Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 1996 Proceedings

Americas Conference on Information Systems (AMCIS)

8-16-1996

Comparison of Analysis Techniques: 3-Dimensional Visualization and Regression.

Nancy J. Lightner
Krannert Graduate School of Management, Purdue University

Susan A. R. Garrod
Krannert Graduate School of Management, Purdue University

Alok Chaturvedi
Krannert Graduate School of Management, Purdue University

Comparison of Analysis Techniques: 3-Dimensional Visualization and Regression.

Nancy J. Lightner, Susan A. R. Garrod, and Alok Chaturvedi
Krannert Graduate School of Management, Purdue University

Introduction

The post-industrial society is driven by economies that are increasingly dependent on information and knowledge (Bell, 1979). The use of computers initially accelerated our ability to collect, present and communicate the vast amounts of data generated by individuals and businesses. As the power of computer hardware increases, our capacity to collect data overwhelms our ability to assimilate it into meaningful information. Additional tools are required to analyze data that we can readily collect and store. Visualization is a tool for both interpreting computerized image data and for generating images from complex, multi-dimensional data sets (McCormick et al, 1978). Its expected value to humans is that it addresses our ability to perceive shapes, patterns, connections and correlations between parameters within large data sets.

Computer graphics began as a method of exploring statistical data (Tukey, 1977). As tools became more advanced, it allowed realistic representation of spatial data and has been used extensively in scientific applications. Commercial packages exist for visualizing physical phenomena like weather, molecular structure and mechanical devices. In contrast to viewing spatial data, information visualization focuses on transforming data that are not inherently spatial into a visual form from which users can perform analysis and form conclusions. Visualization techniques can play a multifaceted role in data analysis. While many analysis techniques focus on results, visualization can be used throughout the process, from problem definition, display of raw input data, interactive analysis, as well as the display of the output results (Jones, 1994). Using visualization, a more complete definition of the problem and data can be presented to allow for better comprehension and more efficient analysis (Chaturvedi and Gupta, 1995).

While visualization techniques are commonly used in command and control situations, they have had limited use in solving managerial problems, especially in the academic research environment (ie Bajaj et al., 1995). The heavy reliance on statistical analysis and hypothesis techniques precludes other techniques from consideration. While some visually-oriented techniques exist, such as cluster analysis, their use is limited and the results are often dependent on idiosyncratic characteristics of the data (Milligan, 1982). In this paper, a business problem is introduced and a comparison of visualization and traditional analysis methods is done. In each case, visualizing the raw data shows patterns that may be further explored using regression. The regression analysis results in values of significance confirming the strength of results from visualization. The problems associated with using current visualization systems are then indicated to direct further research needs.

Hypotheses Selection

To examine the effectiveness of visualization as an analysis tool, IBM's Data Explorer (DX) tool was used to create visualizations about the telecommunications industry. The data was compiled from three sources, "The Network World 200" (Dix, 1995) listing of the top 200 firms in the networking industry, the Standard Industrial Classification Manual (1987) and the Compact Disclosure database of firm financial information. Interrelationships between the firms created a scenario where visualization could enhance the understanding of the relationships and therefore facilitate a better assessment of the problem (Jones, 1994; Chaturvedi and Gupta, 1995). The telecommunication industry offers a rich environment to study. It is relatively new and dynamic and is made up of firms that originated in the traditionally distinct industries of electronics manufacturing, communication services, and business computing software and services. The firms share a great deal of interdependence within the industry, but little is known about the nature of their interdependence. The research question on how the firms differ in their strategy and in their performance as they compete in one industry is translated into two hypotheses.

Hypothesis 1: The level of diversification of firms in the networking industry is not affected by the Primary SIC code of the firm.

Hypothesis 2: Firms whose primary 2-digit SIC designation is communication service are not more likely to have a low level of performance.

The data were analyzed using visualization and regression to show how they perform according to the two hypotheses.

Visualization Technique

The visualization technique applied is that of three-dimensional scatter plots. The exploration of techniques to represent three and more dimensions in data has been under investigation by statisticians (Wegman and Carr, 1993, Young, Faldowski and McFarlane, 1993, Becker, Cleveland and Weil, 1988, Tukey and Tukey, 1981). A three-dimensional scatter plot is a natural extension of histograms and scatter plots; real time rotations give a convincing and accurate perception of three-dimensional shape from the apparent parallax in the motion of the points in the scatter plot. Rotating three-dimensional scatter plots were used in all the exploratory data analysis PRIM systems developed at Stanford and Harvard (McDonald, 1988). In statistics, graphical methods are used primarily for description. It is especially interested in discovering unanticipated kinds of structure in data (Friedman, McDonald and Stuetzle, 1988). This application applies statistical techniques using a Silicon Graphics Indigo^2 platform with the commercially available software package DX from IBM. The raw data accumulated was plotted on three axes and shown from various views to uncover underlying trends and relationships. See figures 1 through 6 for a selection of resultant visuals.

Results of Visualization and Regression Analysis

The visualization and statistical results agreed with each other to support the relationship proposed by hypothesis 1. The communication sector firms have a higher level of diversification than the overall average, while the business services sector has a lower level than the industry average. The manufacturing sector diversifies along the industry average. The two methods also agree on the analysis for hypothesis 2. The communication services firms achieve lower than average growth, while the manufacturing and business service firms achieve higher than average growth.

This study demonstrates that visualization techniques can complement statistical analysis. They can provide a direction for more precise measurement of certain aspects of data relationships. Visualization also sheds more light on the relationships. For example, the visualizations show that a linear model is a poor choice for the data. This is confirmed by the low R^2 (.13 for hypothesis 1 and .22 for hypothesis 2) achieved by the regression models. The combination of visualization and traditional statistical techniques provide more information about the problem under study than either method alone.

Areas for Further Developments

Visualization methods provide an avenue of rapid and information intensive analysis of raw and resultant data. They can be used to examine the effects of outliers, interactions between data and the structural content of data. The technique of three-dimensional scatter plots, commonly used in statistical data analysis, provides a base on which further tools can be developed for use in a managerial setting. With the advancement in desktop computing, managers are now more accustomed to three-dimensional graphical presentations, such as Microsoft Excel provides. Until the area of information visualization is further developed, we must consider the limitations of systems like DX to the managerial world. Because DX was developed for scientific visualization, the needs of information visualization are only partially met. For example, only two glyph shapes are available for scalar data, sphere and diamond. Representing a variable using symbols becomes difficult. In the visuals presented here, a glyph of a smaller size represents a third

shape. This effectively eliminates the use of glyph radius to represent another variable of the data. Direct manipulation of the data is also not possible with DX. Its purpose is to present data collected from physical events. However, altering the data would allow managers to perform 'what if' analyses, a valuable decision making tool.

To produce a visualization in DX, the user must learn the language and prepare the data in a specific format. Although the interface is graphically and object oriented, it would not be intuitive to non-programmers. Once the visualization is presented, the interface provides an easy method of rotatation, the only manipulation possible. This indicates that with minimal training, managers could use them, but likely could not create the visuals they need. With the proliferation of visual means of communication, it is likely that the general audience would readily embrace such techniques once they address managerial needs.

Determining what those specific needs are and how a visualization system can address them are areas in need of further development. What graphical formats are appropriate and under what conditions are they effective continue to be unanswered questions.