

Elucignage: 効率的な情報アクセスのための動向情報可視化インタフェース

内藤 峻

関西大学 総合情報学部

1. はじめに

近年、コンピュータの処理能力の向上やネットワークの発達に伴い、電子化された情報は増加の一途を辿っている。それに伴い、これらの情報を利用して意思決定や問題解決に役立てる試みがなされている。しかし、蓄積された情報は膨大であるうえ、時間の経過に伴って更に増加を続けている。そのため、“情報の在処を見つける”ことを主眼とした検索技術ではユーザの要求に十分に対応することができず、ユーザの関心や興味に合致する情報に直感的かつ簡便にアクセスするための技術、言うなれば“情報の理解”を助ける技術が渴望されている。このような背景の下、新聞記事テキストや統計データといった異なる形式の情報を相補的に用いて編纂し、ユーザの情報アクセス行為を容易にする技術（情報編纂技術）の実現を目指している [7][6]。ネットワーク上にはテキストだけでなく、音声や画像、動画など様々な形式の情報が存在している。情報編纂技術が目指すのは、これらの情報をユーザの興味や関心に基づいて適応的に再構成し一纏まりの情報として提示すると共に、それをトリガとしたインタラクティブな情報アクセスを可能にすることである。このような要求に応える技術の一つとして、松下らは動向情報を対象とし、それらを要約・可視化する技術の研究を行っている [2]。

動向情報とは、ある商品の価格や売上高、台風の進行状況や被害の経過、内閣や政党の支持率の推移など幾つかの統計量や出来事に関する時系列データを基にして、ある観点の下でその変化を通時的に捉えて纏めあげたものである。このような動向情報は単なる一次元の時系列情報ではなく、製品のシェアのように複数の企業が関係したり、地域ごとの土地価格の変動のように空間的な広がりを持ったりするなど、複数主体や空間軸を含んだ多次元情報である。

松下らが目指しているのは、動向情報に対するユーザの関心・質問に、簡潔で平易な文章や可視化表現（グラフなど）を組み合わせで応答するマルチモーダル質問応答システムの実現であり、そのための要約と可視化である。本稿ではその手始めとして統計データなどの数値情報と新聞記事などの言語情報を相補的に用いて編纂し、ユーザの情報アクセス行為を容易にする技術の実現をめざす。

本稿では、まず、動向情報の可視化に関する先行研究を挙げ、本研究に有用な手法について取り上げる。次に、本研究の提案システムのデザイン指針を示し、その後、卒論に向けた進捗状況と今後の展望を報告する。

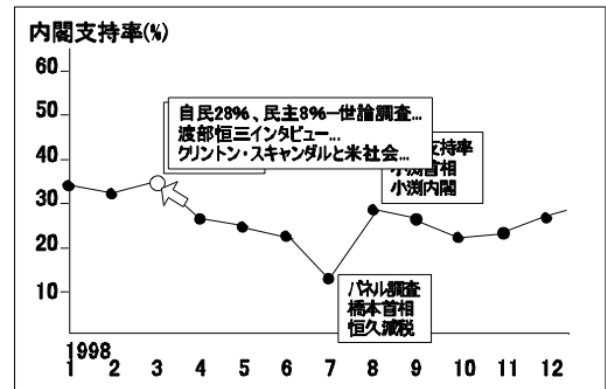


図 1: 山本らのシステムの出力例

2. 先行研究

本章では動向情報を可視化する研究について報告する。

山本らは、ユーザの関心や興味に合致する情報に直感的かつ簡便にアクセスするための技術の一つとして、動向情報の変化とその変化要因とを視覚的に表示するシステムを開発している [1]。システムは、内閣支持率に関連する新聞記事を入力することにより、内閣支持率の推移グラフを出力し、ユーザの興味と見やすさを考慮し、内閣支持率の変化の大きい部分などにその変化の根拠となる要因をグラフ上に配置している。

小泉らはグラフを説明した文章の収集とその文章の適切さを目的として、人がグラフを文章に符号化する際に使用する語彙や文章構造について調べている [8]。その結果、文章の構造を語彙の出現頻度によっておおよそ分類可能であることを示している。分類は5パターンある。

1. グラフを左から右に向かって説明する
2. 全体的な形で説明する
3. 左から右に向かって説明した後に全体的な形を説明する
4. 全体的な形で説明した後に左から右に向かって説明する
5. その他

松下らは、グラフ概形を示唆するシステム STEND を提案している [2]。STEND は時系列数値情報を扱わず、新聞記事のテキストデータのみから取得可能な数値情報と定性情報に着目してグラフ描画を試みたものである (図 2)。テキスト中の「昨年 10 月より約 40 % の下落になっている」「前年同月に比べて 5 ドル上昇した」等の比較表現や「安定傾向にあった」「10 月をピークに下落している」等の定性表現から情報を抽出している。情報提示では、統計グラフを用いず数種類の点や短形、形状

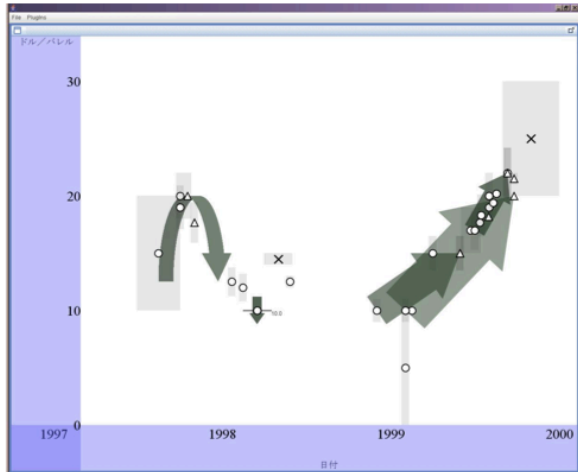


図 2: STEND の描画例

の異なる幾つかの矢印記号を組み合わせることでその代替を試みている。

高間らは空間的な動向情報を含んだ可視化システムについて提案している [5]。動向情報の可視化の対象として選択されているのは、時間情報と数値情報がほとんどである。この研究ではこれらの時系列情報に空間情報を加えた地震情報を対象としている。これは本研究のねらいである多様な種類の情報を編纂するという点で有望であると思われる。

太田らは文書中の数値的特徴を用いた情報可視化について述べている [9]。この論文では、記事から抽出されたデータの特徴を基に適切なグラフの種類を決定し、グラフの作成を行う手法について述べている。折れ線グラフは時間経過による数量の変化を視覚的に表示するときに適していることや、棒グラフは数量の比較を視覚的に表示するときに適しているなど、各グラフの特徴について詳しく述べられている。これは本研究のねらいである異なる形式の情報を相補的に用いて編纂し、ユーザの情報アクセス行為を容易にする技術（情報編纂技術）の実現において有用であると考えられる。

3. デザイン指針

本研究では、既存の検索技術を用いてユーザの意思決定や問題解決に役立つシナリオを想定し、システムに必要な機能を整理する。以下に、複数あるシナリオの中からガソリンに関係するシナリオを示す。

シナリオ 1

工場を営んでいる A さんは最近、自社で扱っているガソリンの価格が上昇していることを知り、不安に思い始めた。そこで、最近のガソリンの価格を調べることにした。すると、ガソリンの価格はある程度上下するものだとわかった。しかし、ある程度上下するものの価格は上昇傾向にあった。何故、上昇傾向にあるのか調べてみると、背景に外国情勢が絡んでいた。一つはウクライナ情勢である。産油国のロシアから欧州への原油供給が止まる可能性が取りざたされていた。さらに、同じく産油国のイラクでも 6 月から武装組織が活動を活性化させ、原油の供給が脅かされていることや円ドル相場が 2013 年初頭と比べ 10% ほど円安になり、輸入コストが増大していること、さらには、経営の厳しさが増す国内の石油元売各社やガソリンスタンドが、収支改善のために原料の上昇分を店頭価格に反映させ始めたとの指摘もあることがわかった。念のため、円ドル相場を調べてみると、2013 年初頭と比べて円安になっていた。調べていくうちに、もしかすると「夏にガソリンの価格が上がるのではないかと」考え、過去の 3 年分のデータを調べることにした。すると、夏と冬をピークに価格が上下することがわかった。また、2008 年に価格が急上昇しているのをみて、何故こうなったのか原因を調べることにした。調べてみると、アメリカの金融危機が原因だとわかった。A さんは不安を解消することができ、安心してガソリンを購入しようと決意した。

シナリオ 2

運送業者の営業をしている C さんは 2013 年 3 月、最近の自社の売上が低下していることに頭を悩ましていた。そこで、何故売上が低下しているのか調べてみることにした。すると、ガソリンの購入額が上がっている事に気づいた。C さんは何故こんなにもガソリンにお金がかかっているのか調べることにした。最近一ヶ月のガソリンの価格を調べてみると、確かに上昇傾向にあった。いつから上昇傾向にあるのか知りたいと思い、もう少しさかのぼって調べてみた。すると、3ヶ月前から上がっていることがわかった。そこで、何故こんなにも上がっているのか疑問に思い、理由を調べることにした。すると、薄水色線のドバイ FOB 価格 (右軸\$/B) が、105 ドルから 113 まで値上がりしたのと円安で、TOCOM 東京ドバイ価格 (橙色) も高騰したためだとわかった。そして、何故ドバイ原油価格が上昇しているのか調べてみると、近年、中国、インド、インドネシアなどアジア諸国での需要の拡大が関係していることがわかった。C さんは自社の売上が低下した原因を知ることができ、しばらくこの調子で経営することに決めた。

シナリオ 3

貨物輸送の経営者の B さんは最近 (2014 年 8 月 4 日)、軽油の価格が下がってきていることを知り、この減少がいつまで続くのか、そして、買い時がいつ頃なのか知りたいと思い始めた。そこで、最近 1 ヶ月の軽油の価格を調べることにした。すると、徐々にではあるが減少していることがわかった。しかし、この情報だけではこれからの価格を推測することができなかった。最近一年間の価格を調べることにした。すると、先月をピークに下がってきていることがわかった。次に、毎年このようなピークがあるのか調べるために、2 年前～1 年前の価格を調べることにした。すると、去年の 3 月にピークがあることがわかった。また、去年の今頃から現在に至るまで上昇傾向にあることがわかった。そこで、何故軽油が去年の 7 月から上昇しているのか気になり、調べてみることにした。すると、軽油は原油価格によって変動するものであり、アメリカの株価の動向、中東情勢の政治的な動き、ヨーロッパの景気動向、特に金融不安の解消、さらには発展途上国の状況によって影響されるものだとなった。また、ドルベースの原油価格の上昇と円安が複合的に作用したからだということもわかった。読み進めると、米国のエネルギー輸出は徐々に許可される方向にあり、それにとまって原油価格も安定してくると思われることがわかった。そこで、B さんは現在の軽油の価格がまだ下がりそうだと考え、もう少し待つことにした。

以上のシナリオから必要な機能のリストを以下に示す。

検索を可能にする

シナリオ 1、2、3 に共通して見られるのが複数の統計情報をまたいで分析していることである。シナリオ 1 ではガソリンの価格から円ドル相場へ、シナリオ 2 ではガソリンの価格からドバイ原油価格へ、シナリオ 3 では軽油の価格から原油の価格へと見たい統計情報が変わっている。

グラフの表示領域の変更を可能にする

シナリオ 1 の「過去の 3 年分のデータを調べることにした」やシナリオ 2 の「もう少しさかのぼって調べてみた」、シナリオ 3 の「2 年前～1 年前の価格を調べることにした」という記述からグラフを拡大、縮小、表示領域を変更可能にする機能が必要だと考えられる。

理由や背景を参照できるインタラクションを付与する

シナリオ 1 の「何故、上昇傾向にあるのか調べることにした」やシナリオ 2 の「何故売上が低下しているのか調べてみることにした」や「何故こんなにも上がっているのか疑問に思い、理由を調べることにした」、シナリオ 3 の「何故軽油が去年の 7 月から上昇しているのか気になり、調べてみることにした」という記述から、理由や背景を参照できるインタラクションが必要であることが考えられる。また、イ

ンタラクションを付与するためのトリガが必要であることが考えられる。

動向情報を予測する

シナリオ 1 の「もしかすると夏にガソリンが上がるのではないか」やシナリオ 3 の「買い時がいつ頃なのか知りたい」という要求から動向情報の予測を望んでいることが考えられる。

最低値と最高値、平均を表示する

シナリオ 1 と 3 に「ピーク」という記述がいくつかされていることから、ユーザはその期間の最低値と最高値が知りたいのではないかと考えられる。また、シナリオ 3 のような「買い時がいつ頃なのか知りたい」という予想を立てる際にはその期間の平均値を知ることができればユーザにとって嬉しいと思われる。

4. Elucignage プロトタイプシステム**4.1 概要**

ユーザは統計グラフの外観を理解するだけでなく、興味を持った箇所についてどのようなことが述べられているかをその要因となる記事にアクセスすることで参照できる。そのため、ユーザの関心がある動向情報を時系列数値情報を用いて統計グラフとして描画し、そのグラフの要因となる記事をその内容に適したアイコンの形式で提示する方法を採用する。このアイコンは要因となる記事のアクセスを可能にする。また、記事の一部を画面の一部に表示し、そこからグラフのどの部分に該当しているかを示す機能を備える。これにより、要因となる記事がグラフのどの部分で記述されたものであるかを確認できる [4][3]。

4.2 進捗状況

試作した Elucignage プロトタイプの外観を図 3 に示す。システムは HTML、CSS、javascript を用いて実装した。JavaScript のライブラリは jquery-1.6.2.js と D3.js を用いている。現在は外観とグラフの描画、新聞記事を提示する機能を実装したところである。このプロトタイプは、記事リストパネルに新聞記事を表示するボタン (図 3 左上画面)、記事のスニペットを表示する記事リストパネル (図 3 左下画面)、グラフとそれに関連する記事へのポインタであるアイコンを表示するグラフパネル (図 3 右上画面) から構成される。

ユーザがボタンをクリックすると、記事リストパネルに記事の一覧がスニペットを伴って表示される。スニペットはユーザに元記事を参照する価値があるかどうかの判断材料として提示されるものである。グラフパネルにはシステム起動時にグラフとアイコンが表示される。

現在の実装はボタンをクリックすると新聞記事の一部を提示するようになっている。また、スニペットは以下に示す xml 形式のファイルから TEXT タグに含まれて

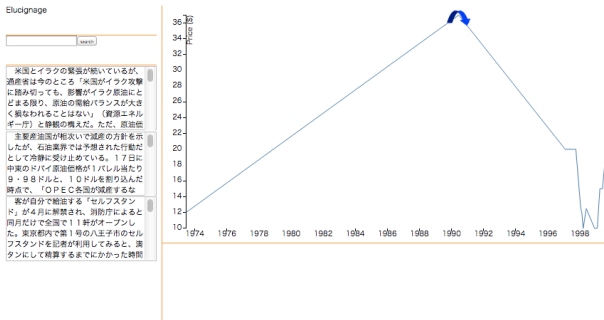


図 3: Elucignage のスナップショット

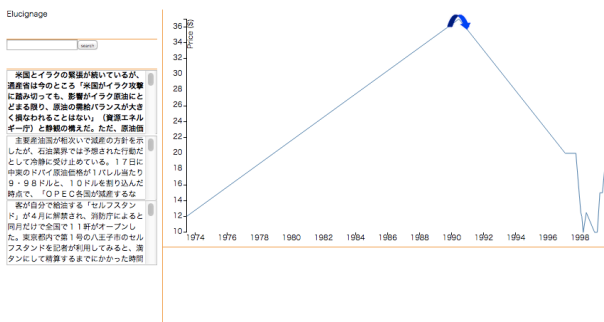


図 4: Elucignage のスナップショット (アイコンをクリックしたとき)

いる本文を表示させている。グラフは D3.js を用いて描画している。

```
<?xml version="1.0" encoding="Shift_JIS"?>
<DOC>
<DOCNO>980214080</DOCNO>
<SECTION>経済</SECTION>
<AE>有</AE>
<WORDS>635</WORDS>
<HEADLINE>「イラク危機、原油の需給バランスに影響
ない」ー通産省は静観の構え</HEADLINE>
<TEXT>
  米国とイラクの緊張が続いているが、通産省は今のと
  ころ「米国がイラク攻撃に踏み切っても、影響がイラク
  原油 …
</TEXT>
</DOC>
```

ユーザがアイコンをクリックすると記事のスニペットの文字が大きくなる (図 4)。これはユーザがグラフ上のある点に興味や関心を持った際にグラフから新聞記事へ視線を誘導するためのものである。

4.3 今後の展望

この節ではこれから実装する予定である機能を示す。まず、グラフの描画範囲を変更することができるコントロールパネルを実装する。次に、検索を可能にする。ボ

タンを検索ボックスにし、統計情報の一部を入力すると、グラフパネルにグラフとアイコンが表示されるようにする。また、同時に記事リストパネルに関連する記事の一覧がスニペットを伴って表示されるようにする。その後は、関連記事の同定や選択、スニペット、アイコンの生成の自動化を考えていく。

5. おわりに

本稿では、ユーザの興味や関心への効率的なアクセスを可能にする可視化インタフェースについて考察した。

参考文献

- [1] 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」, pp. 13-16 (2006).
- [2] 松下光範, 加藤恒昭: 数値情報の補填とグラフ概形の示唆による複数文書からの統計グラフ生成, 知能と情報, Vol. 18, No. 5, pp. 721-734 (2006).
- [3] 松下光範, 加藤恒昭: Elucignage: 探索的データ分析のための動向情報可視化インタフェース, 動向情報の要約と可視化に関するワークショップ第二回成果進捗報告会予稿集, pp. 17-18 (2007).
- [4] 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, 3H8-3 (2007).
- [5] 高間康史, 山田隆志, 中野純: 地震記事からの時空間的動向情報可視化についての取り組み, 第一回 MuST 成果進捗報告会 (2006).
- [6] 加藤恒昭: 情報編纂研究会が目指すもの, 2011 年度人工知能学会全国大会 (2007).
- [7] 加藤恒昭, 松下光範: 情報編纂 (Information Compilation) の基盤技術, 2006 年度人工知能学会全国大会, 1D3-2 (2006).
- [8] 小泉尚之, 松下光範, 松田昌史, 馬野元秀: 言語表現と統計グラフの相互変換に関する基礎検討, 動向情報の要約と可視化に関するワークショップ第 2 回成果進捗報告会予稿集, pp. 57-60 (2007).
- [9] 太田彰, 福本淳一: 文書中の数値的特徴を用いた情報可視化, 動向情報の要約と可視化に関するワークショップ第 2 回成果進捗報告会予稿集, pp. 13-16 (2007).