Interactive Poster: XmdvTool: Interactive Visual Data Exploration System For High-dimensional Data Sets *

Matthew O. Ward, Elke A. Rundensteiner, Jing Yang, Punit R. Doshi and Geraldine Rosario
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609
{matt,rundenst,yangjing,punitd,ger}@cs.wpi.edu

Keywords: Dimension reduction framework, high-dimensional data visualization, interactive dimension reduction, radial space-filling hierarchy visualization, multi-focus distortion, visualization-driven data management.

1 Multi-Display Visual Data Exploration Basics

XmdvTool is a public-domain software developed at WPI for the interactive visual exploration of large-scale high-dimensional data sets [4, 1, 2, 5]. XmdvTool supports a user-driven interactive discovery process of data patterns and outliers. Our multidisciplinary approach of coupling ideas from the fields of data visualization, data mining and database management has allowed us to overcome major hurdles, including those of display clutter due to high dimensions and display clutter due to large number of data items, as well as the intuitive navigation and efficient data access over such database sets to allow near real-time exploration.

XmdvTool incorporates several distinct display methods for multivariate data visualization that allow the users to view as well as manipulate data from different perspectives [5]. The displays are tightly linked, such that visual interactions via one display can be seamlessly refined via other displays. XmdvTool supports a variety of advanced visual interaction tools, including brushing in screen space, data space, and structure space, panning, zooming and distortion [2].

2 Visual Exploration Scale-Up

Conventional multivariate visualization techniques do not scale well with respect to the number of objects in the data set, nor to the number of dimensions, resulting in a display with an unacceptable level of clutter. To overcome this limitation, XmdvTool adopts a hierarchical approach and provides a suite of navigation and filtering tools that facilitate the viewing of objects and of dimensions at different levels of detail. This allows the systematic discovery of data trends.

Our past work on structure-based brushing [2] addressed the hierarchical display of a large number of objects. In this demo, we now instead present our latest work on tackling visual exploration when faced with data sets that consist of large numbers of dimensions. When visualizing data sets with a large number of dimensions, existing multi-dimensional visualization techniques (such as parallel coordinates, glyphs and scatterplot matrices) become cluttered and thus ineffective. In particular, in this demonstration, we will showcase two key techniques towards overcoming this problem, namely, our visual hierarchical dimension reduction framework [7] and our visual exploration tool called the InterRing display [6] which addresses the hierarchical display and manipulation control of a large number of dimensions. To support efficient database access, our latest work on the database management side includes MinMax tree representation of queries, semantic caching, and prefetching [3].

*This work is supported under NSF grants IIS-9732897, IRIS-9729878, and IIS-0119276.

3 Visual Hierarchical Dimension Reduction

To address this problem of high-dimensional clutter, we reduce the dimensionality of the data while maintaining the relationships between data points using a *visual hierarchical dimension reduction approach* [7]. We note that existing dimensionality reduction techniques usually generate lower dimensional spaces that have little intuitive meaning to users and allow little user interaction. In our demonstration, we will instead show a new approach to handling high dimensional data, named Visual Hierarchical Dimension Reduction (VHDR), which addresses these drawbacks.

In VHDR, we construct hierarchical dimension cluster trees based on clustering the dimensions instead of the data points. Once dimensions are grouped into a hierarchy, lower dimensional spaces are constructed using clusters of the hierarchy. Thereafter, we construct low dimensional data spaces guided by user interaction of the hierarchical dimension cluster tree. VHDR not only generates lower dimensional spaces that are meaningful to users, but also allows user interactions in most steps of the process.

Furthermore, we have extended several traditional multidimensional visualization methods, including parallel coordinates, glyphs and scatterplot matrices, to visually convey dimension cluster characteristics when visualizing the data set in lower dimensional spaces. Our case study of applying VHDR to real data sets (Figure 1) confirms that this approach is effective in supporting the exploration of high dimensional data sets.

4 InterRing: Visual Interaction Tool

To allow the user to view, navigate and reconfigure the hierarchy of dimensions, we have developed the InterRing display tool (Figure 2). InterRing is a radial, space-filling (RSF) hierarchy visualization display. As an RSF-based display, it has the ability to efficiently use the display space while effectively conveying the hierarchy structure. Several RSF systems and tools have been developed to date, each with varying degrees of support for interac-

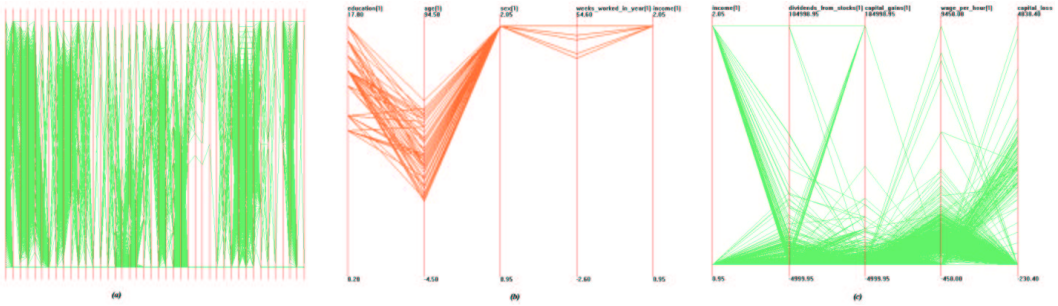

Figure 1: The Census dataset (42 dimensions, 20,000 data items) in parallel coordinates. Figure (a) shows the original high dimensional space. Figure (b) and (c) show two lower dimensional subspaces generated by VHDR.

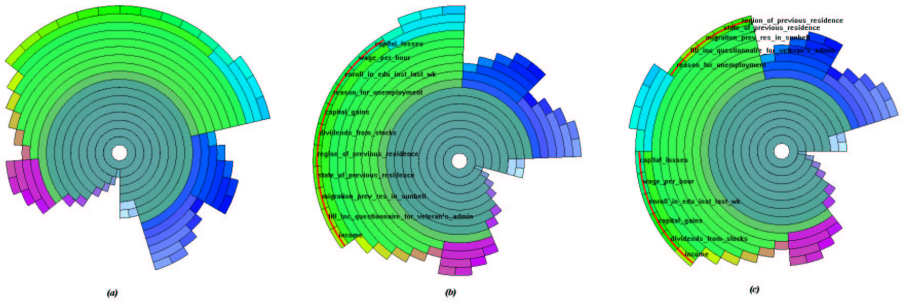

Figure 2: Dimension hierarchy of the Census dataset in InterRing. Figure (a) shows the automatically generated hierarchy. Figure (b) shows the detail of a cluster after brushing and rotation. Figures (c) shows the modified hierarchy after moving some dimensions from that cluster to elsewhere.

tive operations such as selection and navigation. In this demo, we show an extensive set of desirable operations on hierarchical structures, a set which we believe is significantly more complete than what is provided in prior RSF systems. Among other things, InterRing supports multi-focus distortions, interactive hierarchy reconfiguration, and both semi-automated and manual selection. In our demonstration, we show the power and effectiveness of these and other operations.

5 Scalability of Data Access

To allow for the near real-time behavior required among interactive tools, we address how database queries can be internally represented, and also employ data caching and prefetching techniques. To support operations such as visual hierarchical drill-down and roll-up, we represent the hierarchical structure as a Min-Max tree [3], thereby reducing the recursive hierarchy navigation queries into range queries. MinMax has allowed us to achieve performance levels required for interactive visualization even when connecting to large persistent data sets on Oracle.

Exploration via visual interaction tools typically results in predictable user navigation patterns. To take advantage of this, we use customized caching and prefetching techniques which we show to be effective in our tool. In particular, we employ semantic caching principles to maintain relevant results of the previously executed queries indexed by their query specifications in the local client buffer. We also apply a variety of prefetching techniques based on both the analysis of current user navigation patterns as well as the archived user navigation history.

6 XmdvTool Source Code and Homepage

XmdvTool 5.0 is implemented in `C/C++` with TclTk and OpenGL primitives. Interaction to Oracle8i is written in `Pro*C/C++` embedded SQL primitives. The XmdvTool Home Page at `http://davis.wpi.edu/~xmdv` provides downloads of regular releases of our software.

References

[1] Y. Fua, M. Ward, and E. Rundensteiner. Hierarchical parallel coordinates for exploration of large datasets. *Proc. of Visualization '99, p. 43-50*, Oct. 1999.

[2] Y. Fua, M. Ward, and E. Rundensteiner. Navigating hierarchies with structure-based brushes. *Proc. of Information Visualization '99, p. 58-64*, Oct. 1999.

[3] I. Stroe, E. A. Rundensteiner, and M. O. Ward. Scalable visual hierarchy exploration. *Database and Expert Systems Applications, p. 784-793*, 2000.

[4] M. Ward. Xmdvtool: Integrating multiple methods for visualizing multivariate data. *Proc. of Visualization '94, p. 326-33*, 1994.

[5] J. Yang, M. O. Ward, and E. A. Rundensteiner. Interactive hierarchical displays: A general framework for visualization and exploration of large multivariate data sets. *Computer & Graphics, to appear*, 2002.

[6] J. Yang, M. O. Ward, and E. A. Rundensteiner. Interring: An interactive tool for visually navigating and manipulating hierarchical structures. *InfoVis 2002, accepted*, 2002.

[7] J. Yang, M. O. Ward, and E. A. Rundensteiner. Visual hierarchical dimension reduction for exploration of high dimensional datasets. *Technical Report WPI-CS-TR-02-22*, 2002.