

2 変数詳細度制御を用いた大規模データの可視化

Large-Scale Data Visualization with Two-Variate Level-of-Detail Control

長崎 あずさ 伊藤 貴之 伊勢 昌之 宮下 光輔*

Summary. 私たちの身の回りには多数の大規模データが存在する。コンビニでの買い物や、ICチップに記録される行動の経路など、私たちの暮らしは今や大量のデータを収集されることが日常的となっている。しかしそれらのデータは非常に大規模かつ複雑で、解析が困難なものがあまりにも多い。その理由としては、データ自体の莫大さのみならず、特にデータベース化したときの属性数が多すぎるものが挙げられる。たとえばコンビニで買い物をしただけでも、購買者個人情報や商品情報、コンビニの位置情報、購買日時や天気など、得られる情報は非常に多岐に渡りそして膨大である。このような背景からデータの属性数が非常に多くなるために、データ全体の解析への需要は高いにも関わらず、解析が困難な事例はまだ多い。

本報告では、このようなデータの解析を支援する一手段として、2属性詳細度制御を適用した大規模データ可視化手法を提案する。本報告において大規模データとは、レコード数が多いだけでなく、前述のように属性数が多いデータを指す。このようなデータは、そのまま可視化しただけではデータの特徴や傾向を掴むことが非常に難しい。よって本研究では、データの特性を保ったままユーザに提示する情報量を減らして可視化することで、大規模データをユーザにとって解析しやすいものとするを目標とする。

1 はじめに

情報可視化は日常の一般的な情報を可視化する技術である。私達の身の回りの情報は急速に巨大化・複雑化しており、大規模な情報を深く理解することは日々難しくなっている。このことから、複雑な情報の中から特徴的な現象や全体的な傾向をユーザにわかりやすく提示する技術を確立することは、情報可視化の重要な課題の一つであるといえる。

本手法では、情報可視化手法「平安京ビュー」[1]の機能拡張として、2属性による詳細度制御を提案する。2属性による詳細度制御を組み合わせることで、データの特性を保ったままユーザに提示する情報量を減らすことが可能となる。それにより、深い理解が難しいような大規模かつ複雑なデータに対して、概略的な内容把握を支援できる。

2 関連研究

提案手法が前提とする可視化手法「平安京ビュー」は、大規模階層型データ可視化手法である。平安京ビューは、階層型データの葉ノードをアイコンで表示し、枝ノードを長方形の枠で表示することで、階層型データの全体を一画面に表現することができる手法である。階層型データの葉ノードが複数の属性を有するとき、平安京ビューでは色・高さ・グループに割り当てる属性を自由に選ぶことができ、これに

よってデータ中の特定の属性間の相関性などを表現できる。本研究では色とグループに着目し、それらに割り当てる2属性に基づく詳細度制御を実現した。

提案手法にて色の詳細度制御のために適用するslice-and-diceアルゴリズムは、treemaps[2]に採用されている。slice-and-diceアルゴリズムは、2次元平面を複数の長方形で分割することで、帯グラフを入れ子にしたような構造で階層型データを表現する。枝ノード配下における葉ノードの値の大きな分布がわかりやすいことが特徴である。

3 提案手法

提案手法ではまず、特定の属性 A_p でレコードをグループ化することにより、大規模データから階層型データを構築する。そして、これとは別の属性 A_q を平安京ビューの色に割り当てることにより、平安京ビューにて2属性を眺めることが可能になる。続いて、2属性による詳細度制御を適用して、平安京ビューで可視化する。

一つ目の属性の詳細度制御として、特定の階層における属性 A_q の数値分布を表示する。特定階層の枝ノードに属する子ノードを全て結合して一つのノードとして扱い、結合する前のノードの属性 A_q の数値分布を treemaps[2] のように長方形を分割して描くことにより、これを実現する。

二つ目の属性の詳細度制御として、同階層における枝ノード同士の統合により、詳細度制御を実現する(図1参照)。同階層において、属する葉ノード数が一番多い枝ノードに比べて、属する葉ノード数が無視できるくらい少ない任意の枝ノードがあれば、

Copyright is held by the author(s).

* Azusa Nagasaki and Takayuki Itoh, お茶の水女子大学大学院, Masayuki Ise and Kousuke Miyashita, 株式会社インテリジェントウェイブ

その枝ノードを統合する．また，それぞれの枝ノードに属する葉ノード群間の属性 A_q の数値分布が類似していたら，枝ノード同士を統合して一つの枝ノードとして表す．この二つの観点から枝ノードを統合することにより，詳細度制御を実現する．

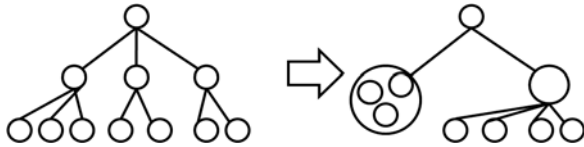


図 1. 2つの属性による詳細度制御

3.1 色に割り当てる属性に基づく制御

色に割り当てる属性 A_q の詳細度制御として，特定階層における属性 A_q の数値分布を表示する．特定の枝ノードに属する子ノードを全て結合して一つのノードとして扱い，結合する前のノードの属性 A_q の値の分布を treemaps のように長方形を分割して描くことにより実現する（図 2,3 参照）．

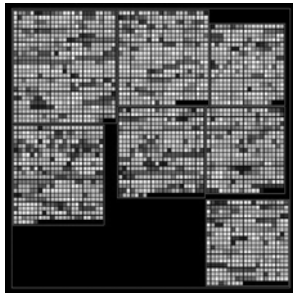


図 2. 平安京ビューによるクレジットカードの購買情報の可視化例

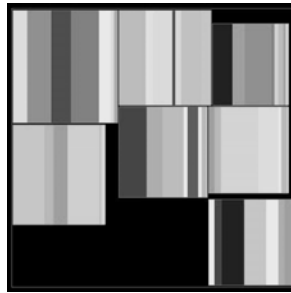


図 3. 図 2 に色による詳細度制御を適用した可視化例

3.2 グループに割り当てる属性に基づく制御

グループに割り当てる属性 A_p に基づく制御では，以下の 2 手法を適用する．また，これによる可視化例を図 4,5 に示す．

3.2.1 ノード数に基づく制御

同階層において，属する葉ノード数が一番多い枝ノードに比べて，属する葉ノード数が無視できるぐらい少ない任意の枝ノードがあれば，その枝ノードを統合する．

3.2.2 枝ノード間の類似性に基づく制御

それぞれのグループに属する葉ノード群間の属性 A_q の数値分布が類似していたら，グループ同士を統合して一つのグループとして表す．類似性の判定方法を以下に示す．

1. データ全体を平安京ビューのグループに割り当てる属性でグループ分けをする
2. それぞれのグループにおいて，色に割り当てる属性の値のヒストグラムを作成する
3. 任意のグループ G_i と G_j のヒストグラムの余弦を計算する
4. 3 の処理を全ての 2 つのグループの組み合わせについて適用する
5. 計算結果の値が高い順にグループを統合する
6. 1~5 の処理を全てのグループに割り当てる属性と色に割り当てる属性の組み合わせについて適用する

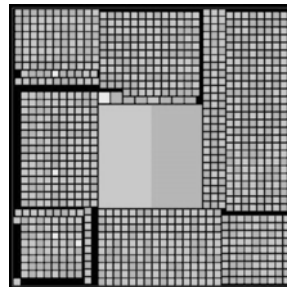


図 4. 平安京ビューによるグループ数が多いデータの可視化例

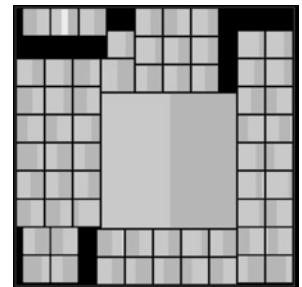


図 5. 図 4 にグループによる詳細度制御を適用した可視化例

4 まとめと今後の課題

本論文では，2変数詳細度制御を用いた大規模データの可視化の一手法として，2変数詳細度制御により大規模データへの理解・認識の向上を目指す手法を提案した．

今後の課題には，詳細度制御を用いることにより失われる情報量の考察，効率的なデータ理解を支援するためのユーザインタフェースの拡充，より多くの変数による詳細度制御の実現が挙げられる．

参考文献

- [1] 伊藤, 山口, 小山田. 長方形の入れ子構造による階層型データ視覚化手法の計算時間および画面占有面積の改善. 可視化情報学会論文集, Vol. 26, No. 6, pp. 51-61, 2006.
- [2] Johnson B., Shneiderman B., Treemaps: a space-filling approach to the visualization of hierarchical information structures. In *Proceedings of IEEE Visualization '91*, pp. 284-291, 1991.