

表と連結図の特徴を組み合わせた時系列データ分析ツール

結城 崇† 三末 和男† 田中 二郎†

† 筑波大学大学院コンピュータサイエンス専攻

1 はじめに

我々が日常的に生活を送る中で、様々な活動がデータとして蓄積されるようになってきた。例えば、商品を購入する時には、誰が、いつ、何を購入したというデータが蓄積される。このように「いつ」を表す時系列情報が含まれるデータを時系列データと呼ぶ。時系列データを分析することで、自身の生活パターンや消費者の購買傾向などの有益な情報が得られる。

データ分析には Excel のような表を使うことが多い。表を用いてデータ範囲を選択しグラフ化することで、傾向やパターンなどのデータの特徴を発見できる。しかし、グラフから読み取った特徴については、表に戻ってグラフと対応するデータを探さなければならない。また、表だけでは、データの傾向やパターンを発見することは難しい。

この問題を解決するために、本研究では表と連結図の特徴を組み合わせた時系列データ分析ツールを開発した。本ツールによって、時系列データの分布と詳細な情報を同時に把握できる。

2 時系列データの分析

時系列データを分析することで、データの傾向やパターンを発見できる。データの時間軸上の分布とデータの内容を理解することで、傾向やパターンの発見が支援されている。例えば、時刻による売上げの推移を棒グラフや折れ線グラフで見ること、大きな売上げのあった時刻を発見できる。また、発見したパターンや傾向について調べたい場合、さらに詳細な情報が必要である。例えば、売上の大きかった理由を調べたい時、その特徴を表す時刻周辺の客、商品などの詳細なデータが必要になる。

また、時系列データを分析する時、複数の視点からデータを眺めることがある。例えば、商品を購入した時刻などの時系列情報や商品の種類などの意味的分類からデータ範囲を選択し、その部分のデータを閲覧する。これは、視点を変更することで、より多くの傾向やパターンを発見できるためである。

3 時系列データ分析ツール

表と連結図の特徴を組み合わせた表現及び時系列データ分析に必要な操作を備えたツールを開発した(図1)。

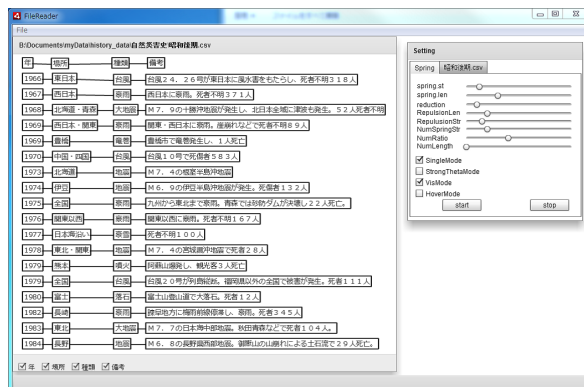


図 1: 開発したツールの外観

3.1 表現方法

表は、縦横の直線とそれによって区切られた矩形領域に文字を表示する表現方法である。例えば、縦方向で属性値の順序を表し、横方向で属性の順序を表す。表はデータを簡潔に表現できることが特徴である。連結図は、頂点(ノード)と頂点の間に結ばれる線(リンク)を用いて関係を表現する方法である。

本手法では、要素と要素のレコードへの所属関係を連結図で表し、属性への所属関係を表の特徴である横方向の位置をなるべく揃えることで表現する(図2)。

本表現の利点として、セルの縦横の位置が厳密に決定される表の配置制約を緩やかにできる。そのため、縦、横の位置をデータの時間軸上の分布の提示や関係する要素の提示に利用できる。

3.2 本表現における操作

本表現手法における時系列データ分析のための操作として、ノードの配置変更及びフィルタリングを開発した。ノードの配置変更では時系列情報の分布を提示し、フィルタリングでは関係するデータの提示を行う。

A tool combining property of table and node-link diagram for analyzing temporal data

†Takashi YUKI †Kazuo MISUE †Jiro TANAKA

†Department of Computer Science, University of Tsukuba

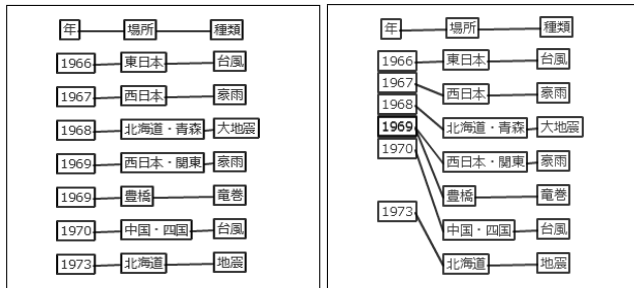


図 2: 表と連結図の特徴を組み合わせた表現手法。日本の自然災害史データを表示している。「年」の属性のノードを(左)時系列の順序、(右)時系列の幅を表すように配置した。

ノードの配置変更

本手法では、属性名を表すノード(第1行のノード)の位置を基準として、同じ属性のノードを配置する。配置方法として、時系列の順序を表す配置と時系列の幅を表す配置の2つを開発した。

時系列の順序を表す配置(図2左)では、各ノードは時系列データの順序に従って並ぶ。この配置では、隣合うノード間の距離が等しくなり、一つ一つのデータを眺める時にノードを辿りやすい。

時系列の幅を表す配置(図2右)では、ノード間の距離が隣合うノードの持つ時刻(例では年)の差に比例する。この配置では、ノードの位置が時間軸上の分布を表現する。これにより、時系列データの分布を視覚的に理解することができる。図2から自然災害が毎年起こるといふ周期性と、1971年と1972年には自然災害が起こらなかったという周期から外れた年をすぐに発見できる。

時系列データのフィルタリング

時系列データのフィルタリングは、データ全体から着目したデータと関係のあるノードを抜き出す操作である。あるノードと関係の持つデータを見たい時に、関係のあるノードを所属している属性から抜き出し、横方向に移動させる(図3)。

この操作によって、着目したノードと関係のあるノードだけを新しい属性のように見ることができる。また、ノードが抜き出されて残った元の属性も着目したノードと関係のないノードを集めた新しい属性として見ることができる。例えば、図3では年と場所のノードの中間に位置する「豪雨」に関するデータと、年のノードの下に位置する「豪雨以外の自然災害」のデータを示している。これによって、データから新たな特徴を発見することが可能になる。

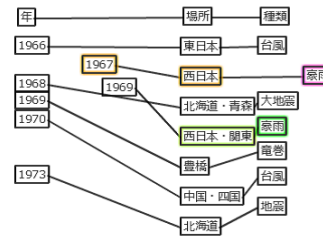


図 3: 自然災害史データから「豪雨」に関するデータをフィルタリングした例。

4 関連研究

時系列データ分析ツールとして、パターンの発見を目的とした研究[1]がある。折れ線グラフを用いて量的データを持つ時系列データを表現し、選択したパターンと似たパターンを探ることができる。また、発見したパターンを集計値として確認することを目的とした研究[2]では、表とタイムラインを組み合わせて時系列データを表現している。時系列データの特徴について調べるためには、集計値ではなく詳細なデータの探索が必要である。

表の特徴を保存した表現として、表と棒グラフを組み合わせた手法[3]がある。これは、詳細なデータを保持しつつ、データの特徴を発見できる点で似ている。本研究では、セルの位置を用いてデータの分布を視覚的に表現し、時間軸上の分布を視覚的に表現できる。

5 まとめ

表と連結図の特徴を組み合わせた表現と時系列データ分析のための操作を統合したツールを開発した。これにより、データの時間軸上の分布からデータの傾向やパターンを発見し、発見した特徴について詳細な情報を取得できるようになった。

参考文献

- [1] Buono P, Aris A, Plaisant C, Khella A, Shneiderman B. Interactive Pattern Search in Time Series. Proceedings of Conference on Visualization and Data Analysis, VDA 2005, pp.175-186.
- [2] Wang T, Plaisant C, Shneiderman B, Spring N, Roseman D, Marchand G, Mukherjee V, Smith M. Temporal Summaries: Supporting Temporal Categorical Searching, Aggregation and Comparison. Proceedings of IEEE Infovis 2009. pp. 1049-1056.
- [3] Ramana Rao, Stuart K Card. The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information, In Proceedings of the SIGCHI Conference on Human Factors in Computing 1994. pp.318-322.