ヒートマップによる時系列データ可視化の一手法
～非類似度と異常値観察を目的として

理学専攻・情報科学コース
熊谷 沙津希　(指導教員：伊藤 貴之)

1　はじめに

ヒートマップは，時系列データ可視化のための効果的な表現の一つであり，折れ線グラフと同様に広く用いられている手法である．しかしながら，入力データが変数や時刻を多数含むとき，表示のための画面領域を大幅に必要とする場合がある．我々は，時系列データ可視化においてヒートマップを適切に形成するとともに，重要ではない変数や時刻を対話的にフィルタリングすることが重要であると考える．またその過程において，変数間の相関関係を考慮することも時系列データ可視化の重要な要素の 1 つであると考える．変数間の相関関係から，変数同士がどの程度似ていないかや，変数群中に他と大きく異なる値が存在するかを導き出し，これを可視化することが重要であると考える．本報告では，変数と時刻の有意義な表示を可能にし，ユーザが対話的に操作できる時系列データのためのヒートマップベースな可視化手法について議論する．また，変数間の非類似度を算出してこれに着目した可視化と，変数が保持するある時刻における異常値を算出してこれを観察する可視化手法について議論する．

2　関連研究

時系列データの可視化には，一般的に折れ線グラフやヒートマップを用いた可視化が利用されている．折れ線グラフはヒートマップに比べて，各変数における正確な数値を読み取りやすい，数値の上昇や下降を読み取りやすい，などの利点がある．しかし，データを構成する変数の増加に伴い，折れ線どうしの絡み合いも増加し，視認性低下の原因となる．これらの問題を解決するために，折れ線の表示数を対話的に調節する手法が提案されている [1]．また，数値の範囲が大きく，かつ数値分布が不均一である場合，画面領域を浪費するような可視化結果になることが多い，という問題もある．

ヒートマップは値の大きさを色で表示する可視化手法である．ヒートマップを用いた時系列データの可視化には，データを構成する変数を縦軸に沿って一列ずつに並べ，横軸に何らかの変数（時系列データの場合には時刻）を割り当てて，両軸を分割して得られる各領域に色付ける．ヒートマップは変数が大きなデータにおいても，数値を表現する形状が画面上で重なり絡み合うことがないため，視認性の維持が容易である．さらに折れ線グラフと比較して，数値の範囲や分布により表示領域を浪費するような可視化結果を生じることもない．

3　非類似度に着目したヒートマップ表示

ヒートマップ型の時系列データ可視化手法の問題点を解決するために，変数間の類似性に限定せずに任意の変数間距離を算出し，それに基づいて生成された距離行列を用い，変数をクラスタリングすることを考える．我々は正と負の相関において変数間の関係を観察したいと考え，相関係数にもとづいて変数間距離を算出することにした．ここで高次元データを構成する変数 $v = v_1 ... v_n$ とし，i 番目の変数は数値 $v = v_{i1} ... v_{im}$ を持つとする．現時点の我々の実装ではケンドール順位相関係数を適用して，i 番目と j 番目の変数間の相関係数 d_{ij} を算出する．また，このとき P は変数 i における任意の隣り合う 2 つの時刻と，変数 i のこれに対応する 2 つの時刻について，それぞれの組の大小関係が一致するとき 1 を加算し，不一致のとき -1 を加算した和とする．ここで与えられた値から $d_{ij} = 1.0 - |c_{ij}|$ を算出し，得られた値を 2 変数間の距離とする．

この実装は，はじめに変数間の距離に基づいて樹形図を構築し，樹状図の構造に基づいてそれらの順序を固定する．この実装では，デンドログラムと閾値距離値に従って変数をクラスタリングする．必要に応じて 1 つの変数のみを含むクラスタの描画を制御し，一連の変数群を観察できる．実装には，閾値をインタラクティブに制御するスライダを設け，ユーザがこれを操作することで重要でない変数，または重要でない時刻をヒートマップから割愛するための閾値を調節することができる．また，この実装は表示スペース内の変数の順序がデンドログラムの構造に従って固定されているため，ユーザーのメンタルマップを保持する．

さらに，ヒートマップのカラーマップを新たに実装し，負の相関を持つ変数の観察理解度を向上させた．我々の以前の実装では，単純に暖色により高い値に割り当てられ，寒色により低い値に割り当てられるように色合いが変化する虹色のカラーマップを適用した．しかしこの実装では，正の相関を有する変数とは異なり，負の相関を有する変数を比較することは容易ではなかった．彩度 s_{ij} と色相 h_i を計算するために次式を算出する．また、我々の実装ではヒートマップの色計算に以下の 2 種類のカラーマップを採用している．なお色計算には HSB 表色系を採用し，入力値には v_{ij} を $[0, 1]$ の区間に正規化した値 v'_{ij} を使用するものとする．

$$s_{ij} = v'_{ij} \; h_i = \frac{160}{240}(1.0 - \frac{1}{m}\sum_{j=0}^{m} v'_{ij})$$

4　異常値観察のためのヒートマップ表示

前章では，同様の相関を持つ変数を可視化した．しかしながら，時系列データ可視化の際には，他の値と異なる値を示す異常値を観察することも重要である．本章では，まず異常値の定義から説明する．本研究では，他の変数や時間と大きく異なる値の出現頻度や，曜日の他の週と大きく異なる値を習慣度として算出し，曜日の値習慣レベルが小さい場所を異常値と定義して，これら異常値を観察するための 3 つのカラーマップを実装する．

1 つ目のカラーマップ (Colormap1) では，全変数の

全時刻における値との相対値を算出する．ここで，色相を計算するために次の値を算出する．

$$h_i = \frac{160}{240}(1.0 - v'_{ij})$$

これは最も一般的なカラーマップの 1 つである．このカラー表示では，全体的な値と比較して他と大きく異なる値をより明確に観察することが可能である．

2 つ目のカラーマップ (Colormap2) では，平均値との差分とその種類に応じて着色方法を区別した．値が平均値を超えた場合値の色を赤，そうでない場合は青と定義した．彩度は平均値との差分に従って定義され，その濃度は放物的に色付けされる．このカラーマップは，変数全体よりも秀でた値を観測する観察には適さないが，変数 i の他の変数と比較した際の数値的差分が全体と比較してどの程度であるかを理解することができる．

3 つ目の算出手法 (Colormap3) では，偏差値 s_{ij} を $v_{i1}...v_{in}$ から得る．ユーザによって閾値 T_1 と T_2 が定義され，閾値を超える値は色付けされる． 値は T_1 より小さい場合は青色に， T_2 より大きい場合は赤色に着色され， T_1 と T_2 の間の値は灰色に着色される．この着色手法は，ある変数の観測期間内の時間的動向に着目して異常値を観測することを可能にした．さらに，[2] に基づいて慣習的傾向をもつ値を定義し，これを反映したヒートマップ可視化を実装した．

5　まとめと今後の課題

本報告では，変数と時刻の有意義な表示と，ユーザの対話的に操作のための，変数間の相関関係を考慮した，ヒートマップベースの時系列データにおける 2 種類の可視化手法について説明した．我々は，相関係数に基づく距離定義と階層的クラスタリング適用した，非類似度に焦点を当てた可視化手法について説明した．また，異常値とその発生頻度を定義した，異常値観測のための可視化手法について説明した．

今後の展望として，異常値観察のための可視化手法において，習慣度を算出するための単位期間を多様にしたい． また，本研究で実装された 2 種類の手法による可視化結果について，ユーザが理解可能なデータ規模 (変数群・時刻数) の上限を検証したい．さらに，これらの 2 種類の可視化を統合して操作する可視化の実装を実現したい．

参考文献

[1] Y. UCHIDA and T. ITOH. A visualization and level-of-detail control technique for large scale time series data. In *13th International Conference on Information Visualisation (IV09)*, pp. 80–85. IEEE, 2009.

[2] A. HAYASHI, M. KOHJIMA, T. MATSUBAYASHI, and H. SAWADA. Regularity measure and influence weight for analysis and visualization of consumer ' s attitude. In *19th International Conference on Information Visualisation (IV2015)*, pp. 290–299. IEEE, 2015.

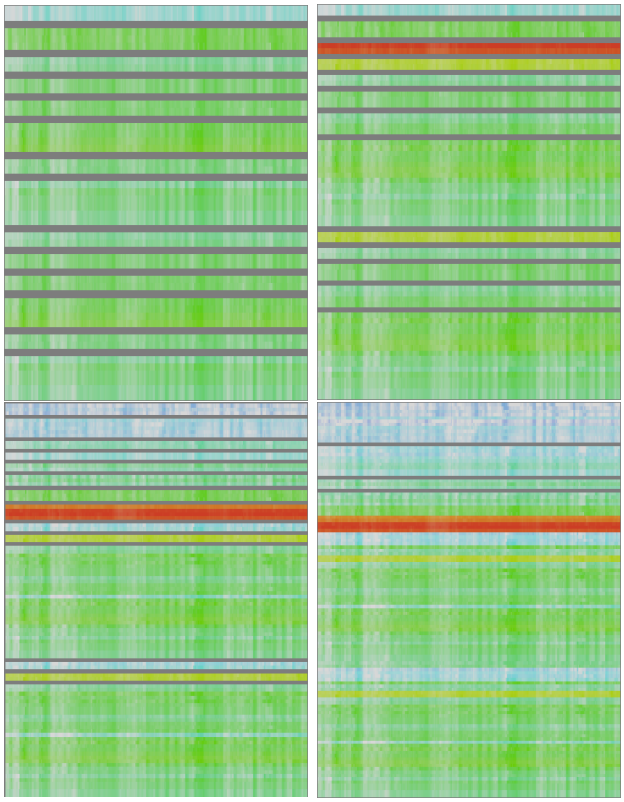

Figure. 1: 新しいクラスタリングと着色手法を適用したヒートマップベースな可視化の実装． アメダス気象データを用いて気候の相関が似ている地域をクラスタリングした．4 枚の図内で変数が同じ順番で配置されている．

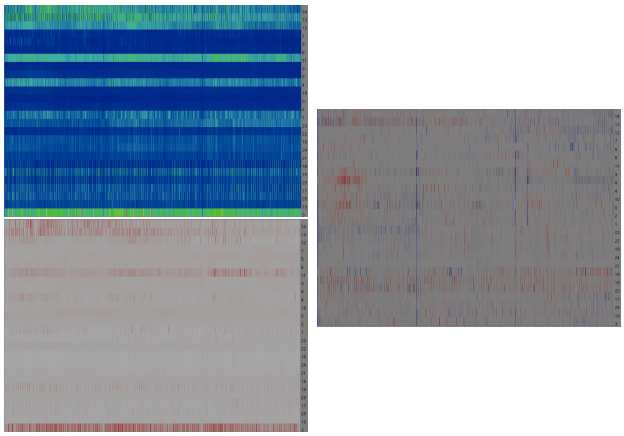

Figure. 2: 3 種類のカラーリングを適用したヒートマップベースの可視化の実装．小売店の売上データにおける異常値を計算した．3 枚の図は，同じ分析結果を別の着色手法を適用した可視化結果例である． 左上の図は Colormap1，左下の図は Colormap2，右の図は Colormap3 の可視化結果を表す．