞれ「比較表現」、「直接表現」と呼ばれる。情報抽出技術を用いて日付(5日)と統計量(14,042円91銭)を抽出し、さらに、この記事が1998年9月6日に書かれている場合には、抽出された「5日」という表現を「1998年9月5日」に変換すれば、 $x = \text{“1998年9月5日”}$ 、 $y = \text{“14,042円91銭”}$ という時間と統計量の対が得られる。さらに、「前日終値比218円33銭安」と「14,042円91銭」から、前日終値が「14,261円24銭」(14,042円91銭+218円33銭)であることが分かるので、 $x = \text{“1998年9月4日”}$ 、 $y = \text{“14,261円24銭”}$ という対も得られる。このような対をテキスト集合全体から抽出し、横軸を日付、縦軸を日経平均株価とするグラフ上に描画すれば、日経平均株価のグラフが得られる。

図1は、実際に1992年3月1日から1992年3月31日の新聞記事から抽出された円相場の推移に関するグラフの例である。図の横軸は年月日、縦軸は円を表す。このようにグラフ化することで、文章で提示するよりも円相場の推移が直感的に分かる。

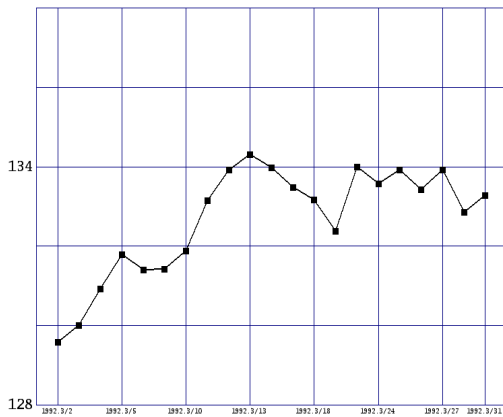


図1 円相場を表すグラフ(システムの出力例)

日経平均株価や円相場のような統計量の推移データの多くは Web などから入手可能であるため、単純にグラフ化するだけであれば、あえてテキスト集合から統計量を抽出する必要はない。しかしながら、テキスト中には「14,042 円 91 銭」といった統計量の周辺に、「なぜ株価が下落したのか」といった要因などが記載されることがある[16]。また、株価について述べた Blog には、「数値の推移が社会にどのような影響を与えているのか」や「世の中の人ほどどのように受け止めているのか」といった情報も書かれることがある[10]。そこで、このような情報を抽出しグラフと共に提示すれば、単純にグラフのみを提示された場合よりも、より深い分析が可能になると考えられる。

さて、前述の日経平均株価の例は、テキスト中にただ一つの統計量が出現するケースであるが、テキスト中に複数の統計量が出現するケースも存在する。例えば、「18日に発表した5月の国内生産の実績によると、日産自動車は前年度比 22.8%減、トヨタ自動車は同 20.4%減となった。」という文において、「5月」、「国内生産」、「日産自動車」、「トヨタ自動車」が統計量名の構成要素である。この場合、統計量名の記載方法は単一の場合と異なり、より複雑な処理が必要となる。この例では、「5月の日産自動車の国内生産」、「5月のトヨタ自動車の国内生産」という2つの統計量名ができる。このような判断を自動的にを行うために、森らは、まず、テキスト中から統計量の構成要素とその役割(例えば、「トヨタ自動車」は「国内生産」という統計量を限定する条件)を認識し[8]、次に、構成要素を組み合わせることで1つの統計量名を作る[9]、という二つの課題を解決する手法を提案している。

上述の日産自動車とトヨタ自動車の国内生産は、直接的に比較可能な統計量の例であるが、この他にも統計量間で因果関係など、何らかの関係が存在する場合がある。河合ら[5]は、新聞記事から統計量名を抽出した後、異なる統計量名が同一テキスト内に出現する度合い(共起関係)に基づいて、関連性の高い統計量を検出・可視化する手法を提案している。また、提案手法を1998年と1999年に毎日新聞に掲載された約22万記事に適用することにより、例えば、「経済成長率」と

「金利」と「マンション価格」の間の関連性を検出できたことを報告している。

複数の統計量を扱ったこの他の研究として、村田ら[6]のものが挙げられる。基本的な方針は河合らのものと共通するが、統計量名間の関係性をテキスト内ではなく文内での共起により測るため、抽出される関係は、例えば台風に関するトピックの場合、「中心気圧」と「最大風速」のように、河合らのものよりも関連性が高い。

本節最後に、テキスト集合から統計量を抽出する研究プロジェクト「動向情報の要約と可視化に関するワークショップ (MuST)」を紹介する。このプロジェクトは、国立情報学研究所(NII)が主催する第6回および第7回評価ワークショップ NTCIR のタスクのひとつとして開催された[3, 4]。評価ワークショップとは、複数の研究グループが協調と競争を通して、問題設定やテストコレクション、評価方法について共同開発していく枠組みである。本節で紹介した関連研究の多くは、MuST で作成されたテストコレクション(評価用ベンチマーク)を用いたものである。なお、このテストコレクションは、NTCIR 事務局を通じて入手することが可能である。詳しくは NTCIR の Web ページ (<http://research.nii.ac.jp/ntcir/index-ja.html>) を参照されたい。また、MuST のオーガナイザや参加者が中心となって、2009年に人工知能学会情報編纂研究会が設立されるなど、この分野の研究はさらなる広がりを見せつつある。

### 3. テキスト集合中の単語出現頻度を用いた動向情報の抽出と要約

2節で述べた統計量そのものをテキストから抽出する研究手法に対し、ある事項が一定期間にテキスト集合中で何回言及されたのかを調べ、その回数に基づいて動向分析を行うものも提案されている。この応用例のひとつとして、世の中ではどのような話題が注目を集めているかを自動検出する研究が挙げられる。

Fujiki ら[1]は、Blog 集合や Web 掲示板を対象にある時点を境に言及が急増する事項に着目し、このような言及の度合いを、その事項の注目度と考えることで、多くの注目を集めている話題を抽出する手法を提案している。類似した考え方に基づいたシステムとして代表的なものに kizasi.jp がある。kizasi.jp とは、一日に約 25 万件のブログエントリを収集・解析し、ブログ上で話題になっている語を約 10 分ごとに集計し、ランキング形式で提示するサービスである。近年では、buzztter(<http://buzztter.com/ja>)など Twitter を対象にした同様のサービスもはじまっている。また、単語の代わりに Twitter の投稿に含まれる URL を収集・集計し、話題のサイトを検出するサービス tweetbuzz(<http://tweetbuzz.jp>)などもある。

言及の度合いは、話題性の検出以外にも書籍の売上予測に利用できる。Gruhl[2]は、ある書籍が Blog で

度々言及されるようになると、連動するようにオンライン書店 Amazon におけるその本の売上が推移することに着目し、ある書籍について言及している Blog のエントリ数から Amazon における本の売上を予測する手法を提案している。

言及の度合いの推移を利用したこの他の研究に、山本ら[17]のものがある。山本らは、ある単語の出現頻度を単位時間ごとに集計し、その単語の頻度の推移を用いた動向分析手法を提案している。具体的には、ピアソンの積率相関係数を用い、単語の頻度の推移という観点から、ある名詞句と関連度の高い別の名詞句を検出したり、ある単語の頻度の推移と類似したある統計量の推移と関連の高い頻度の推移を持つ単語(ある製品の売上げの変動と共に使用されるようになった単語)を検出したりしている。

上述の研究やサービスは、いずれもテキスト集合中の単語の出現頻度に基づいた研究やサービスであるが、この他に評判分析技術を用いた手法がある。評判分析[14]とは、ある対象の評価を記述しているテキスト断片に対して、肯定的な評価であるか、否定的な評価であるかを自動的に推定する自然言語処理技術である。この技術と本節でこれまで述べてきた技術を組み合わせることにより、Blog 上で最近人気を集めている製品を見つける、といったことが可能になる。このようなサービスのひとつに、株式会社ブログウォッチャーの SHOOTI BUZZ レポート(<http://sbr.shooti.jp>)がある。このサービスでは、評判分析技術に加え、Blog の著者の文体や記述内容から自動的に判定した著者の性別や年齢情報を用いた動向情報の分析を行うことが可能である。

この他、Twitter を対象とした評判分析サービスとして、Tweet Sentiments (<http://tweetsentiments.com>)がある。このサービスでは、ユーザが入力したキーワードに関する Twitter の投稿を評判分析し、その結果を時系列のグラフとして提示することができる。また、どの地域(国)のユーザが投稿したのかも併せて提示される。

#### 4. 技術文書からの動向情報の抽出と要約

本節では、特許や論文などの技術文献を対象に、ある分野でどのような研究や技術開発が行われているかを分析する技術について、特に、NII が主催する第 7 回および第 8 回 NTCIR ワークショップにおいて、筆者らが行った特許と論文を対象とした情報処理のためのテストコレクション<sup>1</sup>の構築研究[11, 12]を中心に述べる。

<sup>1</sup> なお、このテストコレクションも MuST と同様、NTCIR 事務局を通じて入手することが可能である。

#### 4.1 研究の背景

近年、大学研究者自身が関連論文だけでなく関連特許について情報を検索したり、特許を出願したりする機会が増えており、2010 年 5 月に政府の知的財産戦略本部が発表した「知的財産推進計画 2010」においても、推進計画 2006, 2007, 2008 および 2009 に引き続き、大学研究における特許情報の重要性が謳われている。

特許と論文を検索するのは、大学研究者に限った話ではない。例えば、特許庁の審査官は出願された技術が特許権の取得に該当するかどうか判断するために、過去に同様の特許が出願されたり論文が発表されたりしていないか調査する。これは一般に先行技術調査と呼ばれている。この他に、サーチャーと呼ばれる専門の担当者が審査官による審査を経た出願技術を再調査し、競合する他者の権利を無効化するために民間企業の社内で行われる無効資料調査でも、論文と特許が検索対象となる。

こうした状況を鑑み、特許と論文を対象にした検索や動向分析など、さまざまな目的に利用可能な言語処理技術の開発を最終目標とし、そのための第一歩として筆者らが位置づけているのが、特許マイニングタスクである。

#### 4.2 特許マイニングタスクの概要

特許マイニングタスクの最終目標は、ある分野の特許と論文から、図 2 に示すような技術動向マップを自動的に作成することである。図は、論文と特許を、「要素技術」と「効果」という観点から分類し、技術動向マップとしてまとめたもので、このような技術動向マップを自動生成するツールは、4.1 節で述べた先行技術調査や無効資料調査の支援ツールとして利用できる。

	効果 1	効果 2	効果 3
要素技術 1	[論文 A] [特許 X]		[論文 B]
要素技術 2	[論文 C]		
要素技術 3		[特許 Y]	[特許 Z] [特許 W]

図 2 特定分野の特許と論文から生成される技術動向マップの例

このようなマップを自動的に生成するためには、以下の 2 つの手順が必要となる。

- (手順 1) ある分野の特許と論文を網羅的に収集する。
- (手順 2) 手順 1 で収集された特許と論文から要素技術と効果の対を抽出し、技術動向マップとしてまとめる。

これらの 2 つの手順について、特許マイニングタスク

では、以下の2つのサブタスクを設定している。

- 学術論文分類サブタスク
- 技術動向マップ作成サブタスク

以下、これらのサブタスクの概要を述べる。

### 学術論文分類サブタスク

このサブタスクでは論文抄録に、特許分類体系のひとつである国際特許分類(International Patent Classification: IPC)のコードを自動的に付与する。IPCは、特許文献の技術内容によって上から順に「セクション」、「クラス」、「サブクラス」、「メイングループ」、「サブグループ」の5階層から構成・分類されており、国際特許分類第6版ではサブグループのレベルで約50,000<sup>2)</sup>のIPCコードが存在する。本サブタスクでは、最下層の「サブグループ」レベルのIPCコードを論文抄録に付与することを目的とする。図2は日本語論文の例である。ここで、<TOPIC-ID>は論文のIDを、<TITLE>と<ABSTRACT>は論文表題と概要を、それぞれ示している。タスクの参加者は、図3のような入力を与えられると、対応するIPCコードを自動的に出力するシステムを構築することが求められる。

```
<TOPIC><TOPIC-ID>312</TOPIC-ID>
<TITLE>二値画像用高速符号化/復号LSI</TITLE>
<ABSTRACT>二値画像データを高速で符号化、復号するLSIを開発した。参照ラインデータ上に「基準色変化点」を探すと並行して、それを参照するランのイメージデータを生成する方式により、復号性能を向上させた。また、符号化時と復号時共に同じ方向にデータが流れるパイプライン構成とし、さらに主な回路は共通化する構成によって回路を簡略化した。
</ABSTRACT>
</TOPIC>
```

図3 学術論文分類サブタスクの入力例

### 技術動向マップ作成サブタスク

このサブタスクは、要素技術とその効果を示す表現を、特許や論文から自動的に抽出することを目的とする。例えば「PM 磁束制御用コイルを設けて閉ループフィードバック制御を施すため、電力損失を最小化できる。」という文が入力されると、図4に示すように、要素技術と効果を示す個所に、それぞれ“TECHNOLOGY”および“EFFECT”タグを自動的に付与する。ここで、“EFFECT”タグの中には、さらに“ATTRIBUTE”と“VALUE”という2種類のタグが付与されている。技術の効果に関する表現は

多様であり、そのすべてを処理対象とするのは、現在の言語処理技術では非常に困難である。このため、例えば、「処理速度(ATTRIBUTE)が向上する(VALUE)」や「ノイズ(ATTRIBUTE)が減少する(VALUE)」のように、技術の効果が「属性(ATTRIBUTE)」と「属性値(VALUE)」の対で表現できるもののみを対象とする。近年の自然言語処理分野では、テキスト中に出現する属性と属性値の対の抽出が活発に研究されており、技術の蓄積が急速に進みつつある。特許や論文中の属性と属性値の対で表現可能な技術の効果に関する表現の抽出も、このような既存の技術の利用が期待できる。こうして、ある分野の論文と特許から、図4に示すような要素技術と効果の対が抽出できれば、図2に示すような技術動向マップの自動作成が実現可能になると考えられる。

```
PM 磁束制御用コイルを設けて<TECHNOLOGY>
閉ループフィードバック制御<TECHNOLOGY>
を施すため、<EFFECT><ATTRIBUTE>電力損失
</ATTRIBUTE> を <VALUE> 最小化
</VALUE></EFFECT>できる。
```

図4 技術動向マップ作成サブタスクに用いるデータの一例 (VALUE タグ内が文字列の場合)

なお、このサブタスクでは、図5のようにVALUEタグが付与される表現が数値となるものも対象としている。

```
<TECHNOLOGY>CRF</TECHNOLOGY> を用いた手法では、<EFFECT><VALUE>0.935</VALUE>の<ATTRIBUTE>精度</ATTRIBUTE></EFFECT>が得られた。
```

図5 技術動向マップ作成サブタスクに用いるデータの一例 (VALUE タグ内が数値の場合)

もし、例えば「形態素解析」や「機械翻訳」などの特定分野の論文や特許から図5に示すような精度値が抽出できれば、精度値を縦軸に、論文の著作年や特許の出願年を横軸にとることにより、精度値の時間的な推移を示すグラフが描画できる。図6は、「形態素解析」に関する複数の論文から、実際に精度値を手で抽出し、グラフにまとめたものである。このグラフから、形態素解析分野では、1994年頃から精度が95%以上に達しており、この分野の技術が成熟しつつあるということがわかる。ここで、2000年に精度が若干低下しているが、これは評価に用いるデータが異なるためである。本来ならば、評価用データや実験条件が違えば、評価値の直接的な比較はできないが、この分野への新規参入を検討している企業にとって、参入する余地があるかどうかの判断材料として利用するという目的であれば、図6のようなグラフは十分に有用である。なお、類似した研究に村田ら[7]のものがある。

<sup>2)</sup> NTCIR-7 特許マイニングタスクでは、これらのうち、学術分野とは関連性の低い分野を除外した30,885のIPCコードを対象とした。

村田らは、自然言語処理分野の論文概要から、「精度表現」、「主要な分野」、「言語名(その研究が対象としている言語)」、「組織・人名」を、情報抽出技術を用いて抽出し、表として出力している。

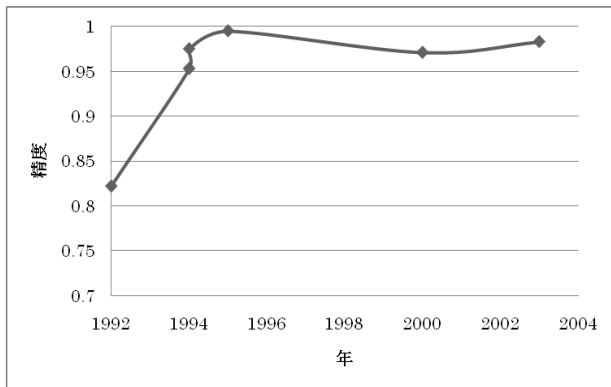


図6 技術動向マップ作成サブタスクの出力例 (解析精度の推移)

### 4.3 システムの動作例

4.2 節で述べたテストコレクションを用いた評価結果の詳細については、文献[11, 12]を参照されたい。本節では、このテストコレクションを用いて著者らが作成したシステムの動作例について説明する。

図7は、「音声認識」という用語をシステムに入力した時の解析結果を示している。図7において、左側に「音声認識」の要素技術が列挙され、各技術の右側にその用語が使われている年が示される。例えば図7中にある要素技術「HMM」の場合、この用語を要素技術に用いた文献が1999年に発表されていることを示している。これらは図中で「●」として表示されており、ユーザが●上にカーソルを重ねることで、その文献の書誌情報がポップアップウィンドウ内に表示される。

また、図7において要素技術として提示されている用語をユーザがクリックすることで、その要素技術が他にどのような分野で利用されているのかを、年代順に一覧表示することができる。図8は、図7中の「HMM」をクリックした結果を示している。学术界では1990年代前半に画像認識(地名認識)の分野で使われていた技術が1990年代後半に入ると動画認識(ジェスチャ認識)の分野でも利用されていることが一覧表示される。

さらに、各要素技術の効果に関する情報が、各図の右端に表示される。図7では、「音声認識」の分野で「モーラ情報」の技術から「精度が向上」という効果があることが分かる。また、図8では、様々な分野においてある要素技術にどのような効果があるのか一覧表示される。

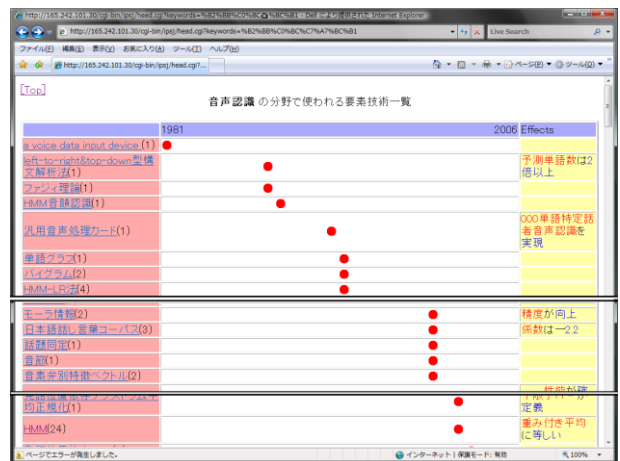


図7 「音声認識」で使われる要素技術と効果の一覧表示

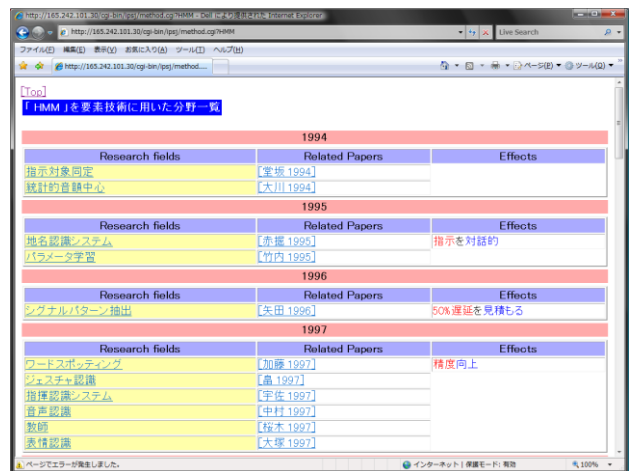


図8 「HMM」を要素技術として用いている分野と各分野における効果の一覧表示

なお、ここで紹介したシステムは、要素技術とその効果を出力しているが、図2とは少し異なっている。図2のような出力を行うためには、複数の技術文書から抽出された要素技術や効果の表記の揺れを同定する必要がある。例えば、自然言語処理分野で頻りに要素技術として利用される「サポートベクトルマシン」は“SVM”や“Support Vector Machine”と表記されることがある。また、効果の表現に関して、「精度が向上」と「解析精度が改善」はほぼ同一内容であると思われるが、現状ではその同定処理を自動的に行うまでには至っていない。なお、このような処理を行う研究もすでに一部が始まっている[13]。

## 5. おわりに

本稿では、新聞記事、Blog、技術文書など、様々なジャンルの文書を対象とした動向情報の抽出と要約に関する研究やサービスについて紹介した。1節でも述べたとおり、動向情報の抽出と要約に関するシステムは、システムを利用するユーザが、そこで得られた結果をどのように分析するか、という目的をもって利用

されているケースが多い。このため、システムがどの程度正しい結果を出力できているのか、という従来の評価に加え、今後は、システムの出力結果を使ったユーザの分析作業をどの程度効率化できたのか、という観点からの評価も、重要になってくると思われる。

## 参考文献

- [1] Fujiki, T., Nanno, T., Suzuki, Y., and Okumura, M., Identification of Bursts in a Document Stream, First International Workshop on Knowledge Discovery in Data Streams, pp.55-64, 2004.
- [2] Gruhl, D., Kumar, R., Novak, J., and Tomkins, A., Predictive Power of Online Chatter, In Proceedings of ACM Special Interest Group on Knowledge Discovery in Data (SIGKDD), pp.78-87, 2005.
- [3] Kato, T., Matsushita, M., and Kando, N., Expansion of Multimodal Summarization for Trend Information - Report on the First and Second Cycles of the MuST Workshop -, In Proceedings of the 6<sup>th</sup> NTCIR Workshop Meeting, 2007.
- [4] Kato, T. and Matsushita, M., Overview of MuST at the NTCIR-7 Workshop - Challenges to Multi-modal Summarization for Trend Information, In Proceedings of the 7<sup>th</sup> NTCIR Workshop Meeting, 2008.
- [5] 河合英紀, 齋藤悠, 土田正明, 水口弘紀, 國枝和雄, 山口敬嗣. 新聞記事における統計量表現の共起ネットワーク, 第 22 回人工知能学会全国大会, 2008.
- [6] 村田真樹, 一井康二, 白土保, 金丸敏幸, 塚脇幸代, 井佐原均. 大規模記事群からの数値情報に関わるテキストマイニング・可視化, 第 21 回人工知能学会全国大会, 2007.
- [7] 村田真樹, Saeger, S.D., 橋本力, 風間淳一, 山田一郎, 黒田航, 馬青, 相澤彰子, 鳥澤健太郎, 論文データからの重要情報の抽出と可視化, 第 23 回人工知能学会全国大会, 2009.
- [8] 森辰則, 藤岡篤史, 村田一郎. 動向情報編纂のためのテキストからの統計量表現の自動抽出, 人工知能学会論文誌, Vol. 23, No. 5, pp.310-318, 2008.
- [9] 森辰則, 上野史紀. 動向情報編纂のためのテキストからの統計量表現の自動抽出, 第 22 回人工知能学会全国大会, 2008.
- [10] Nanba, H., Okuda, N., and Okumura, M., Extraction and Visualization of Trend Information from Newspaper Articles and Blogs, In Proceedings of the 6<sup>th</sup> NTCIR Workshop Meeting, pp.243-248, 2007.
- [11] Nanba, H., Fujii, A., Iwayama, M., and Hashimoto, T., Overview of the Patent Mining Task at the NTCIR-7 Workshop, In Proceedings of

the 7<sup>th</sup> NTCIR Workshop Meeting, pp.325-332, 2008.

- [12] Nanba, H., Fujii, A., Iwayama, M., and Hashimoto, T., Overview of the Patent Mining Task at the NTCIR-8 Workshop, In Proceedings of the 8<sup>th</sup> NTCIR Workshop Meeting, pp.293-302, 2010.
- [13] 西山莉紗, 竹内広宜. 同じ効果を持つ複数技術を同定するための知識抽出, 第 24 回人工知能学会全国大会, 2010.
- [14] 大塚裕子, 乾孝司, 奥村学, 意見分析エンジン - 計算言語学と社会学の接点-, コロナ社, 2007.
- [15] 奥村学, 難波英嗣, テキスト自動要約, オーム社, 2005.
- [16] 山本健一, 殿井加代子, 谷岡広樹, タグ付きコーパスを用いた動向情報とその要因の可視化, 言語処理学会第 12 回年次大会ワークショップ「言語処理と情報可視化の接点」, 2006.
- [17] 山本健一, 谷岡広樹, 藤井加代子, 動向情報の検索による情報編纂, 第 21 回人工知能学会全国大会, 2007.

著者問い合わせ先

住所: 〒731-3194 広島市安佐南区大塚東 3-4-1

電話&FAX: 082-830-1584

E-mail: nanba@hiroshima-cu.ac.jp

著者略歴

1996 年, 東京理科大学理工学部電気工学科卒業. 1998 年, 北陸先端科学技術大学院大学情報科学研究科博士前期課程修了. 2001 年, 北陸先端科学技術大学院大学情報科学研究科博士後期課程修了. 同年, 日本学術振興会特別研究員. 2002 年, 東京工業大学精密工学研究所助手. 同年, 広島市立大学情報科学部講師. 2010 年, 広島市立大学大学院情報科学研究科准教授. 現在に至る. 博士(情報科学). テキストマイニング, 情報検索, 自動要約, 特許情報処理に関する研究に従事. 言語処理学会, 情報処理学会, 人工知能学会, ACL, ACM 各会員.