

HogeHoge における Hoge に関する研究

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を行っている。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した。[3][4][1][2]

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になっていうえ、時間の経過に伴って更に増加を続けている。そのため、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が求められている [?][?]。情報にはテキスト、画像、音声、動画と様々な種類がある。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に着目した。

本稿の構成は以下の通り。まず、デザイン指針について述べる。その後、プロトタイプを説明し、続いて、先行研究を示す。

2. デザイン指針

3. Elucignage プロトタイプシステム

3.1 概要

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。これにより、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。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。

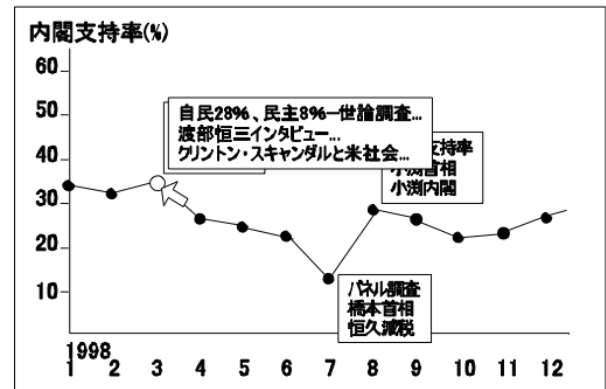


図 1: システムの出力例

3.2 実装

4. 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 [?] 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 [?] 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 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 [?]. 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の代

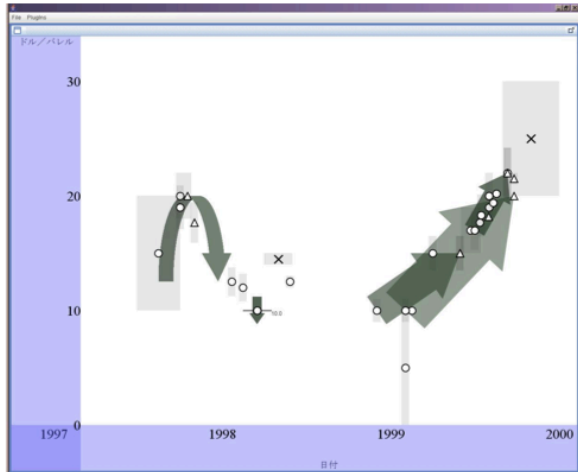


図 2: STEND の描画例

替を試みている。

蓮井らは、動向情報データを対象に、異なるデータ同士の比較とそれに関する詳細な情報へのアクセスの支援を目指したグラフ型インタフェースを提案している[?]。2つのグラフがあり、片方のグラフをドラッグ&ドロップすることにより、比較を実現している。また、このような比較行為によって生じるユーザの興味のきっかけによる取り組みを円滑に行うため、詳細な情報へのアクセスを可能にしている。実験結果から、詳細なアクセス手段は有望であることと重ね合わせによる比較分析手法に有望性があることが述べられている。

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。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3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。

5. おわりに

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] 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



図 3: 文章とグラフを関連付けた視覚オブジェクト

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).

- [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).