

Using Machine Learning Techniques to Capture Engineering Design Behaviors

James P. Bywater, University of Virginia, jpb6qx@virginia.edu
Jennifer L. Chiu, University of Virginia, jlchiu@virginia.edu
Mark Floryan, University of Virginia, mrf8t@virginia.edu
Jie Chao, The Concord Consortium, jchao@concord.org
Corey Schimpf, The Concord Consortium, cschimpf@concord.org
Charles Xie, The Concord Consortium, qxie@concord.org
Camilo Vieira, Purdue University, cvieira@purdue.edu
Alejandra J. Magana, Purdue University, admagana@purdue.edu
Chandan Dasgupta, Indian Institute of Technology Bombay, cdasgupta@iitb.ac.in

Abstract: Engaging students in disciplinary practices can help students but many teachers face barriers implementing practice-based instruction as capturing, assessing, and providing feedback on practices can be labor and time intensive. This working paper reports on our early attempts to leverage machine learning techniques to analyze large process datasets of students engaged in engineering design projects within computer-aided environments. By identifying students' engineering design behaviors, we hope to examine how different sequences of these behaviors can be used provide intelligent feedback and guidance.

Background

As more precollege students engage in engineering design as part of formal schooling, teachers are challenged with supporting engineering design practices in classrooms (Purzer, Moore, Baker, & Berland, 2014). Engineering design projects present unique challenges for K-12 contexts as each student may have a unique solution instead of one “right” answer or explanation. Teachers need to debug and provide feedback on each individual solution while also helping students learn the content and practices of the domain.

Although research identifies various ways to document or capture design behaviors and practices such as video, think-alouds, or reflections, these methods are labor intensive and implemented in undergraduate or professional settings (e.g., Cross, 2011). This research uses Energy3D, a computer-aided design (CAD) program specifically designed for and implemented in K-12 educational settings. Energy3D enables students to construct energy efficient solutions, analyze a variety of performance variables, and learn and reflect upon earth and physical science principles through tutorials and reflective notes. Energy3D collects fine-grained mouse-click level actions that we wish to leverage to support teachers.

This working paper builds on other approaches to capture engineering design practices (e.g., Worsley & Blikstein, 2014) and descriptive analyses of students using Energy3D. To complement these approaches, we present machine learning techniques to identify student design behaviors. We use datasets gathered across multiple sites and multiple projects to explore if there are ways to automatically capture design behaviors that might map onto particular design heuristics or practices (e.g., Crismond & Adams, 2012) and that could be used to: 1) provide descriptive information to teachers about what design practices students are using, and 2) automatically generate support and guidance for students during design projects.

Methods

The data used in this study came from a total of 446 students across three sites within the United States: a mid-western middle school, a New England high school and a mid-Atlantic coast high school. The data were collected while students used Energy3D for different design projects. For example, at times students adjusted premade designs in order to investigate specific science concepts, and at other times students built new designs to meet specifications and constraints. The dataset consisted of a total of 708550 mouse-click initiated actions, including the time each mouse-click occurred and the specific project each student was working on at the time. Within the dataset, there were a total of 175 different mouse-click level *actions*. Examples of these actions include “change the tilt of the solar panel”, “do annual energy analysis” and “animate the sun”. If no clicks were recorded for more than two minutes, an action called “inactive” was added to the log during this time.

To explore design behaviors, we used three different analytic grain-sizes: *Actions*, *action categories*, and then *behaviors*. First, we grouped the 175 fine-grain mouse-click level *actions* into 12 *action categories* using a combination of ‘ground up’ and ‘top down’ procedures. Our ‘ground up’ procedure considered the dataset as sequences of actions ordered by time. We grouped the actions that more frequently transitioned to each other

using the Markov Clustering Algorithm, adjusting the algorithm’s expansion and inflations power parameters manually so as to generate groups that were neither too big nor too small. Our ‘top down’ procedure refined these groups by moving actions that seemed, from the researchers perspective, as incorrectly grouped. For example, when “add wall” was found in a group with actions that edited different types of roof we manually moved it to a group that had other wall-related actions.

Second, we examined the sequences of *action categories* that occurred in the dataset and identified seven distinct states of student design *behaviors*, where each behavior is a set of action categories over a contiguous period of time that represents a distinct goal or task the student is undertaking at that moment. To identify different design behaviors, or design behavior states, we wrote an algorithm that would find the transition points in each student’s sequence of action categories. These transition points were positioned to maximize the difference in characteristics between adjacent behavior states. Having created a total of 8640 states (an average of 19.4 per student), we grouped similar behavior states using a *k*-means clustering algorithm with seven centroids where the distance measure used the difference between the proportions of macro-actions within behavior states. The seven groups that this created were characterized by different distributions of the proportions of action categories (see rows in Figure 1). For example, in one group, 35% of the categories were “edit solar panels”, 20% were “analysis”, and 12% were “edit roof” (see first row of Figure 1). Therefore, we identified this group as containing design behavior states we called “analyze and iterate panels”. In a similar manner, based on the different proportions of macro-actions in each group, we identified other design behaviors such as “analyze sun position” and “edit trees” (Figure 1).

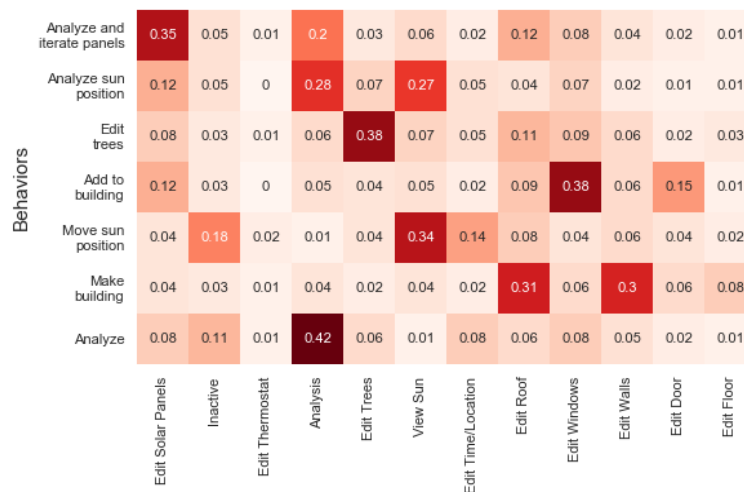


Figure 1. The proportion of each action category (horizontal axis) present in each design behavior state.

Next steps

To meet our goals of finding how different sequences of these behaviors can be used to assess design practices we plan to use a hidden Markov model with the design behaviors as hidden states. In this model, the proportions in Figure 1 give the emission probabilities from each behavior state, and the transition probabilities can be used to examine how students proceed from one design behavior state to the next.

References

- Crismond, D. P., & Adams, R. S. (2012). The informed design teaching and learning matrix. *Journal of Engineering Education*, 101(4), 738-797.
- Cross, N. (2001). Design cognition: results from protocol and other empirical studies of design activity. In C. Eastman, W. Newstatter, & M. McCracken (Eds.), *Design knowing and learning: Cognition in design education* (pp. 79–103). Oxford, UK: Elsevier.
- Osborne, J. (2014). Scientific practices and inquiry in the science classroom. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of Research on Science Education*. Abingdon: Routledge.
- Purzer, S., Moore, T., Baker, D., & Berland, L. (2014). *Supporting the implementation of the Next Generation Science Standards (NGSS) through research: Engineering*. Retrieved from <https://narst.org/ngsspapers/engineering.cfm>
- Worsley, M. & Blikstein, P. (2014). Analyzing engineering design through the lens of computation. *Journal of Learning Analytics*, 1(2), 151-186.