

平安京ビューの詳細度制御における上位階層表現の均一化の試み

山澤舞子 五味愛 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)

Unification of Higher-level Representative Nodes for Level-Of-Detail of Hierarchical Data Visualization Technique “HeiankyoView”

Maiko YAMAZAWA, Ai Gomi, and Takayuki ITOH

ABSTRACT

According to recent growth of information technology, information stored by computers became larger, various, and complicated. Information visualization technique is an active research topic to understand the contents of such information. This thesis discusses the usefulness of JunihitoeView, a technique for visualizing large-scale hierarchical multivariate data. The technique is an extension of HeiankyoView, an existing technique for visualizing large-scale hierarchical data. This thesis shows the results of user tests of JunihitoeView. Based on the discussion of the results of user tests, we found that it may be hard for the users to search the data when the representative icons of the JunihitoeView are unequal. We therefore developed an algorithm to display the equalized representative icons, which show the upper-level of the hierarchy of the data, in HeiankyoView. This thesis presents a technique to display the equalized representative images in CAT, a technique for browsing a large number of images, as an example of the developed algorithm. This technique temporally ignores the numbers of the images in the clusters showed by the representative images, so that it displays all the representative images equally.

Keywords: HeiankyoView, multivariate data, image browser, unification, representative nodes.

1. はじめに

我々は大規模な階層型データの可視化手法として、「平安京ビュー」[1] を発表している。また我々は、平安京ビューの応用手法として、階層型多変数データ可視化手法「十二単ビュー」[2]、大量画像の一覧可視化ブラウザ「CAT」[3] を発表している。我々はこれらの研究過程において、十二単ビューのユーザテストを実施した。その結果、十二単ビューをはじめとする平安京ビューの詳細度制御を活用して、クラスタを代表ノードで表示した際に、代表ノードが不均一なサイズや縦横比で表示されることが、視覚的情報探索において問題となりえることがわかった。

そこで本報告では、CATにおける代表画像の表示手法を例題として、平安京ビューの詳細度制御において、上位階層表現を均一化する手法を提案する。本手法では、まず画像群の階層構造を記録した元データに対して各画像および各画像クラスタの配置結果を求め、その座標値を記録する。それと同時に、元データの下位階層を切り落とした階層型データ（本報告では「階層省略データ」と称する）を生成する。そして、元データの配置結果の座標値を参照し、それに近い配置結果となるように、かつ階層省略データを構成する画像群が均一な大きさに配置されるように、階層省略データの配置を求める。本

手法ではズーム操作に伴い、階層省略データを切り替え表示することで、どの詳細度においても均一に画像群を表示する。

2. 関連研究

2.1 平安京ビュー

平安京ビュー[1]は、階層型データの全貌を一画面に表示する可視化手法である。平安京ビューでは、階層型データを構成する葉ノードを小さなアイコンで、枝ノードを入れ子構造の長方形領域で表現する。そして一階層を表現する葉ノードおよび枝ノードに対して、

- 互いに重ならない
- 画面上の占有領域を小さくする
- 画面上の占有領域の縦横比を良好な状態に保つ
- テンプレート（後述）に記述された座標値に近い場所に配置する

などの条件を満たすような画面配置を求めることで、階層型データ全体を一画面に表示する。

2.2 十二単ビュー

十二単ビュー[2]は、平安京ビューの拡張による、階層型多変数データ可視化手法である。十二単ビューでは、個々のデータ要素が n 個の変数を有するデータに対して、小さなアイコンで表現されている葉ノードを n 個以

上の小領域に分割し、その小領域ごとに固有の色を割り当てることで、各々の葉ノードが有する n 個の変数の値を表現する。さらに枝ノードは、アイコンを囲う入れ子状の長方形の枠で表現する。この考え方により十二単ビューは、数百、数千のデータ要素を有する階層型多変数データを一画面に表現する。十二単ビューによる可視化結果を図1に示す。



図1 (上) 十二単ビューによる階層型多変数データの可視化結果。(下) (上)の一部を拡大した図。

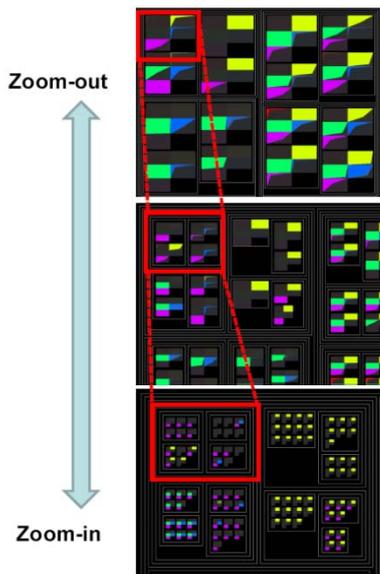


図2 ズーム操作に連動した十二単ビューの詳細度制御。

十二単ビューでは、個々のノードの変数値を正しく視認するためには、ズームイン操作が必要であることがわかる。逆に言えば、階層型データ全体を可視化するためにズームアウトしてしまうと、各々のノードは非常に小さく表示されるので、これらの変数値を視認するのは難しくなる。以上の問題点の解決のために、十二単ビューには詳細度制御機能が備わっている。階層型データ中にて同一クラスタに属する葉ノードを統合し、このクラスタを表す代表ノードとして表示することで、画面上に表示されるアイコンの数や大きさを調節する(図2参照)。

2.3 CAT

CAT[3] は、大量画像のクラスタリング手法、およびその一覧表示と詳細度制御を持ち合わせた可視化手法である。まず前処理として、大量画像を階層型に分類する。階層の上位部分では、あらかじめ画像に付与されているキーワードで分類する。以上の手順によりクラスタリングされた大量画像を、互いに重ならず、等しいサイズで一覧表示することができる。

図3のようにCATでは、画像をサムネイル表示し、サムネイルを長方形の枠で囲うことでクラスタを表現する。さらにCATは、ズーム率に合わせた詳細度制御を設けている。ズームイン時は、低階層クラスタの各々の画像サムネイルを表示する。そしてズームアウト操作に伴って、クラスタを示す長方形領域を、各クラスタの代表画像で置き換えて表示する。このように、CATは階層化された画像群に対するズーム操作によって、直感的に画像を絞り込みながら閲覧できる。



図3 CATにおける画像とクラスタの詳細度制御。

3. 十二単ビューにおける評価

本章では被験者11人により実施した十二単ビューのユーザテスト結果の一例を示し、十二単ビューの有効性について考察する。階層型多変数データを十二単ビューで可視化し、変数値に固有の特性をもつ最下位クラスタを文面で指定し、被験者にその最下位クラスタを探してもらい、その所要時間を測定した。本実験では2種類の階層型多変数データ(以下、データ1、データ2、と称する)を用いた。データ1は、詳細度制御における代表ノードの大きさや形状が均一に表示される特徴を持つデータである。一方のデータ2は、詳細度制御における代表ノードの大きさや形状が不均一に表示される特徴を持つデータである。被験者による所要時間の平均時間を表1に示す。

この結果から、データ1とデータ2を比較すると、データ1の場合の方が短時間で該当する最下位クラスタを見つけられている。つまり、代表ノードが均一に表示されている方が、十二単ビューの詳細度制御を用いて固有の特性をもつ階層を見つけることが容易だった、ということが推察される。

表1 クラスタ探しの所要時間. (単位: 秒)

| | データ1 | データ2 |
|----|------|------|
| 平均 | 21 | 82 |
| 最大 | 60 | 182 |
| 最小 | 2 | 8 |

十二単ビューの詳細度制御の目的に限らず, 平安京ビューや十二単ビューにおいて, 下位階層に属する葉ノードの数にかかわらず, 上位階層を均一なサイズの長方形で表現したい, という要求があると考えられる. 例えば平安京ビューを応用した大量画像の可視化手法CAT[3]においても, 上位階層を表現する代表画像を均一なサイズで表現したい, と考えられる.

4. CAT の代表画像の均一化

CAT の詳細度制御では, 図3(右)をみてもわかるように, 同一クラスタ内に表示される代表画像が不均一に並んでいる. これは, 代表画像のサイズや縦横比が, 各代表画像が示す下位クラスタの画像枚数によって決定されているからである. このように異なるサイズの画像が並んでしまうと, つい大きな画像が気になってしまい, 小さな代表画像で表示されるクラスタ内に属する画像を発見しづらい, という問題点が生じる. この問題点は, 前節で紹介した評価結果からも明らかである. また, この問題では, CAT以外の階層型画像一覧ブラウザ[4]でも全く同様に起こりえる問題である.

そこで本手法では, 以下の処理によって代表画像の均一化を図る (図4参照).

1. 可視化する階層型データ (元データ) の算出結果から, サムネイルの座標値を取得する.
2. 元データから下位クラスタの画像枚数の情報を切り捨てた, 上位階層のみを有する階層省略データを作成する. ここで, 元データの階層の深さが n のとき, 階層省略データとは深さの最大値が k ($0 < k < n$) になるように下位階層を切り取ったデータを指す.
3. 2. の各データにおいて, 1. の座標値を参照して, 代表画像の配置を算出する.
4. ズーム操作に従い, 3. の結果を呼び出し, ズーム率に適応した詳細度の改装詳細データを表示する.

以上の手順により, 下位階層クラスタの画像枚数の情報を切り捨て, 代表画像のサイズを均一化できる.

元データの座標値の参照には, 文献[1]に説明されている「テンプレート」という概念を導入する. 平安京ビューでは, 階層型データを構成する各ノードに対して, 理想的な画面位置を記録した理想座標値をあらかじめ用意する. そしてその理想座標値を参照しながら, 階層型データを画面配置する. この理想座標値を記録したデータを「テンプレート」と呼んでいる. 本手法では図5に示すように, まず元データにおけるノードの配置結果の

座標値をテンプレートに保存する. そして, 記録されたテンプレートの座標値を参照しながら, 階層省略データの可視化において均一なサイズの代表画像を配置する.

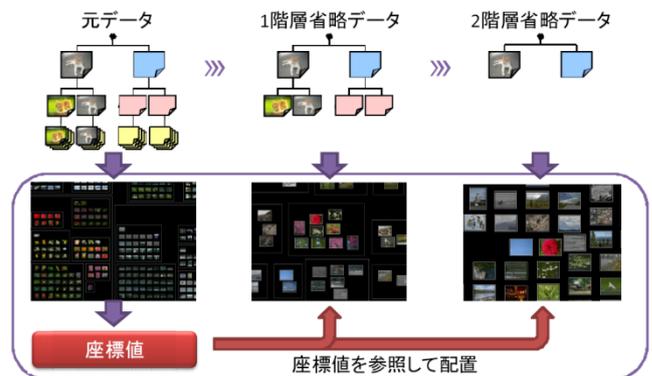


図4 処理手順の概要. 各々の階層省略データの配置は, 元データの配置結果の座標値を参照して得られる.

以上のようにテンプレートを用いて配置することで, 元データの下位クラスタと, テンプレートを参照した階層省略データの代表画像がほぼ近い位置に表示される. したがって, ズームにより表示される階層が切り替わっても, ズーム操作以前に注目していたクラスタがほぼ同じ位置に表示されているため, 注目しているクラスタを見失うことがなくなる. また, 平安京ビューのノード配置アルゴリズムにより, 可視化結果は画面占有面積を最小化し, かつ, クラスタの配置領域を画像の縦横比に近づけることができるようになる.

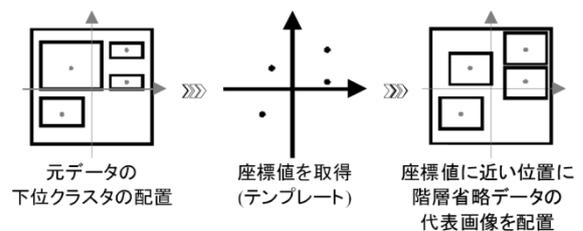


図5 テンプレートを参照して均一なサイズの代表画像を配置する処理.

図6に, テンプレートを使用した場合と使用しなかった場合の配置結果の違いを示す. 図6(中)が元データの配置結果である. それに対して, 図6(左)がテンプレートを使用した場合, 図6(右)がテンプレートを使用しなかった場合の, 階層省略データの配置結果である. ここで元データの9個のクラスタには, それぞれ異なる色で枠が描かれており, 階層省略データの各代表画像には, 元データの対応するクラスタと同じ色で枠が描かれている. この結果を見てわかるように, テンプレートを使用したほうが, 元データのクラスタに対応する階層省略データの代表画像が, クラスタに近い位置に配置されているのが確認できる.



図6 テンプレートを使用した場合と使用しなかった場合の配置結果の違い。

5. ユーザテスト

続いて、本手法の効果をユーザテストによって実証する。本実験では図7に示すように、

- 本手法を適用することで代表画像が均一に表示されたCAT。（以下「新CAT」と称する）
- 代表画像が不均一に表示された既存実装のCAT。（以下「旧CAT」と称する）

の2種類を見比べて、指定したクラスタを見つけやすい方を答えてもらった。被験者には、2種類のCATにおいて上位階層を代表画像で表示し、一画面上でデータ全体を眺められる状態で見比べてもらった。その上で本実験では、以下の質問を提示した。

【問1】 猫の画像があるクラスタを見つけやすいのはどちらか。

【問2】 2人の人間の画像があるクラスタを見つけやすいのはどちらか。

【問3】 湾の画像があるクラスタを見つけやすいのはどちらか。

そして被験者に、各問について以下のいずれかを選択させた。

- 新CAT
- 旧CAT
- どちらも変わらない。



図7 (左) 代表画像が均一に表示されているCAT。(右) 代表画像が不均一に表示されているCAT。

この結果を表2にまとめた。この結果から、全体を見て画像を探索する際には、代表画像が均一に表示された方が画像を見つけやすいといえる。特に旧CATにおいて、

代表画像が比較的小さなサイズで表示され、かつ、代表画像の縦横比が元の画像のものと異なって表示されている画像は、大きなサイズで表示される代表画像と比較して注視しづらい、という問題点があった。新CATでは特に、このような代表画像において、均一なサイズで表示することで画像の内容を解釈しやすくなったといえる。

表2 CATの代表画像の均一化における比較実験結果。(単位：人)

| | 問1 | 問2 | 問3 |
|-------|----|----|----|
| 均一 | 6 | 4 | 7 |
| 不均一 | 1 | 1 | 1 |
| 変わらない | 2 | 4 | 1 |

6. まとめ

本論文では、階層型多変数データ可視化手法「十二単ビュー」の有用性の評価を示すために、ユーザテストを行った。ユーザテストの結果から、平安京ビューにおいて、上位階層を均一なサイズの長方形で表現したい、という要求があると考えられた。そこで上位階層表現を用いている手法の一例として、大量画像の可視化手法「CAT」において、代表画像のサイズや縦横比を均一化する手法を提案した。

今後の課題として、テンプレートを用いた画面配置アルゴリズムの高速化、ズーム率に応じて階層省略データをスムーズに切り替える機能の搭載、多様な縦横比の代表画像が入り混じる画像データにおける実験、などがあげられる。

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] 山澤, 伊藤, 山下, 十二単ビュー:階層型多変数データの可視化と詳細度制御の一手法, 芸術科学会論文誌, Vol. 7, No. 2, pp. 85-96, 2008.
- [3] 五味, 宮崎, 伊藤, Li, CAT:大量画像の一覧可視化と詳細度制御のためのGUI, 画像電子学会誌, Vol. 38, No. 4, pp. 436-443, 2008.
- [4] Bederson B., B., PhotoMesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps, User Interface Software and Technology, pp. 71-80, 2001.