Fast and accurate Mahalanobis brushing in scatterplots, optimized on the basis of a user study

Chaoran Fan, Helwig Hauser
Visualization group, Department of Informatics, University of Bergen, Bergen, Norway
http://www.ii.uib.no/vis/
Chaoran.Fan@uib.no

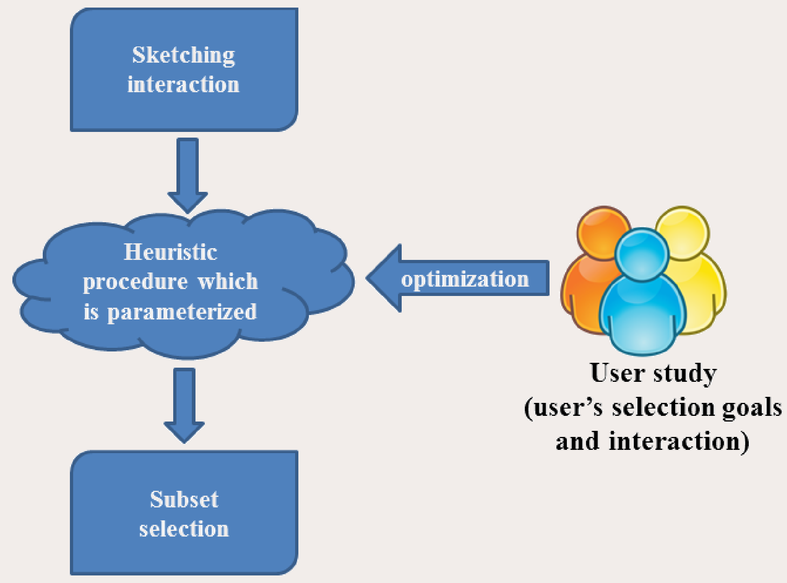
Abstract

Brushing is at the heart of most modern visual analytics solutions with coordinated, multiple views and effective brushing is crucial for swift and efficient processes in data exploration and analysis. Given a certain data subset that the user wishes to brush in a data visualization, traditional brushes are usually either accurate (like the lasso) or fast (e.g., a simple geometry like a rectangle or circle). We now present a new, fast and accurate brushing technique for scatterplots, based on the Mahalanobis brush, which we have extended and then optimized using data from a user study.

Principal approach

To be fast: we use sketching as interaction;

To be accurate: the user study data used to optimize the relevant parameters of selection heuristic.

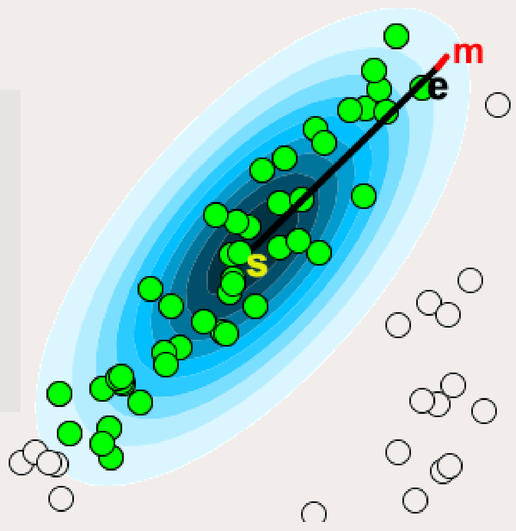


Brushing technique

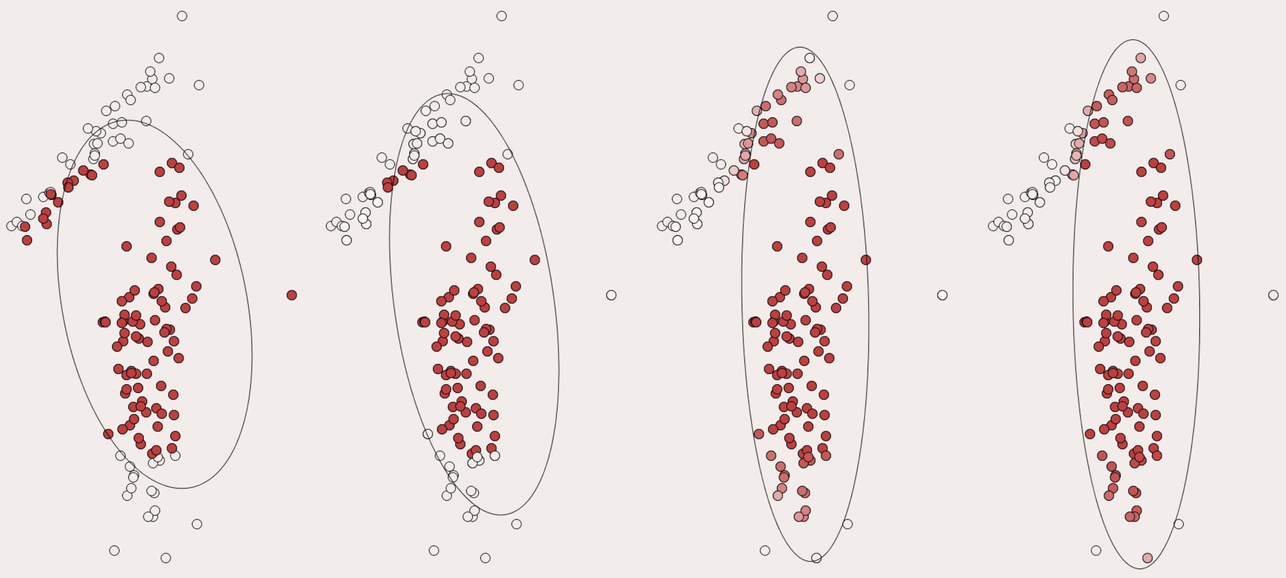
Sketching interaction: the user clicks into the middle of the subset to be selected and then drags to the border of the subset;

Selection is based on the Mahalanobis distance from the click-point.

The user clicks at point **s** and drags to **e**, and then release to finish, the **green points** are selected based on the local covariance information (all points within the Mahalanobis ellipse, which corresponds to location parameter **m**).

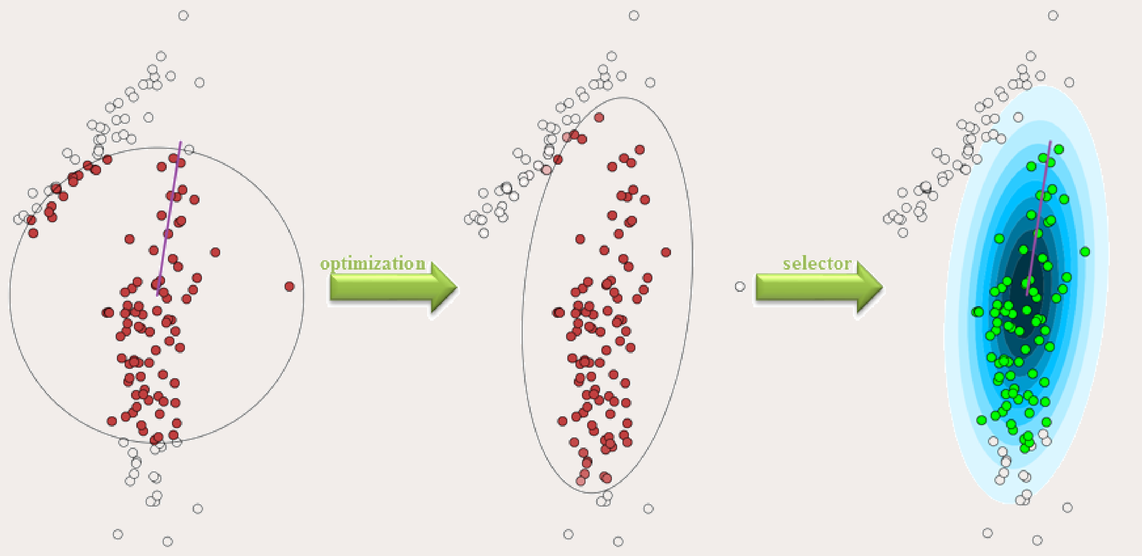


To obtain a more reasonable sample for the covariance computation, we use the weighted covariance to refine the sample iteratively by giving the sample which is stable in Mahalanobis ellipse higher weight during iteration.



Shades of red show points which are weighted, accordingly.

- 1st panel: red points in a circular neighborhood are used to compute the first weighted covariance matrix denoted as Σ_1 (illustrated by one ellipse);
- 2nd panel: the ellipse illustrates Σ_2 (2nd iteration);
- 3rd panel: the ellipse illustrates Σ_{10};
- 4th panel: the ellipse Σ_{100} (converged).



Initial samples with same weight

Converged Mahalanobis ellipse, points frequently in the ellipse have higher weight

Selector contours, brushing results are colored in green

User study and optimization

We collected information about both the brushing goals (which dataset subset did user wish to brush by using lasso) as well as the click- and release- points from the sketching interaction.

- We provided 6 representative datasets and 2 related questions for each dataset;
- 50 participants, and 12 selections are required for each one;
- 400 selections for training and 200 for validation.

Results

We improved the brushing accuracy from 65% (original Mahalanobis brush [1]) to 92% (our technique), when evaluated on the validation data. The average time user spent on the interaction is 41% of Lasso. And our brushing algorithm only cost 20ms for the computation of brushing 2000 points.

References

1. Radoš, Sanjin, Rainer Splechtna, K. Matković, M. Đuras, Eduard Gröller, and Helwig Hauser. "Towards Quantitative Visual Analytics with Structured Brushing and Linked Statistics." In Computer Graphics Forum, 2016.

UNIVERSITY OF BERGEN