# 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. In the current

File MUX v2.01 cars.txt															
File Sho Sho Sho Sho Sho	General Statis w Axis Summar w Axis Box Plot w Query Box Plot w Histograms	tics Proces Y s ots	sses Class	ify +	MPG ↑ R2=82. 46.60 46.6	1 Veight 5140.00		Year A B 82.00	Displaceme		rigin ↑□	Clyinders	Accelerati		HP 30.00
♥ Sho ♥ Sho ♥ Use Correlat Histo	w MLP/SER Hui w Scatterplots w Correlation Ir : Query Correlati ion Threshold: ogram Bin Size: Axis Summary: MATLAB Path	dicators lons 0.5 Median & II /Application	QR Range Is/MATLAB_R	10 (‡) (‡) 2009b.a	9.0	1613.00 F=83	392	70.00 7532 lines se	68.00	a Significant		5.00 5.00 5.00 5.00 5.00 5.00 5.00 5.00	8.00 F=.42		6.00
Enabled	Variable	Continuous	Dependent	Mean	Median St	Dev IQR		Query Max	Query Min	Focus Max	Focus Min	MLR b	SLR R2 S	kewness k	urtosis
	MPG			23.446	5 22.75	7.805	12	46.6	9	46.6	9	•	•	0.455	2.475
<u>রেরেরেরে</u>	Clyinders Displacemen HP Weight Acceleration Year Origin	T SSSS		5.472 194.412 104.469 2,977.584 15.541 75.98 1.577	2 4 2 151 9 93.5 4 2,803.5 1 15.5 8 76 7 1	1.706 104.644 38.491 849.403 2.759 3.684 0.806	4 179.5 52 1,392 3.3 6 1	8 455 230 5,140 24.8 82 3	3 68 46 1,613 8 70 1	8 455 230 5,140 24.8 82 3	3 68 46 1,613 8 70 1	-0.097 0.23 • -0.744 0.055 0.361 0.139	60.469 64.823 60.595 69.263 17.921 33.703 31.946	0.506 0.699 1.083 0.518 0.29 0.02 0.912	1.604 2.216 3.673 2.186 3.423 1.832 2.154

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-correlationclassification-and-regression-analysis/78712

# **Related Content**

### Big Data Analytics in Healthcare Sector

Sheik Abdullah A., Selvakumar S., Parkavi R., Suganya R.and Abirami A. M. (2022). *Research Anthology on Big Data Analytics, Architectures, and Applications (pp. 993-1002).* www.irma-international.org/chapter/big-data-analytics-in-healthcare-sector/291022

### A Brief Survey on Big Data in Healthcare

Ebru Aydindag Bayrakand Pinar Kirci (2020). International Journal of Big Data and Analytics in Healthcare (pp. 1-18).

www.irma-international.org/article/a-brief-survey-on-big-data-in-healthcare/253842

### A Novel Approach for Tenuous Community Detection in Social Networks

Muhammad Asif, Hassan Razaand Muhammad Imran Manzoor (2022). International Journal of Data Analytics (pp. 1-12).

www.irma-international.org/article/a-novel-approach-for-tenuous-community-detection-in-social-networks/297518

#### A Survey on Models and Methods for Preference Voting and Aggregation

Ali Ebrahimnejadand Farhad Hosseinzadeh Lotfi (2017). *Data Envelopment Analysis and Effective Performance Assessment (pp. 57-82).* 

www.irma-international.org/chapter/a-survey-on-models-and-methods-for-preference-voting-and-aggregation/164823

#### Data Analysis and Visualization in Python for Polar Meteorological Data

V. Sakthivel Samy, Koyel Pramanick, Veena Thenkanidiyoorand Jeni Victor (2021). *International Journal of Data Analytics (pp. 32-60).* 

www.irma-international.org/article/data-analysis-and-visualization-in-python-for-polar-meteorological-data/272108