

Elucignage: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

内藤 峻

関西大学 総合情報学部

概要

ネットワーク上には多様な種類の情報が存在しており、それらをユーザの要求に応じて適応的にまとめ上げる技術が渴望されている。その一つとして、テキストなどの言語情報と統計データなどの数値情報の相補的な利用に関する研究が行われている。その一環として、本研究では言語情報と数値情報が密接な関係にある株価などの動向情報に着目しそれらを統一的な枠組みで可視化する手法を提案する。株価などの統計情報の場合、その正確な値を知るには数値情報が適切であるのに対して、変動の大局的な理解や背景となる事象の把握には言語情報が適している。そこで、これらを一つのグラフ上に提示し、その情報源に対話的にアクセスできるようにする。

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。電子化された情報は場所を取らず、劣化もしない、複製も容易であり、どこからでもアクセスできる、といった利点がある。新聞記事や統計データ、書籍など、様々な情報が電子化されている。これらの膨大な情報は単なるアーカイブとしての役割に留まらず、意思決定や問題解決に役立てるために有益な情報や知識を得るためのリソースとして期待されている。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答えることができず、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。

このような要求に答える技術のひとつとして、本提案では統計データ等の数値情報と新聞記事等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。

ネットワーク上にはテキストや統計データ、音声、画像動画など様々な種類の情報が存在している。将来的にはそれらの情報全てを対象とし、状況や目的に応じて取舍選択やモード変換を行い、適切な形態で組み合わせてユーザに提供することが望まれるが、現在の技術レベルではその実現は容易ではない。そこで本研究では、まず時間的変動を伴う統計データ（時系列数値情報）とそれに関連する記事（言語情報）を対象とし、ユーザがそれらの情報にアクセスしたり、その概要を把握したりする際の支援となる可視化手法について議論する [4]。

本稿では、まず、先行研究について述べる。次に、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。

2. 先行研究

動向情報を可視化する枠組みとして、山本らの可視化システム [1] や小泉らの被験者実験 [8]、松下らのSTEND[2]、高間らの地震情報可視化システム [5]、太田らの可視化手法 [9] などがある。

山本らは、記事中に存在するある時点の変化についての定性的な記述や、原因や影響に関する記述を注釈としてグラフに与える方法を提案している [1]。例えば内閣支持率の場合、過去のある時点の支持率と関係が強い出来事を注釈することで、どのような事件が支持率に影響を与えたかをユーザが把握・判断できるようにしている。値の変化の大きな点、減少から増加に転じる極小点など、利用者が関心を持ちそうな点に選択的に情報を付与することによって視認性の向上を測り、ユーザにとって分かりやすい情報提示を目指している。

この方法は本稿の提案方法と問題意識が近く、特に情報提示に関しては参考になる点も多い。しかし、その情報提示がユーザとのインタラクションにおいてどのように作用し適応していくかについては、現状ではあまり深く検討されていない。本提案はユーザのインタラクションに基づく適応的情報提示に大きな関心があり、この点でこれらの方法と方向性が異なる。

小泉らは言語情報と統計グラフの相互変換技術について研究している [8]。その一つとして、被験者実験を通じて、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。その結果、文章の構造を語彙の出現頻度によっておおそ分類可能であることを示している。分類は5パターンあり、グラフを左から右に向かって説明しているパターンが1つ、全体的な形で説明しているのが1つ、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のが1つ、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のが1つ、それ以外が1つの計5パターンある。この分類は、新聞記事に適応でき、分類語を用いることによって文章からどのような内容が述べられているかを把握することができる。そのため、本研究の Elucignage システムの矢印アイコンの選定に有効であると考える。分類語を用いた自然言語処理を行うことによって、分類語に近い数値情報を取得し、それに応じたアイコンを選定することが可能であると考える。以下にこの研究を通して行われた研究を報

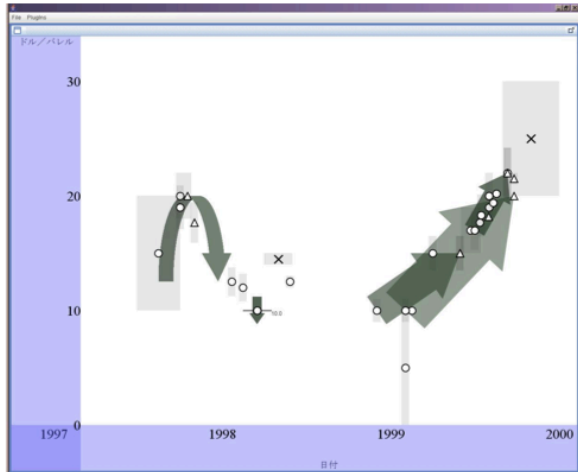


図 1: STEND の描画例

告する。

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り、本提案とは力点が異なる (図 1)。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、形状の異なる幾つかの矢印記号を組み合わせることによってその代替を試みている。

高間らは動向情報だけでなく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意味で有望である。Elucignage システムの展望として、動向情報だけでなく、空間情報もシステムの機能として検討していきたい。

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 Elucignage システムが将来的に状況や目的に応じてやモード変換を行ううえで有用であると考えられる。

3. デザイン指針

本研究では提案システムをデザインするにあたってシナリオ手法を用いる。そして、必要な機能を整理する。以下にシナリオの例を示す。

工場を経営している 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発化させ、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。

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、買い時が何時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。そこで、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かった。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。

運送業者の経営をしている 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。

以上のシナリオから必要な機能を選定する。まず大きな機能として、複数の統計情報をまたいで見る必要があること、グラフの表示範囲を変更可能なことが考えられる。次に、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。最後に、先ほど述べ

たインタラクションを付与するためのトリガーが必要であることが考えられる。

そこで私は、動向情報に対するユーザの関心 (e.g., 最近数ヶ月のガソリン価格の変化、ここ数年の携帯電話加入者数の推移) に応じて数値情報を統計グラフとして描画し、その上に各々の時点の情報をその内容に応じてアイコン形式で重ねて提示する方法を提案する。このアイコンはテキスト情報へのポインタの役割も兼ねている。そのため、ユーザは描画されたグラフ自体を、自分の知りたい情報の外観 (要約) として利用出来るだけでなく、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背景となっている文書群にアクセスすることで確認できる。また、そこからグラフの該当箇所へジャンプする機能を付与する。これにより、注目した記事がどのような状況下で記述されたものであるかをグラフにアクセスすることで確認できる。[3]

4. 卒業研究に向けて

4.1 進捗状況

試作した Elucignage プロトタイプの外観を図 2 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。現在はガワとグラフの描画、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、記事のスニペット (snippet) を表示する記事リストパネル、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されるかどうかを判定し、クリックされ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 xml 形式のファイルから TEXT タグに含まれて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。アイコンはクリックされると記事のスニペットの文字が大きくなる。これはユーザが興味や関心のある新聞記事へ誘導するためにフォーカスするためのものである。

4.2 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。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、

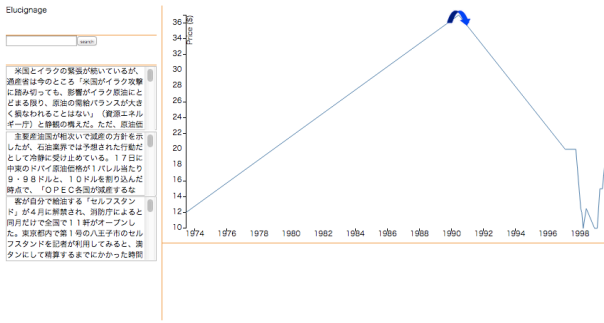


図 2: Elucignage のスナップショット

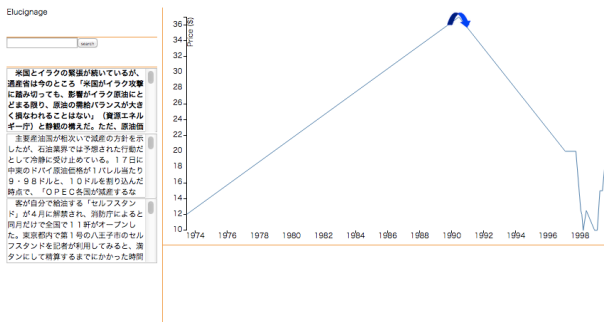


図 3: Elucignage のスナップショット (アイコンをクリックしたとき)

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、動向情報の可視化手法について考察した。先行研究として山本らの可視化システムや小泉らの被験者実験、松下らの STEND、高間らの地震情報可視化システム、太田らの可視化手法を報告した。続いて、本研究のシステムのデザイン指針を説明し、卒業研究への展望を示した。

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).