

多変量データをSPJ質問により統合する 平行座標プロット型情報可視化システムと操作言語

濱崎 裕太^{1,†1,a)} 植村 智明^{2,b)} 富井 尚志^{3,c)}

受付日 2019年3月8日, 採録日 2019年7月2日

概要: センサ・ストレージ技術の発達により, 多くの属性を含む多変量データをすべて蓄積可能になった。それにともない, 多変量データを分析することで, 新たな知見を獲得可能となった。しかしながら, 多変量データの分析は一般に困難であり, 可視化システムによる支援が不可欠である。これに対し本研究では, 可視化手法の1つである平行座標プロット (PCP: Parallel Coordinates Plot) に対し, PCP の操作過程を保存・再現する操作言語を組み込んだ情報可視化システムを提案する。本言語は, 関係代数における SPJ (Selection-Projection-Join) 質問に相当する記述が可能かつ SQL に類似した言語として, 本論文で提案するものである。これらにより, PCP 上での操作と言語を相互変換し, 作業経過を言語として保存することで, データ分析の試行錯誤の支援を実現する。また, 服飾文化財と自動車走行ログという異なる2つのコンテンツに対して可視化を行うことにより, 提案システムの汎用性・有用性を示す。

キーワード: 情報可視化, 平行座標プロット, 多変量データ分析, SPJ 質問

A Visualization System of Parallel Coordinates Plot and an Operation Language Integrating Multivariate Data by SPJ Query

YUTA HAMASAKI^{1,†1,a)} TOSHIKI UEMURA^{2,b)} TAKASHI TOMII^{3,c)}

Received: March 8, 2019, Accepted: July 2, 2019

Abstract: Large amounts of multivariate data can be stored in data storage through the development of sensor and storage technology. Consequently, it is possible to obtain new findings by analyzing multivariate data. However, multivariate data analysis is generally difficult, so support for a visualization system is essential. In this paper, we present new visualization system of parallel coordinates plot (PCP) embedded a feature of snapshotting operation histories in language form. Our proposed language is like SQL, and it can represent SPJ (Selection-Projection-Join) by relational algebra. These make analysts possible to interact both operation and language on PCP, consequently by using proposed system, the users can analyze multivariate data heuristically. In addition, we show the versatility and usefulness of the proposed system by visualizing two different contents of clothing cultural properties and moving logs of cars.

Keywords: information visualization, parallel coordinates plot, multivariate data analysis, SPJ query

¹ 横浜国立大学大学院環境情報学府情報メディア環境学専攻
Department of Information Media and Environment Sciences, Graduate School of Environment and Information Sciences, Yokohama National University, Yokohama, Kanagawa 240-8501, Japan

² 横浜国立大学大学院環境情報学府情報環境専攻
Department of Information Environment, Graduate School of Environment and Information Sciences, Yokohama National University, Yokohama, Kanagawa 240-8501, Japan

³ 横浜国立大学大学院環境情報研究院
Faculty of Environment and Information Sciences, Yokohama National University, Yokohama, Kanagawa 240-8501, Japan

^{†1} 現在, ドコモ・システムズ株式会社
Presently with DOCOMO Systems, Inc.

a) yuuta.hamasaki.yc@nttdocomo.com

b) uemura-toshiaki-xh@ynu.jp

c) tommy@ynu.ac.jp

1. はじめに

センサ・ストレージ技術の発達やオープンデータの公開により、多くの属性を含むデータをすべて蓄積可能になった。それにともない、複数の属性からなる多変量データを分析することで、新たな知見を獲得可能となった。総務省の「情報通信白書」によれば、世界のデータトラフィックは2020年には1カ月あたり228エクサバイトに達すると予想されており、それらのデータ利活用に向けた取り組みの重要性が述べられている [1]。このとき、提供元が異なる複数のデータどうしを結合することによって、新たな知見が産まれることがある。すなわち、複数の多変量データどうしを結合・選択し、試行錯誤しながらデータ分析を行うことは一般的に有用である。

その具体的な手段として、多変量データを n -項関係として解釈し、SPJ (Selection-Projection-Join) 質問 [2] のような関係代数演算を適用する方法が有効であると考えられる。しかしながら、この方法によって出力される分析結果は数値組のリストとして表されることから、データの分布や属性どうしの相関などのデータの特徴を把握することや、分析者が望む適切な演算を行うことができたかを把握することは困難である。

この問題点に対し、我々は多変量データ可視化手法の1つである平行座標プロット (PCP: Parallel Coordinates Plot) [3] に着目した。PCPとは、多変量データの属性を軸として描き、それに対してデータの1タプルを1つの線で表す手法である。この手法の特徴として、データ全体のクラスタの偏りや、属性間の相関を俯瞰的に把握することができる。また、PCPは n -項関係データの属性の並びを変えずに表示することが可能で、さらにはSPJのようなデータ操作をGUI上で容易に再現可能である。そこで、我々はそのPCPとSPJの親和性の高さに着目した。

また、PCPに関しては、これまで様々な論文で議論がされてきた [4]。しかし、PCPの軸の見せ方やクラスタの見せ方に関して議論がされているものがほとんどであり、PCPの操作をSPJのような関係代数演算に明示的に対応させ、データ分析を支援する研究は行われていない。実際のデータ分析においては、リレーションスキーマを適切に定義したうえで多変量データを利用するからこそ、分析者が求めるアドホックなクエリが記述が可能になる。さらに、所望のデータ可視化結果の生成には試行錯誤を繰り返し、時には過去の可視化過程に戻る必要がある。しかし、分析者が支援なしに再び過去の可視化過程を再現することは困難である。したがって、その試行錯誤の過程を保存し分析者を支援することは、情報可視化において重要となる。

そこで本研究では、関係代数演算との親和性が高いと考えられるPCPに対し、PCPの操作過程を保存・再現するSQLに類似した言語を組み込んだ情報可視化システムを

提案する。この言語を $(PC)^2L^{*1}$ と呼ぶ。 $(PC)^2L$ を用いることで、関係代数演算におけるSPJ質問に相当する操作記述をPCPに対して適用することが可能となる。これにより、PCPを用いたデータ分析の試行錯誤の過程を保存することが可能となり、問合せ言語として一般的なSQLに親しみのあるデータ分析者を支援することができる。本研究の目標を次のように定める。

- (1) SPJ質問機能を統合したPCPを構築することによって、可視化システム上で体系的な操作を可能にする。
- (2) 操作結果の保存・再現するSQLに類似した言語を統合した、PCPによる情報可視化システムを構築することによって、データ分析の試行錯誤を支援する。
- (3) 服飾文化財と自動車走行ログという、趣向がまったく異なる2つのコンテンツについて情報提示を行うことにより、本システムの汎用性と有用性を示す。

2. 関連研究

2.1 平行座標プロット (PCP) とデータ操作

1985年、Inselbergによって初めてPCPの概念が定義された [3]。それ以来、PCPに関する様々な議論がされてきており、情報可視化の分野において重要なトピックの1つとなっている [4]。Johanssonらによれば、PCPの研究は次の4つに分類される [4]：(1) PCPの(属性)軸レイアウト、(2) PCPのClutter軽減方法、(3) PCPの実用例の提示、(4) PCPと他のデータ解析手法との比較。上記のとおり、PCPの見せ方に関して議論がされているものがほとんどであり、操作の過程に着目した議論はされていない。また、操作の過程でデータ自体にSPJのような演算を加えながら操作するものは議論されていない。

また、Boualiらは、対話型遺伝的アルゴリズムを使用することで、可視化手法の推薦を行うシステムを構築した [5]。これは、データや利用者の要求に応じてより適切な可視化手法(散布図行列やPCPなど)の選択を支援するものである。我々の提案システムは関係代数におけるSPJ質問に相当する表現力を持つような可視化システムを構築するため、タプルが1つの線で明示され、詳細に参照・分析可能であるPCPが適切である。

一方で、インタラクティブに操作しながらPCPによる分析を支援するシステムの提案もされている。Itohらは、属性軸間の相関に基づいてインタラクティブに次元削減を行い、PCPから所望する情報の発見を支援するシステムを構築した [6]。Zhouらは、エントロピーの概念を導入することで、PCPの属性軸の整列順序をクラスタに基づいて決定する手法の提案をした [7]。

他の研究として、データやシステムの操作過程を管理する研究 (Provenance) が行われている [8]。特にデータや

*1 Parallel Coordinates Plot Commutative Language の頭文字より。

システム、プログラミングコードなどの操作過程や意図を保存することは、複雑なデータ処理を支援するために重要なことであるといわれている。さらには、分析結果データの操作過程や意図を示すことは、SQLのような関係代数演算をサポートする問合せ言語で記述することが有効であるともいわれている。その点で、本論文で提案する操作言語を用いて、可視化システムによるデータの操作過程を保存することは有効な手段であるといえる。

システムの操作過程を保存してユーザのタスクを支援する手法として、Waldnerらは、PCのアプリケーション操作ログを記録し、そのログを時系列が理解できるように可視化することで、ユーザが過去に行った意思決定の検証やデータソースの再取得の支援を行った[9]。Mindekらは、画像データに分析過程に利用する他のデータソースのデータを同時に表示し、分析者の文脈を保存したスナップショットを保存することで、シミュレーションデータの可視化や文書分析の支援を行った[10]。Gratzlらは、PCPやヒートマップ、散布図行列など様々な可視化手法を組み合わせることで複数のデータソースから得られたデータとその分析過程を可視化し、データ分析の支援を行った[11]。これらの手法と比較して我々の提案手法は、「可視化システムのデータ分析過程を可視化して見せる」のではなく、「SQLに類似した言語を用いてデータ分析の途中結果を保存し、問合せ言語として一般的なSQLに親しみのあるデータ分析者を支援する」ものであり、立場が異なる。また、言語を用いて操作過程を保存する利点として、言語の一部を書き換えるだけでデータ分析の改善をすることが容易であり、その点でこれらの研究と比較して優位性を持つ。

一方、多変量データを可視化するその他の有名な手法として、複数の散布図を表示する散布図行列があげられる[12]。しかし、散布図行列は、属性どうしの相関を直感的に把握できるが、散布図数が属性数の2乗に比例して増加する。そのため、データの操作過程でSPJ(特にJoin)を適用することには不向きであるといえる。

2.2 デジタルアーカイブに関連する可視化システム

近年、情報可視化に注目した取り組みが分野を問わず様々に展開されている。情報技術を実際の展示で活用する取り組みに着目すると、次のようなものがある。岡本らは、絵画作品の一部を切り出し、その素材要素群をデジタルミュージアムに展示することで興味喚起を行うインタフェースの開発を行った[13]。Wakitaらは、立体的に織られた大型の織物を計測し、実物大展示システムや視触覚展示システムを構築した[14]。安達らは、大量の写真資料のなかから所望する画像の候補を効果的に提示するシステムの検討を行った[15]。Yoshinoらは、日本の伝統的な書籍に対し直感的に操作可能な体験型鑑賞支援システムを構築した[16]。これらのシステムは、利用者に対しコンテンツ

への興味喚起を行うことを目的としている。一方で、我々はデータ分析を支援することを目的としており、その点で我々の研究とは異なる。

また、利用者へ様々な視点を提供し分析を行えるような枠組みの提案もされている。Thudtらは、デジタルブックコレクションに対してSerendipityに着目した情報可視化システムを構築した[17]。Anらは、自己組織化マップを応用することで、複数博物館の類似点や相違点を視覚的に分析し、観光客のための包括的な博物館ツアーの提案に関する仕組みを提案した[18]。Whitelawは、従来からあるキーワード検索による情報検索に代わり、検索対象全体の概観と個々の詳細とを往来することで検索を行うインタフェースを構築した[19]。これらのシステムは、コンテンツについてあまり詳しくない利用者を想定したものである。一方で、我々の提案システムはコンテンツに対してある程度の知見があるエキスパートを対象としており、その点で我々の研究とは異なる。

さらに、コンテンツについて詳しくない人だけではなく、専門家に対しても有用な情報可視化の検討も行われている。Odijkらは、著者の異なる複数のデータを統合し、文化財に対する様々な視点を提供するシステムを構築した[20]。Hinrichsらは、文学研究を支援するうえで重要な4つの観点を発見し、それに基づいて未知の大規模文学コレクションの可視化システムを構築した[21]。Glinkaらは、作品集の全体を把握しながら個々の作品に着目可能にすることで、文化財の定性分析と定量分析を行うことができる情報提示システムを検討した[22]。これらの提案手法では、それぞれのコンテンツに特化した枠組みの提案にとどまっている。それに対し我々は、コンテンツを問わない汎用的な枠組みの提案をするものであり、その点で我々の研究とは異なる。

2.3 自動車ログに関連する可視化システム

近年、自動車より収集した走行ログを活用する取り組みが展開されている。飯島らは、プローブカーで収集した道路状況を中心に、渋滞状況を地図の色分けで可視化するシステムを構築した[23]。平岡らは、運転者に自発的なエコドライブを促すためのエコドライブ支援システムを提案し、情報提示の重要性を示した[24]。Wilminkらは、情報提示による運転の改善が二酸化炭素の排出を削減することをシミュレートした[25]。Komamizuらは、EVの運転軌跡をドライバの属性情報とともに可視化するフレームワークを提案した[26]。これらの提案手法は、コンテンツに特化した枠組みとなっている。一方我々は、コンテンツを問わない汎用的な枠組みの提案をするものであり、その点で我々の研究とは異なる。

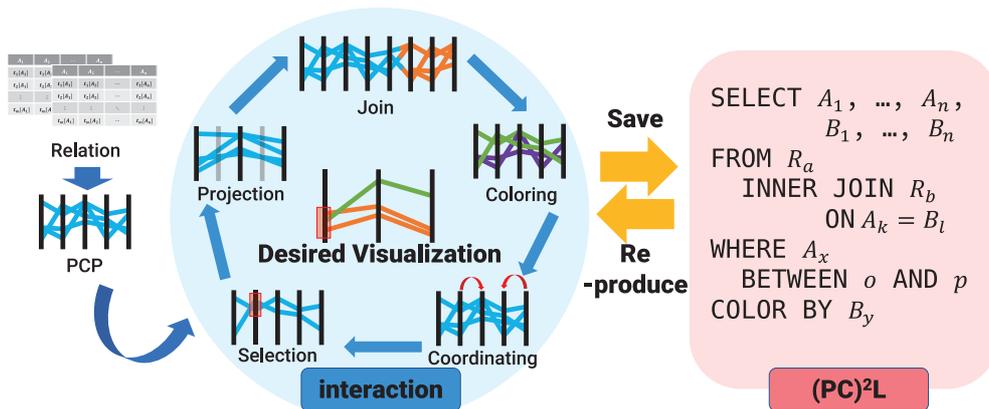


図 1 提案システムの概要図

Fig. 1 A system overview.

3. 情報可視化システムの設計

3.1 提案システム概要

本節で設計する提案システムの概要を図 1 に示す。

提案システムでは、データ分析者が 3.2 節で説明する PCP に対し GUI 上でインタラクションを行いつつ、その操作結果を 3.3 節で説明する (PC)²L という言語で保存・再現する。それにより、試行錯誤を行いながら所望の可視化を獲得することを想定する。以下では、想定する操作の流れについて示す。

- (1) 1 つのリレーションを PCP に適用し可視化する。
- (2) 表示結果をもとに、PCP 上で操作を行い試行錯誤を行う。
- (3) 分析者が所望するときに、(2) で操作した結果のスナップショットを (PC)²L で保存する。
- (4) (2)–(3) を繰り返す。その際、過去のスナップショットに戻る必要がある場合は該当する (PC)²L を入力し、PCP 上に操作結果を再出力する。
- (5) データ分析者が所望の可視化結果を獲得する。

3.2 PCP に対するインタラクション

本節では、提案システムの PCP で利用可能な 5 つの機能について述べる。今回実装するインタラクション機能は、既存の SPJ 質問である「選択 (Selection)」「射影 (Projection)」「結合 (Join)」と PCP 独自の操作として我々が提案する「色分け (Coloring)」「軸配置 (Coordinating)」に分類できる。ここで用いる記号の定義を表 1 に示す。なお、表 1 の記号および「選択 (Selection)」、 「射影 (Projection)」、 「結合 (Join)」に関する定義については、文献 [27] を参考にした。

3.2.1 PCP とリレーション

本項では、提案システムの PCP における関係代数表現について述べる。

本論文では、リレーションの持つ属性を軸として複数並

表 1 記号の定義

Table 1 Definition of symbols.

記号	概要
A_i	属性
$R(A_1, A_2, \dots, A_n)$	リレーション
$t_j = (a_1, a_2, \dots, a_n)$	タブル
$a_i = t_j[A_i]$	タブル t_j の属性 A_i 上の属性値
$D_i = \text{dom}(A_i) = \{\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_m}\}$	属性 A_i のドメイン (定義域)
$\text{dom} : A_i \rightarrow D_i$	属性 A_i のドメイン D_i を導出する関数
$R[A_i \theta A_j]$	R の A_i と A_j 上の θ -選択演算
$R[A_{i_1}, A_{i_2}, \dots, A_{i_k}]$ ($1 \leq i_1 < i_2 < \dots < i_k \leq n$)	R の $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ 上の射影演算
$R[A_i \theta B_j]S$	R と S の A_i と B_j 上の θ -結合演算

列し、タブル 1 つが持つ「属性値」を線で結ぶことでデータ 1 タブルを表現する。この線を複数描き重ねることで線の疎密や偏りが現れるので、ある属性どうしがどのような関係にあるのか、全体を見通してどのような傾向にあるのかを一目で把握することができる。

以下では、PCP とリレーションの関係について定義する。リレーション R の属性 A_i の名称を図 2 の軸名とする。また、 $\text{dom}(A_i)$ の要素、つまり属性 A_i の属性値を図 2 の A_i 軸の中に配置する。その後、タブル t_j を図 2 の線 1 本に相当するような形で描くことで、リレーション R のインスタンスをすべて PCP 上で表す。

3.2.2 PCP と SPJ 質問

本項では、PCP に対するインタラクションとして利用可能な SPJ 質問について述べる。3.2.1 項で示した PCP に対し、関係代数における「選択 (Selection)」、 「射影 (Projection)」、 「結合 (Join)」の機能を忠実に実装し、それらの演算を可視化システム上で簡単に利用可能な仕組みを備

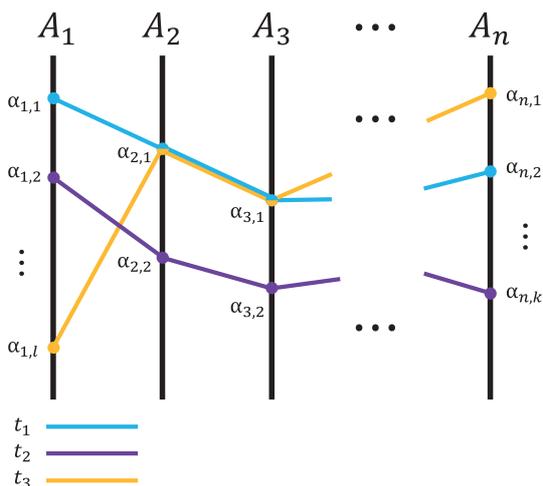


図 2 PCP 上にリレーション R を表した例
 Fig. 2 An example of PCP applied to relation R.

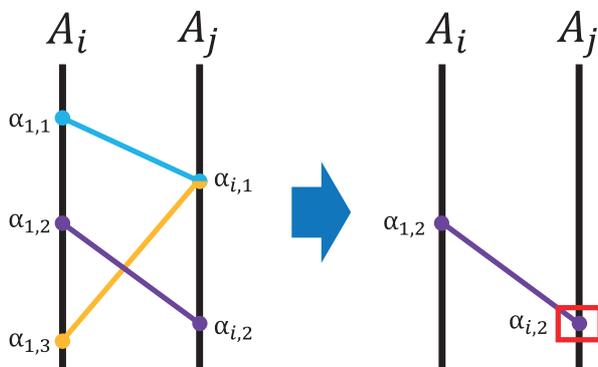


図 3 選択 (Selection) の例
 Fig. 3 An example of "Selection".

える。すなわち、「選択 (Selection)」により属性値による絞込みを行い、「射影 (Projection)」により表示する属性を切り出し、「結合 (Join)」により属性の追加を可能とする。以下では、SPJ 質問と PCP の関係について定義する。
選択 (Selection)

選択操作は、軸に対し図 3 の赤枠のような形で範囲選択をすることで、表示するタブルを絞り込む機能をもって実現する。すなわち、本論文では図 4 のように $R[A_i \geq y_1] \wedge R[A_i \leq y_2]$ を行う操作として定義する。

射影 (Projection)

射影操作は、図 5 のように軸を切り出す機能をもって実現する。すなわち、R の全属性集合 $\{A_1, A_2, \dots, A_n\}$ の部分集合 $X = \{A_{i1}, A_{i2}, \dots, A_{ik}\}$ に該当する属性のみを切り出す操作を $R[X]$ として定義する。また、軸を切り出した後は軸間を対応する各タブルどうしで結合しなおすものとする。

結合 (Join)

結合操作は、図 6 のように軸を追加する機能をもって実現する。すなわち、リレーション $R(A_1, A_2, \dots, A_n)$ と、リレーション $S(B_1, B_2, \dots, B_m)$ それぞれが持つ

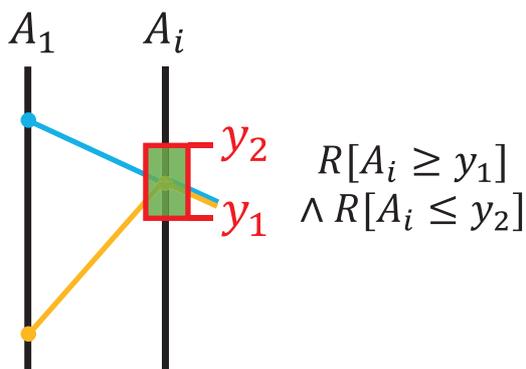


図 4 選択 ($R[A_i \geq y_1] \wedge R[A_i \leq y_2]$)
 Fig. 4 "Selection" ($R[A_i \geq y_1] \wedge R[A_i \leq y_2]$).

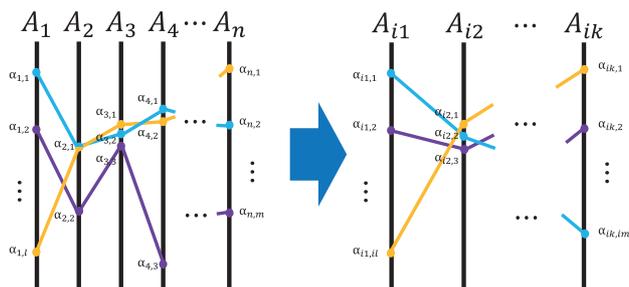


図 5 射影 ($R[X], X = \{A_{i1}, A_{i2}, \dots, A_{ik}\}$)
 Fig. 5 "Projection" ($R[X], X = \{A_{i1}, A_{i2}, \dots, A_{ik}\}$).

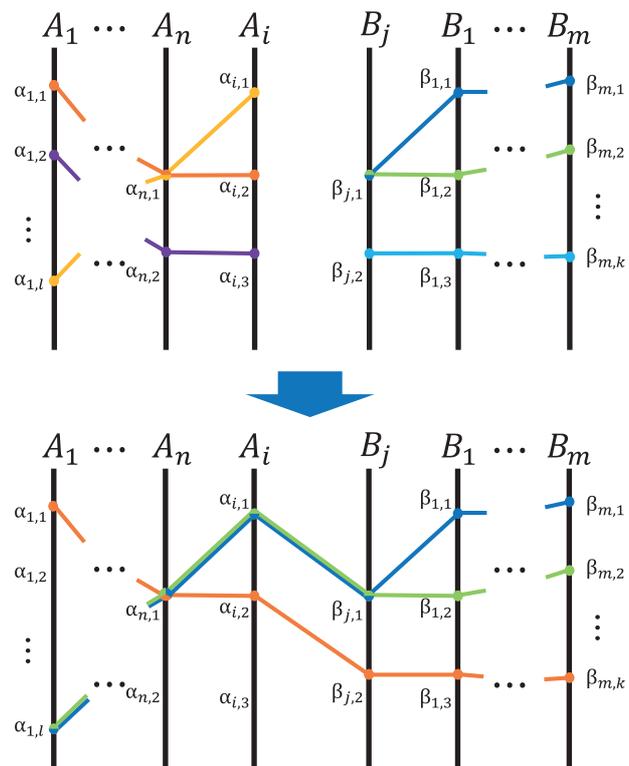


図 6 結合 ($R[A_i \theta B_j]S$)
 Fig. 6 "Join" ($R[A_i \theta B_j]S$).

タブルを、 $A_i \theta B_j$ の条件でつなぎ、軸の追加を行う操作を $R[A_i \theta B_j]S$ として定義する。なお、本論文で扱うことができる θ は $=$ のみ、すなわち等結合のみとする。

厳密には、 $R[A_i]$ と $R[B_j]$ の両方に等値が存在する内部結合 (INNER-JOIN) のほか、片方が空値であっても結合結果を出力する外部結合 (OUTER-JOIN) も可能とする。

3.2.3 PCP の独自機能

本項では、我々が提案する PCP に対するインタラクションとして利用可能な独自機能について述べる。

色分け (Coloring)

Coloring 操作は、任意の属性 A の値を基準に線の色を塗り分けられる機能として定義する。PCP はこの色分けにより、タプルを表す線の色によってデータ全体の見え方が変化し、線の疎密や偏りだけでなく、色によってもデータ全体の傾向を把握することが可能となる。

軸配置 (Coordinating)

Coordinating 操作は、PCP の軸を任意の順番に並べる機能として定義する。PCP は、軸の順番によってクラスタの見え方や隣り合う軸間の相関 [28] が現れるため、軸配置を変更できることは重要な機能である。

そのほかの独自機能として、PCP の各属性の値の並べ替え順序を指定することや集約演算、サブクエリなどが考えられるが、今後の課題とし本論文での範疇外とする。

3.3 (PC)²L

本節では、我々の提案システムにおいて重要な構成要素の 1 つである (PC)²L について述べる。

本節では、3.2 節で定義した、5 種類の PCP に対するインタラクションの過程を保存・再現するための言語である (PC)²L を定義する。すなわち本機能は、PCP における操作結果を言語化し保存する仕組みと、逆にその言語をもとに過去の PCP の操作結果を再現する仕組みを備える機能である。

以下では、(PC)²L の言語仕様の概要を述べる。なお、提案システムが生成および受理すべき (PC)²L の文法の詳細については、付録 A.1 に示す。本言語については、SQL の SELECT, FROM, WHERE 句に加えて COLOR BY 句を新たに利用可能とする。

以下に、(PC)²L においてどのように 5 つのインタラクション機能を表現するかを記述する。

- 選択 (Selection) : WHERE 句の「 A_i BETWEEN y_1 AND y_2 」が「 $R[A_i \geq y_1] \wedge R[A_i \leq y_2]$ 」に相当。
- 射影 (Projection) : SELECT 句に R の部分集合 X の属性名をカンマ区切りで列挙。
- 結合 (Join) : FROM 句の「 R <JOIN 句> S ON $R.A_i \theta S.B_j$ 」が「 $R[A_i \theta B_j] S$ 」に相当
- 色分け (Coloring) : COLOR BY 句に「 A_i 」と記述することで、属性 A_i の値を基準に PCP の線の色分け。
- 軸配置 (Coordinating) : SELECT 句の R の部分集

合 X の属性名の A_1, A_2, \dots, A_n という列挙順番が、PCP 上の軸の配置順に対応。

ただし、使用しているライブラリの制約および実装の都合上、実装システムにおいては以下の制約が存在する。

- WHERE 句において、論理演算は AND および OR のみ使用可能である。
- WHERE 句において、同一軸上の論理演算は OR のみ使用可能である。
- WHERE 句において、軸間の論理演算 (AND/OR) は一括で同一のものだけ使用可能である。
- FROM 句において、INNER JOIN と LEFT OUTER JOIN, RIGHT OUTER JOIN, FULL OUTER JOIN のみ使用可能であり、直積や自然結合などは使用できない。
- JOIN 句において、条件式として記述可能なものは 1 つのみ許容する。

3.4 実装方法

提案システムの画面の例を図 7 に示す。提案システムは環境を問わず利用できるようにするため、Web ブラウザを通して多くの端末から利用できるように構築した。開発言語としては、サーバサイド処理に PHP、クライアント処理に HTML, CSS, JavaScript を使用した。

画面上部 A では PCP を表示している。画面下部 C では、画面上部で操作した結果に対応するデータが抽出および表示されている。なお、画面上部 A の PCP 部分に関しては、「D3.js *2」ならびに「Parallel Coordinates *3」を使用した。

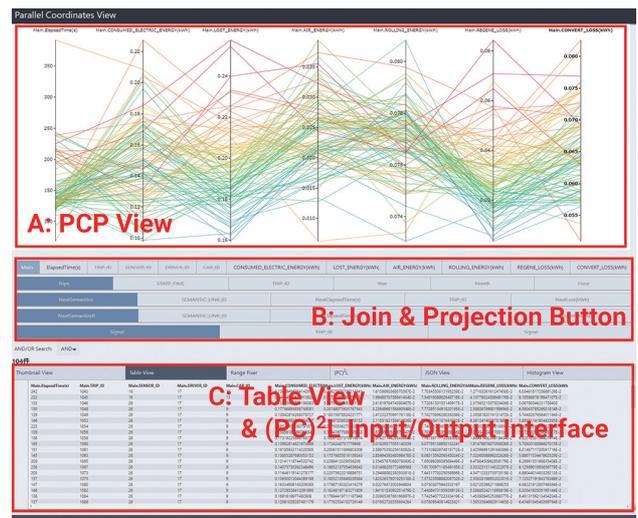


図 7 提案システムの表示例

Fig. 7 An example of system screen.

*2 <https://d3js.org/>

*3 <https://syntagmatic.github.io/parallel-coordinates/>

3.4.1 実装システムにおける SPJ 質問の実行

本画面における操作と 3.2.2 項で述べた選択 (Selection), 射影 (Projection), 結合 (Join) の各操作について述べる。

図 7 において, 画面上部 A の PCP 各軸上をマウスカーソルでドラッグすることで範囲選択が可能となっており, これにより選択 (Selection) が利用可能である。

また, 画面中央 B にあるトグルボタンは, 横 1 行がリレーション 1 つと対応しており, 射影 (Projection) と結合 (Join) が利用可能である。ボタンの最左部には, 結合 (Join) のためのリレーションボタンが配置されており, クリックすることで図 8 の画面が表示され, リレーションの結合条件を選ぶことができる。なお, すでに結合 (Join) されているリレーションボタンをもう 1 度クリックすると, 結合 (Join) を解除できる。

リレーションボタンより右側には, 射影 (Projection) のための軸ボタンが配置されており, クリックすることで軸の表示/非表示を切り替えることができる。なお, これらのトグルボタンに表示する要素の設定については別途設定ファイルで管理しており, データセットに特化した実装にはしていない。

また, 画面上部 A の軸名部分をクリックしながら左右にドラッグすると軸の Coordinating を行うことができる。

3.4.2 実装システムにおける (PC)²L の実行

3.3 節で述べた (PC)²L について説明する。本機能は, 図 7 の画面下部 C の (PC)²L タブから利用可能である。ページ下部に表示されるテキストフィールドに (PC)²L を記述し, PCP 上に反映が可能である。一方で, PCP 上の操作結果に関しても, (PC)²L によって記述されたテキストファイルとしてダウンロード可能である。なお, (PC)²L を解析するパーザに関しては, 「PEG.js *4」を用いて作成した。

画面上部 A の PCP 軸上をクリックすることで, その属性を基準に Coloring が可能である。また, 今回は Coloring の色の順番に関しては, 属性値が数値の場合は昇順に「青, 緑, 黄, 赤」が滑らかに変化するように配色し, 文字列の場合は固定色で配色する。



図 8 結合 (Join) の条件を指定する画面例

Fig. 8 A screen in order to set “Join” condition.

*4 <https://pegjs.org/>

3.5 例題

本節では, 提案システムで想定する操作や見方について, 簡単な例題を用いて説明する。

今回は, 文部科学省より公開されている「学校教育総括*5」から, 年次別・学校種別在学者数のリレーション A を抽出した。リレーション A が持つ属性は, 「年」, 「小学校」, 「中学校」, 「高校」, 「短期大学」, 「大学」がある。

リレーション A を適用した PCP について, 以下 4 項目を行っていくことで図 9 のような表示となる。

- 年度の範囲を指定する選択 (Selection)
- 「年」と「小学校」と「大学」軸を切り出す射影 (Projection)
- 軸の順番を入れ替える Coordinating
- 「年」軸を基準にタプル線を塗り分ける Coloring

PCP は一般的に, 順相関は平行線, 逆相関はクロスした線が引かれる。すなわち, 1990 年~2010 年頃に限っては, 「大学」と「年」が順相関であり, 「小学校」と「年」が逆相関であることが分かる。このことから, 前者に関しては『大学全入時代の影響が表れているかもしれない』, 後者に関しては『少子化の影響が表れているかもしれない』などといった仮説を立てることができる。

そこで今回は, 「小学校」の在学者数と「年」に関して少子化の影響がどう関係してくるのかを分析することを考える。そのために, 厚生労働省より公開されている「人口動態統計*6」から, 人口動態総覧の年次推移のリレーション B を抽出した。リレーション B が持つ属性には, 「年」, 「出生数」がある。

提案システムにおいて, 結合 (Join) を行うことでリレー

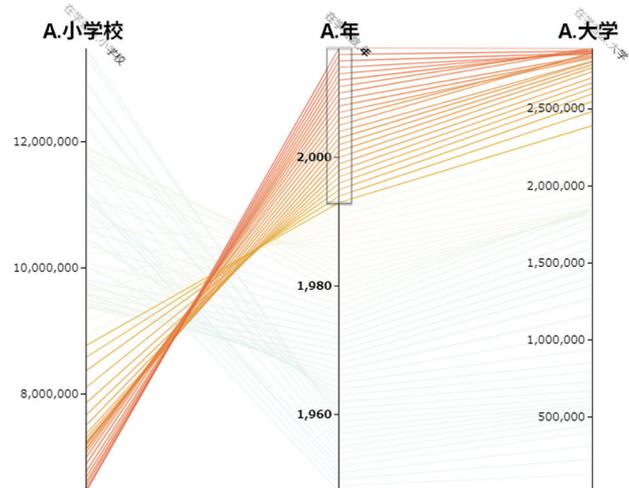


図 9 年次別各学校在学者数に関するデータセットを適用した PCP
Fig. 9 An example of PCP applied a data set which represents the number of students in each schools by year.

*5 http://www.mext.go.jp/b_menu/toukei/002/002b/1403130.htm

*6 <https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/kakutei17/index.html>

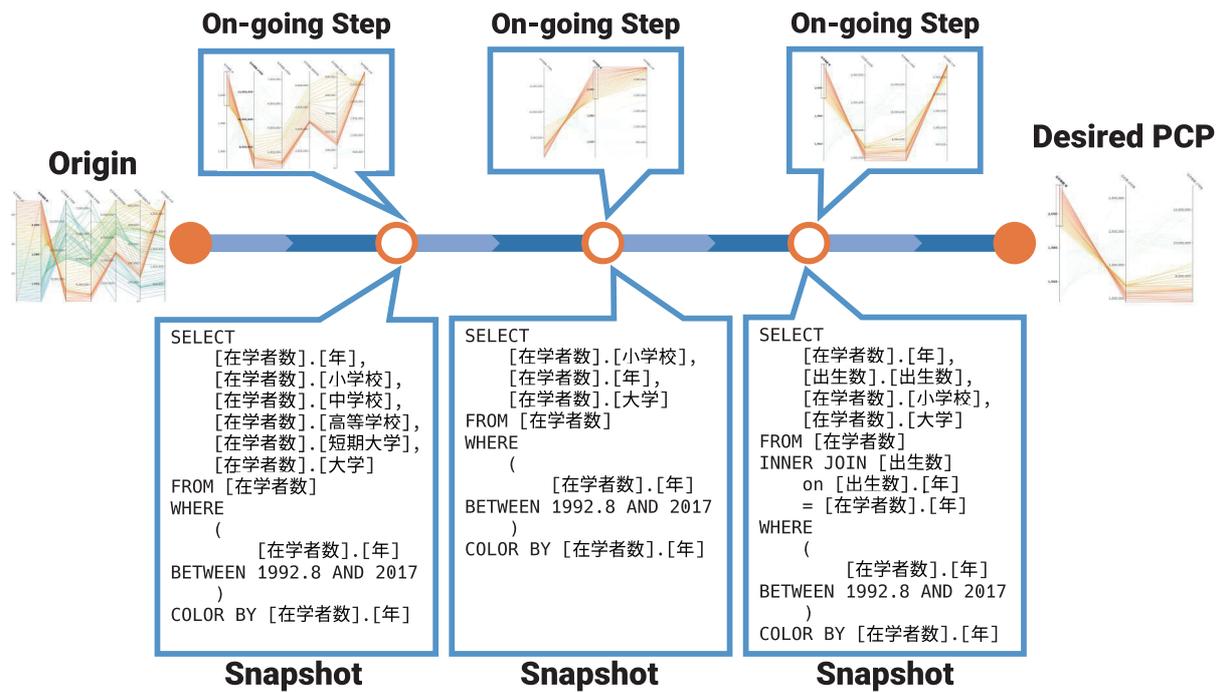


図 11 操作過程と (PC)²L の例

Fig. 11 An example of an operation process and (PC)²L.

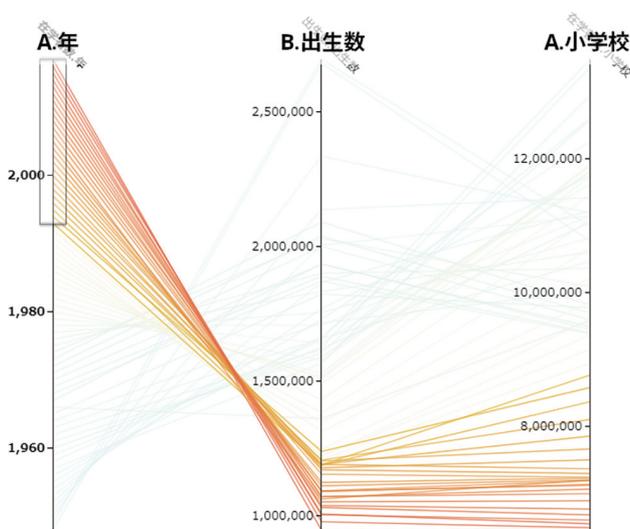


図 10 年度と小学校在学者数、出生数の相関を表す PCP

Fig. 10 Correlations of years, the number of elementary school students, and the number of live births.

ション A とリレーション B を結合することができる。ここでは「年」をキーとして等結合を行うことで、図 10 のような表示となる。図 10 から、「年」と「出生数」が逆相関であり、「出生数」と「小学校」は順相関であることが分かる。このことから、『年が進むにつれて出生数が減少している』ということや、『出生数の減少が小学校在学者数の減少に関係している』といったことが示唆される。すなわち、『小学校の在校生数は出生数の減少に起因している』ということが示唆される。提案システムではこのようにして、定量的・定性的な分析が可能である。

一方で、上記で示したような操作の過程は、図 11 で示すように任意の時点で (PC)²L によるスナップショットを残すことができる。過去に保存したスナップショットはいつでも呼び出すことができ、データ分析の過程を支援する。

4. 情報可視化例

4.1 情報可視化例 1：服飾文化財

これまでに我々は、多様で散在するデジタルアーカイブや書籍をデータベースに統合した「小袖屏風*7データベース」の構築を行ってきた [30], [31]。本データベースにおいては、服飾文化財である小袖屏風 100 隻および小袖 111 着について検索が可能である。本節では、このデータベースを使用して情報可視化例を示す。具体的には、文献 [30], [31] で述べている服飾様式の流行の変遷に基づいて可視化を行う。

ここでは、服飾様式の流行（以下、「モード」という）に着目した組合せを明示する操作について説明する。なお、本節で扱うリレーションおよび属性は表 2 に示す。操作開始時の PCP と適用する (PC)²L、適用結果の PCP について図 13 に示す。この結果を導出するうえでは、2.1 節で示したような PCP 上での操作を試行錯誤しながら所望の結果を獲得した。その手順について以下で説明する。

まず、PCP に対しリレーション R_1 を適用することで、小袖屏風に関するデータの全体像を PCP で表示すること

*7 小袖屏風とは、安土桃山時代から江戸時代ごろに実在した小袖（現在の着物の原型）のうち、状態の芳しくない小袖の裂の実物を金箔地や金砂子地の二曲一隻屏風に貼り付けることで、衣服本来の印象を再現したものである [29]。

表 2 小袖屏風に関するリレーションとそのなかにある属性
 Table 2 Relations related to Kosode Byobu and that attributes.

リレーション	リレーション名	属性
R_1	Main	屏風番号, 資料番号, 年代, 材質, 技法
R_2	モード	資料番号, モード

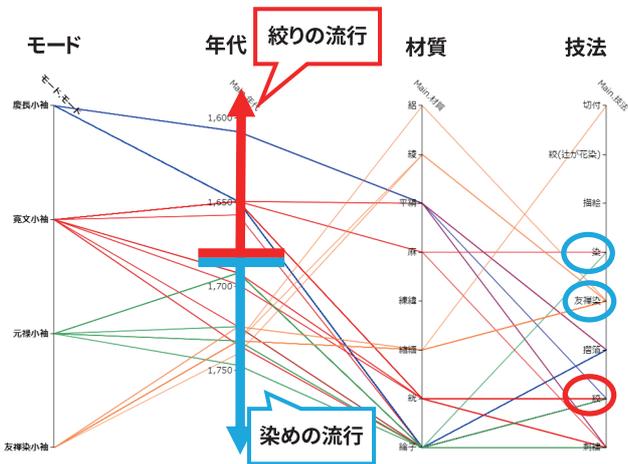


図 12 小袖屏風の流行の変遷と PCP における可視化した例
 Fig. 12 Transition of Kosode Byobu's "Mode" and visualization of PCP.

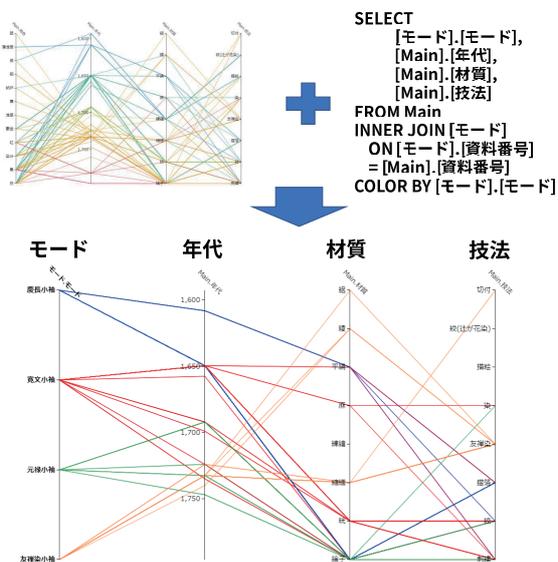


図 13 PCP に小袖屏風の「モード」の変遷に基づいた可視化を行う (PC)²L を適用した例
 Fig. 13 An example of PCP applied (PC)²L in order to visualize transition of Kosode Byobu's "Mode".

ができる。しかしながら、この状態ではどこに着目すればよいのかが不明瞭である。

そこで今回は、専門家の知見の一例として、小袖の流行を示す「モード」に関するリレーション R_2 を結合 (Join) する。すなわち、 R_1 に対し、 R_2 を小袖 1 枚に付与されている「資料番号」をキーとして内部結合を行う。これによ

り、図 12 に示すような表示となる。なお、図 12 の赤丸は絞りの流行を示す要素であり、青丸は染めの流行を示す要素である。この図から、おおよそ「モード」というものは「年代」の変遷と関係があり、「モード」は小袖の流行と関係があるものとみることができる。

加えて、「モード」軸内にある要素「慶長小袖」は、「技法」軸内にある要素「絞」に対して多くのクラスタが存在することが分かる。それに対して、「モード」軸内にある要素「寛文小袖」と「元禄小袖」は、「技法」軸内にある要素「絞」と「染」に対してクラスタが存在することが分かる。一方で、「モード」軸内にある要素「友禅染小袖」は、「技法」軸内にある要素「友禅染」に対して多くのクラスタが存在することが分かる。これらのことから、以下のことが可視化により確認できる。

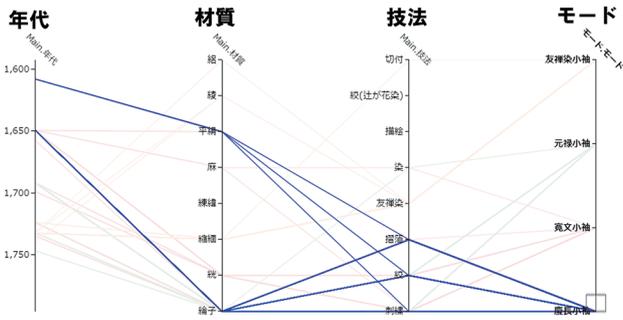
- 「慶長小袖」は絞りを中心とした流行がある
 - 「寛文小袖」「元禄小袖」は絞りを中心としつつ染めの流行がある
 - 「友禅染小袖」は染めを中心とした流行がある
- すなわち、「モード」の変遷にともない、小袖が持つ「技法」の要素が変化の様子を確認できる。

以上から、提案システムによって、小袖や服飾の専門家が持つ「モード」と「技法」との関係性を可視化により明示できた。

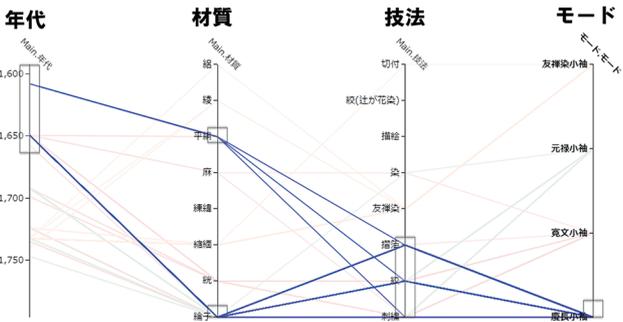
続いて、複数のデータソースを用いた知識の再発見を行うための分析例を示す。ここで、 R_1 、 R_2 は、それぞれ異なる文献から得られた情報を示す。図 14 (a) では、「モード」軸内にある要素「慶長小袖」が選択された状態にある。図 14 (a) を見ると、「慶長小袖」が選択されたことにより、11 件の小袖屏風が表示され、慶長小袖が持つ特徴「年代：1600 年ごろから 1650 年ごろまで」「材質：絹子、平絹」「技法：摺箔、絞、刺繍」が確認できる。これらの特徴の組合せから、明示的にモード「慶長小袖」として示されていない小袖屏風を検索することを考えてみる。

図 14 (b) は、図 14 (a) で得られた「年代」、「材質」、「技法」の値を明示的に選択した状態である。すなわち、 R_2 において、「慶長小袖」と明示された小袖が持つ特徴の傾向を示す具体値を選択した (図 14 (b))。ここから、同一の具体値を持つ別の小袖があるかどうかを検索する。つまり、インスタンスからボトムアップに同一の特徴を有する別のインスタンスを検索することを意図している。

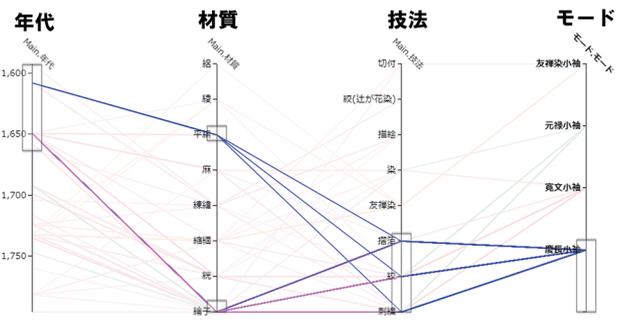
図 14 (b) の状態を (PC)²L を用いて言語記述すると、図 15 のリストが得られる。このリストに対し、 R_1 と R_2 の結合条件を内部結合 (INNER JOIN) から左外部結合 (LEFT OUTER JOIN) に書き換えて検索結果を表示すると、図 14 (c) が得られた。図 14 (a) に現れなかった検索結果が、Coloring 機能によって図 14 (c) ではピンク色の線で示された。図 14 (a) と図 14 (c) の検索結果の件数はそれぞれ 11 件と 20 件であった。この結果は、「慶長小



(a) モード軸の慶長小袖を選択



(b) 慶長小袖の特徴を選択



(c) 結合方法を内部結合から外部結合に変更

図 14 (PC)²L の JOIN 句を変更して可視化した例

Fig. 14 An example of visualization changing (PC)²L JOIN clause.

```

SELECT
  [Main].[年代],
  [Main].[材質],
  [Main].[技法],
  [モード].[モード]
FROM [Main]
INNER JOIN [モード] on [モード].[資料番号] = [Main].[資料番号]
WHERE
  ([Main].[年代] BETWEEN 1593.0191326530612 AND 1663.621173469388)
  AND
  ([Main].[材質] BETWEEN 375 AND 393
   OR
   [Main].[材質] BETWEEN 99 AND 122)
  AND
  ([Main].[技法] BETWEEN 269 AND 393)
  AND
  ([モード].[モード] BETWEEN 367 AND 393)
COLOR BY [モード].[モード]
    
```

図 15 図 14 (b) の PCP を保存した (PC)²L

Fig. 15 (PC)²L preserving Fig. 14 (b) PCP.

袖」と明示的には分類されていないが、同一の特徴を有する小袖が9件存在していることを示す。この結果が妥当であることを検証するため、これらの9件について別の文献

表 3 集計条件

Table 3 Aggregate condition.

集計期間	2017/8/4-2018/12/18
区間 A	岸谷 2 丁目～新子安
区間 B	新子安～七島町
区間 B'	宮ノ下～岸谷 2 丁目
進行方向	横浜方面 (上り)
被験者数	1 名
センサ	Zenfone2
トリップ数	98 件

に記載された解説文*8の確認を行った。すると、3件が慶長小袖、2件が慶長小袖から寛文小袖の移行過程で作成されたものなど慶長小袖と定義するのが難しいもの、残り4件がそもそもモードを定義できるか不明なものの3種類に分かれた。

以上のことより、提案システムにより、専門家の持つ知見が明示的な情報として与えられていない小袖屏風が検索結果に出力された。これにより、小袖屏風の専門家に対しては知識の再確認、専門外の人には新たな知見の発見を支援することができた。加えて、(PC)²Lの簡単な書き換えによるデータ分析に貢献する例を示すことができた。

4.2 情報可視化例 2: 自動車走行ログ

これまでに我々は、自動車走行ログから「その人がEV (Electric Vehicle) に乗り換えて同じ運転をしたときの消費エネルギー」を推定しデータベースに蓄積しておくことで、EVの電力消費に関する多様な検索を実現する ECOLOG (Energy CONsumption LOG) システムを提案してきた [34]。本節では、この ECOLOG システムにおける情報可視化例を示す。具体的には、表 3 の区間 A を対象とした損失エネルギーの要因分析を行うことを考える。なお、詳細な分析に関しては文献 [34] で述べているため、ここでは所望の可視化をどのような手順によって達成したかについて述べる。

ここでは、区間 A に存在する信号 X の影響による、後続する区間 B の損失エネルギーの影響を明示するまでの操作について説明する。本節で扱うリレーション $S_1 \sim S_4$ とその属性を表 4 に示す。なお、「Signal」は区間 A に存在する信号 X の直前に滞在する時間により「通過」、「停止」をラベル付けしたもの、「ElapsedTime (s)」はある区間を通過するためにかかった時間、「LostEnergy (kWh)」はある区間を通過するために損失した電力量を表す。また、分析対象の条件を表 3 に示す。なお、ここでの 1 トリップとは 1 回の運転のことであり、PCP における 1 タプルはこの 1 トリップを単位とした。

操作開始時の PCP と適用する (PC)²L、適用結果の PCP

*8 専門家が執筆した解説文 [32], [33]

表 4 ECOLOG システムに関するリレーションとその中にある属性
Table 4 Relations related to ECOLOG System and that attributes.

リレーション	リレーション名	属性	説明
S_1	Main	TRIP_ID, ElapsedTime (s), LostEnergy (kWh)	区間 A における自動車走行ログ
S_2	NextSemantics	TRIP_ID, ElapsedTime (s), LostEnergy (kWh)	区間 B における自動車走行ログ
S_3	NextSemantics2	TRIP_ID, ElapsedTime (s), LostEnergy (kWh)	区間 B' における自動車走行ログ
S_4	Signal	TRIP_ID, Signal	区間 A にある信号 X について「通過」または「停止」を TRIP_ID ごとに付与したラベル

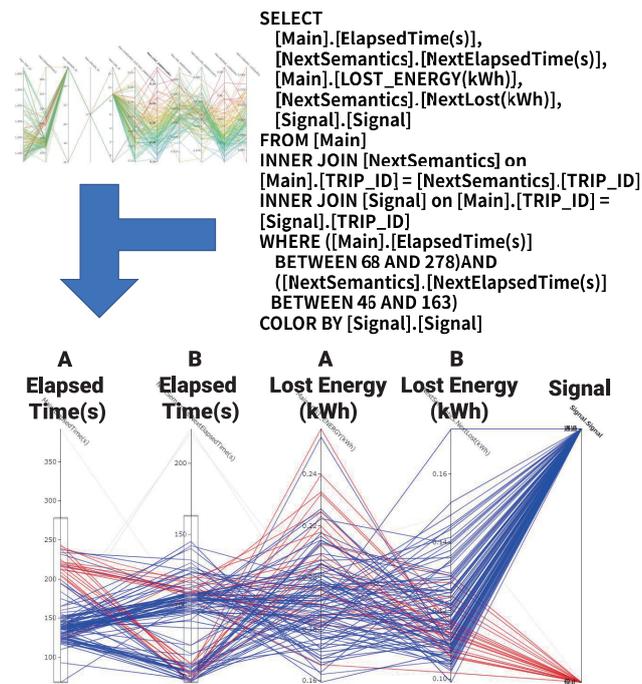


図 16 A 区間にある信号 X による損失エネルギーの影響を表す可視化を行う (PC)²L を適用した例

Fig. 16 An example of PCP applied (PC)²L in order to visualize effects on lost energy by signal X in section A.

について図 16 に示す。この結果を導出するうえでは、試行錯誤しながら提案システム上で操作を行い、所望の結果を獲得した。その手順について以下で説明する。

まず、PCP に対し S_1 を適用することで、区間 A に関するデータの全体像を PCP で表示することができる。しかしながら、この状態ではダブル線が複雑に交差しており、分析が難しい。

そこで今回は、区間 A に新たに設置された信号 X の影響を考えるために、リレーション S_2 を結合 (Join) する。すなわち、 S_1 に対し、 S_4 をトリップに付与されている「TRIP_ID」をキーとして結合 (Join) を行う。これにより、信号 X の「通過」、「停止」を区別しながら、区間 A における「ElapsedTime (s)」「LostEnergy (kWh)」の相関を可視化できる。

さらに分析者は、区間 A に隣接する区間 B や区間 B' と、信号 X の関係の分析を行いたくなった。このような場合でも、提案システム上ではアドホックに操作を実行するこ

とができる。すなわち、 S_1 と S_2, S_3, S_4 のリレーションを「TRIP_ID」をキーとして結合 (Join) を行うことで、PCP 上でそれぞれ属性間の関係を分析できる。これらの操作により分析者は、『信号 X の影響を分析するためには C (= A 区間 + B 区間) という区間に着目するとよい』、という新たな知見をもたらした。

また、4.1 節と同様に、分析者は試行錯誤しながら操作結果を (PC)²L で保存しながら進めことができる。後戻りが発生した際にも (PC)²L クエリを呼び出すことで再現が可能である。

以上から、PCP に対する操作および (PC)²L による操作を行うことで信号 X と各区間との関係を分析できた。

5. まとめと今後の課題

本研究では、可視化手法の 1 つである平行座標プロットを関係度数における SPJ に相当する表現力を持つように拡張した情報可視化システムを提案した。加えて、PCP の操作結果を保存・再現するための言語として (PC)²L の提案も行った。また、本システムを服飾文化財である小袖屏風と、自動車走行ログというまったく異なる 2 つのコンテンツに対して適用し、小袖屏風では服飾のモードの変遷を示し、自動車走行ログでは分析対象の区間 A について新たな知見をもたらした。

(PC)²L の拡張の方向性として、PCP の観点からは、PCP の各属性の値の並び替え順序を指定すること、関係データへの演算という観点では、集約演算、サブクエリなどが考えられる。今後の課題として、以上のような機能を実装することで (PC)²L の機能性と PCP との可換性を高めていくことを目指す。また、可視化事例をさらに積み重ね、より有益な情報可視化が可能であることを示していく。

謝辞 本研究の一部は、JSPS 科研費 (課題番号 18K11750) の助成による。また、一部は国立歴史民俗博物館共同研究 (平成 25 年度~平成 27 年度)「歴史資料デジタルアーカイブデータを用いた知的構造の創生に関する研究—小袖屏風を対象として*9) より着想を得た。

*9) <http://www.rekihaku.ac.jp/research/list/joint/2013/digitalarchive.html>

参考文献

- [1] 総務省：平成 30 年度版情報通信白書，総務省（オンライン），入手先 (<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h30/pdf/index.html>) (参照 2019-07-12).
- [2] 上林弥彦：データベースの基礎理論 (5)：関係データベースにおける質問処理，情報処理，Vol.24, No.3, pp.307-318 (1983).
- [3] Inselberg, A.: The plane with parallel coordinates, *The Visual Computer*, Vol.1, No.2, pp.69-91 (1985).
- [4] Johansson, J. and Forsell, C.: Evaluation of Parallel Coordinates: Overview, Categorization and Guidelines for Future Research, *IEEE Trans. Visualization and Computer Graphics (TVCG)*, Vol.22, No.1, pp.579-588 (2016).
- [5] Bouali, F., Guettala, A. and Venturini, G.: VizAssist: An Interactive User Assistant for Visual Data Mining, *The Visual Computer: Int'l Journal of Computer Graphics*, Vol.32, No.11, pp.1447-1463 (2016).
- [6] Itoh, T., Kumar, A., Klein, K. and Kim, J.: High-dimensional data visualization by interactive construction of low-dimensional parallel coordinate plots, *Journal of Visual Languages & Computing*, Vol.43, pp.1-13 (2017).
- [7] Zhou, Z., Ye, Z., Yu, J. and Chen, W.: Cluster-aware arrangement of the parallel coordinate plots, *Journal of Visual Languages & Computing*, Vol.46, pp.43-52 (2018).
- [8] Herschel, M., Diestelkämper, R. and Lahmar, H.B.: A survey on provenance: What for? What form? What from?, *The VLDB Journal*, Vol.26, No.6, pp.881-906 (2017).
- [9] Waldner, M., Bruckner, S. and Viola, I.: Graphical Histories of Information Foraging, *Proc. 8th Nordic Conf. Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14)*, pp.295-304 (2014).
- [10] Mindek, P., Bruckner, S. and Gröller, M.E.: Contextual Snapshots: Enriched Visualization with Interactive Spatial Annotations, *Proc. 29th Spring Conf. Computer Graphics (SCCG '13)*, pp.49-56 (2013).
- [11] Gratzl, S., Gehlenborg, N., Lex, A., Pfister, H. and Streit, M.: Domino: Extracting, Comparing, and Manipulating Subsets Across Multiple Tabular Datasets, *IEEE Trans. Visualization and Computer Graphics (TVCG)*, Vol.20, No.12, pp.2023-2032 (2014).
- [12] Grinstein, G., Trutschl, M. and Cvek, U.: High dimensional visualizations, *Proc. KDD Workshop on Visual Data Mining* (2001).
- [13] 岡本辰夫, 小山嘉紀, 松田敏之, 池田 隼, 古川 文, 横田一正：美術作品の素材要素検索による興味喚起と鑑賞を支援するパーツミュージアムの開発と評価，日本データベース学会論文誌，Vol.7, No.4, pp.19-24 (2009).
- [14] Wakita, W. and Tanaka, H.T.: A Digital Archiving for Large 3D Woven Cultural Artifacts Exhibition, *ITE Trans. Media Technology and Applications*, Vol.2, No.3, pp.236-247 (2014).
- [15] 安達文夫, 青山宏夫, 田中紀之, 徳永幸生：時間と場所の情報を有する大量の写真資料の提示法，国立歴史民俗博物館研究報告，Vol.189, pp.41-77 (2015).
- [16] Yoshino, K., Obata, K. and Tokuhisa, S.: FLIPPIN': Exploring a Paper-based Book UI Design in a Public Space, *Proc. 2017 CHI Conf. Human Factors in Computing Systems (CHI '17)*, pp.1508-1517, ACM (2017).
- [17] Thudt, A., Hinrichs, U. and Carpendale, S.: The Bohemian Bookshelf: Supporting Serendipitous Book Discoveries through Information Visualization, *Proc. SIGCHI Conf. Human Factors in Computing Systems (CHI '12)*, pp.1461-1470, ACM (2012).
- [18] An, L., Zhou, L., Lin, X. and Yu, C.: Visual Topical Analysis of Museum Collections, *Proc. 17th Int'l Conf. Asia-Pacific Digital Libraries (ICADL2015)*, pp.1-11, Springer-Verlag (2015).
- [19] Whitelaw, M.: Generous Interfaces for Digital Cultural Collections, *DHQ: Digital Humanities Quarterly*, Vol.9, No.1 (2015).
- [20] Odijk, D., Gârbacea, C., Schoegje, T., Hollink, L., de Boer, V., Ribbens, K. and van Ossenbruggen, J.: Supporting Exploration of Historical Perspectives Across Collections, *Proc. 19th Int'l Conf. Theory and Practice of Digital Libraries (TPDL2015)*, pp.238-251, Springer-Verlag (2015).
- [21] Hinrichs, U., Forlini, S. and Moynihan, B.: Speculative Practices: Utilizing InfoVis to Explore Untapped Literary Collections, *IEEE Trans. Visualization and Computer Graphics (TVCG)*, Vol.22, No.1, pp.429-438 (2016).
- [22] Glinka, K., Pietsch, C. and Dörk, M.: Past Visions and Reconciling Views: Visualizing Time, Texture and Themes in Cultural Collections, *DHQ: Digital Humanities Quarterly*, Vol.11, No.2 (2017).
- [23] 飯島護久, 堀口良太：プローブデータに基づくエリア流動性情報提供に関する研究，第 9 回 ITS シンポジウム 2010 論文集，pp.1-4 (2010).
- [24] 平岡敏洋, 西川聖明, 川上浩司, 塩瀬隆之：自発的な省燃費運転行動を促すエコドライブ支援システム，計測自動制御学会論文集 = Trans. Society of Instrument and Control Engineers, Vol.48, No.11, pp.754-763 (2012).
- [25] Wilmlink, I., Olstam, J., van Noort, M. and Choudhury, C.F.: Microscopic Simulation of Eco-driving Systems Using Real-world Vehicle Models, *22nd ITS World Congress, Bordeaux*, pp.1-12 (2015).
- [26] Komamizu, T., Amagasa, T. and Kitagawa, H.: Visual Spatial-OLAP for Vehicle Recorder Data on Micro-sized Electric Vehicles, *Proc. 20th Int'l Database Engineering & Applications Symposium (IDEAS '16)*, pp.358-363, ACM (2016).
- [27] 増永良文：リレーショナルデータベース入門—データモデル・SQL・管理システム，サイエンス社 (2003).
- [28] Lind, M., Johansson, J. and Cooper, M.: Many-to-Many Relational Parallel Coordinates Displays, *13th Int'l Conf. Information Visualisation*, pp.25-31 (2009).
- [29] 国立歴史民俗博物館 (編)：国立歴史民俗博物館資料図録 2 野村コレクション小袖屏風，国立歴史民俗博物館 (2002).
- [30] 萩生田明徳, 田中友章, 大門利都子, 富井尚志：専門家の知見を用いた小袖屏風閲覧システムの構築と展示のストーリーに沿った情報提示，第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM2016)，A4-1, pp.1-8 (2016).
- [31] 濱崎裕太, 大門利都子, 田中友章, 富井尚志：デジタルアーカイブを SPJ 質問により統合する情報可視化システム，第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM2018)，E5-3, pp.1-8 (2018).
- [32] 丸山伸彦：日本ビジュアル生活史 江戸のきものと衣生活，小学館 (2007).
- [33] 国立歴史民俗博物館 (編)：近世きもの万華鏡—小袖屏風展，朝日新聞社 (1994).

- [34] 植村 智明, 濱崎 裕太, 齊藤 祐亮, 吉瀬 雄大, 磯部 康太, 有長 拓海, 富井 尚志: EV 消費エネルギーデータベースの構築と連続する道路間の相関を考慮したデータ分析手法, 第 11 回データ工学と情報マネジメントに関するフォーラム (DEIM2019), E7-5, pp.1-8 (2019).

付 録

A.1 BNF 記法による (PC)²L の文法定義

```

<Query> ::= <SelectClause> <FromClause> <WhereClause>
          <ColorByClause>
          | <SelectClause> <FromClause> <ColorByClause>
          | <SelectClause> <FromClause> <WhereClause>
          | <SelectClause> <FromClause>
<SelectClause> ::= 'SELECT' <ColumnList>
<FromClause> ::= 'FROM' <JoinList>
<WhereClause> ::= 'WHERE' <ConditionList>
<ColorByClause> ::= 'COLOR' 'BY' <ColumnName>
<ColumnList> ::= <ColumnName> ',' <ColumnList>
               | <ColumnName>
<JoinList> ::= <RelationName> <JoinClause>
               | <RelationName>
<ConditionList> ::= '(' <AxisCondition> ')'
                 | <LogicalOperator> <ConditionList>
                 | '(' <AxisCondition> ')'
<JoinClause> ::= <JoinStatement> <JoinClause>
               | <JoinStatement>
<AxisCondition> ::= <Condition> <AxisCondition>
                 | <Condition>
<LogicalOperator> ::= 'AND' | 'OR'
<JoinStatement> ::= <JoinMethod> <RelationName>
                 | 'ON' <JoinCondition>
<Condition> ::= <ColumnName> <Range>
<JoinMethod> ::= 'INNER' 'JOIN'
               | 'LEFT' 'JOIN'
               | 'LEFT' 'OUTER' 'JOIN'
               | 'RIGHT' 'JOIN'
               | 'RIGHT' 'OUTER' 'JOIN'
               | 'FULL' 'JOIN'
               | 'FULL' 'OUTER' 'JOIN'
<JoinCondition> ::= <LeftCondition> <Theta>
                 | <RightCondition>
<Range> ::= 'BETWEEN' <NumberLiteral> 'AND'
           <NumberLiteral>
<LeftCondition> ::= <ColumnName>
<Theta> ::= '=' | '>' | '<' | '>=' | '<=' | '<>' | '!='
<RightCondition> ::= <ColumnName>
<NumberLiteral> ::= DecimalLiteral | FloatLiteral
<ColumnName> ::= <RelationName> '.' <AxisName>
<RelationName> ::= '[' UnescapedIdentifier ']' | Identifier
<AxisName> ::= '[' UnescapedIdentifier ']' | Identifier
    
```

ただし, DecimalLiteral, FloatLiteral, UnescapedIdentifier, Identifier はそれぞれ終端記号を表す. これら終端記号の定義は, 以下の正規表現で表す.

```

DecimalLiteral = [+]?([1-9][0-9]*|0)
FloatLiteral = [+]?([1-9][0-9]*|0)'.[0-9]+
UnescapedIdentifier = [^\[\]x00-x1f\x7f0-9]
                    [^\[\]x00-x1f\x7f]*
Identifier = [^!,<>=().\[\]'\\x00-x20\x7f0-9]
            [^!,<>=().\[\]'\\x00-x20\x7f]*
    
```



濱崎 裕太 (正会員)

2019 年横浜国立大学大学院環境情報学府博士課程前期修了. 2017 年横浜国立大学理工学部数物・電子情報系学科卒業. マルチメディアデータベース, 時空間データベースの研究・開発に従事.



植村 智明 (学生会員)

横浜国立大学大学院環境情報学府博士課程後期在学中. 2016 年横浜国立大学理工学部数物・電子情報系学科卒業. 2018 年横浜国立大学大学院環境情報学府博士課程前期修了. マルチメディアデータベース, 時空間データベースの研究・開発に従事. 日本データベース学会学生会員.



富井 尚志 (正会員)

横浜国立大学大学院環境情報研究院准教授. 1999 年横浜国立大学大学院工学研究科博士課程後期修了. 博士(工学). マルチメディアデータベース, 時空間データベースの研究に従事. 電子情報通信学会, 映像情報メディア学会, 日本データベース学会会員.

(担当編集委員 是津 耕司)