

時系列データ分析における試行錯誤を支援する インタフェースの要素検討

Toward an Interface Supporting User's Trial and Error for Time Series Data Analysis

田中 和広*¹
Kazuhiro Tanaka

蓮井 大樹*²
Daiki Hasui

松下 光範*²
Mitsunori Matsushita

関西大学大学院 総合情報学研究科*¹
Graduate School of Informatics, Kansai University

関西大学 総合情報学部*²
Faculty of Informatics, Kansai University

This paper presents a novel interface that intends to support a user's trial and error for time series analysis. In a process of data analysis, people observe data to be analyzed in multiple perspectives and compare it with another data. By a repetition of such processes, they become to understand the characteristics of the data and find a trend and/or novel aspect of the data. If they don't repeat such an exploratory data analysis sufficiently, they may reach an incorrect decision. To avoid the wrong decision, we propose a system to support user's trial and error in data analysis. This paper presents an interface that make easy browsing time series data with multiple views.

1. はじめに

現在、電子化されて蓄積されるデータの量は時間の経過とともに増加している。それに伴い、この膨大なデータを対象として、その中から有益な情報や知識を得るための手法が研究されている [Fayyad 96]。これらのデータを分析する過程においては、ユーザは様々な視点からデータを多角的に眺める、自らの仮説を支持するデータを探す、複数のデータを見比べて相関を見つける、といった試行錯誤を繰り返しながら、それらのデータに関する理解を深めたり、新たな分析の視点に気づいたりする。このような試行錯誤が十分に行えなければ、拙速な決断や誤った推測をしてしまい、その結果として間違った結論に辿り着く可能性がある。本研究では、このような問題を解消するために、ユーザが試行錯誤を円滑に行えるように支援するシステムの実現を目指している。本稿では、時系列データを対象に、複数のビューを用いて様々なデータの閲覧やそれを用いた分析の試行を行うためのインタフェースのプロトタイプを実装し、必要となる機能についての要素を検討する。

2. プロトタイプシステム

2.1 システムデザイン

我々はユーザのデータ分析行為のうち、特にデータ間の関係を見出す過程に着目している。ユーザがデータ間の関係を見出すには、様々なデータを幅広く俯瞰し、関係のありそうなデータの候補を数多く見つけ出すことが必要となる。また、その中のデータと分析対象のデータがどのような関係にあるのかを調べるためには、データ同士を比較して見比べることが必要となる。また、それらのデータが何故関係しているのかの原因を分析するためには、データに関する詳細な情報を取得する必要がある。

本研究では、このようなユーザの分析行為に基づいて、(1) 様々なデータの俯瞰、(2) 複数データの比較、(3) 異なるデータ間の関係の発見・把握、という行為の支援を試みる。それに向けて、我々は様々なデータを概観したり、比較したりするためのインタフェースのプロトタイプを実装する。

2.2 実装

本稿で実装したプロトタイプシステムは、グラフ表示ビュー、一覧ビュー、テキスト表示ビューの3つから構成される。

グラフ表示ビューは、2つの時系列データを統計グラフへと可視化して表示するためのものであり、ユーザがデータ同士の相関をより分析しやすくすることを想定している。そのため、2つのデータを図1のように並べて表示させたり、図2のように重ね合わせて表示させたりする機能を設けている。グラフの重ね合わせは直接操作によって行い、グラフ全体を重ね合わせたり、図2のようにグラフを部分的に重ねて表示させたりして眺めることで2つのグラフの関係を探索的に把握できる。

一覧ビューは、グラフを一覧として表示させるためのものであり、様々なデータの俯瞰を可能にする。これにより、異なるデータ間での相関の発見や意外な関係への気づきを支援する。一覧ビューの表示例を図3に示す。本システムでは、3×3の合計9個のグラフを表示させて、複数のデータを眺めることを可能にしている。

テキスト表示ビューは、データを抽出した元となる新聞記事テキストを表示させるためのものである。これにより、データが変動している原因を探ることを支援する。このビューを利用して、ユーザは可視化されているグラフからより詳細な情報を得ることが可能となる。

なお、提案するプロトタイプシステムでは、時系列データを抽出する情報源として、動向情報の要約と可視化に関するワークショップ [加藤 04] (以下、MuST と記す) で提供されているタグ付けされた新聞記事コーパス (MuST コーパス) を利用した。

3. 議論

本章では、プロトタイプシステムの実装を通じて、時系列データ分析におけるデータのブラウジングを容易に行うためのインタフェースに必要な要件について検討を行う。

今回提案したプロトタイプシステムでは、2つのグラフを横に並べて見比べたり、マウス操作によって重ね合わせて見たりすることができるように実装した。しかし、図1や図2からも分かるように、本システムでは縦軸や横軸の目盛りの間隔が統一されていないため、値を読み取る際に誤った解釈を与えてしまう可能性がある。そこで、縦軸や時間軸の幅をユーザが統

連絡先: 松下光範 関西大学総合情報学部 〒 569-1095 大阪府高槻市霊仙寺町 2-1-1 Tel: (072) 690-2437 Fax: (072) 690-2491 e-mail: mat@res.kut.ac.jp

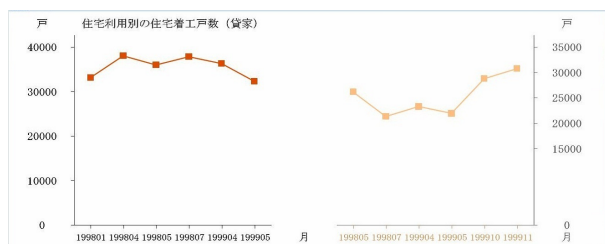


図 1: グラフを並べて表示した例

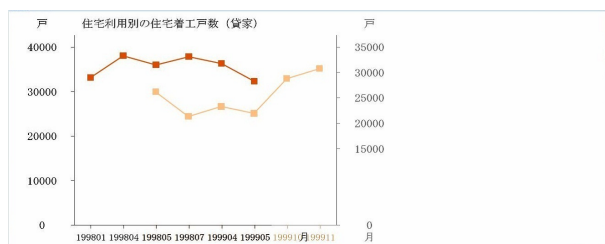


図 2: グラフを重ねて表示した例

一させたり、調節したりすることができるように機能を加えることが必要となる。

また一覧の表示について、本システムは一度に閲覧することができ、かつ出来るだけ多くのデータを見ることができるようにするため、9個のグラフを概観できるように実装した。データの一覧表示は、ユーザの分析の観点を幅広くするために、異なるトピックに関するデータや分析対象とは一見関連のなさそうなデータも含んで表示させることが必要となる。そのため、1つの基準だけではなく、様々な基準に基づいてデータを選び出し、表示させることが求められる。

また、テキスト表示に関しては、分析者がデータに関する詳細な情報を知りたいと思った際に自然な形でそれを得られるようにデザインする必要がある。そのためには、グラフにプロットされている点を選択するとテキストが表示されるなどのような可視化表現と密接に関連したインタラクションが必要であると考えられる。

4. 関連研究

複数のビューを用いてデータ分析を支援するシステムとして、左京と右京 [橘 06]、INFOVISER [磯部 99]、Jigsaw [Stasko 08] が提案されている。

左京と右京は、表形式データに関して、行を構成するデータ要素と列を構成する要素の各々についてクラスタリングを行い、大規模階層型データを可視化する「平安京ビュー [伊藤 06]」を2つ使用して、分析を支援するシステムである。また、INFOVISER は、データを様々な図形表現で一覧表示させることによって、データ分析における試行錯誤の支援を行うシステムである。また、Jigsaw は、文書探索におけるインタラクティブな視覚表現支援の手法として提案されたシステムであり、多数のウィンドウによってデータを多角的に眺めることが可能である。その他、CMV [Roberts 07] でも、Multiple Views を用いたシステムが多数提案されている。我々の提案するシステムは、これらと同じく複数のビューを用いたインターフェースとして位置付けられる。

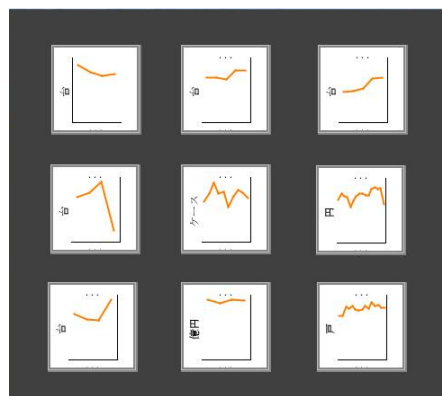


図 3: 一覧表示の例

5. おわりに

本研究では、データ分析における試行錯誤を支援するシステムの実現に向けて、複数のビューを用いて時系列データのブラウジングを容易にするためのインターフェースのプロトタイプを実装し、それを通じて必要な要素について検討を行った。

本稿で実装したインターフェースでは、ユーザがデータ間の関係を発見する過程に着目して、(1) 様々なデータの俯瞰、(2) 複数データの比較、(3) 異なるデータ間の関係の発見・把握、という行為の支援を試みた。今後、プロトタイプの実装を通じて行った要素検討に基づいて、異なるデータ同士の比較や様々なデータの概観を支援する機能について改良を行い、ユーザのデータ分析における試行錯誤をより支援するシステムの開発に取り組んでいく。

謝辞

本研究の遂行にあたり、文部科学省科学研究費 (課題番号: 22300048) の助成を受けた。記して謝意を表す。

参考文献

- [Fayyad 96] Fayyad, U.M., Piatesky-Shapiro, G. and Smyth, P.: "From Data Mining to Knowledge Discovery in Databases", AI magazine, 17(3), pp.37-54 (1996).
- [加藤 04] 加藤恒昭, 松下光範, 平尾努: "動向情報の要約と可視化に関するワークショップの提案", 情報処理学会研究報告, 2004(108), pp.89-94 (2004).
- [橘 06] 橘春帆, 伊藤貴之: "左京と右京:大規模表形式データの可視化の一手法", DBWeb2006, Vol.2006, No.16, pp.127-134 (2006).
- [磯部 99] 磯部成二, 黒川清, 塩原寿子, 飯塚哲也: "視覚化による多次元データ分析システム: INFOVISER", 情報処理学会論文誌, 40(5), pp. 2417-2428 (1999).
- [Stasko 08] J.Stasko, C.Görg, and Z.Liu.: "Jigsaw: supporting investigative analysis through interactive visualization", Information Visualization, 7(2), pp.118-132 (2008).
- [伊藤 06] 伊藤貴之, 山口裕美, 小田山耕二: "長方形の入れ子構造による階層型データ視覚化手法の計算時間および画面占有面積の改善", 可視化情報学会論文集, 26(6), pp.51-61 (2006).
- [Roberts 07] Jonathan C. Roberts: "State of the Art: Coordinated & Multiple Views in Exploratory Visualization", Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization, pp.61-71 (2007).