

人物に着目した動向情報の抽出と提示

A Trend Information Extraction Focused on Person

鈴木 宏哉*1

Hiroya Susuki

平松 慶彦*1

Yoshihiko Hiramatsu

斎藤 博昭*2

Hiroaki Saito

*1 慶應義塾大学大学院理工学研究科

Graduate School of Science and Technology, Keio University

*2 慶應義塾大学理工学部

Faculty of Science and Technology, Keio University

The goal of our research is to construct the trend information extraction system that it is focused on person and it can visualize the relationship between person and topic. In this paper, in order to construct the system, we propose the trend information extraction system focused on person, for our first step.

1. はじめに

近年、電子化された情報の蓄積が進み、その種類や量は既に膨大なものとなっている。しかも、計算機の処理能力の向上やネットワークの高速化により、その増加は今後も継続し、加速していくものと考えられる。

電子化され、蓄積された情報の利用に関して、過去に様々な検索技術が提案されているが、これまでの検索はユーザが求める情報の場所を見つける事を主眼としたものであり、今後も増加を続けるであろう膨大な情報に対して十分とは言い難い。このような環境において、ユーザの関心や興味に応じて情報そのものをまとめ上げ、また、それに対する適切な情報提示の方法を提供可能な技術が必要である。加藤らはこのような技術を情報編纂 (Information Complilation) 技術と呼び [1], 可視化手法 [2] を提案するなどその基盤技術に関する研究を行っている。

このような中で、我々は時間軸を持つような情報、例えば内閣支持率やガソリン価格、地震・台風の発生状況といった動向情報の抽出と提示を行うシステムの構築を行っている。

この動向情報を扱った研究に関しては、「動向情報の要約と可視化に関するワークショップ」[3] において多くの成果が発表されており、我々のグループも参加している。このワークショップではこれまで「携帯電話の加入台数」のような統計量とその統計量の対応付けを行い、グラフ化して提示するような研究が主に行われて来た。すなわち、これまでの研究が対象としてきた動向情報の抽出は、あるトピックを中心としたもので、そのトピックに関連する事象はあくまで付加的な情報でしかなかった。しかし、例えば株価をトピックとして捉えた時、総理大臣の言動は付加的な情報ではあるが、相補的な関係にあると言える。なぜなら、総理大臣の発言により株価が上昇する事もあれば、株価が下落した事により、総理大臣が何らかの行動を行うという事もあり得るからである。

このような人物の動向に注目し、関連する株価動向と総理大臣の動向、更には関連する人物相関のような情報の提示ができれば、統計量だけを提示するよりも、情報編纂研究が目指す目標により近い、多面的な情報の提示が可能になると考えられる。我々は、そのような情報提示が可能なシステムの実現を目指している。今回は、その前段階として人物の動向を抽出するシステムの構築を行った。

2. 人物に着目した動向情報提示システム

2.1 システムの目的

ユーザがある事柄または人物に関心を持っている場合に、システムはその一方の情報だけを抽出するのではなく、双方の情報を抽出し、提示を行う事を目的としている。

2.2 ユーザクエリ

システムはユーザ意図を表すクエリとして、キーワードの入力を受け付ける。ユーザが人物に対する検索要求を持つ場合は直接人物名を検索クエリとして用いる。一方、ユーザが何らかの事柄に対して関心がある場合、ユーザが入力したトピックのキーワードに対して、そのキーワードから検索された記事中の人物を抽出し、抽出された各人物の動向を提示する。今回は事前に作成した人名辞書を元にクエリの処理を行っているため、人物同定の誤りは以後で考慮していない。

2.3 実験データ

今回は、オンラインの新聞記事三紙 (毎日 jp[6], YOMIURI ONLINE[7], asahi.com[8]) を対象に一ヶ月分のデータ抽出を行った。また、関連する付加的な情報として複数のブログ検索サイトでの検索結果から関連度の高いブログ記事を抽出し、提示を行った。

2.4 人物動向の抽出と提示

我々のシステムは時系列の基準となる情報として、新聞記事を利用し、その記事中の人物の動向に関する一連の情報を抽出するという処理を行っている。ユーザがクエリとして入力したトピックに関連する人物の情報は頻度による提示と、一定期間連続して出現している場合はグラフの形式でも提示を行う。

今回、対象としたデータは三紙の新聞記事であるので、図1のような積み上げ棒グラフの形式での提示を行っている。また、同一トピックに対して複数の人物が記事に登場するという場合も多く、この場合は三紙の頻度を足し合わせたものを折れ線グラフで提示するという提示方法も可能である。

3. 結果と考察

3.1 人物間の出現頻度による相違

得られた人物の動向情報は大きく分けて、次の2種類に分類可能である。

- 定期的に一定頻度で登場する人物
- 短期間のみ登場する人物

連絡先: 鈴木 宏哉, 慶應義塾大学大学院, 〒 223 - 8522 神奈川県横浜市港北区日吉 3-14-1, Tel: (045)-563-1141, e-mail: susuki@nak.ics.keio.ac.jp

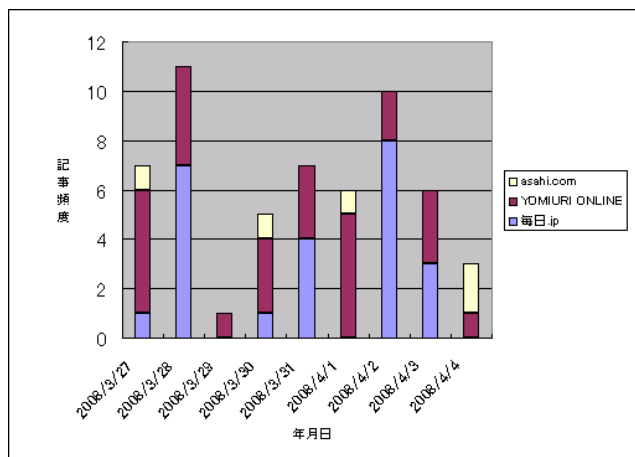


図 1: 福田総理を含む記事数

すなわち、総理大臣のように頻繁に新聞記事に名前が掲載され、複数トピックの記事に登場する人物と、ある特定のトピックの記事にのみ登場する人物の 2 種類に分ける事が可能である。

前者の特徴としては、その人物に関連するトピックそのものに長期的な性質がある場合や、その人物が複数のトピックに登場しているというような傾向が見られた。

一方、後者はある事件の犯人のような人物にあてはまり、その捜査に関する記事にだけ出現するため、その事件が解決するなどして、新聞記事として扱われなくなると出現しなくなるという特徴がある。

ただし、今回の実験では一ヶ月分しかデータを用いていないため、短期間しか登場しなかった人物が再び登場する可能性もある。例えば、前述の犯人の例を考えると、重大事件の場合、数ヶ月後にその裁判の結果が出される前後に再び、その事件に関する記事が書かれるという事があり、そのような場合には再度短期的にその人物が登場するという場合がある。

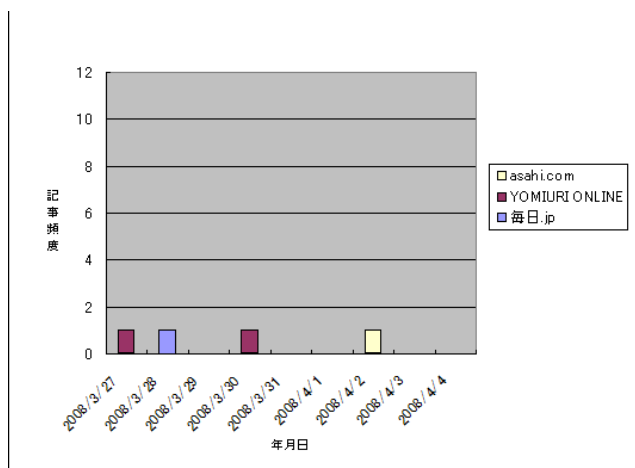


図 2: 東国原知事を含む記事数

例えば、図 1 と図 2 のように、複数のトピックに対して出現していた福田総理と東国原知事とは図からも、その出現頻度の差は明確である。この期間に福田総理は、表 3.2 に示した

表 1: 福田総理に関する記事のクラスタリング結果 (頻度)

	日銀	道路	年金	その他
毎日.jp	12	27	2	21
YOMIURI ONLINE	7	28	1	26
asahi.com	1	4	0	3

ような頻度で「日銀総裁」「道路特定財源」「年金記録」などのトピックに関する記事に登場していた。対して、東国原知事がこの期間に出現した記事は「道路特定財源」に関してのみであった。

3.2 トピックによるクラスタリング

福田総理が出現する記事に対して、クラスタリングを行った結果、表 3.2 のように「日銀総裁」「道路特定財源」「年金記録」などのトピックに分ける事ができた。

なお、ここでは重複を許すクラスタリングを行っており、表 3.2 ラベルはクラスタリング結果から人手で付けたものである。これらのトピックを時系列に並べる事でそれぞれの事象に関する動向情報を知る事ができる。

3.3 リンク情報によるブログ記事との関連度

ブログに新聞記事の感想を書く場合に、その参照情報として情報源となる記事にリンクを張っている場合がある。このリンク情報は、一般に記事のサイズが小さく新聞記事との適切な関連度を求める事が難しいブログ記事と新聞記事の対応付けを行う上で非常に有用な情報となる。

4. おわりに

本論文では、人物の動向に着目した動向情報の抽出と提示を行うシステムを構築について述べた。今後は、人物の共起情報を元に人物相関を求める機能の追加を行い、人物相関とそれぞれのトピックとの関係を提示できるようなシステムの構築に向けて研究を進めていく予定である。

また、今回はブログ記事の情報を各人物動向との関連度のみによって提示する手法を採っていたが、そのブログに書かれた評判情報をシステムの結果に含むような改良も行いたい。

参考文献

- [1] 加藤恒昭, 松下光範, 神門典子: 動向情報の要約・可視化から情報編纂へ, 2007 年度人工知能学会全国大会 (第 21 回) 論文集, 2H5-11, 2007.
- [2] 松下光範, 加藤恒昭: 言語情報と数値情報の相補的利用を目指した可視化手法, 2007 年度人工知能学会全国大会 (第 21 回) 論文集, 3H8-3, 2006.
- [3] 動向情報の要約と可視化に関するワークショップ, <http://must.c.u-tokyo.ac.jp/>
- [4] kizashi.jp, <http://kizashi.jp/>
- [5] blogWatcher, <http://blogwatcher.pi.titech.ac.jp/>
- [6] 毎日.jp, <http://mainichi.jp>
- [7] YOMIURI ONLINE, <http://www.yomiuri.co.jp>
- [8] asahi.com, <http://www.asahi.com>