Exploring High Dimensional Data Through Locally Optimized Viewpoint Selection

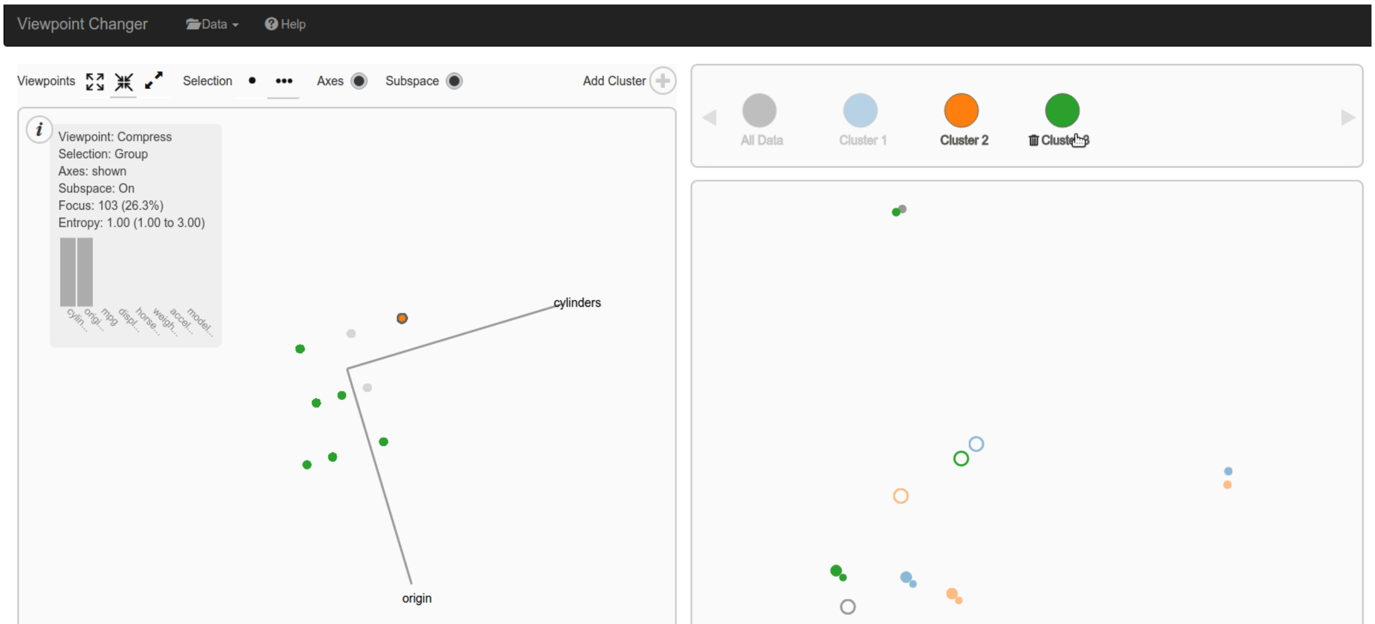

Fig. 1. Teaser here.

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, data relationships 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 distortions are shown in the projection [20] [1], users have no means to control their distribution. In general, globally optimized projections make good overviews, but are not suitable for local data analysis.

One way to alleviate this problem, is to observe the data in a different perspective. Some previous works [9] [16] [13] allow users to change the projection by adjusting dimension weights. It helps to clarify detailed structures and find out more useful information. To help build a targeted analysis, featured clusters are often provided beforehand [16] [14]. However, users know nothing about the given clusters. They have to solve their puzzles by manually searching the enormous data space. It's a blinded and exhausting process, where users have no clue how to steer the dimensions to get a better view. Even if interesting projections are found, it's hard to explain or assess them without distortion information.

Compared to dimension-based exploration, feature-based mining techniques are more efficient in revealing informative projections. Projection pursuit [6] searches for projections to optimize predefined indices. The rank-by-feature framework [18] ranks a series of projections by their interestingness. In more recent works, users are involved to specify desired features [10] and relationships [8] [7]. These methods, though effective, largely depend on predefined metrics or users' prior knowledge. It makes them unsuitable for an interactive exploration starting from scratch. Besides, little attention had been paid to facilitate local data analysis.

In this work, we propose an interactive exploration method, that promotes consecutive local data analyses in locally optimized projections. To be specific, the exploration contains three steps.

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/

(1) First, for any given projection, we help the user find a piece of interesting local data. Projection distortion and cluster suggestions are displayed to indicate potential outliers and clusters. The data chosen by user is called a 'focus', meaning that it's the current focus in local analysis.

(2) After some focus is chosen, we apply projection pursuit to find its most featured projections. By 'features', we refer to three kinds of relationships defined based on data distances. Projections are optimized to show these features with the least distortion.

(3) With the distortion-free projections, users can perceive and analyze the focus more correctly. We provide suggestions to help shape the focus into a complete cluster.

(4) In the featured projections, users can

2 Related Work

2.1 Dimension-Driven Projection Exploration

Dimension Manipulation and Grand Tour [16] [13] [9]

2.2 Data-Driven Projection Pursuit

Projection Pursuit [6] [5] [18] [10] [17]
Projection Pursuit for Classification [19] [4]
Targeted Projection Pursuit [11] [2] [7] [8]

2.3 High-Dimensional Local Data Analysis

Projection Clusters and Error Analysis [12] [15] [20]
Subspace Cluster Estimation [3] [21] [14]

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] M. Aupetit. Visualizing distortions and recovering topology in continuous projection techniques. *Neurocomputing*, 70(7-9):1304–1330, 2007.

[2] 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.

[3] K. M. Carter, R. Raich, and A. O. H. III. On local intrinsic dimension estimation and its applications. *IEEE Trans. Signal Processing*, 58(2):650–663, 2010.

[4] 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.

[5] 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.

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

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

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

[17] D. T. Nhon and L. Wilkinson. Scagexplorer: Exploring scatterplots by their scagnostics. In *IEEE Pacific Visualization Symposium, PacificVis 2014, Yokohama, Japan, March 4-7, 2014*, pages 73–80, 2014.

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

[19] 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.

[20] 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.

[21] 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.

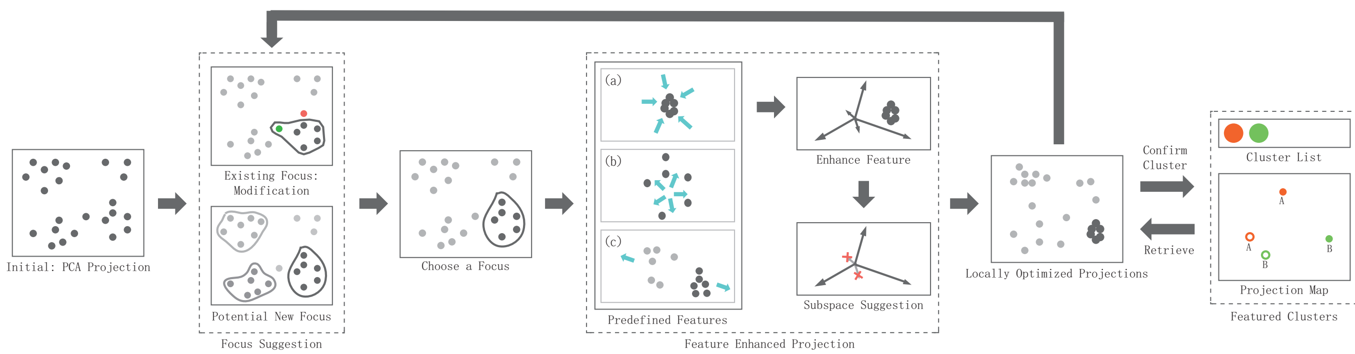

Fig. 2. The overview of delivery system.