



Application of Data Analytics Approach to Spatial Visualization Test Results

Dr. Jorge Rodriguez P.E., Western Michigan University

Faculty member in the Department of Engineering Design, Manufacturing, and Management Systems (EDMMS) at Western Michigan University's (WMU). Co-Director of the Center for Integrated Design (CID), and currently the college representative to the President's University-wide Sustainability Committee at WMU. Received his Ph.D. in Mechanical Engineering-Design from University of Wisconsin-Madison and received an MBA from Rutgers University. His B.S. degree was in Mechanical and Electrical Engineering at Monterrey Tech (ITESM-Monterrey Campus). Teaches courses in CAD/CAE, Mechanical Design, Finite Element Method and Optimization. His interest are in the area of product development, topology optimization, additive manufacturing, sustainable design, and biomechanics.

Dr. Luis Genaro Rodriguez, University of Wisconsin, Waukesha

Faculty member in the Engineering Department at the University of Wisconsin - Waukesha. Received his Ph.D. in Mechanical Engineering - Design from the University of Wisconsin - Madison. His B.S. degree is in Mechanical and Electrical Engineering from ITESM - Monterrey. Teaches courses in the areas of Engineering Graphics, Mechanics and Numerical Methods. His interest are in the area of pedagogical methodology in CAD/CAE and Mechanics.

Application of Data Analytics Methodologies to Spatial Visualization Test Results

Abstract

The field of data analytics has received substantial attention in the past years due to global trend of collecting and analyzing data. Most of the attention and applications relate to consumers behavior, but the applicability of data analytics has extended to processes and market analysis. Data analytics can be considered a generic term used to refer to a set of quantitative and qualitative approaches that are applied to provide the basis for some decision making. The particular objective that is being pursued can be increase in productivity, additional business profit, or expected performance or behavior.

Spatial visualization skills is something that has been linked to abilities to do engineering and technology work. There are several studies that have provided a relationship between the spatial visualization skills of students and their performance in engineering courses, particularly for engineering graphics and design courses. Similarly, there are reports that indicate the value in improving visualization skills when looking at the performance in learning in engineering courses, specifically for female students.

This study is based on the application of data analytics approaches to spatial visualization scores with the goal of obtaining some predictive factors. The data utilized in this study is from the Purdue Spatial Visualization Tests with rotations (PSVT:R), which was administered to groups of first year students, all of them taking a course in engineering graphics. In addition to the scores for the test, demographic data was collected from the students as well as some background curriculum information, and such parameters are used in the data analytics approaches applied. The objective of the study is not to prove a specific trend or hypothesis, but to obtain results from the predictive analytic approach followed, so that, specific behaviors are identified and specific interventions are defined to address any specific behavior/factor. The software used in this study is RapidMiner, and different subsets of data are utilized in the machine learning phase, thus reaching more robust predictive conclusions.

Introduction

It is almost ubiquitous nowadays to read or hear about trends or conclusion(s) that have been established with the application of *data analytics* techniques to consumers' information, from preference in terms of flavor or color, to behavioral trends in terms of purchasing online or tendency to walk short distances [1]. The field of data analytics has received substantial attention in the past years due in part to the competition level in a globalized economy, which has resulted in the almost imperative need for focused or personalized services, thus resulting in this global trend of collecting and analyzing all kinds of data. Even when most of the applications are in the marketplace for consumers, there are significant developments in the scientific community that address this need for studying data. Most of the attention and applications relate to consumers, but the applicability of data analytics has extended to processes and trends analysis. Data

analytics is considered a generic term used to refer to a set of quantitative and qualitative approaches that are applied to provide the basis for some decision making. The specific objectives that are being pursued when using data analytics are increase in productivity, additional business profit, or expected performance or behavior [2].

It is in the expected performance side that data analytics can be used in some academic activities, one of them being prediction of performance or behavior, which relates to the field of spatial visualization skills. There is a variety of tests that have been applied to measure spatial visualization skills of students, and there are numerous studies that have collected and analyzed information regarding demographics, spatial visualization skills, and academic performance [3 - 5]. Of interest is the study where spatial visualization skills is something that has been linked to abilities to do engineering and technology work; subsequent studies have provided a relationship between the spatial visualization skills of students and their performance in engineering courses, particularly for engineering graphics and design courses. Similarly, there are reports that indicate the value in improving visualization skills when looking at the performance in learning in engineering courses, specifically for female students.

This study is based on the application of data analytics approaches to spatial visualization scores with the goal of obtaining some predictive factors. The data utilized in this preliminary study is from the Purdue Spatial Visualization Tests with rotations (PSVT:R) [6], which was administered to groups of first year students, all of them taking a course in engineering graphics. In addition to the scores for the test, demographic data was collected from the students and it is used in the data analytics approach followed. The eventual objective of the study is not to prove a specific trend or hypothesis, but to obtain results from the predictive analytic approach followed, thus having the appropriate information in terms of specific performance and/or behaviors, and therefore specific activities are defined to address any of those performance/behavior factors.

Methodology

For this preliminary study, complete scores for each one of the questions in the PSVT:R test as well as the demographic data collected were used as pilot database for the application of data analytics. The data corresponds to students (Table I) taking a solid modeling course during their first or second year of college, and the demographic data includes: age, gender, major of study, and ethnicity. The dataset contains 40 fields per record: 30 fields corresponding to the answers provided to the standardized test questions, one field corresponding to the test score by the student, another field as indicator of a “top” score (i.e., a score of 25 or above, which can be changed to any value that is desired), four demographic fields (i.e., age, race, gender, and program of study), and four fields with optional answers by the students (e.g., previous CAD experience).

The software used in this study is RapidMiner [7], a commercially available data analytics software that is accepted and used by businesses, usually in a server mode, with the advantage that they do provide a stand-alone license for their standard version to academic users at no cost. The software has several option to analyze and visualize datasets, with various techniques available for statistics and predictive analytics [8]. The main objective of this study is to find some relationships in the datasets that are not as evident from simple observation and analysis.

Initial process applied was some statistical analysis, more than anything for validation purposes [9]. Two processes were used in this pilot study, one for visualization purposes, and the other one to find some relationships that provide insight into the dataset, i.e., decision trees.

Table I. Summary of records in dataset utilized for analysis.

Summary of Dataset		
<i>Demographics</i>	#	%
Number of Students	27	
Female Students	3	11.1
Male Students	24	88.9
Under-represented	5	18.5
Non-traditional (>25)	7	25.9
<i>Test Scores</i>		
Average	22.64	
Standard Deviation	4.62	
Minimum Score	13	
Maximum Score	30	
Median Score	22	

Results

The initial statistical analysis provides information that results in conclusions similar to the ones that have been previously reported [10]. For example, by obtaining the correlation coefficient matrix for the data set it indicates that gender is not a strong factor (Figure 1). Similarly, there is some indication on decreasing correlation between answers for higher number questions to Top25 scores.

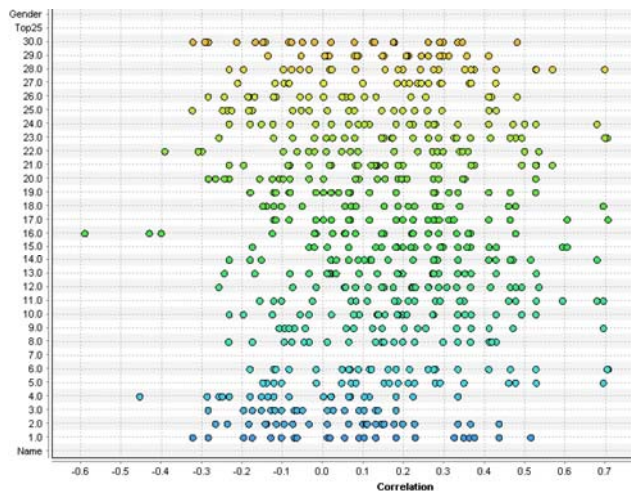


Figure 1. Correlation coefficients for validation.

After this validation processes, RapidMiner was applied to the dataset to look at some of the predictive analytics, the following result were obtained:

- Visualization. The dataset was plotted as bubble plot, mainly with the purpose of visually identifying any outliers (which the software can do as a process as well), or to see the formation of clusters. The plot is shown in Figure 2, and it is important to note that in this case the result is dependent of couple of factors like the size of the bubble that is defined, and the order in which the data is presented. From the plot, there are couple of data points that might be considered out of a cluster.

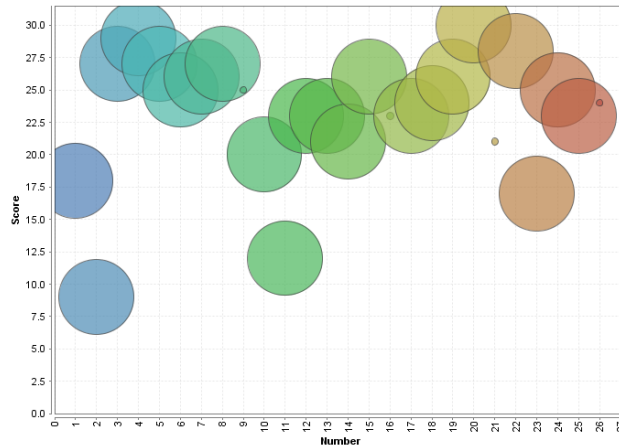


Figure 2. Bubble plot of dataset.

- Relationship. One of the questions that is of interest for some focused teaching in the area of spatial visualization is if any of the question(s) has particular predictive effect. A decision tree is requested as a process in this instance and, as shown in Figure 3, the result is that Question 11 is the one that has the higher effect. It is important to note that there are several parameters that are specified in the requested process, and results might be different depending on the values of the parameters that are specified. Another relationship that was investigated was the predictive effect based on gender. Again, Question 11 was the one that came out as having the most effect.

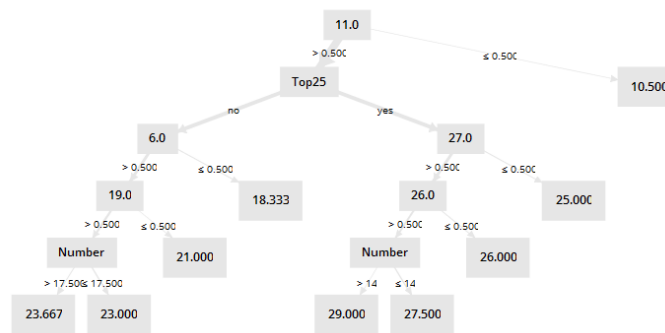


Figure 3. Decision tree for predictive question.

Conclusions

The use of data analytics has great potential, in particular for predictive analytics. The results in this pilot study show some of such potential. The software used is significantly more powerful than the applications presented in here, and there are many alternatives that need to be explored in the use of the data. It is the intention of the authors to explore some of those data manipulations (e.g., grouping) and capabilities (e.g., subset for machine learning) with the intention to have a reliable process that can be applied to different datasets.

Data analytics can include statistical analysis as well, and should not be considered as a substitute for such studies. Actually, data analytics software has the capability to perform some statistical analysis independently, but is not its main focus. The field of data analytics is related to the field of big data as well, and this pilot study utilizes a small dataset, therefore the conclusions should be taken with similar validity than statistical results with reduced sample data, an important issue to be considered. The predictive model developed in this study, without any adjustments, has been applied to a larger dataset with similar overall results.

References

1. Big data analytics: What it is and why it matters, https://www.sas.com/en_us/insights/analytics/big-data-analytics.html, 2017.
2. Data Analytics | Make Better Business Decisions, https://www.qlik.com/us/lp/ppc/qlik-sense-desktop/business-analytics?sourceID1=google&Campaign_Type=Non-Brand&KW=data%20%26%20analytics&k_clickid=24c8bad7-5950-446c-9aa6-68027642b4fd&gclid=EAIaIQobChMIralwe6V2QIVT7nACh39LAGpEAAYAiAAEgJR4_D_BwE, 2017.
3. Sorby, S. A., Developing 3-D Spatial Visualization Skills, *Engineering Design Graphics Journal*, 63(1), 21-32, 1999
4. Sorby, S. A., Assessment of a New and Improved Course for the Development of 3-D Spatial Skills. *Engineering Design Graphics Journal*, 69(3), 6-13, 2005.
5. Samsudin, K., Rafi, A., and Hanif, A. S., Training in Mental Rotation and Spatial Visualization and Its Impact on Orthographics Drawing Performance, *Journal of Educational Technology & Society*, 14(1), 179-186, 2011.
6. Guay, R., Purdue Spatial Visualization Test – Visualization of Rotations. West Lafayette, IN Purdue Research Foundation, 1977.
7. Best Data Analysis Tools, <https://www.softwareadvice.com/bi/data-analysis-comparison>, 2017.
8. RapidMiner User's Manual, <https://docs.rapidminer.com/downloads/RapidMiner-v6-user-manual.pdf>, 2017.
9. Rodriguez, J. and Rodriguez-Velazquez, L. G., Comparison of Spatial Visualization Skills in Two Approaches to Entry-Level Graphic Courses, *Proceedings of ASEE Annual Conference*, 2016.
10. Rodriguez, J and Rodriguez, L. G., Comparison of Spatial Visualization Skills in Courses with Either Graphics or Solid Modeling Content, *Proceedings of ASEE-EDGD Mid-year Conference*, 2016.