

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

山澤舞子¹⁾ 伊藤貴之¹⁾ 山下富義²⁾

1) お茶の水女子大学大学院人間文化創成科学研究科

2) 京都大学大学院薬学研究科

JunihitoeView: A Technique for Visualization and Level-Of-Detail Control of Hierarchical Multi-Variate Data

Maiko Yamazawa¹⁾ Takayuki Itoh¹⁾ Fumiyoshi Yamashita²⁾

1) Graduate School of Humanities and Sciences, Ochanomizu University

2) Graduate School of Pharmaceutical Sciences, Kyoto University

{maiko, itot} @ itolab.is.ocha.ac.jp yama @ pharm.kyoto-u.ac.jp

概要

情報技術の普及に伴い、計算機に蓄積される情報は膨大化・多様化・複雑化の一途を急速にたどっている。このような情報の内容を理解する一手段として、コンピュータグラフィックスを応用した情報可視化技術の研究が進んでいる。

本論文では、階層型多変数データを可視化する一手法「十二単ビュー」を提案する。本手法は、階層型データの可視化手法として既に報告されている「平安京ビュー」を拡張したものである。「平安京ビュー」が階層型データ中の各データ要素をアイコン表示したのに対して、「十二単ビュー」では個々のアイコンに割り当てられた変数の数だけの色を割り振り、この色の濃さによって各データ要素の多変数の値を表現する。

また、「十二単ビュー」の機能拡張として、拡大表示時にはデータの葉ノードを単位として変数値を表示し、縮小表示時には下位階層の変数値を統合し、上位階層のみを表示する、というような詳細度制御を実現する。

本論文では、薬物群の分子構造情報と実験値を格納したデータベースより構築される階層型多変数データを例として、「十二単ビュー」の有用性を議論する。また、いくつかのユーザテストにより、可視化結果の有効性を検証する。

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 paper presents a technique for visualizing large-scale hierarchical multi-variate data, JunihitoeView. The technique is an extension of HeiankyoView, an existing technique for visualizing large-scale hierarchical data. HeiankyoView represents data items as icons, but JunihitoeView represents them by assigning various colors to represent the values of the multi-variate by their brightness.

The paper also presents a level-of-detail control technique for JunihitoeView. The technique displays variables of a leaf node themselves, when a user zooms in the visualization result. Further, it unifies the variables and display typical values of a higher level hierarchy, when the user zooms out the visualization result.

The paper discusses the usefulness of JunihitoeView with the application of the technique for the hierarchical multi-variate data constructed from a drug database. It also shows results of user tests of the presented techniques.

1 はじめに

情報可視化が対象とするデータ構造の分類として Shneiderman は、1 次元、2 次元、3 次元、 n 次元 ($n > 3$)、時系列、階層型、グラフ、の 7 種類を提唱している [1]。これまで情報可視化技術の研究では、各々のデータ構造に対する可視化手法が独立に多数発表されてきた。最近では階層型多変数データ¹に代表されるように、上述の 7 種類のデータ構造を複合した情報を可視化する手法が、活発に提案されている。

本論文では、階層型多変数データの可視化手法「十二単ビュー」を提案する。十二単ビューは、大規模階層型データ可視化手法「平安京ビュー」[2, 3]の拡張手法である。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できる手法である。十二単ビューは、階層型データの最下位階層を構成する葉ノードに多変数が割り当てられていると仮定し、個々の葉ノードのアイコンに変数の個数だけの色相を割り当てることにより、階層構造と同時に多変数情報を表現する手法である。十二単ビューは、大規模な階層型多変数データ全体を一画面に全貌表示し、その階層構造と、変数値の分布を同時に概観できる特徴がある。

さらに本論文では、十二単ビューの機能拡張として、拡大表示時にはデータの葉ノードを単位として変数値を表示し、縮小表示時には下位階層の変数値を統合し、上位階層のみを表示する、というような詳細度制御を実現する手法を提案する。この機能拡張により、画面のサイズやユーザの操作に合った最適な情報量で、階層型多変数データを可視化できると考えられる。

本論文の後半では、薬物群の分子構造情報と実験値を格納したデータベースより構築される階層型多変数データを例として「十二単ビュー」の有用性を議論する。また、いくつかのユーザテストにより、可視化結果の有効性を検証する。

2 関連研究

2.1 多変数データの可視化手法

多変数データの可視化手法は、変数間の相関性の分析に意義のある多くの分野において有用な技術である。例として顧客データの属性間の相関性、測定情報やシミュレーション結果のパラメータ間の相関性、などを視覚的に分析する目的での適用事例が知られている。多変数データの可視化手法の代表例として、以下のようなものがあげられる。

1. 2次元直交座標系の上に変数の数だけ鉛直線を引き、その各々の上に変数値をプロットし、折れ線グラフ状に連結する手法。Parallel Coordinates [4] が特に知られている。
2. 多変数のうち 3 変数を参照して各データ要素を 3次元直交座標系にプロットし、続いて各々のプロットされた点の周囲における小さな 3次元直

交座標系にて、別の 3 変数を参照して各データ要素をプロットし...という操作を反復する手法。Worlds within Worlds [5] が特に知られている。

3. 多次元空間に多変数データ要素をプロットし、主成分分析などの手法を用いてデータの主軸を求め、そこから得られる 2次元平面に各データ要素を投影する手法。一例として Design Galleries [6] という手法がある。
4. Glyph と呼ばれる小片を用いて、その形状や色の変化により、個々のデータ要素の多変数を表現する手法。一例として Ebart らの手法 [7] があげられる。

本論文の提案手法である「十二単ビュー」では、個々のデータ要素が、画面上でクリック可能な状態で可視化され、一種のユーザーインターフェースとして活用できることを目的の一つにしている。上記の関連手法のうち 1.~3. は、画面上でデータ要素間の干渉や重なりが多く発生するため、必ずしも任意のデータ要素をクリック可能な状態で可視化できるとはいえない。

一方で十二単ビューは、個々のデータ要素を表現する小領域の色変化によって多変数を表現するという点において、Glyph を用いた手法に類似している。

2.2 階層型データの可視化手法

計算機のファイルシステム、大会社の人事組織、カテゴリ型の検索ウェブサイト、などに代表されるように、私たちの日常生活には、階層構造で整理されたデータは非常に多く存在する。このような階層型データを可視化する手法は非常に多く報告されており、主に以下のように分類される。

1. 画面空間の再帰分割による手法。代表的な手法として、画面空間を再帰的に長方形領域に分割することで階層型データを表現する Treemaps [8] があげられる。Treemaps の問題点として、画面上で細長くつぶれた長方形領域が発生するケースが多いという点があった。この改良手法は既に多く発表されているが、その中でも個々のデータ要素を同じ大きさのアイコン画像やサムネイル画像で表現する Quantum Treemap [9] は特に有用な手法である。
2. 木構造を描画する手法。代表的な手法として、Cone Tree [10], Fractal Tree [11], Hyperbolic Tree [12] があげられる。
3. 3次元空間で入れ子状に配置された直方体や球状物体により階層構造を表現し、これを半透明表示する手法。代表的な手法として Information Cube [13] や H-BLOB [14] があげられる。
4. 2次元空間で入れ子状に階層構造を構築する手法。後述する「データ宝石箱」[16, 15] およびその改良手法「平安京ビュー」[2, 3] が該当する。

本論文の提案手法「十二単ビュー」は、階層型データの親子関係を表示することよりも、階層構造の最下位に属するデータ要素を限られた画面空間に一覧表示することを目的としている。上述の階層型データ可視化手法のうち、Quantum Treemap, データ宝石箱,

¹ 今後本論文では、 n 次元データを「多変数データ」と称する。

平安京ビューの各手法は、「画面空間にできるだけ隙間無く、しかも重なることなく、等しい大きさでデータ要素を表現する図形を画面に配置する」という点で、この目的に合致した手法である。これらの手法は文献 [16, 2] にて客観評価されており、平安京ビューは計算時間、長方形領域の形状のよさ、類似データの可視化結果の類似性、の各項目において他の手法よりも良好な結果を得ている。

なお、階層構造を構成する多変数データの可視化手法は、我々の調査する限り少数ながら存在する [18]。しかしそれらは木構造を描画しながら多変数を表現するものであり、我々と同じようなアプローチで階層構造と多変数を同時に表現する手法は、我々の調査する限り見当たらない。

2.3 平安京ビュー

1章でも論じたとおり、本論文の提案手法「十二単ビュー」は、平安京ビューという階層型データ可視化手法の拡張手法である。そこで本節では、平安京ビューの概要を紹介する。

平安京ビューは、階層型データを構成する葉ノードをアイコンで表現し、枝ノードを入れ子状の長方形の枠で表現し、階層型データ全体を一画面に配置する手法である。配置結果の例を図 1 に示す。

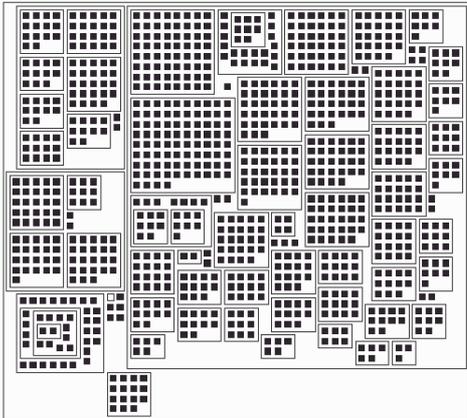


図 1: 平安京ビューによる階層型データの可視化例。

平安京ビューは、以下の条件を満たすような画面配置アルゴリズムの採用により、図 1 に示すような可視化結果を実現している。

- 葉ノードや枝ノードを干渉させないように配置する。これにより、画面上の任意のデータ要素を、クリック操作可能な状態で表示できる。
- 画面空間上の占有面積を小さくするように配置する。これにより、限られた画面空間に多くの情報を表現できる。

これ以外にもいくつかの条件が想定されているが、本論文では説明を割愛する。

平安京ビュー（およびその先行手法である「データ宝石箱」）の適用事例は、すでに多岐にわたって報告されている。主なものとして、ウェブサイトのアクセスログの可視化 [15]、分散計算環境の負荷分布の可視化、ネットワーク不正侵入検出データの可視化 [3]、企業組織のコミュニケーション量の可視化、医療系シミュレーションのパラメータ最適化のための可視化、原子力システムの測定情報の可視化、などが報告されている。

平安京ビューでは、階層型データの葉ノードが有する変数値を、棒グラフ化されたアイコンの色、高さ、形状などで表現することができる。しかしこの表現方法では、5 個、10 個といった多変数を同時に表現することは困難である。また例えば、2 個の変数値をアイコンの色と高さに割り当てたとき、その 2 個の変数を平等に認知できるとは限らない。あるときは色のほうが目につき、あるときは高さのほうが目につき、といった変数間の不平等性が生じる。これらの問題を解消し、葉ノードに属する多変数を平等に可視化する手法を、次章にて提案する。

3 十二単ビュー

本章では、平安京ビューの拡張手法である階層型多変数データ可視化手法「十二単ビュー」を提案する。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、平安京ビューにおい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。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。

十二単ビューが想定する入力データ構造を、図 2 に示す。図 2 において、四角は木構造の枝ノード、丸は木構造の葉ノードを表している。十二単ビューでは、任意の木構造を構成する階層型データの葉ノードが、 n 個の変数を有していると仮定する。本論文では、この変数の値を $t_i (0 \leq i < n)$ と表すものとする。

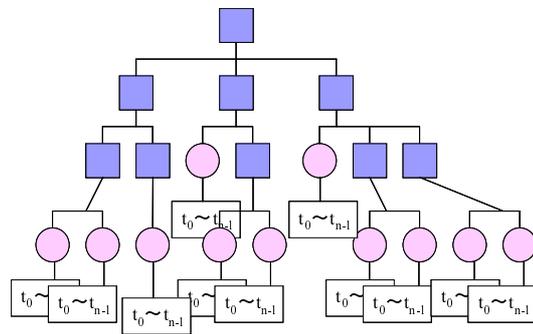


図 2: 十二単ビューが想定する入力データ構造。

続いて、十二単ビューの処理手順について解説する。十二単ビューではまず「平安京ビュー」に用いられている長方形のアイコンを、縦 l 個、横 m 個、の格

子状の小領域に分割する(図3(上)参照)。我々の実装では、以下の式により l および m を算出している。

$$\begin{aligned} l &= \lfloor \sqrt{n} \rfloor + 1 \\ m &= \lfloor n/l \rfloor + 1 \end{aligned} \quad (1)$$

ここで $\lfloor t \rfloor$ は、 t を超えない最大の整数を表す記号である。以上の式によって算出された l と m の積は、常に n 以上になる。このため、以上の方法によって生成された個々の小領域に、各々の変数を割り当てることにより、 n 個の変数値を表現することが可能になる。我々の実装では、 lm 個の小領域に対して、 n 個の変数の各々を、左上から順に割り当てる。 $lm > n$ である場合には、どの変数も割り当てられない小領域が存在するが、我々の実装ではこれらの小領域をブランクのままにしてある。

このような単純な領域分割による多変数の表現を採用した理由は、以下の通りである。十二単ビューでは、数百、数千規模の葉ノードを有する階層型データの一括的な表現を目標としているため、個々の葉ノードの表現に費やせる画面上の面積は小さくなる。一方で、大規模なデータを表現するための領域分割型の可視化手法において、画面上の面積が小さくても視覚的に認識しやすいようにデータ要素を表現するためには、正方形に近い領域を生成することが有利である、という議論が既になされている [17]。十二単ビューにおいても、この議論を参考にして、葉ノードに割り当てられた個々の変数を正方形に近い小領域で表現できるようにした。

続いて、各々の小領域に色を割り当てる。ここでは色合いを表す色相 H 、鮮やかさを表す彩度 S 、明るさを表す明度 I 、の3変数で色を表現するHSI表色系を用いる。

十二単ビューではまず、 lm 個の小領域のうち n 個を選び、これらに固有の色相を与える。我々の実装では単純に、以下の方法により i 番目 ($0 \leq i < n$) の変数の色相 H ($0 \leq H < 2\pi$) を算出する。

$$H = \frac{2\pi i}{n} \quad (2)$$

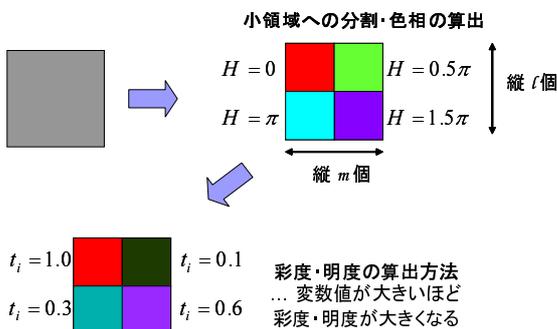


図3: 「十二単ビュー」による多変数データの表現。(上) 長方形アイコンの小領域への分割。(下) 彩度・明度の算出方法。

十二単ビューでは続いて、各データ要素が有する n 個の変数値を用いて、各小領域の彩度 S ($0 \leq S \leq 1$)

および明度 I ($0 \leq I \leq 1$) を算出する。なお、変数値 t_i ($0 \leq i < n$) は $0 \leq t_i \leq 1$ の範囲に正規化されていると仮定する。

情報可視化技術では、変数値の大きいデータ要素への注目を促すような表現を求められる場合が多い。そこで我々の実装の一例として、変数値の大きいデータ要素の彩度・明度を大きくすることで、これらのデータ要素を注目しやすくするよう算出方法を実装している。我々の実装では、以下の数式

$$S = I = 0.2 + 0.8t_i \quad (3)$$

によって彩度 S および明度 I を算出している(図3(下)参照)。

4 十二単ビューにおける詳細度制御



図4: (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4は十二単ビューを用いて階層型多変数データ全体を表示した結果、およびその一部を拡大した図である。この例より、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本論文では、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、この階層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ことを考える。

図5(左上)は、5変数を有する6個の葉ノードをアイコンで表現している例である。十二単ビューにおける詳細度制御では図5(右上)のように、これら6個の葉ノードを統合して1個のノードとみなし、これを1個のアイコンで表現する。この複数のノードが統合

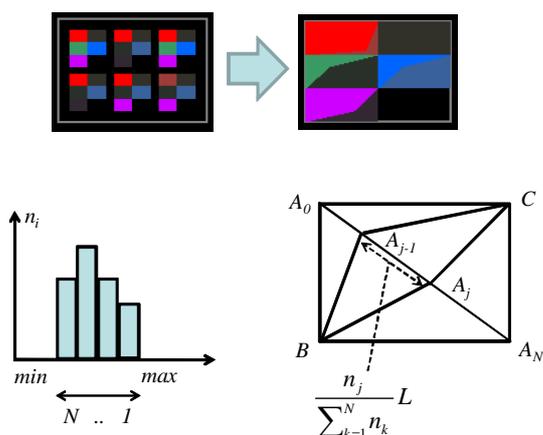


図 5: 同一階層下にあるノードの統合表示.

されてきたノードを, 提案手法では代表ノードと呼ぶ. なお, 図 5(右上) に示す代表ノードの例は, 縦 3 列, 横 2 列の小領域に分割されており, 各々の小領域が各々の変数を表現している. この例では赤, 緑, 青, 紫, 灰の 5 色で塗られた 5 個の小領域が, 各々の変数の値を表現しており, 右下の黒い小領域だけは変数を表現していない.

代表ノードの作成にはまず, 図 5(左下) に示すように, ノード変数値の範囲を N 個に分割したヒストグラムを構築する. このようなヒストグラムを用いることで, このクラスタに属する下位ノード群の変数値分布を, 代表ノードで表現できる.²

ヒストグラムを使用して, 代表ノードの小領域を生成する方法は, 以下の通りである. まず図 5(右下) に示すように, 代表ノードの 4 頂点を A_0, B, C, A_N とし, A_0 から A_N へ対角線をひく. そして A_0A_N 間に, A_1 から A_{N-1} の点を配置することで, この対角線を N 個に分割する. ここで, 以下の式によって $A_{i-1}A_i (i = 1..N)$ 間の長さ d_i を算出する. なお L は A_0A_N 間の長さ, n_i はヒストグラムで i 番目の領域に含まれるノードの個数を示す.

$$d_i = \frac{n_i}{\sum_{j=1}^N n_j} L \quad (4)$$

さらに, これによって生じた, i 番目の領域を示す 2 つの三角形, $A_{i-1}A_iB$ と $A_{i-1}A_iC$ を, 同じ色で描画する. ここで 2 つの三角形の明度と彩度を算出するために, 以下の式によって算出される t_i を式 (3) に代入する.

$$t_i = \frac{(j - 0.5)t_{min} + (N + 0.5 - j)t_{max}}{N} \quad (5)$$

ここで t_{min} は変数値の最小値を, t_{max} は変数値の最大値を示す.

² 本論文の実用例では, N の値を $N = 3$ に固定して実装した例を示している.

図 6 は, 各クラスタを代表ノードを用いて表示した例である. 図 6(右上) のように, 代表ノードの各小領域が長方形に見える場合, この代表ノードが示すクラスタでは, 全てのアイコンの明彩度がほぼ同一であることがわかる (図 6(左上) 参照). 一方, 図 6(右下) のように, 小領域内に異なる明彩度の領域が共存して, 小領域が単色の長方形に見えない場合, この代表ノードが示すクラスタに属する各アイコンは, 明彩度に基づきがあることがわかる (図 6(左下) 参照).

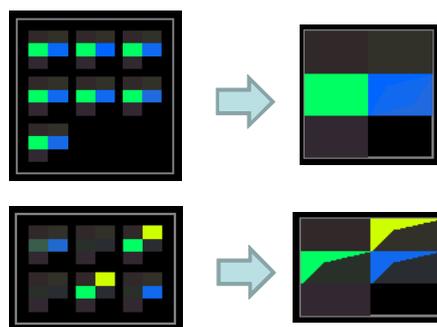


図 6: クラスタを代表ノードを用いて表示した例 (左上) 分子構造と実験値の関連性が高い階層の例. (右上) ズームアウトした際に (左上) の階層を表す代表ノード. (左下) 分子構造と実験値の関連性が低い階層の例. (右下) ズームアウトした際に (左下) の階層を表す代表ノード.

以上の表示方法に加えて十二単ビューでは, ユーザによるズーム操作に伴って詳細度を自動制御する機能を実装する. この機能では図 7 に示すように, ズームアウト操作に伴い, その縮小率に合わせて, 下位階層のノードを統合して上位階層全体をアイコン表現する, という操作を再帰的に反復する. またズームイン操作に伴い, 自動的に上位階層のアイコン表示を省略して下位階層をアイコン表現する, という操作を再帰的に反復する. 以上の機能により, ズームイン操作によって個々の葉ノードの変数値を読み取れるようになり, ズームアウト操作によってデータの全体像と数値的傾向を見渡せるようになる.

5 薬物情報の可視化への適用

本論文では十二単ビューの適用事例として, 階層型多変数データを構成する薬物情報の可視化を試みた例を示し, 十二単ビューの有用性を議論する.

我々は 161 種類の薬物情報を収集し, その各々に対して, 5 種類の酵素への代謝感受性という実験値を収集した. 本実験ではこれらの薬物群を, 5 個の変数を有する 161 個の葉ノードとして扱った. 続いて本実験では, これらの薬物群を, 分子構造的特徴に基づいて階層的に分類することで, 階層型多変数データを構築した. なお, この階層構造の具体的な構築方法については, 文献 [19] に詳しく記述されているので, ご参照いただきたい.

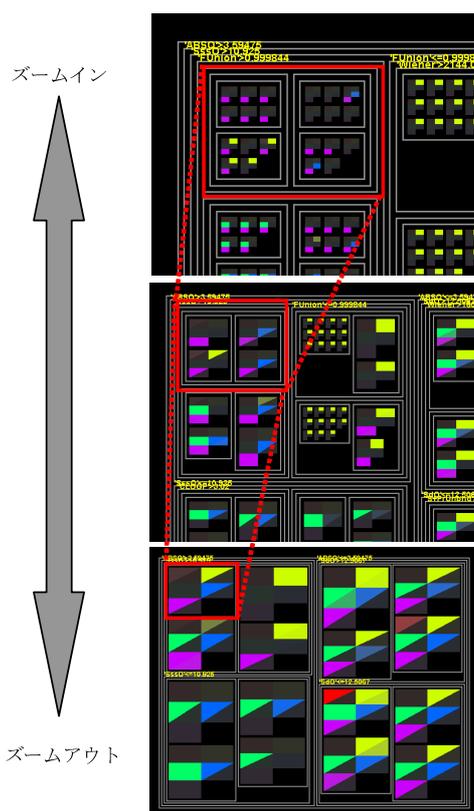


図 7: ズーム操作に連動した詳細度制御 .

この薬物情報の可視化に関する有用な利用方法の一例に、新しい薬物に対する実験値の推定が考えられる。仮に、分子構造が既知であり、また実験を施されていない薬物があるとすると。もし、この薬物群の分子構造と実験値に、強い相関性が発見できるなら、この新しい薬物に対する実験結果は、実験を施す前からある程度の予測が可能であり、ひいては実験コストの削減に貢献できると考えられる。提案手法は既に、製薬会社の研究開発の現場において、薬物実験のコスト削減を目標として試用されている。

上述の手順によって薬物情報から構築した階層型多変数データの可視化結果を、図 15 に示す。なお図 15 において、5 種類の酵素の名前、およびそれらの酵素に対する代謝感受性を表現する色相は、以下の通りである。

- CYP1A2: 赤
- CYP2C9: 黄
- CYP2C19: 緑
- CYP2D6: 青
- CYP3A4: 紫

本適用事例における薬物の分類において、最上位に選ばれた条件文は「Sanderson 原子電気陰性度の総和 (Se) が 44.89 以上か否か」であった。図 15 に示す可視化結果では、 $Se \leq 44.89$ である薬物が左半分、それ以外の薬物が右半分に表示されている。左半分に赤

および緑の明るいアイコンがほとんど見られないことから、Se 値と CYP1A2, CYP2C19 の 2 種類の酵素には高い相関性があることが、この可視化結果から視覚的に確認できる。以上の結果から十二単ビューが、薬物の分子構造上の特徴と、酵素に対する実験値の相関性を、全体的に概観する目的に有用であることがわかる。

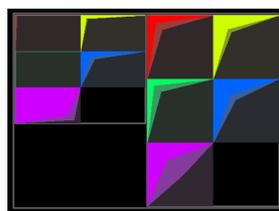


図 8: 詳細度制御の例 (1) .

図 8 は、図 15 をズームインし、詳細度制御によって 2 つの代表ノードが表示されている例である。ここで、右の代表ノードには、赤と緑の小領域中に明るい色で表示された領域があることがわかる。一方、左の代表ノードの赤と緑の小領域には、明るい色の領域はほとんど見られない。このことから、最上位の条件文から得られる「Se 値と CYP1A2, CYP2C19 の 2 種類の酵素と高い相関性がある」という情報が、図 8 では図 15 より容易に発見できることがわかる。また左の代表ノードでは、紫の小領域の明るい領域がほぼ長方形に近い形で現れていることが視認できる。これは、左側のクラスタに属する薬物の大部分が、酵素 CYP3A4 に対して敏感な反応を示していることを表現している。一方の右の代表ノードでは、紫の小領域の約半分が明るい色で表示されている。このことから、右側のクラスタに属する薬物の約半数が、CYP3A4 に敏感な反応を示していることがわかる。これらの特徴も、図 15 よりも図 8 から容易に得ることができる。以上の結果から、ユーザがデータの特徴を容易に把握するためにも、クラスタのノード変数値の傾向を代表ノードで表現する詳細度制御が有効であるといえる。

続いて図 15 の一部分を拡大表示した例を、図 9 に示す。図 9(上) では、灰色の長方形の枠の中にあるアイコンを比較すると、その配色パターンが非常に類似していることがわかる。このことから、図 9(上) に表示される薬物群は、分子構造上の特徴による分類結果と、酵素に対する代謝感受性という実験値に、高い相関性を有することがわかる。一方、図 9(下) では、灰色の長方形の枠の中にあるアイコンを比較すると、その配色パターンに差異が大きいがわかる。このことから、図 9(上) に表示される薬群は、分子構造上の特徴による分類結果と、酵素に対する代謝感受性という実験値に、相関性があまり見られないことがわかる。以上の結果より十二単ビューが、薬物の分子構造上の特徴と、酵素に対する実験値の相関性を、局所的に細かく眺める目的でも有用であることがわかる。

図 10 は、図 9 と同じ部分について、中間層のクラスタを代表ノードで表示した例を示している。図 10(上) では大半の代表ノードにおいて、1 つまたは 2 つの小領域のみが明るい色で表示されている。この図

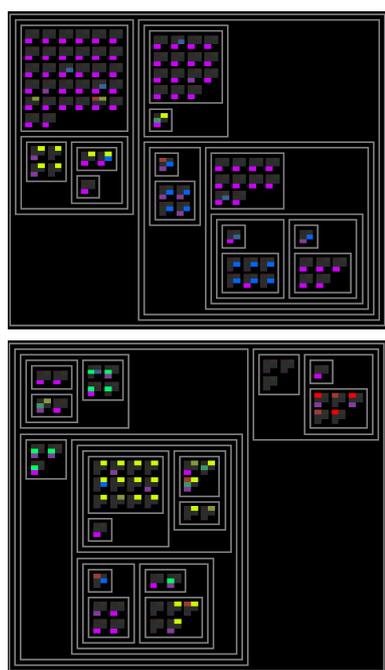


図 9: 薬物データの可視化の例 (2) . (上) 分類結果と実験値の相関性の高い部分 . (下) 分類結果と実験値の相関性の低い部分 .

における最も右下の代表ノードが、青と紫の小領域が長方形でないということを除けば、大半の代表ノードの小領域が長方形に近い形で表現されている。ゆえに、図 10(上) におけるクラスタの多くにおいて、薬物の分子構造上の特徴と、酵素に対する実験値の相関性が高いことが、詳細度制御によって明確に表現されているのがわかる。一方、図 10(下) における 3 つの代表ノードでは、長方形というより三角形に見える小領域が多数ある。これらの結果により、図 10(上) におけるクラスタの大半は、図 10(上) のクラスタよりも相関性の低いクラスタであることが、詳細度制御によって明確に表現されているのがわかる。

この適用事例における課題は、各薬物の実験値の傾向という低レベル情報の可視化だけでなく、その上位概念にあたるクラスタを単位とした実験値の統計的分布という高レベル情報も可視化したい、という点にある。この課題に対して我々は、まず上位階層を単位として実験値のばらつきを確認し、続いて実験値がばらついている階層をズームインすることで個々の薬物の実験値を確認する、というズーム操作に連動した詳細度制御が有効であると考えている。このことから、薬物の分子構造上の特徴と実験値の相関性を、高レベルからも低いレベルからも眺められる十二単ビューは、薬物を扱う各種の現場において有効に活用できる可能性があると思われる。

一方で薬物情報以外の可視化に適用する際には、このままでは十二単ビューが有効に働かない場合があるかもしれない。あるいは、対角線を引いて変数値の統計的分布を表現するよりも良好な表現があるかもしれない。今後の課題として、薬物情報以外の適用事例を

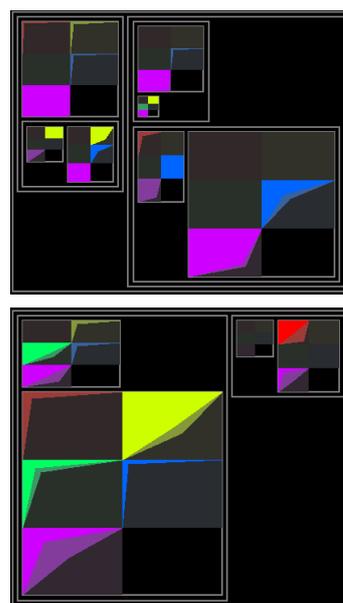


図 10: 薬物データの可視化の例 (2) . (上) 分類結果と実験値の相関性の高い部分 . (下) 分類結果と実験値の相関性の低い部分 .

視野に入れて、提案手法をどのように汎用化するか、という点があげられる。

6 ユーザテストと考察

本章では、20 代の大学生 11 人を被験者として実施した十二単ビューのユーザテスト結果を示し、十二単ビューの有効性について考察する。

6.1 平安京ビューとの比較

5 変数を有する階層型多変数データを用意し、これを十二単ビューで可視化した図 11 と、変数ごとに 5 個の平安京ビューで可視化して並べた図 12 を用意した。なお本実験で用いた平安京ビューでは、変数値の大きさから色相を算出し、最大値を赤で、最小値を青で表現した。この 2 つの図を被験者に見比べさせた上で、以下の質問を提示した。

- 問 1 変数値が大きいノードを見つけやすいのはどちらか。
- 問 2 変数値が小さいノードを見つけやすいのはどちらか。
- 問 3 1 つのノードの、すべての変数値を読み取りやすいのはどちらか。
- 問 4 1 つの階層に含まれるすべてのノードの、変数値の傾向 (全体的に値が大きい、値にばらつきがある、等) が読み取り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．

- 十二単ビュー
- 平安京ビュー
- どちらも変わらない

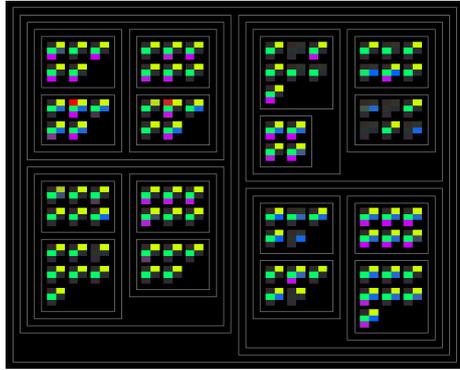


図 11: 十二単ビューで可視化した結果．

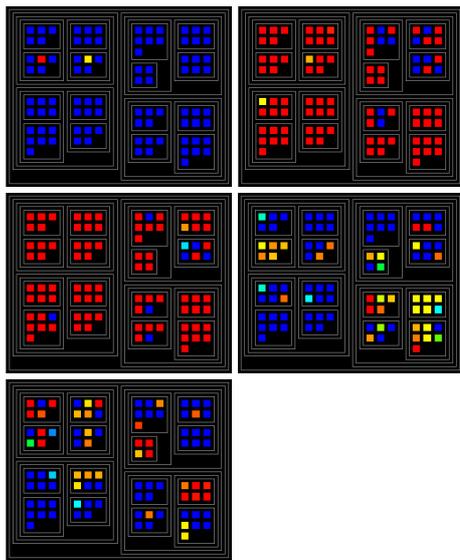


図 12: 平安京ビューで可視化した結果．

この結果を表 1 にまとめた．問 1 と問 2 の結果から，特定の変数について値の大きい，または小さいアイコンを見つける目的には，平安京ビューが優位である，という傾向が見られた．この一因として，値の大小を見比べるためには，明彩度の違いよりも色相の違いのほうが効果が高いからではないか，と考えられる．

一方，問 3 と問 4 の結果から，1 ノードのすべての変数値を読み取るには，あるいは 1 つの階層に含まれるすべてのノードの変数値の傾向を読み取る目的には，十二単ビューが優位である，という傾向が見られた．十二単ビューでは平安京ビューと違って，1 つの

可視化結果の 1 つのアイコンからすべての変数値を読み取ることができる．このため，すべての変数値が一画面上に表現されている十二単ビューのほうが，すべての変数値を読み取るのに適していると考えられる．

前章で論じた薬物情報の可視化の事例では，問 4 に示した「1 つの階層に含まれるすべてのノードの変数値の傾向」を視認することが重要であるため，平安京ビューよりも十二単ビューの利用が適切であると考えられる．

表 1: 平安京ビューとの比較実験結果．(単位：%)

	平安京ビュー	十二単ビュー	どちらも変わらない
問 1	63.6	9.1	27.3
問 2	63.6	9.1	27.3
問 3	0	90.9	9.1
問 4	18.2	72.7	9.1

6.2 より多くの変数への対応

続いて，変数の数が 10, 20, 30, 40, 50 個の階層型多変数データを，被験者に十二単ビューで可視化させた．この実験では，詳細度制御を利用せず葉ノードを表示した場合と，詳細度制御を利用して下位階層を代表ノードで表示した場合のそれぞれについて，被験者には各データの可視化結果を自由に操作してもらい，小領域を判別できる最大の変数の個数を回答してもらった．この回答結果を集計した結果を，表 2 に示す．

この結果から，詳細度制御を利用した方が，40～50 変数といった多変数を視認できる被験者が多いことがわかった．このことから，詳細度制御を利用して，画面の解像度，データの大きさ，ユーザのズーム操作に応じて，適切な大きさで代表ノードを表示することが，より変数の数が多い階層型多変数データにおいて有利に働いている，という傾向を示唆しているといえる．

なお本実験において，30 変数まで判別できる，と回答した被験者はなく，20 変数以下，または 40 変数以上に回答が偏った．その理由は不明だが，本実験において被験者の色彩識別能力が大きく二分されている可能性が考えられる．

表 2: 小領域を判別できる変数の数．(単位：%)

変数の数	10	20	30	40	50
詳細度制御無	27.3	54.5	0	9.1	9.1
詳細度制御有	0	27.3	0	36.4	36.4

6.3 一定時間における階層探索

階層型多変数データを十二単ビューで可視化し、被験者には30秒間で「すべての変数において、階層内のデータ要素の変数値がほぼ同じ値をもっている」という特性をもつ最下位クラスタを探してもらった。データの葉ノードを表示した場合と、最下位階層を代表ノードで表示した場合の2種類の場合において実験を行い、被験者が30秒間で見つけ出した最下位クラスタの数を比較した。

この結果を表3に示す。一定時間において、代表ノードを表示した場合は、葉ノードを表示した場合の約2倍の数のクラスタを見つけて出すことができる。詳細度制御を用いることで、葉ノードをすべて表示した状態よりも短時間で、代表ノードから情報を得ることが可能であると考えられる。この結果から、詳細度制御は、注目したいクラスタをより短時間で見つけるためにも有利に働くといえる。

表3: 30秒間で見つけ出したクラスタの数。(単位:個)

	葉ノードを表示	代表ノードを表示
平均	6.4	12.8
最大	11	15
最小	4	11

6.4 階層探索時間の計測

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階層を文面で指定し、被験者にその最下位階層を探してもらい、その所要時間を測定した。本実験では以下のような固有の特性をもつ階層を指定した。

- 特定のいくつかの変数の値は葉ノード間で大小ばらつきのある値をもち、その他の変数の値はすべての葉ノードにおいてほぼ同じ値をもつ。

本実験では4種類の階層型多変数データ(以下、データ1、データ2、データ3、データ4、と称する)を用いた。4種類のデータにおける葉ノード数は、それぞれ9140、9216、727、および729であった。ここでデータ1、3は、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。その一例として図13に、本実験におけるデータ1の可視化結果の一部を示す。またデータ2、4は、代表ノードの大きさや形状が均一に表示される特徴をもつデータである。その一例として図14に、本実験におけるデータ4の可視化結果の一部を示す。

被験者による所要時間の平均時間を、表4に示す。この結果からわかるように、すべてのデータにおいて詳細度制御の効果があらわれた。葉ノード数が多いデータ1、データ2と、葉ノード数が小さいデータ3、データ4とを比較すると、詳細度制御機能はデータサイズによらず、効果的であるといえる。

さらに、代表ノードが不均一であるデータ1、データ3と、代表ノードが均一であるデータ2、データ4

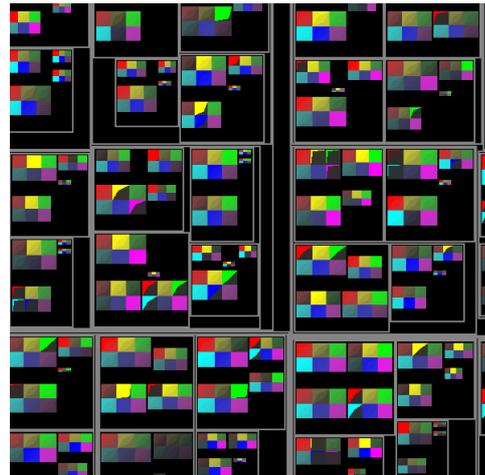


図13: 下位クラスタを統合した代表ノードが不均一な大きさ・形状で表示された例

を比較すると、代表ノードが均一なデータの方が、短時間で該当する最下位階層を見つけれられている。

データ1やデータ2のように、代表ノードの大きさや形状が不均一である場合、被験者は固有の特性をもつ階層を見つけにくくなることが予想される。特に代表ノードが小さく(または細長く)表示された代表ノードを被験者が見落とす、という現象が容易に推察される。一方でデータ3やデータ4は、葉ノード数が均一に割り振られた木構造を構成しているために、代表ノードの大きさや形状も均一になる。その結果として、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。

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、上位階層を均一な大きさの長方形で表現したい、という要求は今後も多くあると考えられる。例えば「平安京ビュー」を応用した大量画像の可視化手法[20]においても、上位階層を表現する代表画像を均一な大きさで表現したい、という要求は考えられる。今後の課題として、上位階層を表現する代表ノードの大きさや形状の均一化があげられる。

表4: 最下位階層を見つけるまでにかかった時間。(単位:秒)

	詳細度制御	データ1	データ2	データ3	データ4
平均	無	403	252	186	92
	有	82	21	29	94
最大	無	1076	517	524	294
	有	182	60	92	573
最小	無	77	95	27	10
	有	8	2	5	3

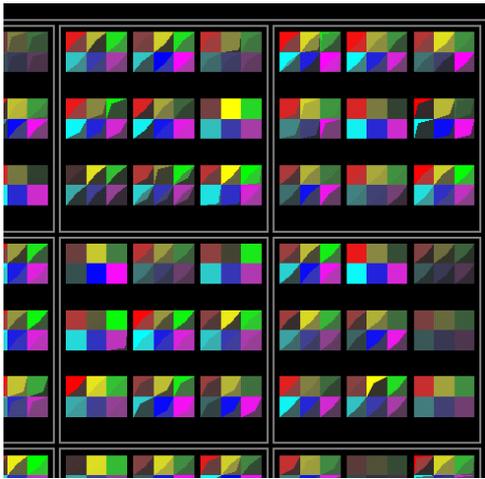


図 14: 下位階層を統合した代表ノードが均一な大きさ・形状で表示された例

7 まとめ

本論文では、階層型多変数データを可視化する一手法「十二単ビュー」を提案した。十二単ビューは階層型データ可視化手法「平安京ビュー」の拡張手法であり、データ全体の画面配置は平安京ビューと同じ原理を踏襲しつつ、個々のデータ要素を表現するアイコンを小領域に分割し、変数ごとに小領域の色を算出することで、階層構造と同時に多変数を表現する。

また本論文では、下位階層は変数値を統合した代表ノードで表現して、階層構造の上位階層だけを表示する、といった詳細度制御により、一画面に表示する情報量を制御する手法を提案した。

また本論文では、十二単ビューを用いた薬物情報の可視化に関する適用事例を示し、この事例における十二単ビューの有用性を議論した。さらにユーザテストによって、十二単ビューの性能や有用性を検証した。

十二単ビューは、薬物に限らず、幅広い分野に適用が可能であると考えられる。そこで今後の課題として我々は、十二単ビューをさらに多くの分野に適用し、その有用性の実証を進めたいと考えている。

参考文献

[1] Shneiderman B., The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization, *Proceedings of IEEE Symposium on Visual Languages '96*, pp. 336-343, 1996.

[2] 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可視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.

[3] Itoh T., Takakura H., Sawada A., Koyamada K., Hierarchical Visualization of Network Intrusion Detection Data in the IP Address

Space, *IEEE Computer Graphics and Applications*, Vol. 26, No. 2, pp. 40-47, 2006.

[4] Inselberg A., Dimsdale B., Parallel Coordinates: A Tool For Visualizing Multidimensional Geometry, *IEEE Visualization '90*, pp. 35-38, 1990.

[5] Feiner S., Beshers C., Worlds within Worlds: Metaphors for Exploring n-Dimensional Virtual Worlds, *ACM SIGGRAPH Symposium on User Interface Software and Technology (UIST'90)*, pp. 76-83, 1990.

[6] Marks J., et al., Design Galleries: A General Approach to Setting Parameters for Computer Graphics and Animation, *ACM SIGGRAPH '97*, pp. 389-400, 1997.

[7] Ebert D. S., et al., Automatic Shape Interpolation for Glyph-based Information Visualization, *IEEE Visualization '97*, 1997.

[8] Johnson B., et al., Tree-Maps: A Space Filling Approach to the Visualization of Hierarchical Information Space, *IEEE Visualization '91*, pp. 275-282, 1991.

[9] Bederson B., Schneiderman B., Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies, *ACM Transactions on Graphics*, Vol. 21, No. 4, pp. 833-854, 2002.

[10] Carriere J., et al., Research Paper: Interacting with Huge Hierarchies beyond Cone Trees, *IEEE Information Visualization 95*, pp. 74-81, 1995.

[11] Koike H., Fractal Views: A Fractal-Based Method for Controlling Information Display, *ACM Transactions on Information Systems*, Vol. 13, No. 3, pp. 305-323, 1995.

[12] Lamping J., Rao R., The Hyperbolic Browser: A Focus+context Technique for Visualizing Large Hierarchies, *Journal of Visual Languages and Computing*, Vol. 7, No. 1, pp. 33-55, 1996.

[13] Rekimoto J., The Information Cube: Using Transparency in 3D Information Visualization, *Third Annual Workshop on Information Technologies & Systems*, pp. 125-132, 1993.

[14] Sprenger T. C., et al, H-BLOB: A Hierarchical Visual Clustering Method Using Implicit Surfaces, *IEEE Visualization 2000*, pp. 61-68, 2000.

[15] 山口, 伊藤, 池端, 梶永, 階層型データ可視化手法「データ宝石箱」とウェブサイトの可視化, *画像電子学会論文誌 Visual Computing 特集号*, Vol. 32, No. 4, pp. 407-417, 2003.

[16] Itoh T., Yamaguchi Y., Ikehata Y., and Kajinaga Y., Hierarchical Data Visualization Using a Fast Rectangle-Packing Algorithm, *IEEE Transactions on Visualization*

and *Computer Graphics*, Vol. 10, No. 3, pp. 302-313, 2004.

- [17] Bruls D.M., Huizing K., Wijk J. J., Squarified Treemaps, *Data Visualization 2000 (joint Eurographics and IEEE TCVG Symposium on Visualization)*, pp. 33-42, 2000.
- [18] Sifer M., User Interfaces for the Exploration of Hierarchical Multi-Dimensional Data, *IEEE Symposium on Visual Analytics Science and Technology*, pp. 175-182, 2006.
- [19] 伊藤, 山下, 階層型多変数データ可視化手法「十二単ビュー」による薬物データベースの可視化, 情報処理学会データベースと Web 情報システムに関するシンポジウム (DBWeb2006), pp. 119-125, 2006.
- [20] 五味, 伊藤, Li, CAT:大量画像の一覧可視化と詳細度制御の一手法, 画像電子学会 Visual Computing / 情報処理学会グラフィクスと CAD 合同シンポジウム, pp. 1-6, 2007.

山澤 舞子



2007 年お茶の水女子大学理学部情報科学科卒業。現在お茶の水女子大学大学院人間文化創成科学研究科理学専攻在学中。情報処理学会会員。

伊藤 貴之



1990 年早稲田大学工学部電子通信学科卒業。1992 年早稲田大学大学院理工学研究科電気工学専攻修士課程修了。同年日本アイ・ピー・エム (株) 入社。1997 年博士 (工学)。2000 年米国カーネギーメロン大学客員研究員。2003 年から 2005 年まで京都大学大学院情報学研究科 COE 研究員 (客員助教授相当)。2005 年日本アイ・ピー・エム (株) 退職, 2005 年よりお茶の水女子大学理学部情報科学科助教授。ACM, IEEE Computer Society, 情報処理学会, 芸術科学会, 画像電子学会, 可視化情報学会, 他会員。

山下 富義



1988 年京都大学薬学部薬学科卒業。1990 年京都大学大学院薬学研究科修士課程修了。1992 年同大学院博士後期課程退学。同年京都大学薬学部助手。1995 年博士 (薬学)。1995 年~1996 年南カリフォルニア大学薬学部客員研究員。1996 年京都大学大学院薬学研究科助手 (配置換)。1996 年より京都大学大学院薬学研究科助教授。日本薬学会, 日本薬物動態学会, 日本薬剤学会, 他会員。

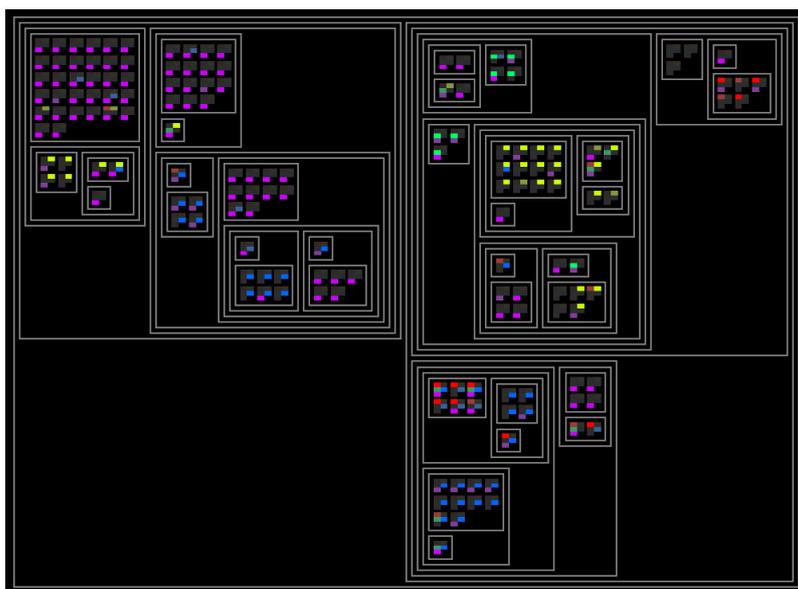


図 15: 薬物データの可視化の例 (1) .