DEIM Forum 2021 G11-3

多次元データ可視化のための重要度・類似度にもとづく散布図の選択・描画手法の評価実験

中林明日香† 伊藤 貴之†

† お茶の水女子大学大学院人間文化創成科学研究科 〒 112–8610 東京都文京区大塚 2–1–1

E-mail: †{g1520533,itot}@is.ocha.ac.jp

あらまし 多次元データの可視化手法としてよく知られるものに散布図行列や平行座標法などがあるが，これらの手法では膨大な次元数を有するデータにおいて非常に大きな画面空間を必要とする問題点がある．この問題を解決するための一手段として我々は，多次元データ中の任意の 2 変数を 2 軸とする散布図の中から限られた数の散布図を選出し，さらにその散布図を「例外点群」および「例外でない点群の包括領域」の 2 種類に分けて描画する手法を提案している．本報告では散布図の選出方法として，より重要度が高いと思われるものから選出するか類似していないと思われるいくつかの散布図を選出する手法を紹介する．本報告では，小売店の気象と売上の関係のデータを適用した可視化の評価実験を行い，本手法の有用性を示す．

キーワード 多次元データ，可視化手法，散布図，次元選択，例外抽出

1 はじめに

多次元データは我々の日常生活や専門業務に幅広く存在している．このような身近な存在である多次元データから重要な知識を得るための方法論として，データの特徴や規則性を観察することがあげられる．その観察手段として多次元データの可視化が有効であると考えられる．

多次元データの可視化の代表的な手法に，散布図行列や平行座標法 (Parallel Coordinate Plots; 以下 PCP) があげられる．多次元データを構成する次元数を m としたときに，散布図行列は図 1(左) のようにデータから生成しうる全ての 2 次元ペアの散布図を作成し $m \times m$ の格子状に並べることで表現し，平行座標法は図 1(右) のように m 本の平行な座標軸に変数の値をプロットしそれを折れ線で結ぶことで表現する．これらの手法は多次元データを構成する全ての次元を可視化するものであるため，膨大な次元数を有するデータにおいては非常に大きな画面空間を必要とする問題点がある．また一方で，多次元データの全ての次元に興味深い特徴や規則性が見られるとは限らない．言い換えれば，多次元データを構成する全ての次元を可視化するより，多次元データの中から興味深い特徴や規則性を有する次元だけを事前選択して可視化したほうが，効率よくデータを観察できる場合もある．そこで近年では，多次元データから可視化する意義の高い次元だけを選択して表示する手法が多く提案されている．

我々は次元選択手法を搭載した多次元データ可視化手法として，低次元 PCP の集合で多次元データを可視化する Hidden [1] を提案した．さらに我々は，低次元 PCP の代わりに選択的な散布図集合による多次元データの可視化手法 [2] [3] を提案し，様々なデータに対し本手法を適用してきた [4]．この手法は具体的には以下の 2 つの処理工程から構成されるものである．

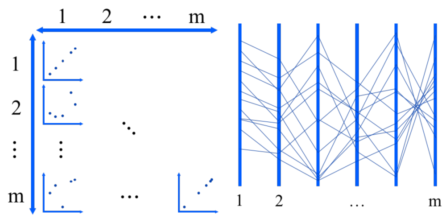

図 1 (左) 散布図行列の例，(右)PCP の例

- 多次元データ中の任意の 2 変数を 2 軸とする散布図の中から重要ないくつかを，または類似していないいくつかの散布図を，単純かつ対話的なスライダー操作によって選出する．
- 散布図に表示される点群を「例外点群」および「例外でない点群の包括領域」の 2 種類であるとして描画する．

本手法は複数の説明変数と目的関数を持ち，カテゴリ変数によってラベル付けされているようなデータに対しての適用が有効である．このような多次元データ可視化手法の有効活用が期待される場面の例として，多次元データのモデル化があげられる．多次元データを構成する数値群の中にどのようなノイズや例外値が含まれているかを理解し，適切なスクリーニング処理によってこれらを除去したのちに，どのようなモデルを適用できるかを検討する処理が必要となる場面が多い．例えば機械学習の訓練データに多次元データを利用する際に，このような工程が重要な意味を持つことが多い．このような工程にも多次元データの可視化手法が貢献できる可能性が期待される．

一方で，我々のこれまでの発表 [2] [4] は可視化の実行例を示したにとどまり，その有効性はまだ検証されていなかった．そこで我々は，これまで提案してきた散布図の選択・描画手法に対するユーザ評価実験を実施した．本報告ではその実験結果を示し，提案手法の有効性について議論する．

本報告の構成は以下の通りである．2 章では関連研究につい

て，3 章では提案手法について述べる．そして 4 章で本手法の実行結果と考察について，5 章で評価実験の内容と結果について，6 章で本報告のまとめと今後の課題について述べる．

2 関 連 研 究

2.1 次元選択を用いた多次元データの可視化

多次元データの中から重要な部分だけを可視化するためのアプローチとして，可視化する意義の高い低次元部分空間を事前に抽出する手法は従来から数多く提案されている．例として，多次元データから所定の基準を満たす複数の 2 次元ペアの散布図を生成し，各散布図間の類似度距離に基づいて配置する手法 [5] や，所定の基準を満たす次元間の低次元 PCP を生成し，各 PCP 間の次元の共有率から算出される類似度距離に基づいて配置する手法 [6] などがある．しかしこれらの手法による可視化結果は固定的なものであり，PCP や散布図の表示数を対話的に調節することができなかった．

この問題点を解決する多次元データ可視化手法として Itoh らは Hidden [1] を発表した．Hidden [1] は画面右部の次元散布図上を対話的に操作することによって選択される低次元部分空間群を，画面左部で複数の PCP によって表示する．図 2 に Hidden による可視化の例を示す．また Hidden を拡張し PCP と散布図を併用して可視化した手法 [7] も発表されている．この手法では原則として PCP で多次元データを可視化しつつ，PCP では視認しにくい数値分布を有する 2 軸のみに対して散布図を適用して表示する．

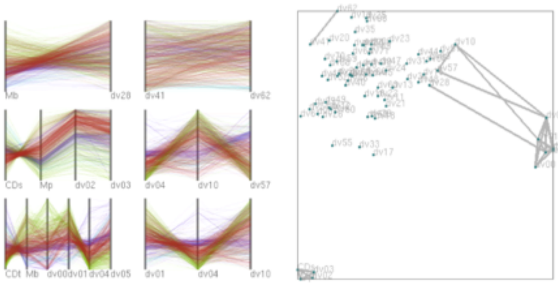

図 2 Hidden による可視化画面 [1]

2.2 散布図集合による多次元データの可視化

散布図を用いて多次元データを可視化する手法として，Wilkinson らの手法 [8] や Dang らの手法 [9] が挙げられる．Wilkinson らの手法では，多次元データの各 2 次元ペアから生成される散布図の形状などから Scagnostics と呼ばれる 9 種類の特徴を定量評価し，特定の傾向を持った散布図を生成する 2 変数を推薦する．Dang らの手法では，散布図行列から Scagnostics の基準により各散布図を特徴づけてクラスタリングし，リーダーとなる散布図を選出して類似するものを近くに配置する．これらの手法では，膨大な次元数の多次元データでも非常に大きい画面空間を使うことなく可視化することができる．我々の手法では 3 章にて後述する基準で散布図を選出しているが，Scagnostics を適用することも可能である．

2.3 密度を考慮した散布図による可視化

散布図を塗りつぶすように表現する手法の例として，Continuous Scatterplots [10] は，n 次元の入力データセット上で定義された任意の密度を考慮して，この密度値を次元散布図にマッピングする手法である．この手法は散布図のような統計的可視化と流体の可視化のような科学的可視化を組み合わせたものである．我々の手法も散布図を連続的に描画するものであるが，密度を直接考慮して散布図を描画するわけではない．

2.4 点群を囲むように描画する可視化

Schreck らは 2 次元空間上の点群を囲むように描画する手法を提案している [11] [12]．これらの手法では，多次元データを主成分分析などの手法により 2 次元空間で表現し，クラスごとにデータ点を凸包やよりコンパクトな形状で囲むように描画することによって，各クラスの分布を視認しやすくしている．また Mayorga らが提案した Splatterplot [13] では，点の密な領域を塗りつぶした輪郭として描画し，疎な領域は点をサンプリングして描画している．これらの従来手法は囲み領域のみ，あるいは囲み領域とサンプリングされたいくつかの点のみを描画するものであり，全てのデータ点を描画するものではない．本手法は全てのデータ点を描画しつつも例外点や囲み表示の描画を実現するものであり，従来手法に比べデータ点の情報を見落としにくい．

3 選択的な散布図集合による多次元データの可視化

本章では我々が提案した多次元データ可視化手法 [2] [3] について紹介する．1 章でも述べた通り，この手法では，多次元データ中の任意の 2 変数を 2 軸とする散布図の中から重要なもの，あるいは類似していない散布図の組み合わせを選出し，さらにその散布図を構成する点群を「例外点群」および「例外でない点群の包括領域」の 2 種類であるとして描画する．

3.1 重要度を考慮した散布図の選出手法

現時点での我々の実装では，相関係数またはエントロピーまたは点群領域の細長さによる基準を採用している．

相関係数による基準を適用した際には，各散布図を生成する 2 次元間の相関係数を計算し，その絶対値の大きい散布図を優先的に表示する．任意の 2 次元 d_1, d_2 を与えられたとき，d_1, d_2 間の相関係数の絶対値 d_{d_1,d_2} はスピアマンの順位相関係数 $f_c(d_1, d_2)$ を用いて以下の式のように表せる．

$$d_{d_1,d_2} = |1.0 - f_c(d_1, d_2)| \tag{1}$$

エントロピーによる基準を適用した際には，多次元データ中のカテゴリ型変数が各個体のラベルに相当するとみなして，点群がラベルごとによく分離されている散布図を優先的に表示する．具体的には以下の式より各次元ペアのエントロピー $H(d_1, d_2)$ を算出する．

$$H(d_1,d_2) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} p(y_i = c|x_i^{d_1,d_2}) \log p(y_i = c|x_i^{d_1,d_2}) \quad (2)$$

ここで入力データセット Ds を $Ds = \{x_1, \ldots, x_N\}$ としたとき，x_i は i 番目のプロット，y_i は i 番目のプロットに割り当てられたラベルを表し，N と C はそれぞれプロットの数とラベルの数，$p(y_i = c|x_i^{d_1,d_2})$ は c 番目のクラス Y_c が i 番目のプロット x_i に割り当てられている確率を表す．また $x_i^{d_1,d_2}$ はプロット x_i について次元 d_1, d_2 を取り出してできる 2 次元ベクトルを表す．この式によって算出されるエントロピー H は，次元 d_1 と d_2 によって生成される散布図において，クラスラベルが全体的にどの程度分離されているかを示すものである．

点群領域の細長さによる基準を適用した際には，Wilkinson ら [8] の実装に準じて各散布図の基準値を算出する．具体的には 3.3 節で後述するように，散布図の点群を全て連結するような Delaunay 三角メッシュを生成し，ユーザ指定の閾値を超える長さの辺を削除するという操作を適用した後に，以下の値 S_{d_1,d_2} を算出する．

$$S_{d_1,d_2} = 1 - \frac{\sqrt{4\pi A_{rea}(T_{d_1,d_2})}}{P_{erimeter}(T_{d_1,d_2})} \quad (3)$$

ここで T_{d_1,d_2} は，次元 d_1, d_2 間の散布図の点群を連結する三角メッシュを表す．また，$A_{rea}(T_{d_1,d_2})$ は三角メッシュT_{d_1,d_2} の総面積を，$P_{erimeter}(T_{d_1,d_2})$ は三角メッシュT_{d_1,d_2} の外周となる辺の長さの合計値を表す．これにより，点群領域の面積に対し点群領域の外周長が長くなる散布図を，つまり点群が細長く分布する領域を有する散布図を優先的に表示する．

3.2 類似度を考慮した散布図の選出手法

3.1 節では 3 つの基準で重要度を算出し，重要度の高いものから散布図を選出する手法を紹介したが，これでは似た傾向の見られる散布図ばかり表示してしまうこともある．そこで前述した 3 つの基準を利用して散布図間の類似度を算出し，これにもとづいて多様な組み合わせの散布図を選出する [3]．具体的には，各散布図の 3 つの基準値 d_{d_1,d_2}, $H(d_1,d_2)$, S_{d_1,d_2} を持つ 3 次元ベクトル $\{d_{d_1,d_2}, H(d_1,d_2), S_{d_1,d_2}\}$ を生成し，各 3 次元ベクトル間のコサイン類似度を算出する．そして，ユーザ指定の閾値よりもコサイン類似度が高い散布図群を同時に選出しないという制約を設けることで，多様な組み合わせの散布図を選出する．

3.3 点群領域の囲み表示と例外点の描画手法

現時点での我々の実装では，「例外点群の抽出」および「例外でない点群の包括領域の生成」に Delaunay 三角分割法を利用している．Delaunay 三角分割法は与えられた点群を連結して三角メッシュを生成する手法であり，三角メッシュを構成する三角形の最小角度が最大になるように三角メッシュを生成するものである．本手法では，各散布図に対して，散布図中の全ての点群を包括する大きな四角形を作成し三角形に分割し，散布図中の点群を 1 つずつ追加して頂点として連結していくことで三角メッシュを逐次的に更新し，全ての点群を追加したら最初に作成した大きな四角形とその頂点に連結される辺を削除する，というインクリメンタルなアルゴリズムを採用している．

このような処理によって生成された三角メッシュから，ユーザ指定の閾値を超える長さの辺を削除することで，図 3 のようにどの点群とも連結されていない点を例外点として抽出する．ユーザによる対話操作で閾値を調節することで，例外点と判定された点の数を調節できる．そして，例外点以外の点で構成される三角形群の領域境界を構成する辺のみを濃い色で描画し，三角メッシュを薄い色で塗りつぶすことによって，点群の包括領域を表示する．

図 3 散布図を「例外点群」と「例外でない点群の包括領域」の 2 種類として描画した例

また，図 4 のようにユーザが入力データセットから任意の点を削除し散布図を再描画することもできる．図の左画像の右上の例外を削除すると，右画像のように縦軸の縮尺が更新され，例外でない点群の領域も再描画される．このように離れた位置にある点を削除することで例外でない点群の領域を拡大表示できる．これにより，それまでは発見しにくかった新しい特徴や傾向，あるいは新しく例外となり得る点を発見できることが期待される．

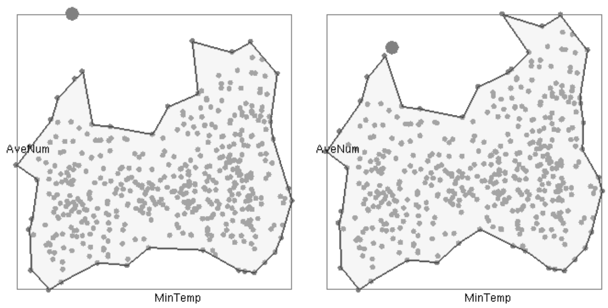

図 4 例外削除の例 (左は例外削除前，右は例外削除後)

4 実 行 例

本報告では，2016 年 5 月 1 日から 2017 年 7 月 31 日までの 457 日間のアパレルの小売店における各日の来客数や売上と，その各日の気象値との関係のデータを題材にして，本手法を用いた可視化の実行例を示す．データ中の説明変数 (気象の数値・横軸) と目的関数 (売上の数値・縦軸) の対応表を表 1 に示す．

表 1　データの説明変数と目的関数の対応表

説明変数 (気象数値)		目的関数 (売上数値)	
MinTemp	最低気温	Revenue	売上
MaxTemp	最大気温	Guest1	購入人数
SumRain	降水量	Guest2	来客人数
SumSnow	降雪量	Ratio	買上率
SumSnowC	積雪量	PerGuest	客単価
SumSunTime	日照時間	AveUnit	平均買上商品単価
MaxWind	最大風速	AveNum	平均買上点数

なお本章で用いるデータは現実のデータに乱数を加算したものであり，現実の数値をそのまま可視化したわけではない点に注意されたい．

図 13(左) はデータを散布図行列で可視化したものであり，図 13(右) は本手法の相関係数による基準を適用した可視化の例である．右下がりに点が分布しているような負の相関が見られる散布図が優先的に選出されていることがわかる．

図 14(左) はデータを散布図行列で可視化したものであり，図 14(右) は本手法のエントロピーによる基準を適用した可視化の例である．カテゴリ変数は平日か否かを用いており，水色は平日，赤色は休日を示している．赤色の点と水色の点が上下によく分離しているような散布図が優先的に選出されていることがわかる．

図 15(左) はデータを散布図行列で可視化したものであり，図 15(右) は本手法の点群領域の細長さによる基準を適用した可視化の例である．点群領域が細長く分布しているような散布図が優先的に選出されていることがわかる．

図 16(左) はデータを散布図行列で可視化したものであり，図 16(右) は本手法の類似度による基準を適用した可視化の例である．カテゴリ変数は平日か否かを用いており，水色は平日，赤色は休日を示している．多様な形の散布図が選出されていることがわかる．

5　評 価 実 験

本章では本手法を用いた可視化の有用性を示すための評価実験について述べる．情報科学を専攻する 20 代女性の被験者 8 名に，説明資料を配布して本手法を用いた可視化ツールの使い方と評価実験の方法を理解してもらった後，散布図行列と本手法を用いた可視化のプログラムをダウンロードしてもらった．その後各自の所有する PC を使って散布図行列と本手法を用いた可視化の両方でデータを観察してもらい，そこから読み取れる情報について回答してもらった．評価項目は以下の 5 つである．

（1） 負の相関が見られる散布図はどれか．

（2） 平日・休日のカテゴリ変数を選択した際，散布図中の点がラベルごとに分離している傾向がある目的関数はどれか．

（3） 月ごとのカテゴリ変数を選択した際，2 月・7 月のデータがそれぞれ 2 つずつクラスタを作っているような散布図はどれか．

（4） 外れ値となるような点はどれか．

（5） 形や傾向が類似していると感じる散布図の組み合わせはどれか．

また (1)〜(3) については散布図行列と本手法を用いた可視化のどちらが使いやすかったか，(4) についてはどの機能を使用したかを回答させた．比較のために，被験者のうち半数の 4 人には，散布図行列を使用して (1)〜(3) の問いに回答してもらってから，本手法を用いた可視化を使用して (1)〜(3) の問いに回答してもらった．一方で残りの 4 人には，本手法を用いた可視化を使用して (1)〜(3) の問いに回答してもらってから，散布図行列を使用して (1)〜(3) の問いに回答してもらった．その後に，使用するツールに制限のない状態で (4) と (5) の問いに回答してもらった．問いに回答する際の制限時間は設けなかった．表 2 に (1)〜(3) の正答率と使いやすさについての回答をまとめる．(3) については正解だと想定していた散布図以外にも該当すると回答された散布図が多くあったため，正答・不正答は定めていない．

表 2　評価実験の結果 ((1)〜(3) の正答率と使いやすさ)

		(1)	(2)	(3)
散布図行列	正答者	6 人	4 人	-
	不正答者	2 人	4 人	-
	未回答者	0 人	0 人	-
本手法を用いた可視化	正答者	5 人	6 人	-
	不正答者	3 人	1 人	-
	未回答者	0 人	1 人	-
散布図行列の方が使いやすかった		1 人	0 人	1 人
どちらかというと散布図行列の方が使いやすかった		1 人	1 人	0 人
どちらも同じくらいの使用感だった		0 人	3 人	2 人
どちらかというと本手法を用いた可視化の方が使いやすかった		1 人	3 人	1 人
本手法を用いた可視化の方が使いやすかった		5 人	1 人	4 人

「負の相関が見られる散布図」については，図 5 の 4 つの散布図を正解とした．散布図行列を使用した場合も本手法を用いた可視化を使用した場合も正答率はほぼ変わらなかったが，本手法を用いた可視化の場合は囲み表示を描画した場合に図 6 のような散布図も負の相関があると認識させてしまう場合もあった．表 2 にも示されているように，本手法を用いた可視化の方が使いやすかったと回答する人が多かった，

「平日・休日のカテゴリ変数を選択した際の，散布図中の点がラベルごとに分離している傾向がある目的関数」については，図 7 のような散布図を生成する Revenue, Guest1, Guest2 の 3 つの目的関数を正解とした．散布図行列を使用した場合よりも本手法を用いた可視化を使用した場合の方が正答率は高かったが，使用感はどちらも同じくらいであったとの回答が多かった．

「月ごとのカテゴリ変数を選択した際の，2 月・7 月のデータがそれぞれ 2 つずつクラスタを作っているような散布図」については，図 8 の散布図を正解とした．しかし回答を集めてみると，図 9 のような散布図も該当するという意見が多かった．

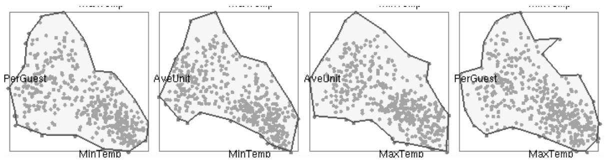

図 5　負の相関が見られる散布図の例

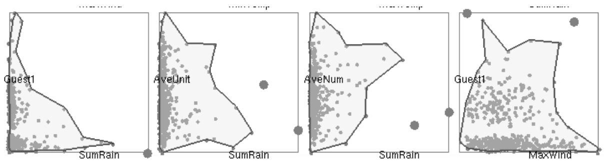

図 6　負の相関があると誤認させてしまう散布図の例

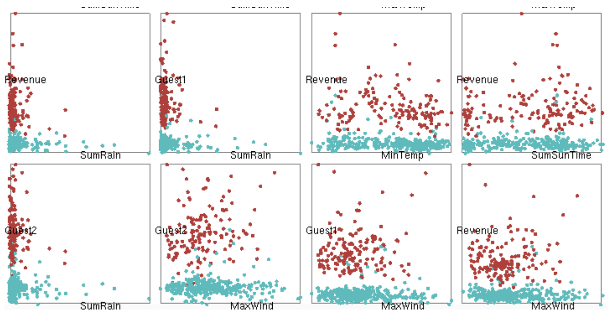

図 7　点がラベルごとに分離している傾向がある散布図の例

本手法を用いた可視化の方が該当する散布図を見つけやすくなるが，囲み表示によって図 10 のように描画されたときにクラスタが分離していないと認識させる場合も多かった．

「外れ値となるような点」については，同じ点を外れ値であると答える人は少なく，人によって外れ値であるとみなすものは異なることがわかった．被験者 8 人中 6 人が本手法を用いた可視化の点群領域の囲み表示と例外点の描画機能を使用したため，この機能が外れ値を探索する助けになると考えられる．

「形や傾向が類似していると感じる散布図の組み合わせ」については，図 11 のような散布図の組み合わせが類似していると回答され，本手法においても類似しているとみなされた．一方で，図 12 のような組み合わせも類似していると回答されたが，本手法においては類似しているとみなされなかった．

以上の評価実験によって，相関の見られる散布図やクラスタを作るように分布した散布図といった，特徴が見られる散布図を探索するためには，本手法を用いた可視化はある程度の有用性があるということがわかった．また散布図中の例外点を探索するためにも本手法は有用であることがわかった．一方で，本手法における点群領域の囲み表示と例外点の描画機能は，散布図の点群の分布によってはユーザに誤認を与えてしまうこともあった．これは，本手法の現時点の実装では散布図ごとに点群の密度が異なることを考慮していないため，密度の薄い領域の三角形が削除されないことがあり，必要以上に広い領域が塗りつぶされてしまうことがあるからである．また，人間が類似していると知覚できる散布図の組み合わせと，本手法で類似しているとみなす散布図の組み合わせが必ずしも一致するとは限らないことがわかった．これらの点について改良していく必要が

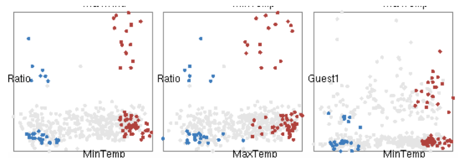

図 8　2 月・7 月のデータがそれぞれ 2 つずつクラスタを作っているような散布図の例

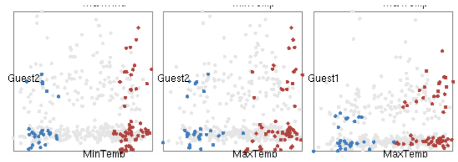

図 9　2 月・7 月のデータがそれぞれ 2 つずつクラスタを作っていると回答された散布図の例

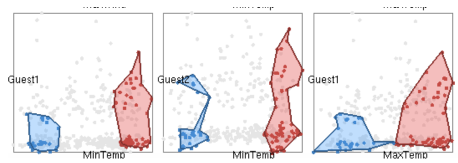

図 10　2 つのクラスタに分離していないと誤認させる散布図の例

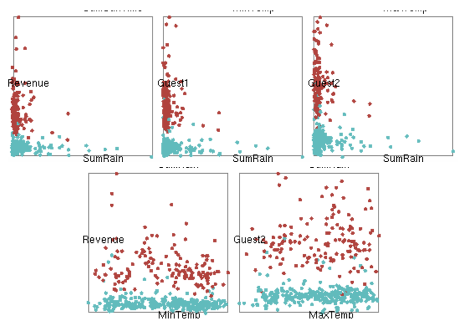

図 11　類似している散布図の組み合わせの例 (上の 3 つ，下の 2 つがそれぞれ類似している)

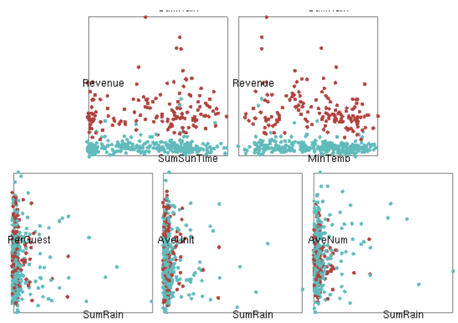

図 12　類似していると知覚できる散布図の組み合わせの例 (上の 2 つ，下の 3 つがそれぞれ類似している)

ある．

6　まとめと今後の課題

本報告では，我々がこれまで提案してきた散布図の選択・描画手法に対するユーザ評価実験を実施し，提案手法の有効性について議論した．我々の可視化手法では，多次元データ中の任

意の２変数を２軸とする散布図の中から重要なもの，あるいは類似していない散布図の組み合わせを選出し，さらにその散布図を構成する点群を「例外点群」および「例外でない点群の包括領域」の２種類であるとして描画する手法を提案している．我々の可視化手法は，例外点をデータから削除するか否かの判断や，例外でない点群のモデル化手法の検討などに有用であると考えられる．

今後の課題として，まず点群領域の囲み表示と例外点の描画機能において，削除する辺の閾値を全ての散布図で一定にするのではなく，散布図ごとに調節できるようにしたい．現在の実装では散布図ごとに点群の密度が異なることを考慮していないため，密度の薄い領域にて三角形が削除されないことがあり，これによって必要以上に広い領域が塗りつぶされることがある．この現象がユーザに誤認を与える可能性があることが評価実験からわかった．そこで，点群の密度に応じて削除する辺の閾値を散布図ごとに自動的に調節できるようにしたい．

また，散布図の選出基準についても引き続き検討したい．評価実験の結果から人間が類似していると知覚できる散布図の組み合わせと，本手法で類似しているとみなす散布図の組み合わせが一致するわけではないことがわかった．そこで，人間が類似していると知覚できる散布図の組み合わせが選出できるように，類似度の基準について再考したい．

そしてこれらの機能を実装した後に，さらに大きな次元を持つデータセットを本手法に適用し，さらに汎用性に富んだ実装になるように開発を進めたい．

謝　　辞

小売店の気象と売上に関するデータセットを提供して頂いた株式会社 ABEJA 様に感謝いたします．

文　　献

[1] Takayuki Itoh, Ashnil Kumar, Karsten Klein, and Jinman Kim, “ High-Dimensional Data Visualization by Interactive Construction of Low-Dimensional Parallel Coordinate Plots,” Journal of Visual Languages and Computing, Vol. 43, pp. 1–13, 2017.

[2] Asuka Nakabayashi, Takayuki Itoh, “ A Technique for Selection and Drawing of Scatterplots for Multi-Dimensional Data Visualization,” Proceedings of 23rd International Conference on Information Visualisation (IV2019), pp. 62–67, 2019.

[3] 伊藤貴之, 中林明日香, 萩田真理子, “ 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,” 第 28 回インタラクティブシステムとソフトウェアに関するワークショップ (WISS), 2020.

[4] 中林明日香, 伊藤貴之, “ 多次元データの散布図表示のための次元選択手法の応用例,” 第 12 回データ工学と情報マネジメントに関するフォーラム (DEIM), A3–4, 2020.

[5] Yunzhu Zheng, Haruka Suematsu, Takayuki Itoh, Ryohei Fujimaki, Satoshi Morinaga, and Yoshinobu Kawahara, “ Scatterplot layout for high-dimensional data visualization,” Journal of Visualization, Vol. 18, No. 1, pp. 111–119, 2015.

[6] Haruka Suematsu, Yunzhu Zheng, Takayuki Itoh, Ryohei Fujimaki, Satoshi Morinaga, and Yoshinobu Kawahara, “ Arrangement of Low-Dimensional Parallel Coordinate Plots for High-Dimensional Data Visualization,” Proceedings of 17th International Conference on Information Visualisation (IV2013), pp. 59–65, 2013.

[7] Ayaka Watanabe, Takayuki Itoh, Masahiro Kanazaki, and Kazuhisa Chiba, “ A Scatterplots Selection Technique for Multi-Dimensional Data Visualization Combining with Parallel Coordinate Plots,” Proceedings of 21st International Conference on Information Visualisation (IV2017), pp. 78–83, 2017.

[8] Leland Wilkinson, Anushka Anand, and Robert Grossman, “ Graph-Theoretic Scagnostics,” Proceedings of IEEE Symposium on Information Visualization, pp. 157–164, 2005.

[9] Tuan Nhon Dang and Leland Wilkinson, “ ScagExplorer: Exploring Scatterplots by Their Scagnostics,” Proceedings of IEEE Pacific Visualization Symposium (PacificVis 2014), pp. 73–80, 2014.

[10] Sven Bachthaler and Daniel Weiskopf, “ Continuous Scatterplots,” IEEE Transactions on Visualization and Computer Graphics, Vol. 14, No. 6, pp. 1428–1435, 2008.

[11] Tobias Schreck and Christian Panse, “ A new metaphor for projection-based visual analysis and data exploration,” Proceedings of IS&T/SPIE Conference on Visualization and Data Analysis, 2007.

[12] Tobias Schreck, Michael Schüßler, Katja Worm, and Frank Zeilfelder, “ Butterfly plots for visual analysis of large point cloud data,” Proceedings of the 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG’08), pp. 33–40, 2008.

[13] Adrian Mayorga and Michael Gleicher, “ Splatterplots: Overcoming overdraw in scatter plots,” IEEE Transactions on Visualization and Computer Graphics, Vol. 19, No. 9, pp. 1526–1538, 2013.

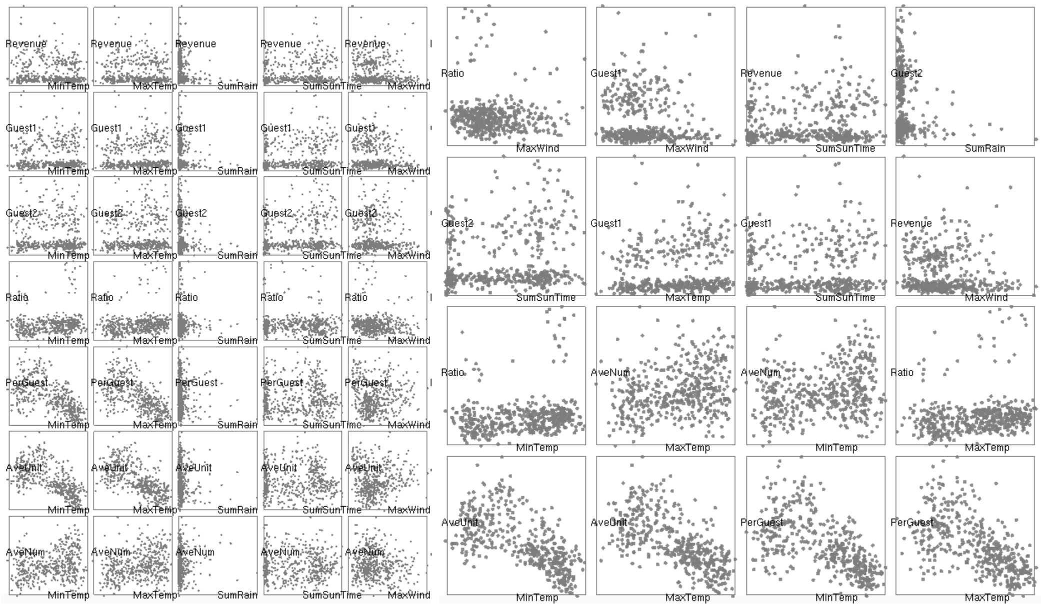

図 13　(左) 散布図行列による可視化，(右) 本手法の相関係数による基準を適用した可視化

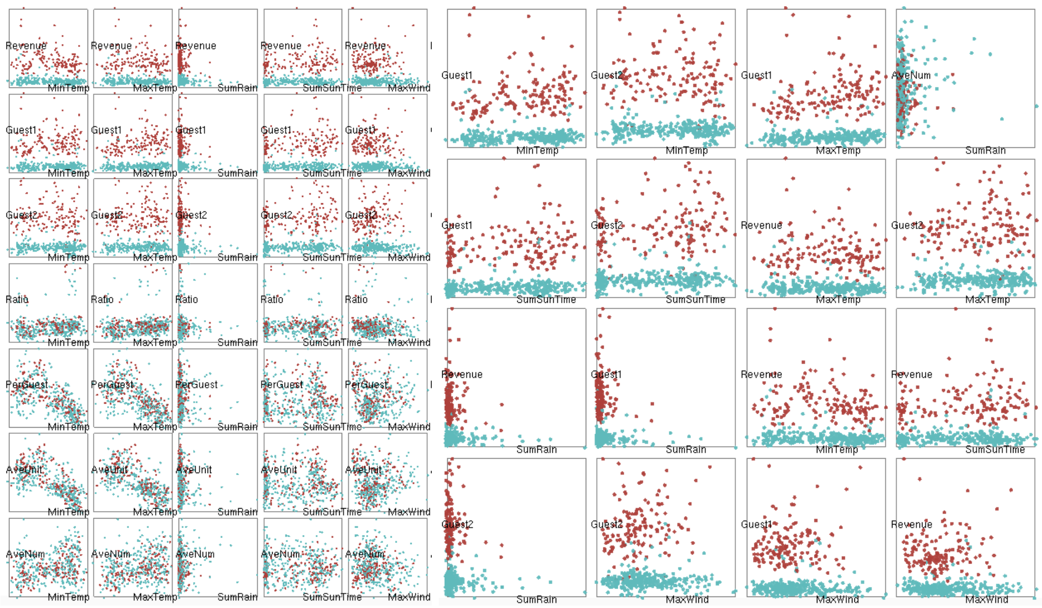

図 14　(左) 散布図行列による可視化，(右) 本手法のエントロピーによる基準を適用した可視化

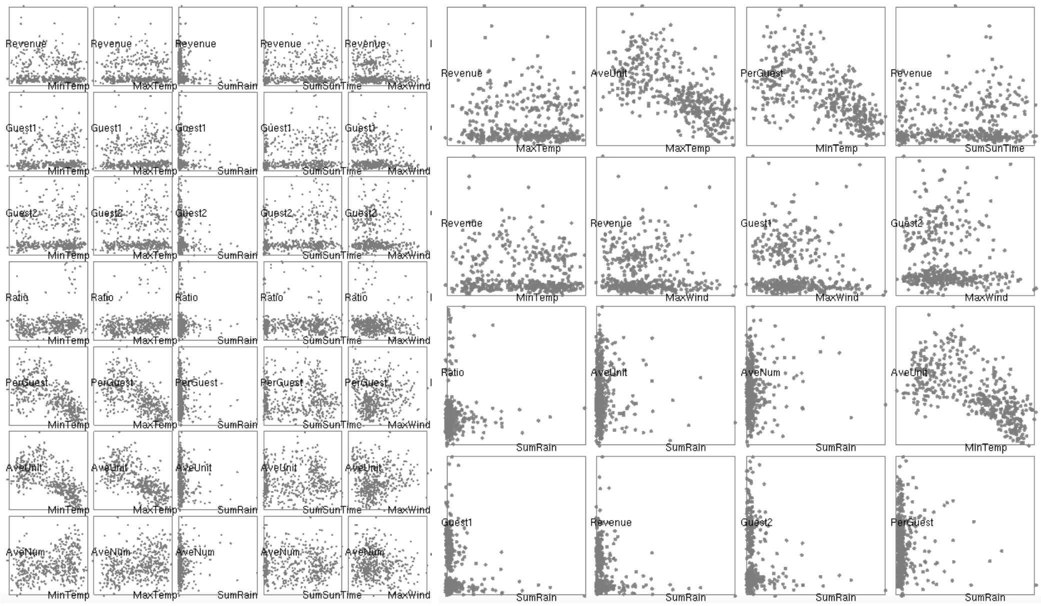

図 15　(左) 散布図行列による可視化，(右) 本手法の点群領域の細長さによる基準を適用した可視化

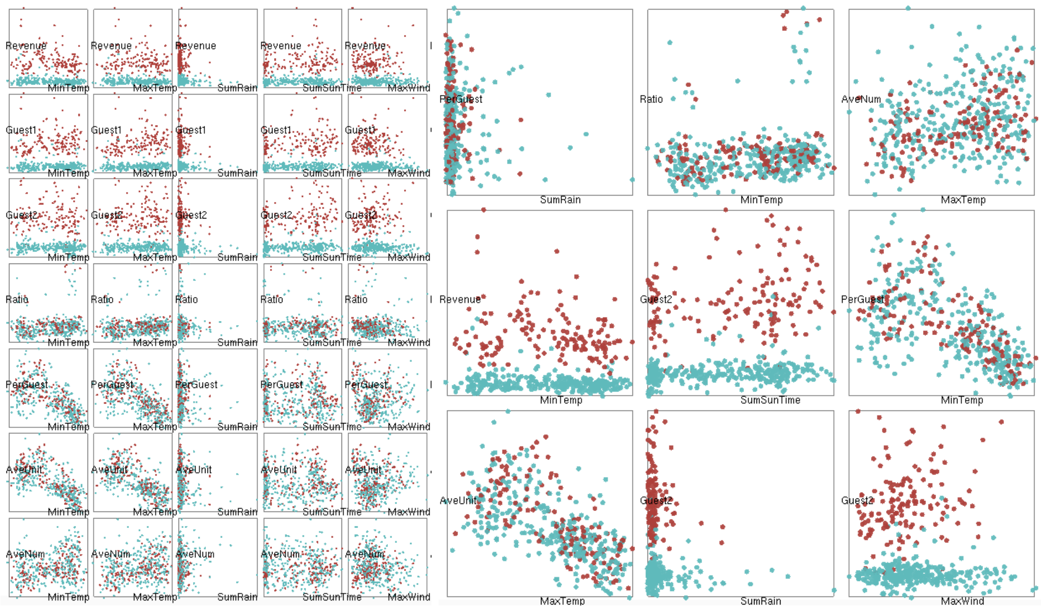

図 16　(左) 散布図行列による可視化，(右) 本手法の類似度による基準を適用した可視化