

## 散布図選択による多次元データ可視化へのグラフ彩色問題の適用

伊藤貴之\* 中林明日香\* 萩田真理子\*

**概要.** 現実社会のデータの多くは多変数・多次元である．限られたディスプレイ空間に多次元データの特徴を表現する手法として，選択的な散布図表示が有効である．閲覧する価値のある重要な散布図を選択するための指標は既にいくつか提案されており，これらの多様な指標を同時に参照して多様な散布図を選択することが，データの特徴を網羅的に理解する上で有効であると考えられる．この考えにもとづいて本報告では，多数の指標を同時に参照した散布図選択手法を提案する．本手法では多数の指標の各々に沿って散布図のスコアを算出し，スコアの類似度が高い散布図を接続することで，散布図群をノード群としたグラフを生成する．このグラフに彩色問題を適用することで，非常に類似した複数の散布図を同時に選択しないように散布図を選択する．本報告ではアパレル小売店の販売情報と気象情報の関係をあらわす多次元データへの適用事例から提案手法の有効性を議論する．

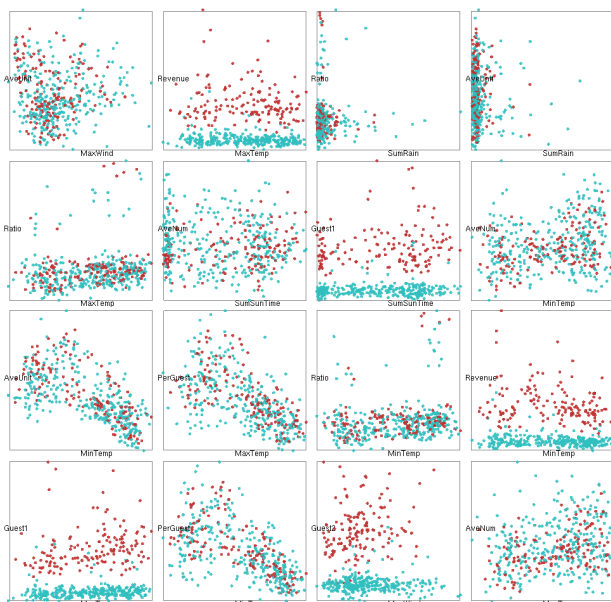


図 1. 提案手法による散布図選択結果の例．次元間の相関，クラスターや例外点の分離性，クラス間の分離性，といった多様な指標に沿って単一の多次元データから多様な散布図を選択している．

### 1 はじめに

タッチパネルデバイスの普及や，ウェブブラウザ上での対話操作技術の発達などに影響を受けて，情報可視化システムのデザインにも変化が生じている．ダッシュボードと呼ばれる情報一覧型のデザインがその典型例である．ダッシュボードでは広大な表示空間に多数の可視化画面を並べて表示することで，

単純なスクロール操作だけで多様な情報を閲覧できる仕組みを提供している．このようなデザインの普及により，情報可視化技術においても，従来の探索的・反復的なインタラクションだけでなく，閲覧する価値のある多数の可視化結果を最初から全部選出しておき，スクロール操作だけでそれを全部閲覧する，という手軽なインタラクションの実現が容易になった．逆に言えば，このような単純なインタラクションだけで重要な現象を全て閲覧できるようにするためには，多数の有用な可視化結果を適切に自動選択することが重要な研究課題となる．

多次元データ可視化は情報可視化の最重要課題の一つである．多次元データ可視化手法には，散布図行列 (SPM: ScatterPlot Matrix) や平行座標プロット (PCP: Parallel Coordinate Plots) のように座標軸を明示した幾何学的手法の他に，アイコンベースの手法や画素ベースの手法が知られている．幾何学的手法の多次元データ可視化手法の研究では近年，次元選択を施すことで重要な次元だけを重点的に可視化する手法 [5, 19, 21] が多数発表されている．

一方で，多数の散布図を用いた多次元データ可視化手法の研究の一環で，散布図の価値を数値評価する手法がいくつか発表されている．その代表例である Scagnostics [18] は，これまで多くの散布図選択手法 [17, 10, 22] に適用されてきた．Scagnosticsには多数の数値指標が提案されているが，その中の単一の指標にもとづいて散布図を選択しただけでは，データに潜む多様な現象を網羅的に発見できるとは限らない．ある時は次元間の相関性に興味深さが潜んでいるかもしれないし，またある時にはクラスターやクラス間の分離性に興味深さが潜んでいるかもしれない．そこで，多様な指標に沿って多様な散布図を選択することで，データ中の多様な現象を網羅的に可視化できる可能性が高まると考えられる．

Copyright is held by the author(s).

\* お茶の水女子大学

本報告では、多様な指標にもとづいて高速に散布図を選択する一手法を提案する。提案手法では任意の2変数を2軸とした多数の散布図を生成する。続いて、各散布図について、複数の指標に沿って複数のスコアを算出し、そのスコア群を並べたベクトルを生成する。そしてベクトル間の類似度が一定値以上である散布図を連結することで、散布図のグラフを生成する。このグラフに彩色問題を適用して、エッジで連結された隣接散布図には異なる色を割り当てる。その結果として、同一色が割り当てられた散布図群は「互いに一定以上の非類似度を有する多様な散布図群」を構成する。そこで、同一色が割り当てられた散布図の中から、ユーザ指定の個数の散布図を選出して画面表示することで、図1に示すように多様な散布図をユーザに提示することができる。

本報告では、アパレル小売店の販売実績と気象情報から構成される多次元データを用いた適用事例を紹介し、提案手法の有効性を議論する。

## 2 関連研究

### 2.1 多次元データ可視化のための次元選択

次元選択は多次元データの中から興味深い部分のみを可視化する手段として広く議論されてきた。Claessen ら [2] は多次元データを散布図と平行座標プロットで、Suematsu ら [15] は低次元平行座標プロットの集合で、Zheng ら [22] は少数の散布図の集合で可視化する手法を提案し、その中で次元選択を適用してきた。しかしこれらの手法では次元選択のための対話操作を搭載していないため、可視化結果を動的に調節することはできなかった。

多次元データ中の低次元空間を適応的に可視化するための対話操作を備えた手法も近年になっていくつかが報告された。Lee ら [6] と Liu ら [7] は多次元データから抽出した低次元空間に次元削減を適用する手法を提案した。Nohno ら [11] は平行座標プロットにて表示する軸の数を適応的に調節する手法を提案した。Itoh ら [5], Watanabe ら [17], Nakabayashi ら [10] の一連の手法では、画面に表示される低次元平行座標プロットの規模や散布図の数を単純なスライダー操作で調節する手法を提案した。また、これらの操作を支援するために次元間の関係を可視化する機能を備えた多次元データ可視化手法 [5, 19, 20] も報告されている。

以上のように、多次元データ可視化のための次元選択手法は多数発表されている。しかし、多様な指標を同時に参照することで有限個の多様な散布図を選ぶ、という考え方にもとづいて多次元データ中の多様な現象を表現する手法は見当たらない。本報告はこの点に注目した手法を提案する。

### 2.2 散布図の数値評価

散布図の特徴を数値評価する手法も既にいくつか発表されている。その中でも最も有名な Scagnostics [18] では、散布図を構成する点群の分布から9種類の指標を算出する手法を提案している。Scagnosticsの改良手法として、次元間の相関に着目した改良手法 [4, 13] や、クラス分離に着目した改良手法 [1, 12, 14] が報告されている。また Wang ら [16] は、視覚認知を考慮した Scagnostics の改良手法を提案している。

多次元データから生成した多数の散布図群の概観と探索を目的とした手法もいくつか発表されている。Dang ら [3] は多数の散布図群から類似する散布図を発見するための探索的な可視化手法を提案している。Matute ら [8] は多数の散布図群の分布を一覧するための可視化手法を提案している。本論文の提案手法はこれらの手法と目標設定に近いが、提案手法はディスプレイ空間の制約を考慮して有限個の散布図を選択表示するという点でこれらの手法と異なる。

## 3 散布図選択へのグラフ彩色問題の適用

本章は提案手法による散布図選択の処理手順を示す。本手法では多次元データから生成された散布図群の各々に対して、複数の指標に沿って複数のスコアを算出し、それをベクトルとして扱う。提案手法による散布図選択の概念を図2に示す。提案手法では以下の要件を満たすようにベクトル群を選択し、それらに対応する散布図群を表示する。

要件1: 距離の近すぎるベクトル群を選ばずに、距離の離れたベクトル群を選ぶことで、多様な散布図を選ぶことができる。

要件2: 長いベクトルを選ぶことで、特徴的な散布図を選ぶことができる。

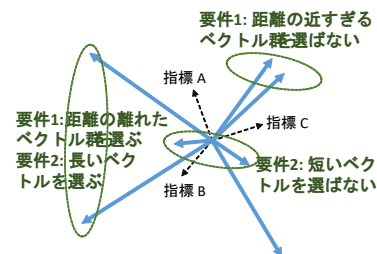


図2. 提案手法による散布図選択の概念を示すために、複数の指標を座標軸とした多次元空間を描いた図。青い矢印で描かれたベクトルが各散布図に対応する。提案手法では要件1,2を満たすように有限個の散布図を選択する。

### 3.1 データ構造

本報告では提案手法が解くべき問題を以下の通り定式化する． $m$ 次元データ  $A$  は  $n$  個の標本  $A = \{a_1, a_2, \dots, a_n\}$  を持っており， $i$  番目の標本  $a_i$  は  $m$  次元ベクトル  $a_i = (a_{i1}, a_{i2}, \dots, a_{im})$  を有するとする．また各標本には1個以上のクラスが割り当てられている場合もあるものとする．そして，このデータから  $N$  個の散布図  $S = \{s_1, s_2, \dots, s_N\}$  が生成されるとする．また，各散布図について  $M$  種類の指標に沿ってスコアを算出し，そのスコアを  $M$  次元ベクトル  $s_i = (s_{i1}, s_{i2}, \dots, s_{iM})$  として格納する．本章では  $i$  番目と  $j$  番目の散布図のコサイン類似度を  $d_{ij} = (s_i \cdot s_j) / (|s_i| |s_j|)$  と記述する．

### 3.2 散布図選択のための指標

提案手法では複数の指標に沿って散布図のスコアを算出する．現時点で実装されている指標を以下に示す．

#### 3.2.1 相関

筆者らの実装では  $k$  番目の散布図の1番目のスコアを単純に以下の式で算出する．

$$s_{k1} = |S_{pear}(i, j)|^2 \quad (1)$$

ここで  $S_{pear}(i, j)$  は散布図の2軸となる  $i$  番目と  $j$  番目の次元間の Spearman 順位相関係数である．この値が大きいほうが正または負の相関が高く，閲覧するに値する散布図であると評価される．

#### 3.2.2 細さ

散布図上で点群が細長い領域を構成していると，その点群はモデル化が容易である場合が多い．そこで点群が集中する領域の細長さを指標とする．筆者らの実装は Wilkinson ら [18] の実装に準じるものである．具体的には，点群を接続する Delaunay 三角メッシュを生成し，閾値以上の長い辺をもつ三角形を削除した上で，以下の値を算出する．

$$s_{k2} = 1 - \sqrt{4\pi A_{rea}(T) / P_{perimeter}(T)} \quad (2)$$

ここで  $A_{rea}(T)$  は三角メッシュ  $T$  の総面積であり， $P_{perimeter}(T)$  は  $T$  は外周を構成する三角形辺の長さの合計値である．

#### 3.2.3 クラスタや例外点の分離性

散布図を構成する点群が有限個のクラスタに明確に分離する場合や，あるいは明らかに例外とみられる点が存在している場合も，閲覧することで知見が得られる場合が多い．これらを検出するために，筆者らの実装では分離性に関する Wilkinson ら [18] の Clumpy という指標に沿って，以下の値を算出する．

$$s_{k3} = \max(1 - \text{length}(e_i) / \text{length}(e_{i+1})) \quad (3)$$

ここで前項で論じた通り，筆者らの実装では点群を連結する Delaunay 三角メッシュを生成し，メッシュを構成する三角形辺をその長さでソートする．長い順に  $l$  番目と  $l+1$  番目の辺の長さから，上述の式が最大となる値を算出する．

#### 3.2.4 クラスの分離性

特定のクラスを有する点群がほかの点群から明確に分離している散布図も，閲覧することで知見を得られる場合が多い．筆者らの実装では情報エントロピーにもとづいてクラス分離性を数値化している．具体的には， $i$  番目と  $j$  番目の次元から構成される散布図について，以下の値を算出する．

$$H(i, j) = -\frac{1}{n} \sum_{k=1}^n \sum_{c=1}^C p(y_k = c | (a_{ki}, a_{kj})) \log p(y_k = c | (a_{ki}, a_{kj})) \quad (4)$$

ここで  $y_k$  は  $k$  番目の標本のクラス， $(a_{ki}, a_{kj})$  は  $k$  番目の標本の座標値， $C$  はクラスの数である．筆者らの実装では散布図を  $L$  個の部分領域に分割し， $l$  番目の部分領域におけるエントロピー  $H(i, j)_l$  を上述の式により算出する．そして最終的に，以下の式により  $k$  番目の散布図のスコアを算出する．

$$s_{k4} = (H_{max} - \sum H(i, j)_l) / H_{max} \quad (5)$$

ここで  $H_{max}$  は  $\sum H(i, j)_l$  の最大値である．

### 3.3 グラフ彩色問題の適用

グラフ彩色問題は組み合わせ最適化に関する多様な問題に適用されており，筆者らも写真選択の組み合わせ最適化のためにグラフ彩色問題を適用している [9]．提案手法では散布図選択のためにグラフ彩色問題を適用する．

散布図の集合  $S$  と散布図ペアを接続するエッジの集合  $E$  で構成されるグラフ  $G = \{S, E\}$  があるとする．提案手法では， $i$  番目と  $j$  番目の散布図が有するベクトルのコサイン類似度  $d_{ij}$  が閾値  $d_{thres}$  よりも大きいときに，この2つの散布図をエッジで接続する．

続いて  $G$  を構成する各散布図に固有の色を割り当てる．この時，エッジで隣接する2つの散布図には必ず別々の色を割り当てるという制約を与える．この制約により，類似度の高い散布図には必ず異なる色が割り当てられる．結果として，同一の色を割り当てられた散布図群は互いに一定以上の非類似度を有する散布図で構成される．図3にこの工程を示す．提案手法ではまず  $|s_k|$  の値が最大である散布図の1つを選出し，この散布図に色識別子  $c_k = 0$  を与える．続いて提案手法では，幅優先探索によって散布図を再帰的に探索し，各散布図に色識別子を割り当てる．このとき， $j$  番目の散布図について，そ

れとエッジで接続している隣接散布図群に既に割り当てられている色識別子を羅列し、どの隣接散布図にも割り当てられていない色識別子のうち値が最小であるものを散布図  $j$  に割り当てる．例えば散布図  $j$  の隣接散布図に色識別子  $0, 1, 3$  が割り当てられていれば、散布図  $j$  の色識別子は  $c_j = 2$  となる．

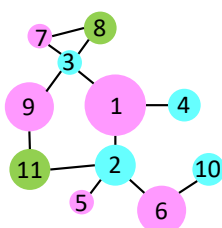


図 3. グラフの彩色．エッジで接続された隣接散布図には必ず別の色を割り当てる．

続いて提案手法では、ユーザが指定した個数の散布図を自動選択して表示する．ここで提案手法では、同一の色識別子が割り当てられた散布図群の中から所定の個数を選択する．同一の色識別子が割り当てられた散布図群は互いに一定以上の非類似度を有する．よって、この中から散布図を選択することで要件 1 を満たす．また、同一の色識別子が割り当てられた散布図群の中から、スコアの最大値  $\max(s_{k1}, s_{k2}, s_{k3}, s_{k4})$  が大きい順に散布図を選ぶことで要件 2 を満たす．

以上をまとめると、提案手法は以下の処理手順で散布図を自動選択する．

1. コサイン類似度  $d_{ij}$  が一定値以上である散布図  $i$  と  $j$  をエッジで連結する処理を反復することで、グラフ  $G$  を生成する．
2. ベクトルの長さ  $|s_k|$  が最大である散布図の 1 つを選択し、探索の出発点とする．
3. グラフ  $G$  を構成する散布図を幅優先探索し、その各々に色識別子を割り当てる．このとき色識別子には、エッジで連結された隣接散布図に割り当てられたいずれの色識別子とも異なるものを選ぶ．
4. 同一の色識別子を有する散布図を集めて、 $\max(s_{k1}, s_{k2}, s_{k3}, s_{k4})$  の値が大きい順に、ユーザ指定の個数だけ散布図を選ぶ．

#### 4 適用事例

本章では、アパレル小売店の販売情報と気象情報を提案手法で可視化した事例を示す．本事例では説明変数（気象値）を散布図の横軸に、目的変数（販売値）を散布図の縦軸に割り当てている．表 1 に各変数を示す．この可視化結果では、2016 年 5 月 1

日から 2017 年 7 月 31 日までの記録を、散布図中に 457 個の点群で示している．点群のうち青色は平日の記録を、ピンク色は休日の記録をあらわしている．なお、このデータは現実のデータに微小な乱数値を加算したものであり、現実の販売実績とは異なる値を可視化している点に注意されたい．

表 1. 説明変数（気象値）と目的変数（販売値）．

説明変数 (気象値)	
MinTemp	最低気温
MaxTemp	最高気温
SumRain	降水量
SumSunTime	日照時間
MaxWind	最大風速

目的変数 (販売値)	
Revenue	総売上額
Guest1	販売者数
Guest2	訪問者数
Ratio	販売率
PerGuest	販売者あたり売上額
AveUnit	商品あたり金額
AveNum	販売者あたり商品数

図 1 は提案手法による散布図選択結果の例である．いくつかの散布図からは次元間相関が見受けられ、またいくつかの散布図からはクラスや例外が明確に見受けられる．またいくつかの散布図は青色とピンク色の点群が明確に分離している．この図から、提案手法が多様な散布図を自動選択していることがわかる．

図 4, 5, 6 はそれぞれ、相関、クラス、クラスタの分離性、クラスや例外の分離性、の各々のスコアが高い値を有する 4 つずつの散布図を示したものである．図 4 に示す散布図は、横軸が MinTemp または MaxTemp で縦軸が PerGuest または AveUnit である．この散布図から当該小売店では、夏場には単価の安い商品が売れるか 1 人当たりの購買額が低い傾向があり、年の平均よりやや寒い時期（3 月や 12 月頃）に単価の高い商品が売れるか 1 人当たりの購買額が高い傾向がある、という明確な相関が見受けられる．図 5 に示す散布図は、縦軸が Revenue, Guest1, Guest2 のいずれかである．つまり当該小売店では、総売上額、販売者数、訪問者数において平日と休日に明確な差があることがわかる．図 6 ではクラスや例外が明確にみられる散布図が選ばれている．この中で左から 2 番目の散布図では、縦軸に Ratio が割り当てられており、購買率が異常に高い日（購買せずに帰った人の割合が低い日）を明確に見つけることができる．提案手法による図 1 の散

布図選出結果は、このような多様な特徴を有する散布図をバランスよく選出した結果であることが見受けられる。

一方で、4種類のいずれのスコアも高くない散布図の例を図7に示す。これらの散布図を注意深く観察しても、当該小売店の販売実績に関する興味深い現象を示しているようには見受けられない。提案手法はこのような散布図を選ばないことに成功している。

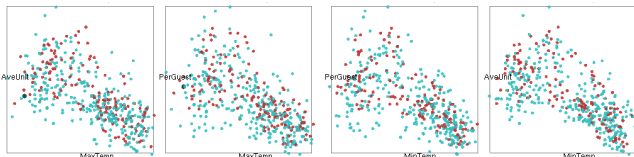


図4. 「相関」のスコアが高い4つの散布図。

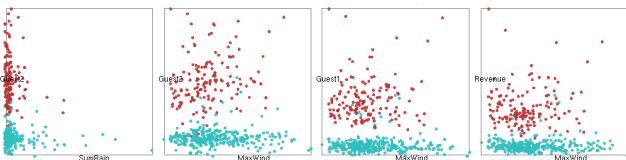


図5. 「クラス分離性」のスコアが高い4つの散布図。

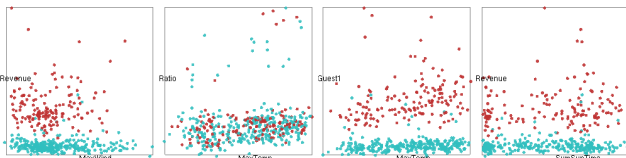


図6. 「クラスタや例外の分離性」のスコアが高い4つの散布図。

図1に示された16個の散布図について、正規化した4つのスコアを可視化した結果を図8に示す。この結果からも、提案手法が多様なスコアを有する16個の散布図を選択できていることがわかる。

提案手法による散布図選択結果は閾値  $d_{thres}$  の選択に大きく依存する。閾値  $d_{thres}$  を小さくすると多数のエッジが生成され、結果として色数も多くなる。図1に示したデータにおけるエッジ数と色数の変化を表2に示す。ここで、色数が多くなるほど「非常に類似した複数の散布図」の選択を避けることが容易になるが、代わりに同一の色を割り当てられた散布図の数が減少するので、スコアの低い散布図も選出せざるを得なくなる可能性が高まる。閾値  $d_{thres}$  の調節に対する  $\max(s_{k1}, s_{k2}, s_{k3}, s_{k4})$  の変化を図9に示す。この図からも、提案手法では閾値  $d_{thres}$  を注意深く調節する必要があることが示唆される。

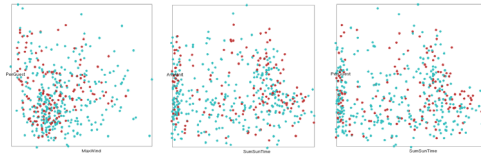


図7. いずれのスコアも高い値を有さない散布図の例。

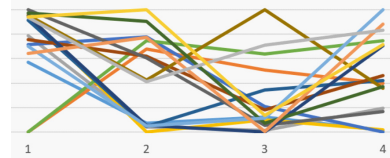


図8. 図1に示された16個の散布図のスコア分布。

## 5 まとめ・今後の課題

本報告ではグラフ彩色問題を用いた散布図選択手法を提案した。提案手法では複数の指標に沿って算出された散布図のスコアからベクトルを構成し、そのベクトルの類似度にもとづいて散布図のグラフを生成する。そしてこのグラフに彩色問題を適用し、同一の色を割り当てられた散布図の中から所定の個数の散布図を選択する。本報告ではアパレル小売店の販売実績のデータに提案手法を適用した事例を紹介し、提案手法の有効性を議論した。

今後の課題として、まずスコア算出のための指標の改善や追加があげられる。2.2節で論じた通り、Scagnosticsには多くの指標があり、いくつかの改良手法が提案されており、これらを追加実装することが可能である。一方で、適切な散布図の個数は入力データの次元数とスコア算出のための指標数に依存する。そこで、スコア算出の指標数が少ないために十分な個数の散布図を選べないという状況が生じるのを避けるためにも、スコア算出の追加実装が重要であると考えられる。また、スコア算出の各指標の重要度はデータの数値分布や可視化の用途に依存することから、スコア算出に際して指標に重みをつけてデータに応じて調節することも有用であると考えられる。

さらに、提案手法の拡張性を検証したい。特に、より多くの次元数や標本数を有するデータでのテストを試みたい。また、適切な散布図の個数と配置について、データの数値分布の観点だけでなく、インタラクションの観点からも検証したい。これらの改良と検証を試みたのちに、より多様なデータで提案手法の適用事例をつくり、さらにユーザ評価実験にも着手したい。

## 参考文献

- [1] M. Aupetit and M. Sedlmair. Sepme: 2002 new visual separation measures. In *IEEE Pacific Vi-*

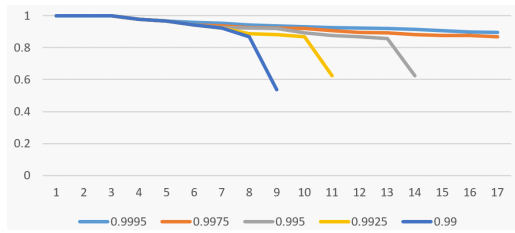


図 9. 閾値  $d_{thres}$  を調節した際の順位の変動 .

表 2. 生成されたグラフのエッジ数と色数 .

$d_{thres}$	0.9995	0.9975	0.995	0.9925	0.99
エッジ数	5	19	39	60	83
色数	2	3	5	6	7

ualization Symposium 2012, pages 43–52, 2016.

- [2] J. H. T. Claessen and J. J. van Wijk. lexible linked axes for multivariate data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2310–2316, 2011.
- [3] T. N. Dang and L. Wilkinson. Scagexplorer: Exploring scatterplots by their scagnostics. In *IEEE Pacific Visualization Symposium 2014*, pages 73–80, 2014.
- [4] L. Harrison, F. Yang, S. Franconeri, and R. Chang. Ranking visualizations of correlation using weber’s law. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1943–1952, 2014.
- [5] T. Itoh, A. Kumar, A. Klein, and J. Kim. High-dimensional data visualization by interactive construction of low-dimensional parallel coordinate plots. *Journal of Visual Languages and Computing*, 43(1):1–13, 2017.
- [6] J. H. Lee, K. T. McDonnell, A. Zelenyuk, D. Imre, and K. Muller. A structure-based distance metric for high-dimensional space exploration with multidimensional scaling. *IEEE Transaction on Computer Graphics*, 20(3):351–364, 2013.
- [7] S. Liu, B. Wang, P.-T. Bremer, and V. Pascucci. Distortion-guided structure-driven interactive exploration of high-dimensional data. *Computer Graphics Forum*, 33(3):101–110, 2014.
- [8] J. Matute, A. C. Telea, and L. Linsen. Skeleton-based scagnostics. *IEEE Transaction on Computer Graphics*, 24(1):542–552, 2017.
- [9] N. Morishita, M. Hagita, H. Shioya, and T. Itoh. Graph coloring algorithms for photo selection. In *Conference on Applied Mathematics (in Japanese)*, pages 106–109, 2016.
- [10] A. Nakabayashi and T. Itoh. A technique for selection and drawing of scatterplots for multi-dimensional data visualization. In *23rd International Conference on Information Visualisation (IV2019)*, pages 62–67, 2019.
- [11] K. Nohno, H.-Y. Wu, K. Watanabe, S. Takahashi, and I. Fujishiro. Spectral-based contractible parallel coordinates. In *18th International Conference on Information Visualisation*, pages 7–12, 2014.
- [12] M. Sedlmair, A. Tatu, T. Munzner, and M. Tory. A taxonomy of visual cluster separation factors. *Computer Graphics Forum*, 31(3):1335–1344, 2012.
- [13] L. Shao, A. Mahajan, T. Schreck, and D. J. Lehmann. Interactive regression lens for exploring scatter plots. *Computer Graphics Forum*, 36(3):157–166, 2017.
- [14] M. Sips, B. Neubert, J. P. Lewis, and P. Hanrahan. Selecting good views of high-dimensional data using class consistency. *Computer Graphics Forum*, 28(3):831–838, 2009.
- [15] H. Suematsu, Y. Zheng, T. Itoh, R. Fujimaki, S. Morinaga, and Y. Kawahara. Arrangement of low-dimensional parallel coordinate plots for high-dimensional data visualization. In *17th International Conference on Information Visualization*, pages 59–65, 2013.
- [16] Y. Wang, Z. Wang, T. Liu, M. Correll, Z. Cheng, O. Deussen, and M. Sedlmair. Improving the robustness of scagnostics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):759–769, 2020.
- [17] A. Watanabe, T. Itoh, M. Kanazaki, and K. Chiba. A scatterplots selection technique for multi-dimensional data visualization combining with parallel coordinate plots. In *21st International Conference on Information Visualisation (IV2017)*, pages 78–83, 2017.
- [18] L. Wilkinson, A. Anand, and R. Grossman. Graph-theoretic scagnostics. In *IEEE Symposium on Information Visualization*, pages 157–164, 2005.
- [19] X. Yuan, D. Ren, Z. Wang, and C. Guo. Dimension projection matrix/tree: Interactive subspace visual exploration and analysis of high dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2625–2633, 2013.
- [20] Z. Zhang, K. T. McDonnell, E. Zadach, and K. Muller. Visual correlation analysis of numerical and categorical data on the correlation map. *IEEE Transactions on Visualization and Computer Graphics*, 21(2):289–303, 2015.
- [21] Z. Zhang, K. T. McDonnell, and K. Mueller. A network-based interface for the exploration of high-dimensional data spaces. In *IEEE Pacific Visualization Symposium 2012*, pages 17–24, 2012.
- [22] Y. Zheng, H. Suematsu, T. Itoh, R. Fujimaki, S. Morinaga, and Y. Kawahara. Scatterplot layout for high-dimensional data visualization. *Journal of Visualization*, 18(1):111–119, 2015.