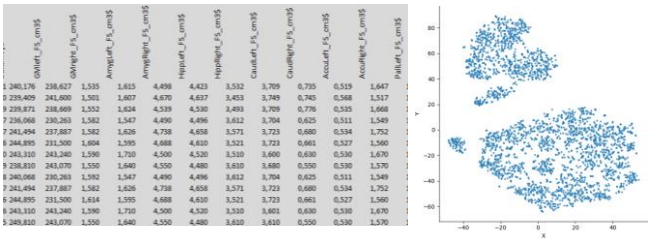
VISAnt: Unsupervised Data Exploration with Chernoff Faces

Ivana Sixtova, Ladislav Peska, Jakub Lokoc, David Bernhauer, Tomas Skopal
SIRET research group, Faculty of Mathematics and Physics, Charles University, Prague, Czechia

Motivation

Problem: exploratory analysis of large multidimensional datasets



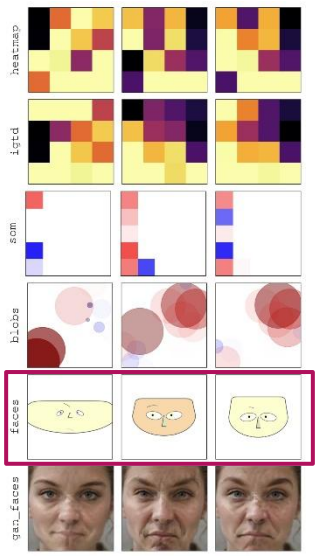
"zoomed in" view

"zoomed out" view

- How to bridge the gap between a **"zoomed in" view** on the raw data and the **"zoomed out" view** (dimensionality reduction + scatterplot), where we see the big picture?
- How can a user without a data analytics background easily assess **the similarity/dissimilarity of two objects** (represented by table rows) or clusters of objects in the plot?
- How to choose the most effective dimensionality reduction method?

👉 We can employ **an intermediate visualization** on the level of object instances and use the user's visual perception, which is inherently trained for pattern discovery [1].

Why Chernoff faces?



Class 1 Class 2 Class 2

Visualizations of three objects belonging to two different classes.

There are many ways to encode an object into a visual representation, including bar charts, heatmaps, and parallel coordinates.

One of them is **Chernoff face [2],** which is a cartoon-like face generated through configurable attributes such as eye shape, nose length, and eyebrow angle.

The idea behind this visualization is that humans are highly sensitive to subtle facial differences.

In our two user studies, which involved **classification and similarity search tasks** across various datasets, Chernoff faces:

- Outperformed other visualization methods
- Received positive feedback from users

In our tool, we used an implementation that can accommodate up to **22 features**. For datasets with more features, a feature combination technique is necessary.

VISAnt

Our tool combines dimensionality reduction methods with Chernoff faces in an interactive manner.

Objects are projected to the 2D (by either PCA, t-SNE, or PHATE), and then they are **implicitly grouped by their position in a regular grid**. Each grid cell is represented by one face glyph.

Users can navigate through individual groups, explore their characteristics in the GROUP tab, and also view the raw data within a group in the DATA tab.

References

[1] Keim, Daniel, et al. Visual analytics: Definition, process, and challenges. In: Information visualization: Human-centered issues and perspectives. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. p. 154-175.
[2] Chernoff, Herman. The use of faces to represent points in k-dimensional space graphically. Journal of the American statistical Association, 1973, 68.342: 361-368.

Acknowledgements

The database of patients was provided by MS center, Department of Neurology and Center of Clinical Neuroscience, First Faculty of Medicine, Charles University and General University Hospital in Prague.
The study was supported by SVV project number 260 821.