

動向情報の可視化手法

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。[5]

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。[1][6]。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。

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 [3] や松下らの STEND [4]、高間らの地震情報可視化システム [2] などがある。

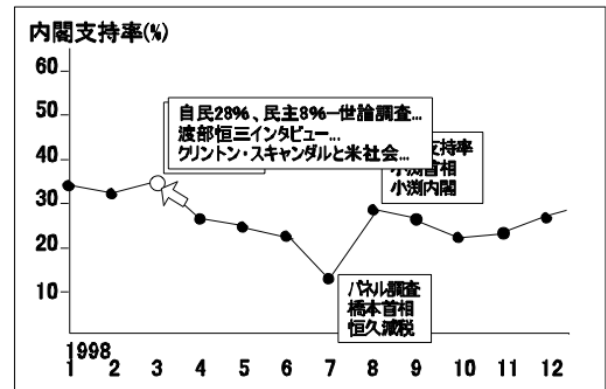


図 1: 山本らのシステムの出力例

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 1) [?]. 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 [4]。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。

3. デザイン指針

まず、本研究で扱う数値情報と言語情報の特徴について述べる。統計情報等の数値情報は正確である。変わって、言語情報はあいまいであるものの、その時点の背景等を理解することができる。本研究では、数値情報には

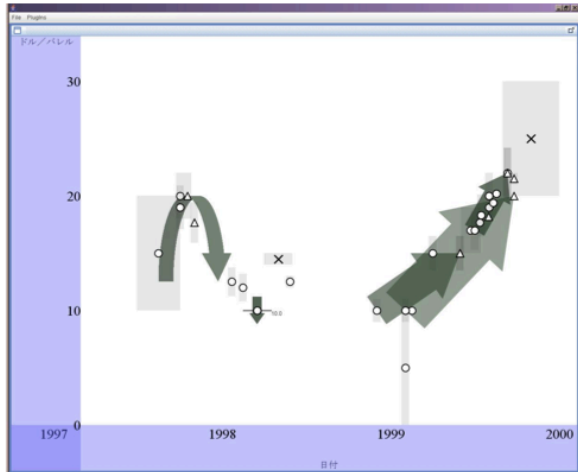


図 2: STEND の描画例

統計 DB を、言語情報には新聞記事を用いて、ユーザの効率的なアクセスを促すシステムを提案する。ユーザの効率的なアクセスを支援するためにはユーザのふるまいを想定をする必要がある。ユーザはグラフを見て、そのときの変化における要因や背景を知りたいと思う。そこで、本提案ではユーザの興味や関心を想定したインタクションを考える。ユーザはグラフの変曲点になる部分に興味を持つ。また、そこで、要因と背景を知りたいと思う。以下の様なシナリオを想定している。

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。

4.2 実装

現段階の実装に関して述べる。検索ボックス記事リストパネルグラフパネルコントロールパネル

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] 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 (第 20 回) 論文集, 1D3-2 (2006).
- [2] 高間, 山田, 中野: 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).
- [3] 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).
- [4] 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).
- [5] 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).
- [6] 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).