

Chapter 2

A Visual Analytics Approach for Correlation, Classification, and Regression Analysis

Chad A. Steed

Oak Ridge National Laboratory, USA

J. Edward Swan II

*Bagley College of Engineering, Mississippi
State University, USA*

Patrick J. Fitzpatrick

*Northern Gulf Institute, Mississippi State
University, USA*

T.J. Jankun-Kelly

*Bagley College of Engineering, Mississippi
State University, USA*

ABSTRACT

New approaches that combine the strengths of humans and machines are necessary to equip analysts with the proper tools for exploring today's increasingly complex, multivariate data sets. In this chapter, a visual data mining framework, called the Multidimensional Data eXplorer (MDX), is described that addresses the challenges of today's data by combining automated statistical analytics with a highly interactive parallel coordinates based canvas. In addition to several intuitive interaction capabilities, this framework offers a rich set of graphical statistical indicators, interactive regression analysis, visual correlation mining, automated axis arrangements and filtering, and data classification techniques. This chapter provides a detailed description of the system as well as a discussion of key design aspects and critical feedback from domain experts.

INTRODUCTION

A byproduct of continued technological advances is increasingly complex multivariate data sets, which, in turn, yield information overload when explored with conventional visual analysis techniques. The ability to collect, model, and store

information is growing at a much faster rate than our ability to analyze it. However, the transformation of these vast volumes of data into actionable insight is critical in many domains (e.g. climate change, cyber-security, financial analysis). Without the proper techniques, analysts are forced to reduce the problem and discard layers of informa-

DOI: 10.4018/978-1-4666-4309-3.ch002

tion in order to fit the tools. New techniques and approaches are necessary to turn today’s flood of information into opportunity.

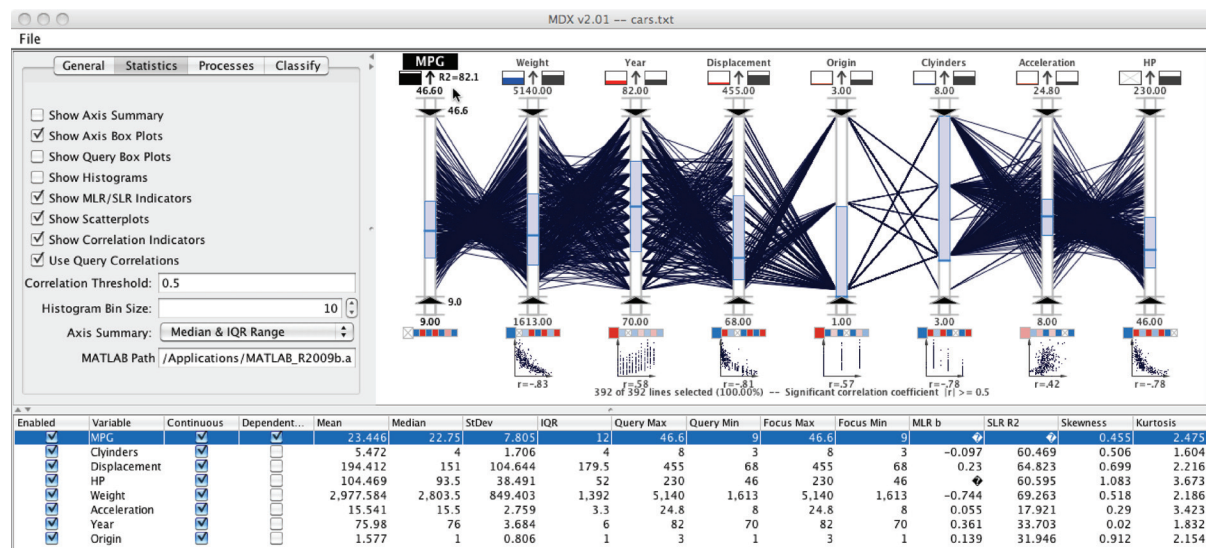
One of the most promising solutions for the so-called big data challenge lies in the continued development of techniques in the rapidly growing field of visual analytics. Visual analytics, also known as visual data mining, combines interactive visualizations with automated analytics that help the analyst discover and comprehend patterns in complicated, heterogeneous data sets. In general, visual analytics can be described as “the science of analytical reasoning facilitated by interactive visual interfaces” (Thomas, 2005). Visual analytics seeks to combine the strengths of humans with those of machines. While methods from knowledge discovery, statistics, and mathematics drive the automated analytics, human capabilities to perceive, relate, and conclude strengthen the iterative process.

In this chapter, a novel visual data mining framework—called the Multidimensional Data Explorer (MDX)—is presented that utilizes statistical analysis and data classification techniques in an interactive multivariate representation to improve

knowledge discovery in the complex multivariate data sets that characterize today’s data (see Figure 1). In addition to intuitive interaction capabilities, this framework introduces a rich set of graphical statistical indicators, automated regression analysis, visual correlation indicators, optimal axis arrangement techniques, and data classification algorithms. These capabilities are combined into a parallel coordinates based framework for enhanced multivariate visual analysis.

This chapter features an expanded version of MDX that builds on recent efforts in which MDX was applied to tropical cyclone climate studies. In Steed, Fitzpatrick, Jankun-Kelly, Yancey, and Swan II (2009b), the initial version of MDX, which lacked integrated statistical processes, was introduced and the system was demonstrated in a case study with a set of tropical cyclone predictors. Follow-on work by Steed, Fitzpatrick, Swan II, and Jankun-Kelly (2009a) and Steed, Swan II, Jankun-Kelly, Fitzpatrick (2009c) presented an enhanced version of MDX that included statistical analytics and deeper analysis of the previously analyzed tropical cyclone predictors, as well as analysis of a new set of predictors. In the current

Figure 1. The Multidimensional Data Explorer (MDX) consists of a settings panel (upper left), a data table panel (bottom), and an interactive parallel coordinates panel (upper right)



19 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/a-visual-analytics-approach-for-correlation-classification-and-regression-analysis/78712

Related Content

Blended Approach: Efficient Track Structure for High Speed Rail

Richard Harnish and F.K. Plous (2018). *Intelligent Transportation and Planning: Breakthroughs in Research and Practice* (pp. 1042-1056).

www.irma-international.org/chapter/blended-approach/197174

A Multi-Objective Ensemble Method for Class Imbalance Learning: Application in Prediction of Life Expectancy Post Thoracic Surgery

Sajad Emamipour, Rasoul Sali and Zahra Yousefi (2017). *International Journal of Big Data and Analytics in Healthcare* (pp. 16-34).

www.irma-international.org/article/a-multi-objective-ensemble-method-for-class-imbalance-learning/197439

Spatial and Temporal Distribution of Pedestrian Crashes

(2019). *Big Data Analytics in Traffic and Transportation Engineering: Emerging Research and Opportunities* (pp. 122-151).

www.irma-international.org/chapter/spatial-and-temporal-distribution-of-pedestrian-crashes/220533

Fuzzy Logic-Based Predictive Model for the Risk of Sexually Transmitted Diseases (STD) in Nigeria

Jeremiah A. Balogun, Florence Alaba Oladeji, Olajide Blessing Olajide, Adanze O. Asinobi, Olayinka Olufunmilayo Olusanya and Peter Adebayo Idowu (2020). *International Journal of Big Data and Analytics in Healthcare* (pp. 38-57).

www.irma-international.org/article/fuzzy-logic-based-predictive-model-for-the-risk-of-sexually-transmitted-diseases-std-in-nigeria/259987

Statistical Features for Extractive Automatic Text Summarization

Yogesh Kumar Meena and Dinesh Gopalani (2016). *Enterprise Big Data Engineering, Analytics, and Management* (pp. 126-144).

www.irma-international.org/chapter/statistical-features-for-extractive-automatic-text-summarization/154558