Visualization and 3D Printing of Multivariate Data of Biomarkers

Thrun, M. C., Lerch, F., Lötsch, J., Ultsch, A.

Conference: International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)
Overview

- Finding structures in high-dimensional data
- Visualization of structures
- How to explain the visualization to topical experts?
Discovering patterns in data

- Searching for similarities
- Clustering: process of finding groups of similar objects in high-dimensional data
  - e.g. labelling data points with colors
- Example: N=3 dimensions in Hepta
  - Equidistant clusters
  - One Cluster has higher density
  - Small distances in each cluster
Discovering patterns in data

- Searching for similarities
- Clustering: process of finding groups of similar objects in high-dimensional data
- Problem: high-dimensional: $N \gg 3$

How are we able to find similarities?
Finding similarities

1. Use a clustering algorithm
   - Every algorithm has a geometric model for a cluster
   - How to define a cluster? -> application-specific

2. Dimensionality Reduction (DR)
   - In general, *manifold learning*, see [Lee, Verleyson 2007]
   - Special case: *projections* into two dimensions
Problems

- Problems of high-dimensionality
  - Empty space and concentration of measure phenomena, curse of dimensionality

- Choice of dissimilarity (distance)

- Projection $\mathbb{R}^N \rightarrow \mathbb{R}^2$
  - Various methods
  - Choice of parameters
  - Problem of structure preservation: no projection method is able to preserve all distances
Various projection methods

- **linear**
  - PCA
  - ICA
  - ...

- **non-linear**
  - deterministic
    - distance-error
      - MDS
      - ...
  - graph based
    - ISOMAP
    - LLE
    - ...
  - stochastic
    - focusing
      - CCA
      - t-SNE
      - ...
    - databionic
      - ESOM
      - ...

- projections
Example:

- Choice of projection method
- Choice of parameters for the projection

Structure preservation:
projection

\[ R^N \rightarrow R^m \] with \( N > M \) cannot preserve all distances!
Emergent Self-Organizing Map (ESOM)

- Unsupervised neural learning algorithm
- Many neurons => projection method with emergent properties
  - Emergent: novel and unforseen properties of a multi-agent system

Why ESOM?
- For high-dimensional data, the SOM remains a reference tool for 2D visualizations. [Lee/Verleysen, 2007, p. 227]
ESOM result of Hepta

(⁻) Projected points uniformly distributed
(⁻) toroidal grid of neurons
(⁻) structure preservation hidden

Toroid:
- borders are cyclically connected
ESOM result of Hepta

- Projected Points uniformly distributed
- toroidal grid of neurons
- structure preservation hidden

Toroid: borders are cyclically connected

How can we solve this?
Solution: 3D visualization

3D landscape allows the human eye to detect meaningful cluster structures

e.g. all black points equal one cluster
Display information in an easily understandable way

- Blue colors indicate small distances (sea level)
- Green and brown colors indicate middle distances (small hilly country)
- White colors indicate high distances (snow and ice of high mountains).
- The valleys and basins indicate clusters
- Watersheds of hills and mountains indicate borderlines of clusters
Comparison
**U*matrix**

- 3D landscape defined by **U*matrix** [Ultsch et al., 2016]
  - Combines Umatrix and Pmatrix

- Umatrix: folding of high dimensional space
  - [Ultsch/Siemon, 1990]
  - In literature cited as grey-scaled 2D visualization (e.g. [Kadim Tasdemir/Merényi, 2012 ])
    - Precise colored definition required

- Pmatrix: high dimensional density estimation technique [Ultsch, 2003]
  - Estimation for hypersphere of radius is trying
    - use ABCanalsys [Ultsch/Lötsch, 2015]

- **U*matrix represents distance and density based structures!**
U*matrix -> 3D landscape

- use colors proposed in [Ultsch, 2003]

=> **hypsometric tints**: surface colors which depict ranges of elevation

- Calculate contour lines
  - Normalization of U*heights, define height intervals, ...
- Combine specific color scale with contour lines
- Create island of toroid map (rectangular)

=> topographic map with hypsometric tints
Example: pain biomarkers

- Data: responses to different types of nociceptive stimuli [Flühr et al., 2009]
  
  ESOM -> U*matrix -> 3D visualization -> Clustering

- Result: three main pain sensitivity groups [Lötsch/Ultsch, 2013]
  - High-pain sensitivity (HPS)
  - Average pain sensitivity (APS)
  - Low-pain sensitivity (LPS)
  - Subclusters...
Example: pain biomarkers

- Data: responses to different types of nociceptive stimuli [Flühr et al., 2009]
  
  \[ \text{ESOM} \rightarrow \text{U*matrix} \rightarrow \text{3D visualization} \rightarrow \text{Clustering} \]

- Result: three main pain sensitivity groups [Lötsch/Ultsch, 2013]
  - High-pain sensitivity (HPS)
  - Average pain sensitivity (APS)
  - Low-pain sensitivity (LPS)
  - Subclusters...

Problem:
complexity for domain experts
Complexity for domain experts

- 3D visualizations [Jansen et al., 2013] cites [Shneiderman, 2003]
  - Have to be viewed from multiple viewpoints
  - Are often subject to serious occlusion and navigation issues
Solution: colored 3D printing

To our knowledge, first application of 3D printing techniques used directly for data mining and knowledge discovery

(+) Provide topical experts a haptic grasp of high-dimensional structures

(-) Technical limitation of height-dependent colors in 3D printers

(-) Automatically cutting a non-rectangular island defined by curved borders remains a problem
Summary

- Structure preservation: projections do not preserve all distances
  1. Calculate U*matrix to represent distance and density based structures
  2. Generate 3D landscape: topographic map with hypsometric tints
- Solve complexity for domain experts by colored 3D printing

For (1) and (2) use our freely available R package Umatrix

[Version 2.0.0, www.uni-marburg.de/fb12/datenbionik/software]
Thank you for listening! Any questions?
Sources


