

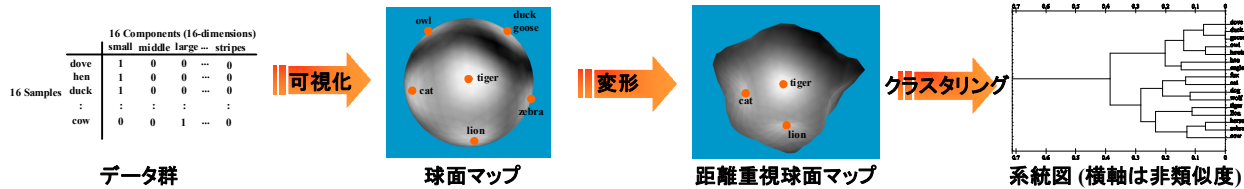
球面SOM 究極のクラスター分析へ

1. テーマ 大量のデータから隠れた背後関係・法則を導き出す

数が膨大で、個々のサンプルの特徴が多様なデータ群から、有益な情報を得るにはどうすれば良いでしょう。ここでは、そのようなデータを可視化して人間に理解しやすくする方法として、球面自己組織化マップ (球面SOM) を用いた手法を提案し、それを実装したアプリケーションソフトについてご紹介します。

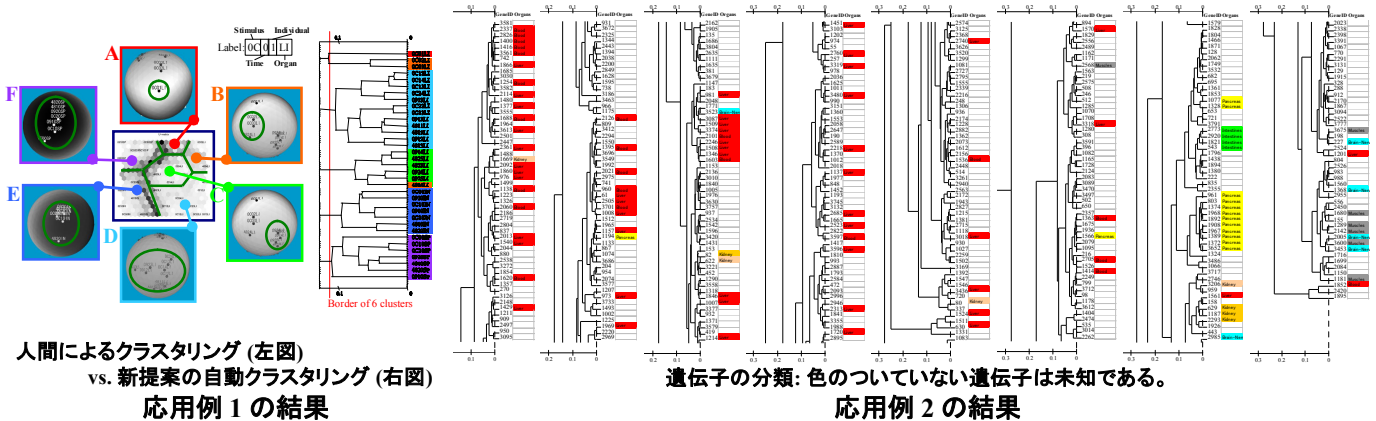
2. 新発明 球面SOMを使って、可視化とクラスタリングを同時に

自己組織化マップ (SOM) はデータ群の可視化の手法としてよく知られています。具体的には平面上にサンプルを一定のルールに従って配置し、サンプルのマップを作ります。SOMジャパンのアプリケーションでは、平面よりも歪みの出にくい球面を使った球面SOMによる可視化と、系統図を用いた定量的な分析方法を使用することができます。データ群を球面SOMを用いて可視化し、系統図を作成するまでの流れは次のようになります：



3. 応用例 遺伝子情報 (DNAマイクロアレイデータ) の解析

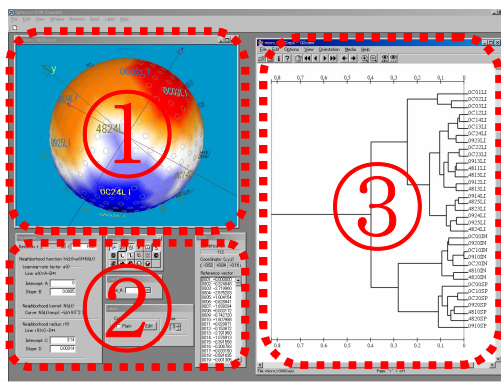
例 1: ラットに薬を与えたときの3824遺伝子の反応から、35の実験条件・測定時点での結果を一度に比較する。
例 2: 35の条件下でテストされた425の遺伝子を分類する。そして、未知遺伝子を推定する。



4. ソフトへの実装

球面SOMを用いたクラスタリングツール、そのコンセプトはマップと系統図の両側面からデータを解析できること、そしてシンプルでいて柔軟なインターフェイスです。

①ここにデータ群が可視化されます。これは立体的な図形なので回転させることで様々な視点を取ることができます。



②インターフェイスは至ってシンプルです。表示方法の変更や可視化の実行などをここからコントロールします。

③系統図はポストスクリプト形式なので、美しく印刷することができます。図ではRussell Lang氏 (Ghostgum Software Pty Ltd) のGSviewで表示しています。