Multidimensional Visual Analysis

Survey of Multidimensional Visual Analysis Methods and Software

Ožbej Golob

Graz University of Technology

11 Jan 2023

Abstract

Writing a survey can be a traumatic endeavour. It might be a student's first foray into academic research. There are often obstacles and false dawns along the way. This survey paper takes a fresh look at the process and addresses new ways of accomplishing this daunting goal.

The abstract should concisely describe what the survey is about. State the areas which are covered and also those which are not covered. Market your survey to your readership. Also, make sure you mention all relevant keywords in the abstract, since many readers read *only* the abstract and many search engines index *only* the title and the abstract.

This survey explores the issues concerning the writing of an academic survey paper and presents numerous novel insights. Special attention is paid to the use of clear and simple English for an international audience, and advice is given regarding the use of technical aids to production.

Contents

List of Figures

List of Tables

List of Listings

Chapter 1

Introduction

Multidimensional Visual Analysis (MVA) is a field that focuses on the use of visual representations to explore and analyze complex data sets. This approach aims to provide a more intuitive and interactive way to understand and extract insights from large and complex datasets.

One of the main challenges in MVA is the effective representation of high-dimensional data in a way that is meaningful and intuitive for the user. To address this challenge, a wide range of visual encodings and interaction techniques have been developed, including scatter plots, parallel coordinates, and multidimensional scaling.

One of the key benefits of MVA is its ability to reveal patterns and relationships in the data that may not be apparent through traditional statistical analysis methods. This is particularly useful in the exploratory phase of data analysis, where the aim is to gain a better understanding of the data and identify potential areas of interest for further investigation.

In addition to its use in exploratory data analysis, MVA also has a range of applications in areas such as data mining, machine learning, and business intelligence. By enabling users to interact with and visualize data in a more intuitive way, it can help to improve decision-making and facilitate the discovery of insights and trends that may not have been identified through other means.

Overall, MVA is an important tool for understanding and making sense of complex data, and has the potential to greatly enhance our ability to extract knowledge and insights from large and complex datasets.

In this survey, we will review multiple popular MVA approaches and popular MVA software. See Chapter 2 for MVA approaches and Chapter 3 for MVA software.

Chapter 2

Multidimensional Visual Analysis (MVA)

Multidimensional Visual Analysis (MVA) approaches are methods and techniques used to analyze and understand complex data sets using visual representations. These approaches typically involve the use of specialized software or tools that allow analysts to create and manipulate graphical representations of the data in order to uncover patterns, trends, and relationships. In this section, we will review some popular MVA approaches.

2.1 Scatter Plots

A scatter plot is a type of graph that is used to display the relationship between two numerical variables, to identify any potential trends or patterns in the data, and to identify outliers. It uses dots or markers to represent the values of the two variables, and position of each dot on the graph indicates the value of the two variables in a single observation.

To create a scatter plot, the values of one variable are plotted on the x-axis (horizontal axis) and the values of the other variable are plotted on the y-axis (vertical axis). The resulting graph will show a set of dots, with each dot representing a single observation. If there is a positive relationship between the two variables, the dots will tend to form a diagonal line that slopes upwards from left to right. If there is a negative relationship, the dots will tend to form a diagonal line that slopes downwards from left to right. If there is no relationship between the two variables, the dots will be scattered randomly across the graph.

2.2 Similarity Maps

Similarity maps are projections of high-dimensional datasets to two (or sometimes three) dimensions. Projection techniques are used to reduce the number of dimensions in the data, while attempting to preserve distances between items as far as possible. Items which are close in the high-dimensional space should also be close in the resulting two-dimensional projection space. Such projections can be further split into *linear* and *non-linear* projections.

2.2.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a linear projection. PCA uses linear algebra to identify the underlying dimensions or factors in the data and project the data onto a lower-dimensional space [Abdi and Williams 2010].

2.2.2 Multi-Dimensional Scaling (MDS)

Multi-Dimensional Scaling (MDS) is a non-linear projection. MDS uses a distance metric to preserve the distances between data points in the high-dimensional space and project the data onto a lower-dimensional space [Morrison et al. 2003].

2.2.3 t-Distributed Stochastic Neighbor Embedding (t-SNE)

t-distributed stochastic neighbor embedding (t-SNE) is a non-linear projection. t-SNE uses a probabilistic model to preserve the local structure of the data in the high-dimensional space and project the data onto a lower-dimensional space [Van der Maaten and Hinton 2008].

2.3 Parallel Coordinates

Parallel coordinates are a type of chart that is used to visualize multi-dimensional data. In this type of chart, each data point is represented by a vertical line that extends across multiple axes. The axes are typically arranged in parallel. Each axis represents a different variable, and the position of a point on that axis indicates the value of that variable for the data point. This allows multiple variables to be compared and analyzed simultaneously [Inselberg and Dimsdale 1990].

2.4 Brushing and Linking

Brushing and linking are techniques used in multidimensional visual analysis to allow the user to interact with a visualization and explore data in greater depth.

Brushing refers to the process of selecting data points or regions in one visualization and highlighting those data points in other visualizations. This allows the user to see how the data points or regions of interest are related to other variables in the dataset.

Linking refers to the process of synchronizing the views of multiple visualizations, such that a change made to one visualization is reflected in the other visualizations. This allows the user to explore the data from different perspectives and understand how different variables are related to one another.

Both brushing and linking are useful for helping users to identify patterns and relationships in the data and for facilitating the process of data exploration and analysis.

2.5 Grouping and Labelling

2.5.1 Manual Grouping

2.5.2 Automated Clustering

Cluster analysis is a type of statistical technique used to identify groups of similar objects within a data set. It is a commonly used method for exploratory data analysis, and is often used as a way to gain insight into the underlying structure of the data. In cluster analysis, the data is divided into groups, or clusters, based on the similarity of the objects within each group. This allows analysts to identify common patterns and trends within the data, and to gain a better understanding of the relationships between the objects in the data set [Duran and Odell 2013].

Chapter 3

Multidimensional Visual Analysis Tools

MVA software is a type of computer program that is designed to help analysts visualize and analyze complex data sets. These programs typically include a wide range of tools and features that allow users to create and manipulate graphical representations of the data, such as scatter plots, parallel coordinates, and heat maps. By using these tools, analysts can quickly and easily explore the data and gain insights that might not be immediately apparent from looking at the raw data. MVA software is commonly used in fields such as business, finance, and marketing to help make data-driven decisions and uncover hidden trends in the data. In this section, we will review some popular MVA software programs.

3.1 InfoScope

InfoScope [Macrofocus 2007] is an interactive visualization tool to access, explore, and communicate large or complex datasets. InfoScope is available as free software with the last update issued on 09 Feb 2007. InfoScope is available for Windows, macOS, and Linux.

InfoScope can visualize a collection of selected publicly available datasets, mainly from the finance sector. InfoScope provides an overview of global relationships between objects by using multiple views to show different aspects of the data at the same time. InfoScope provides the following views: *Carto Plot*, *Similarity Map* as high-dimensional projections, *Parallel Coordinates*, and *Table View*. InfoScope supports brushing and linking, so all views are highly interactive and tightly linked. Specific numeric values of attributes can be obtained by probing the objects in the views, and dynamic queries on a combination of attributes can be executed using range sliders. All actions are accompanied by visual feedback within a common frame of reference. The visual representations make it easy to identify outliers, patterns, or anomalies.

3.2 High-D

High-D [Macrofocus 2022] is the successor to InfoScope. It offers similar functionality with some improvements and added views. High-D is a versatile tool for revealing hidden features, highlighting trends and relationships, and finding anomalies in datasets of any size. At its heart is a powerful interactive parallel coordinates plot for quick data access, analytical, and presentation purpose. High-D is available as paid software for 199 USD and with 30-day free evaluation period. The last update was issued on 05 Dec 2022. High-D is available for Windows, macOS, and Linux.

High-D can visualize a collection of selected publicly available datasets as well as custom datasets. High-D provides an overview of global relationships between objects by using multiple views to show different aspects of the data at the same time. High-D provides the following views: *Parallel Coordinates*, *Table Plot*, *Distributions*, *Scatter Plot Matrix*, *Parallel Coordinates Matrix*, *Scatter Plot*, *Similarity Map* as high dimensional projections (Sammon, Spring, t-SNE, and PCA), *Tree Map*, and *Carto Plot*. High-D

also supports clustering with k-means++ algorithm. High-D supports brushing and linking, so all views are highly interactive and tightly linked. It is possible to obtain specific numeric values of attributes by examining the objects within the views, and dynamic queries on a combination of attributes can be performed using range sliders. All actions are accompanied by visual feedback within a common frame of reference. The visual representations make it easy to spot unusual data points, patterns, or anomalies.

3.3 GGobi

GGobi [Cook et al. 2007] is a visualization program for exploring high-dimensional data. GGobi is written in C and is available as a free and open-source software with the last update issued on 10 Jun 2012. GGobi is available for Windows, Apple macOS, and Linux.

GGobi can visualize custom datasets and provides a dynamic and interactive graphics as tours, where data is displayed in an animation. The data is also available in the following views: *Scatter Plot*, *Scatter Plot Matrix*, *Parallel Coordinates*, *Time Series*, *Distributions* as bar charts. GGobi supports limited brushing. Views offer limited interactivity and interpretability and are not closely connected.

3.4 mVis

mVis [Chegini et al. 2019] is a visual analytics tool for visualizing multi-dimensional data. mVis is written in Java and is available as a free and open-source software with the last update issued on 20 Jan 2021. mVis is available for Windows, macOS, and Linux.

mVis can visualize custom datasets and provides an overview of global relationships between objects by using multiple views to show different aspects of the data at the same time. mVis consists of four data visualization views: *Scatter Plot Matrix*, *Scatter Plot*, *Similarity Map* as high dimensional projections (t-SNE, PCA, and MDS), and *Parallel Coordinates*. mVis also consists of a panel for controlling data partitions. All of the visualizations are interconnected through standard brushing and linking, so that any changes or selections made in one view are reflected in all of the other views. Additionally, the user has the ability to close, rearrange, or expand any view as needed.

mVis also supports creating and modifying ML models, specifically introducing an interactive visual labeling technique that allows an analyst to build and iteratively improve a ML classification model for multi-dimensional data sets. This technique combines linked visualizations, clustering, and active learning to allow the analyst to interactively label a multi-dimensional dataset in an efficient manner.

3.5 Improvise

Improvise [Weaver 2014] is a program that allows users to create and interact with visualizations that are linked together in various ways. Improvise is written in Java and is available as a free and open-source software with the last update issued on 28 Oct 2020. Improvise is available for Windows, macOS, and Linux.

The program uses a shared-object coordination model and a declarative visual query language to give users control over how data is displayed in multiple views. This allows users to create visualizations with a variety of coordination patterns, such as synchronized scrolling, overview and detail, drill-down, and semantic zoom. Improvise also has a user interface that allows users to build and explore visualizations in a live environment, making it easy to modify visualizations as needed. The goal of Improvise is to provide a high level of coordination flexibility while also being easy to use. It is designed to make it simple to create basic coordination patterns and also possible to create more complex ones.

3.6 MyBrush

MyBrush [Koytek et al. 2017] is an application that allows users to customize and control the brushing and linking process in their visualizations. It provides flexibility by allowing users to specify the source, link, and target of multiple brushes, and supports a variety of visualization types and multiple simultaneous brushes. Improvise is written in JavaScript and is available as a free and open-source web application with the last update issued on 22 Sept 2017.

MyBrush serves as experimental software and offers limited functionality. Its purpose is to explore the implemented brushing and linking functionality. A user can explore a predetermined set of data with the following views: *Scatter Plot*, *Parallel Coordinates*, and *Bar Plot*. Any changes or selections made in one of the visualizations will be reflected in all of the other views because they are all interconnected through standard brushing and linking.

3.7 XDAT

XDAT [XDAT 2020] is a multidimensional data analysis tool designed to help users quickly and easily extract valuable insights from large, complex data sets with many variables. XDAT is written in Java and is available as a free software with the last update issued on 26 Aug 2020. XDAT is available for Windows, macOS, and Linux.

XDAT can visualize custom datasets and displays data in separate views. XDAT displays the data with the following views: *Parallel Coordinates*, *Table View* and *Scatter Plot*. All of the visualizations are interconnected through standard brushing and linking, so that any changes or selections made in one view are reflected in all of the other views.

3.8 TabuVis

TabuVis [Nguyen et al. 2013] is a flexible and customizable visual analytics system that is optimized for analyzing multidimensional data. Its visualizations can be customized by domain experts to suit the specific needs of the data being analyzed. TabuVis is written in Java and is available as a free software with the latest update issued on 19 Feb 2022. TabuVis is available for Windows, macOS, and Linux.

TabuVis can visualize custom datasets and displays data in separate views. TabuVis includes various features for analyzing data, such as the ability to process data, add automatic marks, create custom interactive visualizations, and filter the data. These features are designed to support the entire data analysis process. TabuVis displays the data in the following views: *Scatter Plots*, *Parallel Coordinates*, and *Star Plot*. TabuVis doesn't support brushing and linking.

3.9 Comparison of Tools

A comparison of general information of MVA software can be seen in Table 3.2. A comparison of features of MVA software can be seen in Table 3.3.

Software	Last update	Licence	Systems	Language
InfoScope	09 Feb 2007	Free demo	Win, macOS, Linux	?
High-D	05 Dec 2022	Commercial	Win, macOS, Linux	?
GGobi	10 Jun 2012	Free, open-source	Win, macOS, Linux	C
mVis	20 Jan 2021	Free, open-source	Win, macOS, Linux	Java
Improvise	28 Oct 2020	Free, open-source	Win, macOS, Linux	Java
MyBrush	22 Sept 2017	Free, open-source	Web Browser	JavaScript
XDAT	26 Aug 2020	Free	Win, macOS, Linux	Java
TabuVis	19 Feb 2022	Free	Win, macOS, Linux	Java

Table 3.1: Overview of MVA tools.

Info	InfoScope	High-D	GGobi	mVis	Improvise	MyBrush	XDAT
Last Update	09 Feb 2007	05 Dec 2022	10 Jun 2012	20 Jan 2021	28 Oct 2020	22 Sept 2017	26 Aug 2020
Licence	Free demo	Commercial	Free, open-source	Free, open-source	Free, open-source	Free, open-source	Free

Table 3.2: Overview of MVA tools.

Software	Custom Datasets	Brushing	Linking	Table View	Scatter Plot	Scatter Plot Matrix	Parallel Coordinates	Parallel Coordinates Matrix	Similarity Map	Time Series	Distributions	Table Plot	Tree Map	Carto Plot
InfoScope	✓	✓	✓	✓			✓		✓					✓
High-D	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓
GGobi	✓				✓	✓	✓			✓	✓			
mVis	✓	✓	✓		✓	✓	✓		✓					
Improvise	✓			✓	✓	✓			✓	✓	✓	✓	✓	✓
MyBrush		✓	✓		✓		✓				✓			
XDAT	✓	✓	✓	✓	✓		✓							
TabuVis	✓				✓		✓							

Table 3.3: Comparison of MVA Tools.

Chapter 4

Concluding Remarks

At the end of your survey, give a clear recommendation as to which approach or tool to use in which situation.

Bibliography

Abdi, Hervé and Lynne J Williams [2010]. *Principal component analysis*. Wiley interdisciplinary reviews: computational statistics 2.4 (2010), pages 433–459 (cited on page 3).

Chegini, Mohammad, Jürgen Bernard, Philip Berger, Alexei Sourin, Keith Andrews, and Tobias Schreck [2019]. *Interactive labelling of a multivariate dataset for supervised machine learning using linked visualisations, clustering, and active learning*. Visual Informatics 3.1 (2019). Proceedings of Pacific-VAST 2019, pages 9–17. ISSN 2468-502X. doi:https://doi.org/10.1016/j.visinf.2019.03.002. https://www.sciencedirect.com/science/article/pii/S2468502X19300178 (cited on page 6).

Cook, Dianne, Deborah F Swayne, and Andreas Buja [2007]. *Interactive and dynamic graphics for data analysis with R and GGobi*. Volume 1. Springer, 2007 (cited on page 6).

Duran, Benjamin S and Patrick L Odell [2013]. *Cluster analysis: a survey*. Volume 100. Springer Science & Business Media, 2013 (cited on page 4).

Inselberg, Alfred and Bernard Dimsdale [1990]. *Parallel coordinates: a tool for visualizing multi-dimensional geometry*. Proceedings of the First IEEE Conference on Visualization: Visualization90. IEEE. 1990, pages 361–378 (cited on page 4).

Koytek, Philipp, Charles Perin, Jo Vermeulen, Elisabeth André, and Sheelagh Carpendale [2017]. *Mybrush: Brushing and linking with personal agency*. IEEE transactions on visualization and computer graphics 24.1 (2017), pages 605–615 (cited on page 7).

Macrofocus [2007]. *Macrofocus: InfoScope*. 09 Feb 2007. http://www.trenddisplay.com/public/products/infoscope.html (cited on page 5).

Macrofocus [2022]. *Macrofocus: High-D*. 05 Dec 2022. https://www.high-d.com/ (cited on page 5).

Morrison, Alistair, Greg Ross, and Matthew Chalmers [2003]. *Fast multidimensional scaling through sampling, springs and interpolation*. Information Visualization 2.1 (2003), pages 68–77 (cited on page 4).

Nguyen, Quang Vinh, Yu Qian, MaoLin Huang, and JiaWan Zhang [2013]. *TabuVis: a tool for visual analytics multidimensional datasets*. Science China Information Sciences 56.5 (2013), pages 1–12 (cited on page 7).

Van der Maaten, Laurens and Geoffrey Hinton [2008]. *Visualizing data using t-SNE.* Journal of machine learning research 9.11 (2008) (cited on page 4).

Weaver, Chris [2014]. *Chris Weaver: Improvise*. 21 Sep 2014. https://cs.ou.edu/~weaver/improvise/index.html (cited on page 6).

XDAT [2020]. *XDAT*. 26 Aug 2020. https://www.xdat.org/ (cited on page 7).