Multidimensional Data Visualization

Dmitri Eidenzon
NovoSpark Corporation, Canada

Olga Pilipczuk
University of Szczecin, Poland

INTRODUCTION

The visualization methods of multidimensional data have been studied in various research areas for many years. Today data visualization is a powerful tool for business analytics. Using the different visualization methods they are able to transform the data into graphical primitives of different colors, shapes, sizes, and locations.

The goal of using the visualization methods in business analytics is to show the input data in a dimensional space so that properties of the data set structure are preserved as clearly as possible and to observe the insight into the specific properties of the data by analyzing the records distribution and finding clusters of records. Other advantages of multidimensional visualization are possibility to visually summarize a data set and find key trends and relationships among the variables. The large data sets become ubiquitous in business and bear an enormous potential. Thus, it gets very important to find ways to transform and present data adequately. However, the screen space for displaying multidimensional data is limited and the size of the data sets often exceeds the number of pixels on the screen. Another problem occurs when we have the deal with more than three dimensions. Displaying such data sets in two or three dimensions, which is the normal for displaying tools, become a real challenge.

The visualization specialists try to overcome these problems. They are still looking for the novel methods and tools for multidimensional data visualization.

BACKGROUND

The history of the multidimensional visualization development is very long and can be divided into four stages:

1. **The beginning stage:** From 1700 to 1970 - then the point symbols were used to show the geographical distribution of 56 commodities in Europe, scientists started to research on visual texture perception, and texture mapping works were created (Wong, & Bergeron, 1997);
2. **The forming stage:** From 1970 to 1985 – the stage of new methods creation, such as Chernoff faces, parallel coordinates, iconography, words within words and many others; some of them were computerized (Wong, & Bergeron, 1997);
3. **The mining stage:** From 1985 till now - modification and combination the well-known visualization methods with new elements; attempts to combine different types of expression, for example of sound with graphics, were made;
4. **The cognitive visualization stage:** From 2000 till now – the dissemination of cognitive visualization conception and new cognitive multidimensional visualization methods creation.

Visualization methods can be divided into three groups (Wong, & Bergeron, 1997):

1. **Methods based on two-variate display:** Include the fundamental two-variate displays and simultaneous views of two-variate displays. Most of them have been used in statistics. These techniques applied to data that have relatively small size, about hundreds of items.

DOI: 10.4018/978-1-4666-5888-2.ch153

Copyright © 2015, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

2. **Methods based on multivariate display:** Use colorful plots created by high-speed computer graphics computation. These techniques applied to data that are larger and more complicated than those used in techniques based on two-variate display.
3. **Animation methods:** Are a multifunctional tool for visualizing multidimensional data. They include various movie animation models and techniques.

The traditional methods of multidimensional data visualization are: parallel coordinates, line graphs, survey plots, scatter plot matrix and it's variations, star glyphs, treemaps, Sammon's mapping, self-organizing map, dendrogram, radar chart, Voronoi diagrams, parallel glyphs, Bertin's permutation matrices, Chernoff faces, worlds within worlds, table lens, VisDB, dynamic queries, attribute explorer et.

Very popular are the methods of dimensionality reduction (van der Maaten, Postma, & van der Herik, 2009). The most widespread of them is the Principal Component Analysis (Jolliffe, 2002). Other methods of dimentionality reduction are: Linear Discriminant Analysis, Simple PCA, Probabilistic PCA, Factor Analysis, Sammon Mapping, Isomap, Landmark Isomap, Locally Linear Embedding, Laplacian Eigenmaps, Hessian LLE, Local Tangent Space Alignment, Diffusion Maps et.

The most common visualization techniques to present multidimensional data are the scatter plot, scatter plot matrix, parallel coordinates, Voronoi diagram, Chernoff faces, principal component analysis, self-organizing map et.

Scatter plot presents the data as a set of points, each having the value of horizontal and vertical axes to show the relationship between two sets of data.

The scatter plot matrix is a matrix where each column and row contains the same X-axis and Y-axis, respectively. Each cell in the matrix is presented as a scatter plot.

Parallel coordinates is a very popular method for visualizing and analyzing the multivariate data. In this method, each dimension is related to an axis. A point in *N*-dimensional space is presented as a line with vertices on the parallel axes. The position of the vertex on the axis corresponds to the coordinate of this point.

A very common technique for reducing data dimensionality is principal component analysis (PCA). It finds the data subspaces concentrated in the main space of a coordinate system. Size reduction is accomplished via removal of the small variance of the characteristics.

Another method of transforming multidimensional data into two-dimensional space is Kohenen Network. This network presents the data in the form of self-organizing map (SOM) units (neurons) and has the shape of a two-dimensional grid. It is very useful for detection and visualization of clusters in a data set.

Voronoi diagram consists of cells (fields), which have only one entry point inside. The edge of the diagram is equidistant from two adjacent input points. It means that the dividing line between the two points located midway between them.

Another method of multi-dimensional data visualization is Chernoff Faces. "Face" is a picture that reproduces each observation separately. Individual part of the face attributed to the values of the tested variables by their shape, size, placement and orientation.

However, all the traditional methods and techniques of multidimensional data visualization have some disadvantages or weaknesses. For example, a problem using the star glyphs is in organizing the star glyphs on the screen in a meaningful manner. Furthermore, a limitation of the radar chart is a lack of transparency in the analysis in cases of a large number of items. Restrictions for using parallel coordinates are the same, i.e. in cases where a large number of visual characteristics are used the chart loses its clarity and the analysis becomes impossible. The second drawback is that the variables must be in common scale. The disadvantage of the PCA method is that it is unsuitable for the analysis of a non-linear data structure. A key limitation of this method is that it is based only on the quantitative data and cannot be applied in cases when the only available data is information about object similarity. Kohonen maps are mostly useful for the detection and visualization of clusters in a data set; they do not provide the ability of dynamic data visualization.

Thus, using Voronoi diagrams, we have difficulties in considering dynamic changes. The second drawback is the lack of tolerance for sensory faults. The analysis of images created by using the Chernoff faces technique poses many problems; one of them is that it is hard to compare too many features of the face.

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/multidimensional-data-visualization/112564

Related Content

Methodologies of Damage Identification Using Non-Linear Data-Driven Modelling
Miguel Angel Torres Arredondo, Diego Alexander Tibaduiza Burgos, Inka Buethe, Luis Eduardo Mujica, Maribel Anaya Vejar, Jose Rodellar and Claus-Peter Fritzen (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 978-991).*
www.irma-international.org/chapter/methodologies-of-damage-identification-using-non-linear-data-driven-modelling/112491

Modified Distance Regularized Level Set Segmentation Based Analysis for Kidney Stone Detection
K. Viswanath and R. Gunasundari (2015). *International Journal of Rough Sets and Data Analysis (pp. 24-41).*
www.irma-international.org/article/modified-distance-regularized-level-set-segmentation-based-analysis-for-kidney-stone-detection/133531

Medco: An Emergency Tele-Medicine System for Ambulance
Anurag Anil Saikar, Aditya Badve, Mihir Pradeep Parulekar, Ishan Patil, Sahil Shirish Belsare and Aaradhana Arvind Deshmukh (2017). *International Journal of Rough Sets and Data Analysis (pp. 1-23).*
www.irma-international.org/article/medco/178159

A Tool for Creating Community Knowledge Objects
Zbigniew Mikolajuk (2019). *Handbook of Research on the Evolution of IT and the Rise of E-Society (pp. 49-66).*
www.irma-international.org/chapter/a-tool-for-creating-community-knowledge-objects/211610

Hybrid Data Mining Approach for Image Segmentation Based Classification
Mrutyunjaya Panda, Aboul Ella Hassanien and Ajith Abraham (2016). *International Journal of Rough Sets and Data Analysis (pp. 65-81).*
www.irma-international.org/article/hybrid-data-mining-approach-for-image-segmentation-based-classification/150465