

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等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、様々な情報が電子化されネットワーク上に蓄積されている。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[3]。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[7]。

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捨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、デザイン指針を説明し、その後、先行研究を示す。

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]や松下らのSTEND[2]、加藤らの

Elucignage[6]、高間らの地震情報可視化システム[5]などがある。

山本らは、動向情報の変化要因に関する重要語を注釈としてグラフに表示する方法を提案している(図??)[?]。要因の抽出では、動向情報が記載された新聞記事と文章の類似度が高い新聞記事を要因としている。また、要因とされた新聞記事からのキーワード抽出は、1月分の新聞記事を1ドキュメントとみなしたTF・IDF値に基づき行っている。動向情報とその要因の表示方法に関するアンケート調査からは、ユーザが動向情報に関する要因を知りたいことは、動向情報を示したグラフ中の変化が大きい部分とその前後、最大位置と最小位置、及び最初と最後の3つに分類できることが分かったと述べられている。また、要因の表示方法としては、1ウィンドウに3つ程度のトピックを簡潔に示したものや、要因をタイトルやキーワードとともに表示すれば良いことが分かったと述べられている。

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??)。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でその代替を試みている。加藤らは、グラフに文章を関連付けた視覚オブジェクトを提案している[6]。2つの画面を上下に配置しており、上画面にグラフ概形(図??上画面)、下画面に状況記述を表現している(同下画面)。また、下画面ではグラフ上のアイコンに対応する記事の一部が表示されている。この記事とグラフは一方を選択すると他方の対応する部分がハイライトする等の対応付けがなされている。さらに、視覚オブジェクトの操作として、グラフ表示を操作するためのパネルがグラフの右下に配置されている。これを使って、表示する期間の変更や、拡大縮小が行える。そのため、これらの視覚オブジェクトを介して特徴点や文書の閲覧が行われる必要があるとされている。これら全般を扱うインタラクションの設計が必要となると述べられている。

小泉らは言語情報と統計グラフとを意味のレベルで処理するために、被験者実験を通じて実際に使われる語彙

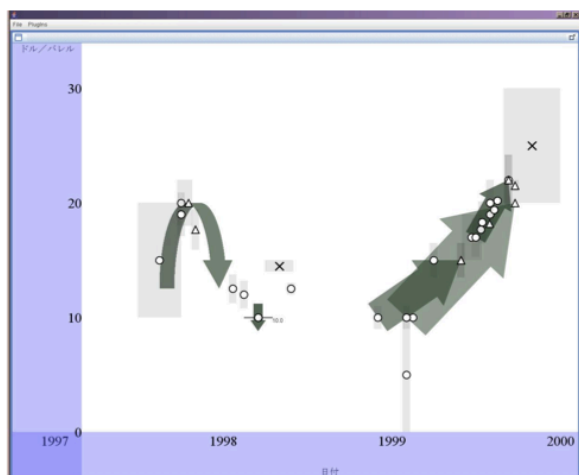


図 1: STEND の描画例

の分析と文書構造の解析を行っている [8]。グラフをどのように言語として表現しているのか分類している。グラフを左から右だと「左右」、全体だと「全体」、左右の後に全体を示している場合は「左右+全体」など計5つの分類を行っている。その結果、文章の構造を語彙の出現頻度によっておおよそ分類可能であることを示している。これは本研究の Elucignage の矢印アイコンの選定に有効であると考えられる。

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]。地震や台風などは時間的な動向情報だけでなく空間的な動向情報を含んでおり、両者を考慮した可視化が必要となる。地図とグラフ、それに対応する新聞記事を一面に提示している。複数の地震について同じ空間へ描画することを可能にしている。本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付与したい。

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てきしていることや、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おり、参考になった。

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。

4. 今までやったこと

試作した Elucignage プロトタイプの外観を図1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ト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²。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。

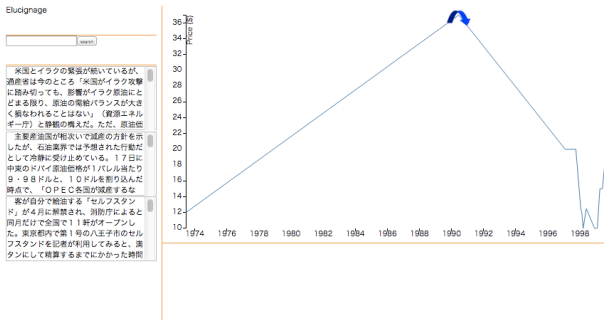


図 2: Elucignage のスナップショット

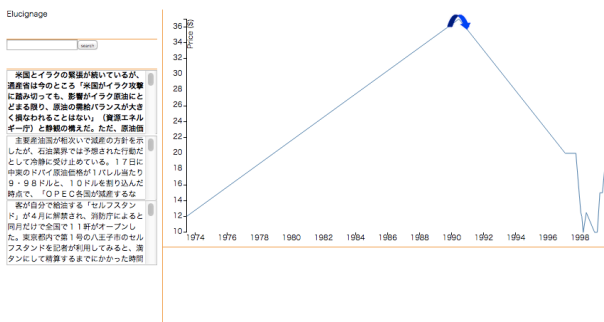


図 3: Elucignage のスナップショット (アイコンをクリックしたとき)

5. これからやること

この章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、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。

6. 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松下らの STEND、蓮井らのグラフ型インタフェース、加藤らの視覚オブジェクトを報告した。続いて、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).