

修了年月 (西暦): 2015 年 3 月

専攻名: 複雑理工学専攻

氏名: 能野 琴

学生証番号: 47136110

論文題目: Contractible Parallel Coordinates via Correlation Analysis

相関分析を用いた軸縮約可能平行座標系表示

キーワード: 平行座標系, グラフスペクトル解析, 軸縮約, 相関グラフ, 系統樹

指導教員: 高橋 成雄 准教授

## ABSTRACT

High-dimensional datasets occur frequently in various fields. Parallel coordinates is well-known as a popular tool for visualizing the underlying relationships among variables in high-dimensional datasets. This is accomplished by plotting each sample in terms of multiple vertical axes as a polyline. However, this representation still suffers from visual clutter arising from intersections among polyline plots especially when the number of data samples and their associated dimension become high. This thesis presents a method of alleviating such visual clutter by contracting multiple axes through the analysis of correlation between every pair of variables. In this method, we first construct a graph by connecting axis nodes with an edge weighted by data correlation between the corresponding pair of dimensions, and then reorder the multiple axes by projecting the nodes onto the primary axis obtained through the spectral graph analysis. This allows us to compose a dendrogram tree by recursively merging a pair of the closest axes one by one. The proposed visualization platform helps the visual interpretation of such axis contraction by plotting the samples along the composite axis by referring to the principal component analysis and linear discriminant analysis. Smooth animation of the associated axis contraction and expansion has also been implemented to enhance the visual readability of behavior inherent in the given high-dimensional datasets. A user study has been also conducted to investigate how the correlations among coordinate axes are better visualized using my approach, and this experimental results demonstrated the capability of the proposed approach.

## 論文要旨

様々な分野において情報が高次元データとして表現される現在、可視化は、そのようなデータから有益な情報を取得するために、一般的に普及している。高次元データの可視化で広く用いられ、次元間の相関を効果的に把握できる手法のひとつに、平行座標系がある。平行座標系は、データにおける各変量の値を縦に平行に並んだ座標軸上にとる表現方法であり、各データサンプルは、平行軸と交差する折れ線として表現される。しかしながらこの表示方法では、データのサンプル数および次元数の増加に伴い、データに対応する折れ線同士が重なり、視覚的にデータの振る舞いの把握が困難になる視覚的乱雑さが生じる。そこで本論文では、この問題を解決するために、任意の次元間の相関に基づいて複数の座標軸を縮約してひとつにまとめ、全体のデータの次元数を減らして視覚的にわかりやすく平行座標系を描画する手法を提案する。具体的には、以下の手順で処理を行う。まず、各次元軸同士の相関係数を類似度とみなし、軸をノードとして、類似度の高い軸同士を結ぶ完全グラフを構築する。次に、グラフスペクトル解析を用いて各軸の位置を1次元数直線上に写像し、得られた座標軸の順番に平行座標軸を並び替える。その後、写像した1次元数直線上において、最近接する軸の併合を適宜行い、軸の系統樹を構成して、階層クラスタリングを行う。また、上記の操作時に、平行座標軸や系統樹にアニメーション効果を導入し、座標軸の縮約および展開の過程も視覚的に確認しながら、次元数を適宜増減させることを可能にする。このように、相関に基づいて平行座標系の軸を縮約して、高次元データをより少ない座標軸で表示することで、視覚的乱雑さを軽減すると共にデータに内在する大局的な振る舞いの視覚的な理解を実現している。さらに、次元間の相関がより明確に可視化されるかを調べるためユーザスタディを実施し、本手法の有効性を示した。