Chapter XI
Visualizing Multi Dimensional Data

César García-Osorio
University of Burgos, Spain

Colin Fyfe
The University of Paisley, UK

ABSTRACT

This chapter gives a survey of some existing methods for visualizing multi dimensional data, that is, data with more than three dimensions. To keep the size of the chapter reasonably small, we have limited the methods presented by restricting ourselves to numerical data. We start with a brief history of the field and a study of several taxonomies; then we propose our own taxonomy and use it to structure the rest of the chapter. Throughout the chapter, the iris data set is used to illustrate most of the methods since this is a data set with which many readers will be familiar. We end with a list of freely available software and a table that gives a quick reference for the bibliography of the methods presented.

INTRODUCTION

The advent of the personal computer has provided mankind with enormous benefits: we have access to more information than ever before and have such access virtually 24 hours per day, 7 days per week thanks to the Internet. However, it is in the nature of humankind that we always wish for more: in this case, we have all this data but actually finding information within the data is often an extremely complex task. This chapter will devote itself to this problem. We will restrict ourselves to numeric data: this is the simplest and perhaps most frequent form of data in which we seek information. The numeric data will often be multi dimensional: we might envisage information about an individual consisting of height, weight, bank balance, and so on, so that each of these fields constitutes one dimension of a long description of the individual. Thus, if we have 10 fields, we have a 10-dimensional vector describing each individual. The task then might be to identify groups of individuals, all of whom share some common characteristic. For example, it is known that tall people tend to earn more than their shorter brethren (a fact which causes some of us some disquiet). In order to ascertain this fact, we must have a way of identifying structure across dimension boundaries. This chapter will investigate methods for performing such identifications in a semi-automated manner.

We state semi-automated since we believe that the computer and the human both have roles to play in identifying structure: the computer is very good at handling large volumes of data and manipulating such data in an automatic manner; but humans are very good at pattern identification much better than computers (consider how face recognition systems have failed to live up to our expectations, even now). Therefore, we envisage a partnership between the human and the computer software with each performing the role to which they are best suited. The computer software will manipulate the multi dimensional data and present it to the human in a way which facilitates the human's pattern matching. An example of this is Exploratory Projection Pursuit (see later) in which the high-dimensional data is projected onto a two dimensional subspace in such a way that the structure of the data (for example, clusters) is most easily identified by the human eye. Another example is the use of Andrews' Curves in which each data point is represented by a curve. A human can run an eye along a set of curves (representing the members of the data set) and identify particular regions of the curve which are optimal for identifying clusters in the data set.

However, interestingly, several of the methods which we use to find structure (i.e., the computer software part of the partnership) are based on neural networks, the network of neurons which we have in our brains. Does this suggest that humans might be able to dispense with the computer software and perform the whole task themselves? It is a nice thought but common experience suggests that it is not so; we require intelligent software to help us find structure in multi dimensional data sets.

HISTORY

Let us see first a quick review of the history of visualization techniques. The roots of information visualization as a practical field can be established in the works of Tukey (1977), Bertin (1981, 1983), and Tufte (1983) who focused on 2D and 3D visualization and produced general rules for the plane, colour composition, and attribute mapping, among others. The use of the attributes of a database as dimensions was the rationale behind the study of the multi dimensional techniques. The contents of this chapter can be classified into this last category. However, the study of multivariate visualization began some centuries before; following Wong and Bergeron (1997), the evolution of the field can be divided into four periods:

1. **The searching stage (from 1782 to 1976):** Characterized by relatively small sized data and tools for data visualization that usually consisted of colour pencils and graph paper.
2. **The awakening stage (from 1977 to 1985):** Two and three dimensional spatial data were the most common data types being studied, although multivariate data started gaining more attention.
3. **The discovery stage (from 1986 to 1991):** The limited availability of high speed graphics hardware during the previous stage was gradually conquered. Most of the visualization methods presented in this chapter were developed in this period.
4. **The elaboration and assessment stage (from 1992 to present):** This period has been a retrenchment in the development of

Related Content

Development of a Framework for Preserving the Disease-Evidence-Information to Support Efficient Disease Diagnosis
Venkatesan Rajinikanthand Seifedine Kadry (2021). *International Journal of Data Warehousing and Mining (pp. 63-84).*
www.irma-international.org/article/development-of-a-framework-for-preserving-the-disease-evidence-information-to-support-efficient-disease-diagnosis/276765

A Look back at the PAKDD Data Mining Competition 2006
Nathaniel B. Norieland Chew Lim Tan (2007). *International Journal of Data Warehousing and Mining (pp. 1-11).*
www.irma-international.org/article/look-back-pakdd-data-mining/1780

Legal Issues in E-Healthcare Systems
Jawahitha Sarabdeen (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications (pp. 1752-1777).*
www.irma-international.org/chapter/legal-issues-healthcare-systems/73521

Robust Classification Based on Correlations Between Attributes
Alexandros Nanopoulos, Apostolos N. Papadopoulos, Yannis Manolopoulosand Tatjana Welzer-Druzovec (2007). *International Journal of Data Warehousing and Mining (pp. 14-27).*
www.irma-international.org/article/robust-classification-based-correlations-between/1787

Application of Big Data in Healthcare: Opportunities, Challenges and Techniques
Md Rakibul Hoqueand Yukun Bao (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 1189-1208).*
www.irma-international.org/chapter/application-of-big-data-in-healthcare/150211