Exploring High Dimensional Data Through Locally Optimized Viewpoint Selection

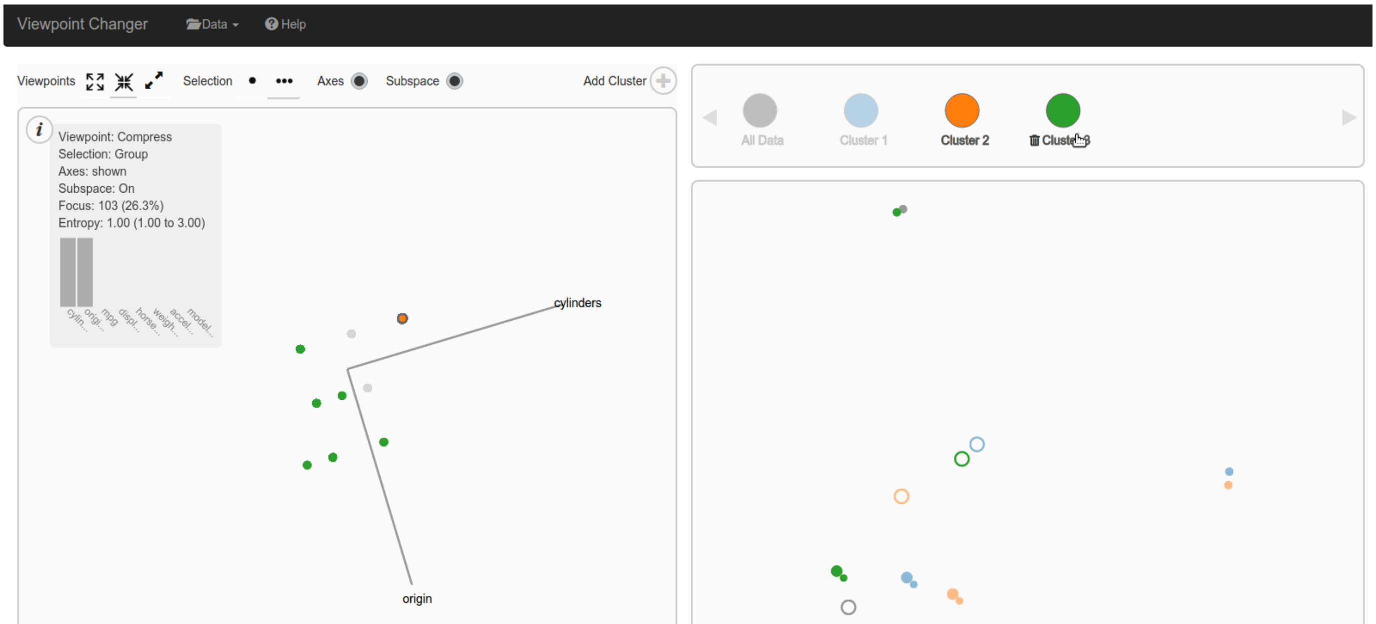

Fig. 1. Teaser here.

Abstract—Dimension reduced projection is usually obtained via a global optimization. It gives a good overview of the data, but cannot satisfy different users with different focuses. That's because any local data could be distorted and misleading due to projection errors. To address this problem, we propose an interactive visualization method, to customize projections for better local analysis. First, we allow users to define their point of interest (POI) data. Then we generate local projections to minimize distortions regarding the POI. Multipe optimizations are provided for different analyses. We also reveal relationships among different POIs, by comparing their local projections. At last, our method is proved effective via case studies with a real-world dataset.

Index Terms—Dimension-reduced projection, local analysis, high-dimensional data

1 Introduction

Dimension-reduced projection is widely used for high-dimensional data analysis. It seeks to approximate the original distribution in a low-dimensional space. Such approximation is often globally optimized to make a good overview of the data. But due to approximation errors, relationships among data will inevitably be distorted. The distortions are hard to ignore when it comes to local data. They largely harm the perception of local structures, yet are often transparent to users. Even if they are shown in the projection [17], users have no means to control their distribution. In general, dimension-reduced projections make good overviews, but are not suitable for local data analysis.

Nevertheless, local analysis is not only necessary, but also an efficient way to explore high-dimensional data. On one hand, a sole overview cannot show all aspects. After a quick glance, users tend to pick up some local structure (e.g. clusters, outliers) and go into details. It helps them fully understand the data and discover more useful information. On the other hand, featured local data are good breakthrough points for an efficient exploration. Traditional methods [14] [7] [11] allow users to manipulate projection dimensions. But such exploration is blind and exhausting, since the data space is huge while users have no clue where to go. Keeping some cluster However, such Lots of traditional methods provide clustering results as starting points.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org.
Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx/

2 Related Work

2.1 Dimension-Driven Projection Exploration

Dimension Manipulation and Grand Tour [14] [11] [7]

2.2 Data-Driven Projection Pursuit

Projection Pursuit [4] [3] [15] [8]

Projection Pursuit for Classification [16] [2]

Targeted Projection Pursuit [9] [1] [5] [6]

2.3 High-Dimensional Local Data Analysis

Subspace Clusters [18] [12]

Projection Clusters and Error Analysis [10] [13] [17]

3 Overview

4 High-Dimensional Exploration Through Locally Optimized Viewpoints

4.1 Find a Focus

4.1.1 Focus in Different Granularities

4.1.2 Distortion Hints

4.1.3 Focus Suggestion

4.2 Featured Viewpoints for the Focus

4.2.1 Desirable Local Features

4.2.2 Feature Enhanced Projections

4.2.3 Subspace Suggestion

4.3 From Focus to Cluster

4.3.1 Focus Improvement

4.3.2 Cluster Comparison via Viewpoint Map

5 Case Study

6 Discussion

7 Conclusion

Acknowledgments

The authors wish to thank A, B, C. This work was supported in part by a grant from XYZ.

References

[1] E. T. Brown, J. Liu, C. E. Brodley, and R. Chang. Dis-function: Learning distance functions interactively. In *2012 IEEE Conference on Visual Analytics Science and Technology, VAST 2012, Seattle, WA, USA, October 14-19, 2012*, pages 83–92, 2012.

[2] J. Choo, H. Lee, J. Kihm, and H. Park. ivisclassifier: An interactive visual analytics system for classification based on supervised dimension reduction. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology, IEEE VAST 2010, Salt Lake City, Utah, USA, 24-29 October 2010, part of VisWeek 2010*, pages 27–34, 2010.

[3] D. Cook, A. Buja, J. Cabrera, and C. Hurley. Grand tour and projection pursuit. *Journal of Computational and Graphical Statistics*, 4(3):155–172, 1995.

[4] J. H. Friedman and J. W. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Computers*, 23(9):881–890, 1974.

[5] M. Gleicher. Explainers: Expert explorations with crafted projections. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2042–2051, 2013.

[6] X. Hu, L. Bradel, D. Maiti, L. House, C. North, and S. Leman. Semantics of directly manipulating spatializations. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2052–2059, 2013.

[7] D. H. Jeong, C. Ziemkiewicz, B. D. Fisher, W. Ribarsky, and R. Chang. ipca: An interactive system for pca-based visual analytics. *Comput. Graph. Forum*, 28(3):767–774, 2009.

[8] S. Johansson and J. Johansson. Interactive dimensionality reduction through user-defined combinations of quality metrics. *IEEE Trans. Vis. Comput. Graph.*, 15(6):993–1000, 2009.

[9] P. Joia, D. Coimbra, J. A. Cuminato, F. V. Paulovich, and L. G. Nonato. Local affine multidimensional projection. *IEEE Trans. Vis. Comput. Graph.*, 17(12):2563–2571, 2011.

[10] E. Kandogan. Just-in-time annotation of clusters, outliers, and trends in point-based data visualizations. In *2012 IEEE Conference on Visual Analytics Science and Technology, VAST 2012, Seattle, WA, USA, October 14-19, 2012*, pages 73–82, 2012.

[11] D. J. Lehmann and H. Theisel. Orthographic star coordinates. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2615–2624, 2013.

[12] S. Liu, B. Wang, J. J. Thiagarajan, P. Bremer, and V. Pascucci. Visual exploration of high-dimensional data through subspace analysis and dynamic projections. *Comput. Graph. Forum*, 34(3):271–280, 2015.

[13] R. M. Martins, D. B. Coimbra, R. Minghim, and A. C. Telea. Visual analysis of dimensionality reduction quality for parameterized projections. *Computers & Graphics*, 41:26–42, 2014.

[14] J. E. Nam and K. Mueller. Tripadvisor$^{\mathrm{n\text{-}d}}$: A tourism-inspired high-dimensional space exploration framework with overview and detail. *IEEE Trans. Vis. Comput. Graph.*, 19(2):291–305, 2013.

[15] J. Seo and B. Shneiderman. A rank-by-feature framework for interactive exploration of multidimensional data. *Information Visualization*, 4(2):96–113, 2005.

[16] M. Sips, B. Neubert, J. P. Lewis, and P. Hanrahan. Selecting good views of high-dimensional data using class consistency. *Comput. Graph. Forum*, 28(3):831–838, 2009.

[17] J. Stahnke, M. Dörk, B. Müller, and A. Thom. Probing projections: Interaction techniques for interpreting arrangements and errors of dimensionality reductions. *IEEE Trans. Vis. Comput. Graph.*, 22(1):629–638, 2016.

[18] X. Yuan, D. Ren, Z. Wang, and C. Guo. Dimension projection matrix/tree: Interactive subspace visual exploration and analysis of high dimensional data. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2625–2633, 2013.

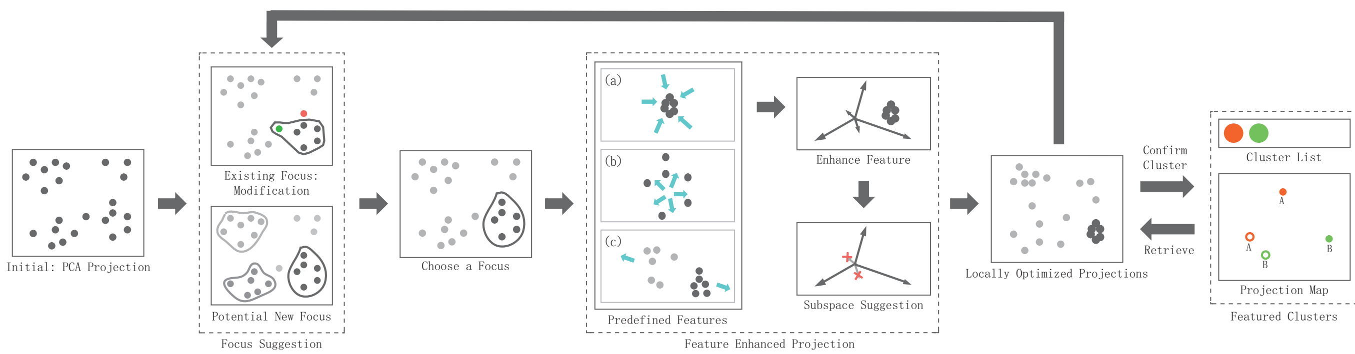

Fig. 2. The overview of delivery system.