

データバイアスを俯瞰するための階層型データ可視化

空調の温感の男女差への応用

中井 祐希[†] 伊藤 貴之[†] 高橋 秀和^{††} 中島 哲^{††} 山本 哲^{††}

[†] お茶の水女子大学 〒112-8610 東京都文京区大塚 2-1-1

^{††} 富士通 AI 倫理研究センター

E-mail: †{g1820524,itot}@is.ocha.ac.jp

あらまし データのバイアスは定量的な手法のみでの判断が難しい場合もあり、またその解決には人間による意思決定が必要である。可視化技術により人間がデータのバイアスを理解することで、この課題を支援できると考えられる。本報告では、複数の属性がもたらす交差バイアスを発見するための可視化手法を提案する。複数の属性によってデータを分類して階層構造を生成し、帯グラフを搭載した階層型データ可視化手法を適用することで、データ全体を俯瞰しつつ、データ中の特定の部分集合に見られるバイアスを発見しやすい可視化を実現する。本報告では空調の温感の男女差に関する適用事例を紹介するとともに、ベースライン実装に対する優位性を検証するための被験者実験結果を示す。

キーワード 可視化, データバイアス, 男女差, 階層型データ

1 はじめに

機械学習やデータサイエンスの普及にともない、データのバイアスがもたらす問題が注目されるようになった。例えば機械学習に用いる訓練データにバイアスがあることで、学習結果にもバイアスが生じることがあり、これが社会的に大きな問題をもたらすことがある。その例として、企業の採用人事に AI を導入したら女性が不利な結果が生じた、クレジットカード発行や住宅ローン貸付が許可される割合に男女差が生じた、といった問題が公表されている。

データのバイアスを発見するには、データが有する属性ごとの数値分布の違いを観察することが重要である。ここでいう属性とは例えば、性別、地域、人種、世代などを含む。そして特定の属性に起因する不利益を解消することは、公平性の高い社会の構築のためにきわめて重要である。特に性別に起因する利益の不均衡は近年非常に重要視されており、これを解決するための取り組みとし Gendered Innovation という新しいキーワードが注目されている。

データのバイアスはデータ全体に見られるとは限らない。むしろ特定の条件を満たす部分集合にバイアスが見られる事例が多い。端的に言えば、データを構成する人物の中で、特定の属性を有してかつ別の要因を満たす人物で構成される部分集合の中にバイアスが見られることが多い。具体的な例として「職業が〇〇である女性は住宅ローン貸付が許可されない割合が高い」といったことが起こり得る。このように、複数の属性・要因が絡んで生じるバイアスを交差バイアスという。

このようなバイアスの基準には定性的・主観的な問題を含んでおり、必ずしも定量的に判断できる問題ばかりではない。また、バイアスの解決策の模索には人間による意思決定が必要で

ある。つまり、データに潜むバイアスの発見や理解を機械に任せるのではなく、分析と解析を担当する人間がデータを理解し、データに潜むバイアスを人間が発見することが有用である。可視化技術はデータの理解を促進する技術として活発に研究されており、データのバイアスの発見の目的にも有用であると考えられる。特に、データ中の特定の部分に含まれるバイアスを発見するには、データ全体の俯瞰と、その中の特定の部分集合の発見を両立できる可視化手法が有用である。この観点から我々は、以下の要件を満たす可視化手法の開発を目標とした。

要件 1: データの全体像を俯瞰できる。

要件 2: データ中の特定の属性に起因する部分集合に交差バイアスが生じているのを発見しやすい。

要件 3: バイアスを発見しやすい視覚表現を実現できる。

本報告では、多数の人物を対象としたデータから、データの分布のバイアスを可視化する一手法を提案する。本研究では、人物をデータ要素とし、各々のレコードが各種の属性値を有するデータを仮定する。そして提案手法では、このデータ要素を属性で階層的に分割し、著者らが提案している「平安京ビュー」[1] という階層型データ可視化手法を用いて可視化する。ここで、階層を表現する長方形領域を複数の帯グラフで塗りつぶし、その帯グラフによって所定の属性値に関する数値分布を表現することで、特定の属性に起因する数値分布の違いを可視化する。例えば男女間のバイアスを可視化する場合には、男性用・女性用の帯グラフを 2 列に並べて塗りつぶす。「平安京ビュー」はデータ全体の俯瞰を目的とした手法であることから要件 1 を満たし、データ要素の属性で階層的にデータを分類して表示することから要件 2 を満たす。また、帯グラフを並べた視覚表現によって属性間のバイアスを発見しやすくすることから要件 3 を満たす。

本報告では、空調の温感に関する評価値の男女差を可視化し

た事例を報告し、本手法の有効性を議論する。

2 関連研究

2.1 データのバイアスの可視化

データのバイアスの可視化は最近 3,4 年で急速に研究が増えている課題である。機械学習の公平性が大きな社会問題として注目される中で、データのバイアスがその大きな要因であり、機械学習のプロセスに人間が積極的に介入することでその解決を図ることが有用であるとされているためである。データのバイアスの可視化に関する近年の研究の例を以下にあげる。Cabrera らによる FairVis [2] は、人種や性別などのセンシティブな属性を複合的にグループ化し、グループ間で発生する交差バイアスに注目することで、差別や不平等な学習結果を防ぐ可視化解析システムである。Ahn らによる FairSight [3] は、データ処理 (Data)、学習モデルの選択 (Model)、学習結果となるランキングの生成 (Outcome) の 3 つのフェーズを可視化することで、公平性に着目した可視化を実現する。Wang らによる DiscriLens [4] は、差別発見モジュールと可視化モジュールを組み合わせることで、機械学習におけるバイアスの包括的な視覚分析を支援する。また柘木ら [5] は、機械学習における映画推薦システムデータにおいて鑑賞履歴と推薦結果の差異を表示することで、推薦システムにおける機械学習のバイアスを可視化している。

これらのシステムでの可視化表現は単純な棒グラフ・折れ線グラフ・散布図を中心とした複数のビューで構成されており、ユーザは反復的な操作を繰り返すことでデータのバイアスを探す必要がある。一方で、情報可視化システムの多くは Schneiderman [6] が提唱した「Overview first, zoom and filter, then details on demand」という操作シナリオに則って設計されており、Overview(俯瞰)のための最初の一画面がその後の適切な操作を導く重要な要素であるとされている。この考えに則り、データを俯瞰する最初の一画面でデータ中の特定の部分に潜むバイアスを発見させることに重点を置いた、という点で本研究は上述の従来研究と大きく異なる。

2.2 交差バイアス

バイアスは必ずしもデータ全体に存在するとは限らない。むしろデータ中の部分集合にバイアスが見られる事例も多い。その中でも特に、特定の複数の属性を満たす部分集合に見られるバイアスを交差バイアスと呼ぶ。交差バイアスは見落としやすい上に、その緩和も簡単ではない。この問題への対策の一例として小林ら [7] は、One-vs-One Mitigation という手法を提案している。この手法では部分集合ペア間の比較処理を適用した機械学習により、2 値分類の公平性の改善を試みている。

2.3 階層型データ可視化手法「平安京ビュー」

伊藤らによる「平安京ビュー」[1] は、階層型データの葉ノードを長方形のアイコンで、枝ノードを長方形の枠で表現し、階層構造を 2 次元の長方形群の入れ子構造で表現し、これらができるだけ小さい画面空間に配置することで、階層型データ全

体を一画面にする。図 1 に「平安京ビュー」による可視化例を示す。

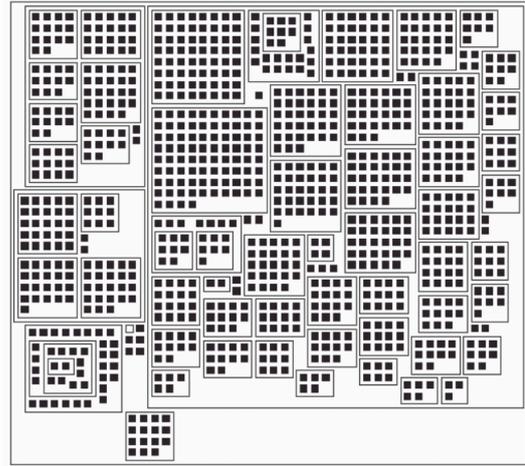


図 1 「平安京ビュー」による階層型データの可視化。

「平安京ビュー」は長方形領域とアイコン群で画面を埋め尽くすタイプの階層型データ可視化手法である。このように、長方形やアイコンなどの視覚要素で画面を埋めるタイプの可視化手法を「空間充填型 (Space-filling)」と呼ぶ。空間充填型の階層型データ可視化手法の代表例として Treemaps [8], [9] があげられる。「平安京ビュー」は Treemaps と比べて、アルゴリズムが生成する長方形領域の形状の点で優位であるという実験結果があり [1], その点で「平安京ビュー」は Treemaps と比べて本研究の目的にも合致していると考えられる。

また、長方形以外の形状領域で画面を埋めるタイプの可視化手法の例として Voronoi Treemap [10] が知られている。一方で本手法は後述する通り、帯グラフでバイアスを表現することから、長方形領域で画面を埋める必要がある。その意味からも本手法では「平安京ビュー」を採用している。

3 階層型データとしてのバイアスの可視化

3.1 想定する入力データ

提案手法では以下のデータが与えられることを前提とする。ここで A は人物集合によるデータ全体を表し、 a_i は i 番目の人物を表し、 n はデータ中の人数を表す。

$$A = \{a_1, a_2, \dots, a_n\}$$

また、 i 番目の人物に相当する a_i は以下の変数を有するものとする。ここで e_i は可視化の対象となる実数値、 g_i は i 番目の人物の性別、 r_{ij} は j 番目の実数型変数の属性値、 c_{ik} は k 番目のカテゴリ型変数の属性値である。

$$a_i = \{e_i, g_i, r_{ij}, \dots, c_{ik}, \dots\}$$

3.2 階層構造の生成と帯グラフ表示

提案手法では、属性値 r_{ij} または c_{ik} のうちユーザが選んだ複数の属性値を用いて人物群を階層的に分類し、木構造を構成する。この木構造の特定のノード配下の実数値 e_i の偏りが見られるようであれば、その偏りはユーザが選択した複数の属性値

がもたらす交差バイアスに起因する偏りであることが示唆される。提案手法ではこの木構造を「平安京ビュー」によって可視化する。図2にその構造を示す。提案手法では属性を1つずつ選び、その属性にもとづいて人物を分類し... という処理を何度か反復することで木構造を生成する。その木構造を「平安京ビュー」で可視化した結果において、外側の長方形領域が最初に選んだ属性での分類結果に対応し、内側の長方形領域が後から選んだ属性での分類結果に対応する。

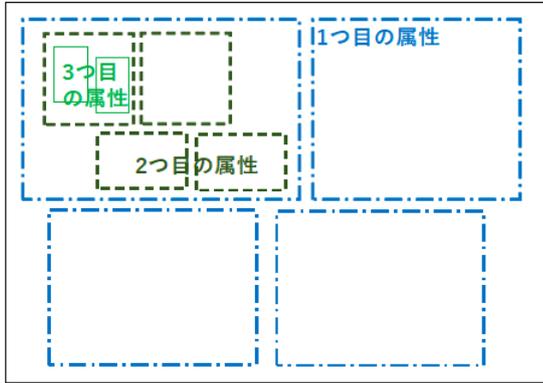


図2 複数の属性にもとづく人物群のデータの木構造表現。

ここで「平安京ビュー」では葉ノードを正方形のアイコンで表現したのに対して、提案手法では葉ノード群に相当する人物群が有する e_i の分布を複数の帯グラフで表現する。具体的には、特定の属性で人物群を分類し、各属性に対して実数値 e_i をいくつかの階級に分割し、各階級に該当する人数を集計する。そして、各属性に関する各階級の集計結果を帯グラフで表示する。現段階の我々の実装では階級数を7に固定している。

以下、男女間のバイアスの可視化を例にして、図3に帯グラフの例を示す。この例では、左側が男性の e_i の数値分布、右側が女性の e_i の数値分布を示す。帯グラフの各領域の色算出にはHSI表色系を採用し、以下の原則に沿って数値分布の各階級の色を算出する。

- 色相 (H): 属性ごとに値を割り当てる。図3の例では、男性の色相を青に、女性の色相を赤にしている。
- 彩度 (S): 各階級の中央値が全体の平均値に近いほど低い値を、最大値/最小値に近いほど高い値を割り当てる。
- 明度 (I): 各階級の中央値が大きいほど高い値を割り当てる。

図3の例では、男性より女性のほうが総数が多いこと、男性のほうが暑いと感じる人の比率が高いこと、女性のほうが寒いと感じる人の比率が高いこと、が帯グラフから読み取れる。

3.3 インタラクション

我々の実装では、階層構造を生成するための複数（現時点での実装では3種類）の属性を選択するためのラジオボタンを搭載している。ここで3つの属性を選択して「更新」ボタンを押すたびに、階層構造を再生成して表示を更新する。また、帯グラフの上にカーソルをホバリングすることで、対応する3つの属性を文字表示する。

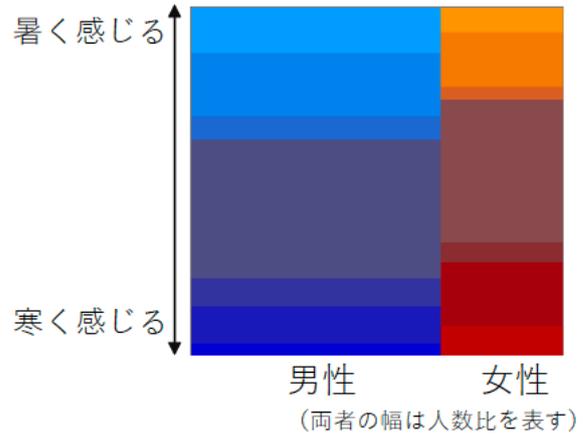


図3 複数の帯グラフによる属性間のバイアスの表現。

4 空調温感データでの適用事例

本報告では空調の温感に関するオープンデータ [11] を適用した事例を示す。このデータから著者らは 32,373 人を対象として以下の属性値を抽出した。

TS: 温感に対する評価値。0 がちょうどよい、正値が暑い、負値が寒い、という回答を表す。

Sex: 生物学的な意味での性別。なお「男性」「女性」以外で回答した人物は本事例では対象外とした。

Age: 年齢。

Cloth: 服装の厚さに関する実数値。大きいほど厚い。

Metab: 代謝量に関する実数値。

Season: 春夏秋冬の4種類のカテゴリ値。

Building: オフィス・教室・住居・高齢者施設・その他の5種類のカテゴリ値。

Strategy: エアコン・換気・混合の3種類のカテゴリ値。

以上をまとめると、TS が e_i に、Sex が g_i に、Age, Cloth, Metab が r_{ij} に、Season, Building, Strategy が c_{ij} に対応する。

4.1 可視化例 (1)

図4は Building, Season, Cloth の順に属性値を参照して人物を分類した可視化結果である。可視化結果の外側には Building の5種類のカテゴリ値に対応する長方形の枠があり、その内側には春夏秋冬の4種類のカテゴリ値に対応する長方形の枠があり、さらにその内側には代謝量の4つの区間に対応する長方形領域がある。可視化結果上部にあるオフィスの枠の内側の帯グラフはいずれも、濃い青よりも濃い赤のほうが大きい領域を占めていることから、男性よりも女性のほうが「寒い」と判断している人が軒並み多いことがわかる。その内側の (a) の枠は夏に対応する枠であり、さらにその内側にある4組の領域は (1) が最も薄着で (4) が最も厚着であることを示す。この帯グラフの太さから、夏に薄着である人物は女性に多く、厚着である人物は男性に多いことがわかる。また配色の比率から、特に (3)(4) では水色の領域が大きいことから男性が「暑い」と判断したのに対して、濃い赤の領域が大きいことから女性が「寒い」

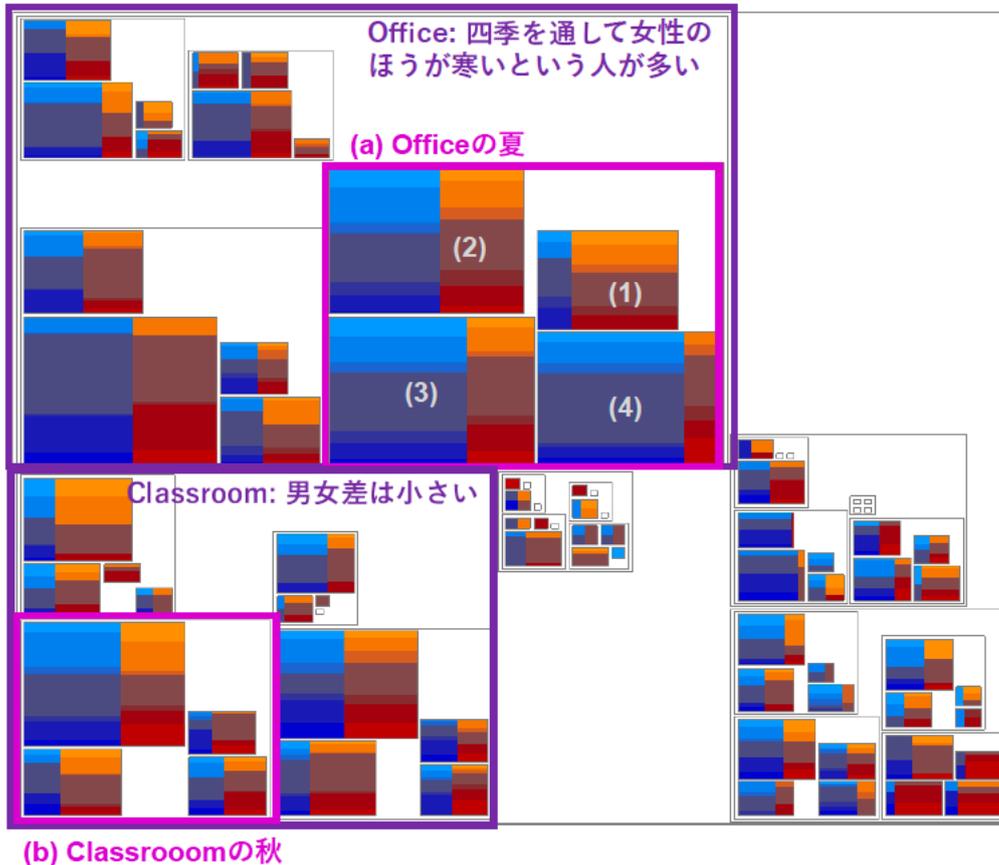


図4 可視化例 (1). Building, Season, Cloth の順に人物を分類した.

と判断しており、両者の判断が食い違う傾向にあることがわかる。一方で可視化結果左下分にある教室の枠の内側の帯グラフは、水色と橙色の面積、濃い青と濃い赤の面積に大差がないことから、教室では温感の男女差があまり大きくないことがわかる。ただしその内側の (b) の枠に対応する秋では、服装の厚さによっては男女間で温感の判断が複雑に食い違っていることがわかる。

4.2 可視化例 (2)

図5はStrategy, Season, Buildingの順に属性値を参照して人物を分類した可視化結果である。可視化画面左上部にあたる換気の枠の内側をみると、夏に対応する(a)の枠では水色や橙色の領域が濃い青や濃い赤の領域よりも軒並み大きく、男女ともに「暑い」と判断している人が多いことがわかる。逆に、冬に対応する(b)の枠では水色や橙色の領域よりも濃い青や濃い赤の領域のほうが軒並み大きく、男女ともに「寒い」と判断する人が多いことがわかる。可視化画面左下部にあたるエアコンの枠の内側をみると、ほぼ全ての帯グラフにおいて濃い青より濃い赤の領域のほうが大きく、季節や場所を問わず女性のほうが「寒い」と判断する人が多いことがわかる。可視化画面右下部にあたる混合の枠の内側をみると、秋に対応する(c)の枠と春に対応する(d)の枠において濃い青より濃い赤の領域のほうが大きい帯グラフのほうが多く、女性のほうが「寒い」と判断する人が多いことがわかる。一方で夏と冬では男女間の判断の

差が小さい傾向があることから、夏と冬ではエアコンと換気の混合が望ましいことが示唆される。

5 被験者実験による評価

提案手法の有効性を検証するために被験者実験を実施した。被験者は情報科学を専攻する学生12名であった。本実験では提案手法(以下Pと称する)による可視化結果の他に、以下の2種類のベースライン実装(以下B1, B2と称する)による可視化結果を用意した。

B1: 平安京ビュー [1] をそのまま適用した可視化結果。帯グラフは表示されず、かわりにデータ中の人数だけアイコンが表示される。

B2: 帯グラフの集合のみを表示し、階層構造を表す長方形の枠を一切表示しない可視化結果。

可視化例の生成には前章で紹介した空調温感データを用いた。

被験者にはP, B1, B2の3種類の可視化結果をランダムな順番で提示した後、各々の可視化結果について以下の質問Q1~Q8に対して7段階のリッカート尺度(1:全くそう思わない~7:強くそう思う)で回答してもらった。このうちQ1, Q2は要件3を、Q3, Q4は要件2を、Q5は要件1を検証するために設定した質問である。またQ6~Q8はNASA Task Load Index (NASA-TLX) から抜粋した質問である。またQ6~Q8は回答値が大きいほどネガティブな結果を示すことから上付きの-をつけて表記した。

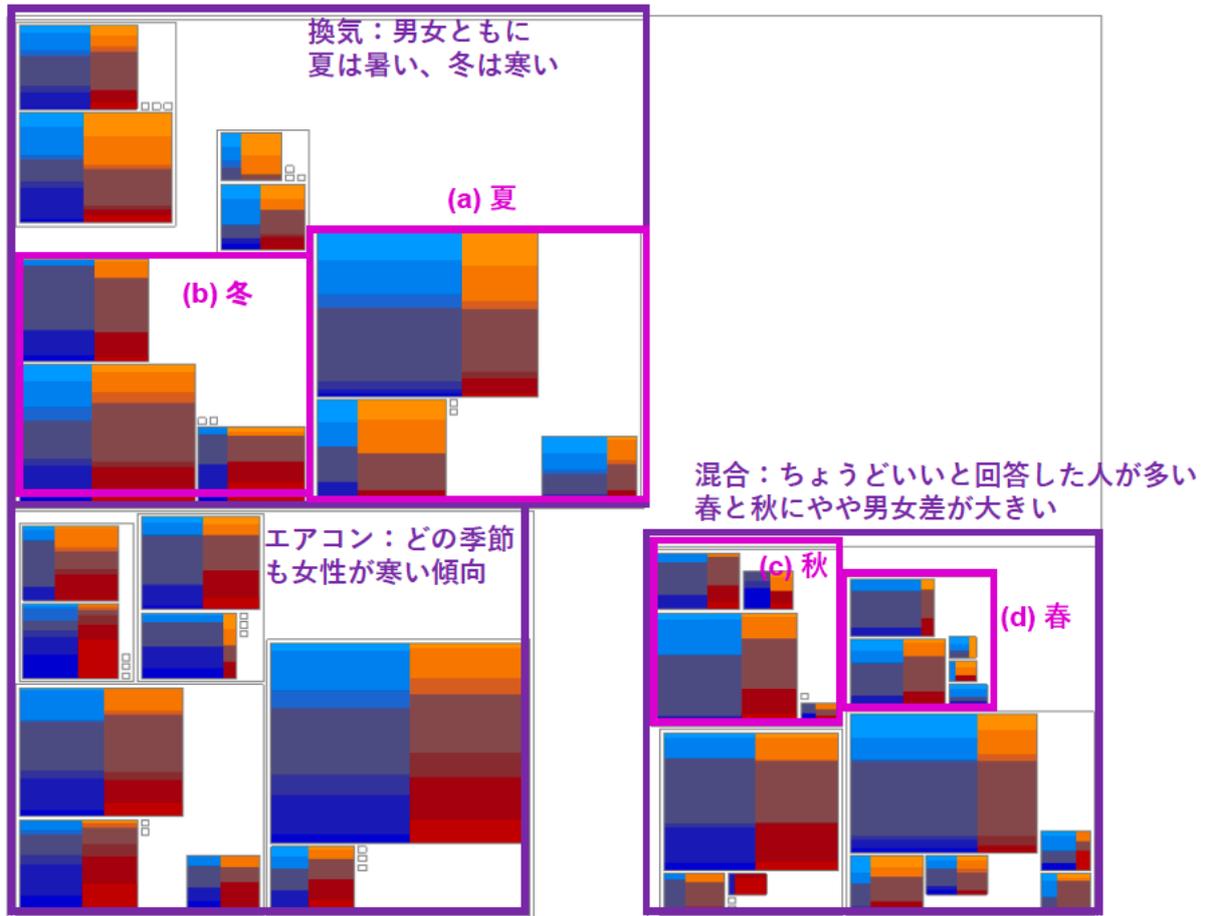


図5 可視化例(2). Strategy, Season, Building の順に人物を分類した。

Q1: どの3つの属性の組み合わせが「男性が多い」「女性が多い」かを見つけやすい

Q2: どの3つの属性の組み合わせが「暑い/寒いという男性が多い」「暑い/寒いという女性が多い」かを見つけやすい

Q3: 3つの属性のうちどの単一の属性が男女間の回答差の要因となっているかを理解しやすい

Q4: 3つの属性のうちどの「複数の属性の組み合わせ」が男女間の回答差の要因となっているかを理解しやすい

Q5: 温感の男女差の俯瞰と要因分析のための可視化として満足できる

Q6⁻: この可視化を利用する際に熟考や記憶などの知覚的活動が必要である

Q7⁻: この可視化を理解するために大きな努力が必要である

Q8⁻: この可視化を眺めている間にフラストレーションを感じる

回答を間隔尺度とみなして平均を算出した結果を表1に示す。またPとB1, PとB2に対してWilcoxonの順位検定を適用し、 $p < 0.05$ が成立した結果に対して*印をつけた。

Q1~Q5の結果はいずれも、提案手法(P)がベースライン(B1, B2)に比べて有意に高評価であった。このことから、本稿で掲げた要件1,2,3を満たすために「階層構造の表現」「帯グラフ」の両者が有効であることが示唆される。

一方でQ6~Q8の結果はいずれも、提案手法はベースライン

表1 ベースラインとの比較のための被験者実験結果。

	P	B1	B2
Q1	6.33	3.25*	5.42*
Q2	5.17	2.33*	4.50*
Q3	4.92	2.25*	3.33*
Q4	4.42	2.00*	3.33*
Q5	5.67	2.50*	3.92*
Q6 ⁻	4.33	4.83	4.42
Q7 ⁻	3.92	5.33	4.83
Q8 ⁻	2.92	5.50*	4.00

より高評価であるものの、有意とは言い切れない結果となった。本研究は日常業務などでデータ分析に従事するユーザを主たるターゲットとすることから、その意味では「知覚的活動が必要である」「努力が必要である」という点は大きな問題ではないと考える。一方で、この可視化結果を一般消費者などに広く提示する目的で利用することがあるとしたら、知覚的活動や努力の必要性がその普及を妨げる可能性がある。言い換えれば、一般消費者などにもターゲットユーザを広げる場合には、可視化デザインに対して一層の工夫が必要であろう。

6 まとめ・今後の課題

本報告では多数の人物を対象としたデータ中に潜む交差バイ

アスを階層型データとして可視化する一手法を提案した。本手法では、データを構成する人物群を属性に沿って階層的に分類し、可視化手法「平安京ビュー」を用いて可視化する。この際に、階層構造の特定のノード配下に属する人物群が有する数値を複数の帯グラフで表現することで、数値分布のバイアスを表現する。本報告では空調の温感データを題材として可視化結果を示すとともに、被験者実験結果を示した。

今後の課題として、非常に多くの属性を有するデータへの対応が考えられる。提案手法では階層構造を生成するための属性の選択操作が必要であるが、非常に多くの属性を有するデータではこの操作が非常に煩雑になる。そこで、可視化する価値のある属性の組み合わせを自動選出する手法を開発したい。

続いて、空調の温感以外の多様なデータに対して可視化を試みると同時に、性別以外の属性（例えば世代、地域、人種、職業など）に起因する数値分布の偏りの可視化に取り組みたい。さらに、機械学習の訓練データの偏りを是正するためのツールとして提案手法を活用し、機械学習の現場における有効性についても検証したい。

本稿で紹介した被験者実験結果は、可視化結果を静止画で提示したものであり、可視化システムを実際に操作させてはいない。そこで、何らかのタスクを与えて可視化システムを実際に操作してその達成度を図るユーザ実験にも取り組みたい。

文 献

- [1] Takayuki Itoh, Yumi Yamaguchi, Yuko Ikehata, and Yasumasa Kajinaga. Hierarchical data visualization using a fast rectangle-packing algorithm. *IEEE Transactions on Visualization and Computer Graphics*, 10(3):302–313, 2004.
- [2] Ángel Alexander Cabrera, Will Epperson, Fred Hohman, Minsuk Kahng, Jamie Morgenstern, and Duen Horng Chau. Fairvis: Visual analytics for discovering intersectional bias in machine learning. In *2019 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 46–56. IEEE, 2019.
- [3] Yongsu Ahn and Yu-Ru Lin. Fairsight: Visual analytics for fairness in decision making. *IEEE transactions on visualization and computer graphics*, 26(1):1086–1095, 2019.
- [4] Qianwen Wang, Zhenhua Xu, Zhutian Chen, Yong Wang, Shixia Liu, and Huamin Qu. Visual analysis of discrimination in machine learning. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1470–1480, 2020.
- [5] Ami Tochigi, Takayuki Itoh, and Xiting Wang. Visualization of bias of machine learning for content recommendation.
- [6] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*, pages 364–371. Elsevier, 2003.
- [7] Kenji Kobayashi and Yuri Nakao. One-vs.-one mitigation of intersectional bias: A general method for extending fairness-aware binary classification. In *International Conference on Disruptive Technologies, Tech Ethics and Artificial Intelligence*, pages 43–54. Springer, 2021.
- [8] B Johnson and B Shneiderman. Tree-maps: A space filling approach to the visualization of hierarchical information space. In *IEEE Visualization*, volume 91, pages 275–282, 1991.
- [9] Benjamin B Bederson, Ben Shneiderman, and Martin Wattenberg. Ordered and quantum treemaps: Making effective use of 2d space to display hierarchies. *AcM Transactions*

on Graphics (TOG), 21(4):833–854, 2002.

- [10] Michael Balzer and Oliver Deussen. Voronoi treemaps. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 49–56. IEEE, 2005.
- [11] Ashrae global thermal comfort database ii, <https://www.kaggle.com/datasets/claytonmiller/ashrae-global-thermal-comfort-database-ii>.