Integrating Data Quality Data into Decision-Making Process: An Information Visualization Approach

Bin Zhu, G. Shankar, and Yu Cai

Information Systems Department
School of Management
Boston University
{bzhu,gshankar,ycai}@bu.edu

Abstract. Poor data quality (DQ) has been a troubling issue within many organizations. Thus it is important for a decision-maker to be aware of the data quality of the information based on which he/she makes a decision. However, the mental integration of DQ metadata in to a decision process poses a significant cognitive challenge. A decision-maker could be overloaded and may not effectively take advantage of the DQ metadata [2]. This paper thus proposes a visualization approach to reduce the amount of cognitive resource required for such mental integration.

Keywords: Information visualization, data quality data.

1 Introduction

Data quality (DQ) management is becoming a critical activity in organizations [1, 5]. It could lead to increased cost, poor execution of organizational strategies, and poor decision-making [5]. To manage quality effectively, it is crucial to provide decision-makers with DQ metadata, *data that describes the quality of the data used in the decision-process*. In the form of a [0, 1] ratio or as a percentage representing the quality, DQ metadata permits decision-makers to gauge data quality in the context of the decision-task [4].

However, the mental integration of DQ metadata in to a decision process poses a significant cognitive challenge. A decision-maker could be overloaded and may not effectively take advantage of the DQ metadata [2]. This paper thus proposes a visualization approach to reduce the amount of cognitive resource required for such mental integration. We developed a prototype system, SPIDEV (Special Purpose interface for Interactive Decision Evaluation using Visualization), that not only facilitates the easy integration of DQ metadata into decision-making process, but also enables decision-makers to incorporate contextual factors for better decision outcomes.

This study focuses on supporting individual analytical decision tasks [3]. Analytical decision-making is a computational approach applied to well-structured decision tasks and usually involves a mathematical model [3]. A decision-maker

M.J. Smith, G. Salvendy (Eds.): Human Interface, Part I, HCII 2007, LNCS 4557, pp. 366–369, 2007.

interacts with the model to answer his/her "what-if" questions and seeks an optimal solution. An information visualization system thus could be applied to visualize the decision outcomes of the different decision alternatives. Decision makers could then easily identify a decision alternative leading to a desired decision outcome by browsing visual representations of possible decision outcomes.

2 The Visualization Approach Proposed

The SPIDEV system visualizes decision outcomes that incorporate DQ metadata to help identify optimal solutions. Figure 1 displays the interface of the prototype system. The visualization displayed adopts the spoke chart. The spokes are used to visualize both the data and its associated quality. Each spoke represents a decision variable in a decision task (i.e. price, size, or quantity). The length of the spoke represents the importance of that decision variable. To make the display more comparable, the importance of all decision variables is standardized to fit in the wheel. Such standardization is achieved according to the importance and value of the decision variable and is also determined by the data quality of the data associated with each decision variable. The DQ metadata, defined as a percentage, is also represented along with the regular data in the spoke chart. It is visualized by the length of the colored part of the spoke. The color code also indicates the quality of the data for each decision variable. For instance, color green represents higher data quality than color red.

Since we implemented all the decision rules at the backend of this visualization, a decision-maker could immediately see the decision outcomes of different decision alternatives when using the sliders in the interface. And the "web" created by linking the points on the spokes shows the solution space. A decision-maker could choose to maximize the overall value of the decision outcome, measured by the area of the solution space. He/she could also maximize the importance of decision variables with high data quality (in color green) to get as much overall value as possible while simultaneously ensuring the quality of decision outcome.

3 Summary

The prototype system presented in this short paper proposes an approach that allows easy integration of DQ data into decision-making process. It also provides decision-makers the flexibility to customize their decision criteria for the decision context. Our future research includes a series of empirical studies to validate the performance of the system. Such studies include measuring its effect on decision outcomes (with and without SPIDEV), the reduction in differential performance caused by the task complexity and individual differences, and measuring the increased usage of DQ metadata. Future work also includes understanding the decision circumstances under which SPIDEV is likely to be more effective.

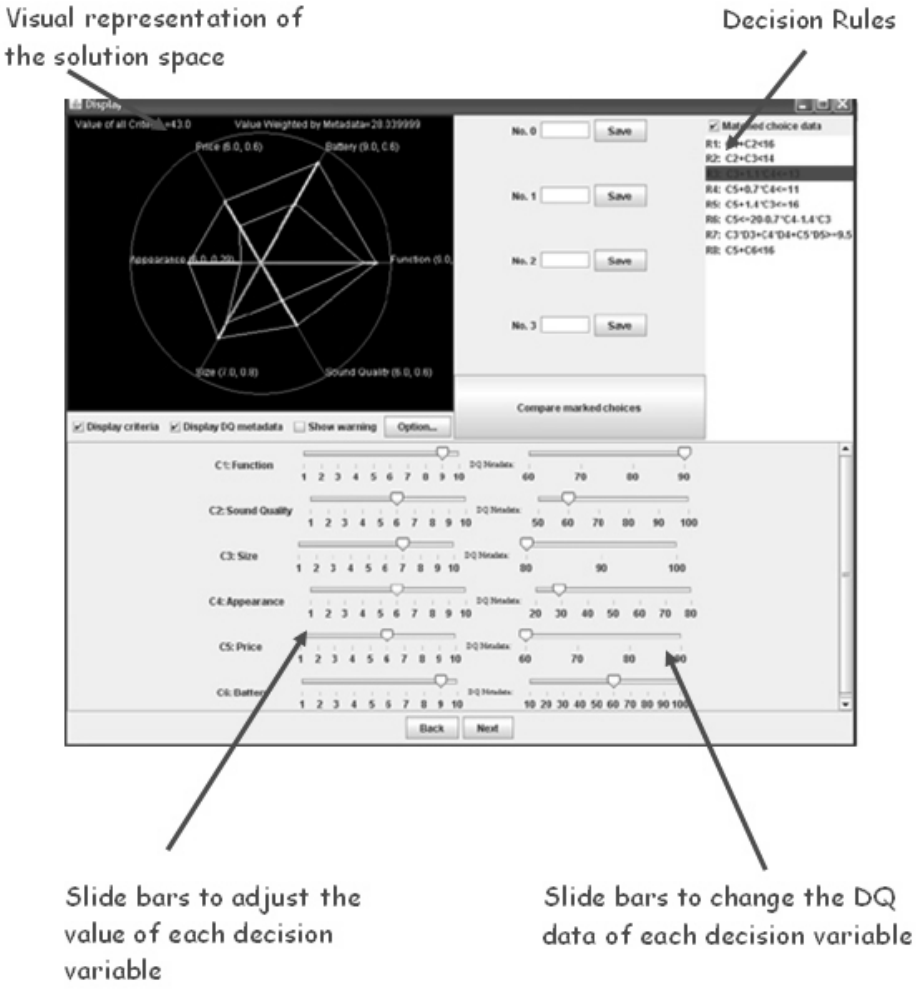

Fig. 1. The Interface of the SPIDEV Prototype System

References

1. Ballou, D.P., Wang, R.Y., Pazer, H., Tayi, G.K.: Modeling information manufacturing systems to determine information product quality. Management Science 44(4), 462–484 (1998)
2. Chengalur-Smith, I.N., Ballou, D., Pazer, H.: The impact of data quality information on decision making: an exploratory analysis. IEEE Transactions on Knowledge and Data. Engineering 11(6), 853–864 (1999)
3. Nutt, P.C.: Selecting tactics to implement strategic plans. Strategic Management Journal 10(2), 145–161

4. Shankaranarayanan, G., Ziad, M., Wang, R.Y.: Managing Date Quality in Dynamic Decision Environment: An Information Product Approach. Journal of Database Management 14(4), 14–32 (2003)
5. Wang, R.Y.: A Product Perspective on Total Data Quality Management. Communications of the ACM 41(2), 58–65 (1998)