© Nottingham University Press
Proceedings of the International Conference on Computing in Civil and Building Engineering
W Tizani (Editor)

Evaluation of toolkits for visualizing information about performance of earthwork activities on a construction job site

G Atasoy & B Akinci
Carnegie Mellon University, USA

Keywords: information visualization, HCI toolkits, performance monitoring, productivity analysis

In order to monitor the performance of construction projects, project team members are dealing with large amounts of data that are in various forms. In order to acquire insights and make timely decisions, they need to locate, synthesize, visualize and analyze such data to identify correlations, and trends; hence perform multivariate analyses. The traditional methods used for analyzing multivariate data are the usage of statistics and visualization of trends by graphing the data on x-y coordinates, to see the correlation of variables. However, these methods are disadvantageous compared to multi-dimensional information visualization, which can provide patterns of unprocessed data in a more descriptive and clear way.

Within the domain of information visualization, a large number of information visualization tools are generated to support handling and interpreting data. These toolkits vary in terms of their inputs, outputs and interaction functionalities; hence, might not be easily applicable to each specific domain and require adjustment and modifications. Within the construction management domain, a limited number of studies focus on the visualization and interaction of data collected on a job site to support project monitoring (Russell et al., 2009). Our research builds on the existing studies and extends them by investigating the support of toolkits and parallel coordinates graphic in visualizing a multivariate dataset for construction productivity analyses.

The goal of the study described in this paper is to evaluate the potential benefits and limitations of available visualization toolkits in analyzing multivariate data for construction. We used three information visualization toolkits (released in 2004, 2006 and 2007) developed within the HCI domain. The characteristics that we used in the assessment are the size and variety of inputs, form of representation schemas, variety of outputs for multivariate datasets, and interaction functionalities enabled. Inputs, outputs and representations of all theses toolkits are limited in terms of not supporting visualization of CAD, Building Information Models (BIM) and site maps, which are paramount for construction projects. Some of the toolkits do not handle inputting of large data sets (e.g., for our case 12000 values) and some do not provide visualization of all of the inputted data and focus only on one aspect of the data. All three toolkits provide dynamic functionalities to support user interaction with both the data (e.g. filter and search) and the graphics (e.g. zoom, pan and color). However, some functionality that project members might require, such as classification, segmentation, projection, calculation of new values, retrieval of statistics and finding of anomalies, are not supported by these toolkits.

We used two of the toolkits, which generate parallel coordinates, to assess the support that multivariate graphs can provide for construction productivity analyses. Parallel coordinates is a graph composed of parallel columns representing data dimensions plotted with a common scale, enabling the visualization of all dimensions in 2D. Each column represents a variable (dimension) of a dataset and each value of an instance (row) of that dataset is marked as a point on corresponding columns and connected with lines. It provides insights about the relationship of the variables of a dataset according to the pattern of the lines connecting two columns such as being cluttered, crossed and revealing a linear or nonlinear pattern. As a test-bed, we used earthwork productivity analysis dataset used in another study, within which the productivity and its variation according to weather are evaluated (Pradhan and Akinci, 2007). We performed several manipulations prior to using the data in the toolkits such as data cleaning, data type changing, and manually integrating the payload and weather data. The resulting integrated dataset is composed of 23 columns and 375 rows.

In assessing the value of multivariate data visualization support, we performed similar queries that Pradhan and Akinci (2007) had and applied on the same dataset. We investigated both productivity and payload variations via detecting the patterns of the lines and the usage of color-coded mapping and reordering of dimensions techniques. We compared our results obtained from visualization of multivariate data with the ones obtained by Pradhan and Akinci (2007) through normalization and utilization of scatter graph.

By means of visualizing our dataset, we observed some of the findings that Pradhan and Akinci (2007) presented. In contrast to these supporting results, there are some other contradicting ones. We identified a number of cases where although the averaged and normalized values of some variables were revealing a relationship and comparison, either the data was observed to be uncorrelated or another trend at the extreme or inter-quartile ranges of these variables was detected. Moreover, visualizing the whole dimensions of our dataset and the correlation between these dimensions provided us with identifying some variables affecting productivity, which could not be found with the usage of scatter graph.

The visualization of the test-bed dataset using parallel coordinates showed that information visualization provides some detailed insights for multivariate datasets by enabling the observation of the trends and correlations between the variables which may not be realized through analyzing averages. Based on the findings of this research, the advantages of parallel coordinates include; (1) revealing the general trends and showing the correlations of variables via the utilization of color maps or pair-wise comparisons; (2) highlighting the patterns of data for average, highest and lowest quartile values; and (3) not requiring normalizing the data set. A major disadvantage of using parallel coordinates is that the generated graphics are not intuitive and hence might require some user training.

Acknowledgements

The project is funded by a grant (# 0448170) from the National Science Foundation. NSF's support is gratefully acknowledged. Any opinions, findings, conclusion or recommendations presented in the paper are those of authors and do not necessarily reflect the views of the National Science Foundation.

References

PRADHAN, A. AND AKINCI, B, 2007. A planning based approach for fusing data from multiple sources for productivity monitoring. *In: Construction Research Congress,2007, Grand Bahama Island.*

RUSSELL, A. D., CHIU, C. Y. AND KORDE, T, 2009. Visual representation of construction management data. *Automation in Construction*, 18, 1045-1062.