Macro Data for Micro Learning: Developing the FUN! Tool for Automated Assessment of Learning

Taylor Martin

O'Reilly Media
Sebastopol, CA, USA 95472
tmartin@oreilly.com

Sarah Brasiel

Utah State University
Logan, UT, USA 84322
sarah.brasiel@usu.edu

Soojeong Jeong

Utah State University
Logan, UT, USA 84322
sadsjs@gmail.com

Kevin Close

Utah State University
Logan, UT, USA 84322
kevin.j.close@gmail.com

Kevin Lawanto

Utah State University
Logan, UT, USA 84322
kevin.lawanto@aggiemail.usu.edu

Phil Janisciewicz

Agile Dynamics
pjanisiewicz@gmail.com

Abstract

Digital learning environments are becoming more common for students to engage in during and outside of school. With the immense amount of data now available from these environments, researchers need tools to process, manage, and analyze the data. Current methods used by many education researchers are inefficient; however, without data science experience tools used in other professions are not accessible. In this paper, we share about a tool we created called the Functional Understanding Navigator! (FUN! Tool). We have used this tool for different research projects which has allowed us the opportunity to (1) organize our workflow process from start to finish, (2) record log data of all of our analyses, and (3) provide a platform to share our analyses with others through GitHub. This paper extends and improves existing work in educational data mining and learning analytics.

Author Keywords

Micro learning; Digital learning environments;
Educational data mining; Assessment

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).

L@S 2016, April 25–26, 2016, Edinburgh, Scotland UK
ACM 978-1-4503-3726-7/16/04.
<http://dx.doi.org/10.1145/2876034.2893422>

Design goals for the FUN! tool

1. To create easily reproducible analyses
2. To create easily reusable analyses (for different data sets)
3. To record details about every procedure
4. To be able to adapt a wide variety of data
5. To scale well as data sets become larger
6. To be a model automated analysis tool for education researchers

Background

Our work is informed by research in Educational data mining (EDM) and learning analytics (LA). Researchers [4] have discussed the analysis of learning events at the level of log data to understand learning that is ubiquitous, heterogeneous, and complex. For example, such research has shown a relationship between engagement and learning, and the prevalence of specific affective states [1] through the use of automated detection techniques and data mining. This fine-grained detection allowed the discovery of a relationship between frustration and confusion, and learning [1,6]. Beyond detection, this work is important because engagement has been shown to be predictive of significant long-term education outcomes [6,8]. Another example of research in EDM/LA that has informed our research is [9] to understand learning from game play in a reliable and meaningful way using “stealth assessment.” There are other tools being used for real time streaming data [2,3] that show promise for adding to our understanding in education data science. Unfortunately, while the field of education data mining supplies methodologies for analyzing these data, the analytic techniques are often difficult for researchers without advanced data science experience. In addition, we have experienced the challenge of conducting complex analyses for multiple projects in an inefficient manner repeating similar processes with each new data set for different projects. We have also suffered from the consequences of either lack of documentation or incomplete documentation of research analysis processes. This inefficiency is compounded with the rotation of graduate and affiliated researchers on projects. These challenges led us to consult with our data scientist to explore domain engineering with reusable

software [7]. This approach takes a domain and analyzes the based function and processes and attempts to automate them. One researcher [5] identified three steps in this process: (1) identification of reusable entities, (2) abstraction or generalization, and (3) classification and cataloging for further reuse. Using a similar approach we believed that there were processes that we repeated for multiple projects that could be automated and abstracted so that they could be used for similar projects. We also saw how these processes could be shared with others on our research team as well as more broadly through workshops and other dissemination opportunities.

Design FUN! Tool

We created the FUN! tool because we anticipated that researchers, who have access to an increasing amount of data from digital learning environments, would benefit from a tool to help manage and analyze all of this data. We hope the tool is flexible enough to be used with a variety of data sets and research questions, but easy enough to use, that a researcher (or team member) would only need an intermediate amount of programming experience to repurpose the tool. Research questions drive every analysis process, but some elements of the analysis process (such as cleaning data and reporting results) can be repeated regardless of which research questions are driving the analysis. The FUN! tool automates the elements of the analysis that are repetitive and also makes the other elements (such as adapting to different data sources, applying different analysis techniques, and reporting results in a variety of ways) simple to change. The FUN! tool is essentially an engine that automatically moves data through a workflow process which is

How to use the FUN! tool

To access the open source FUN! tool, researchers should follow these steps:

1. Download python 3 from <https://www.python.org/downloads/>. Windows: Download pyyaml from <http://pyyaml.org/download/pyyaml/PyYAML-3.11.zip>. Run "py -3 setup.py --without-libyaml install"
2. Windows: Open the command prompt; run "py -3 -m install funtool". Mac: Open the terminal; run "pip3 install funtool"
3. Windows: Run "py -3 -m install funtool_scratch_processes". Mac: Run "pip3 install funtool_scratch_processes"
4. Download the example analysis template at <https://github.com/activelearninglab/funtool-analysis> and clone the funtool-analysis to your desktop.
5. Download the Userguide for the FUN! tool from <https://github.com/activelearninglab>. Follow the directions in the Userguide.

divided into four parts: adaptors, selectors, measures, and reporters

Data are adapted from a raw form into something usable by adaptors, selected based on various measures, analyzed, and reported to the user. (see Figure 1).

Figure 1. Diagram showing how data ideally flows through the FUN! tool.

Though the data appear to travel on a single path through the FUN! tool, the processes (see Table 1) can feed between each other in a non-linear pathway depending on the complexity and stage of analysis.

Process	Definition
Adaptors	Transform datum from source into a useful form for the engine
Grouping Selector	Creates groups of states
Analysis Selector	Always internal to the measure process, creates an analysis
State Measures	Creates a measure on a state
Group Measures	Creates a measure on a group
Reporters	Outputs information (could be intermediate or end results)

Table 1: FUN! Tool Processes

For example, researchers dealing with new data may not know what measures to apply or what analyses to run. In this case, the data may be adapted, selected, measured, then re-selected, and re-measured before being analyzed and reported. The processes themselves are run by the engine of the FUN! tool.

Preliminary Results and Conclusions

We created FUN! tool and applied our initial version of it to three real project data sets from several different digital learning environments (e.g., Scratch). We then found that the FUN! tool allowed us to (1) organize our workflow process from start to finish, (2) record log data of all of our analyses, and (3) provide a platform to share our analyses with others through GitHub. Although automated analysis tools, like the FUN! tool, are already being used in the business sector (see Spark, Oozie, Hamake), the FUN! tool is part of the first generation of tools to anticipate and address the challenges that will come with the incoming data deluge in education [10,11].

In fact, the FUN! tool is deliberately open source so that others in the community, researchers, educators, or otherwise, can continue to adapt the tool as data sources change and data analysis techniques advance. The FUN! tool is not designed to be a final product that addresses every need, but rather a first generation tool that models, not only a useful tool, but also a high quality data framework for those interested in analyzing educational data. Data sources change, but the approach for how to handle the data analysis process stays, more or less, the same. Features, such as logging each step and separating concerns, are key elements for a responsible and efficient data analysis process. When these elements are compromised, the analysis process often becomes convoluted which can

undermine the validity of the results. Researchers should spend their time defining complex and robust measures to apply to data, not worrying about how to clean, analyze, and report them. Furthermore, researchers should not consider the FUN! tool to be a cure-all to their data afflictions. The FUN! tool is best used with quality data that already has the capacity to answer the desired research questions (e.g., timestamps are necessary for temporal analyses). Researchers should develop a relationship with the developer of the data source to make sure the data is rich enough for their research and be prepared to translate event data into state data if necessary. The FUN! tool is best used in conjunction with rich data and quality data scientist support.

Acknowledgements

This material is based upon work supported by the National Science Foundation (Grant IIS-1319938). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. 2010. Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4), 223-241.
2. Dumbill, E. 2013. Making sense of Big Data. *Big Data*, 1(1): 1-2.
3. Hare, J. 2014. Bring it on, Big Data: Beyond the Hype. *Big Data*, 2(2): 73-75.
4. Martin, T., & Sherin, B. 2013. Learning analytics and computational techniques for detecting and evaluating patterns in learning: An Introduction to the special issue. *Journal of the Learning Sciences*, 22(4), 511-520.
5. McCain, R. 1985. "Reusable Software Component Construction: A Product Oriented Paradigm", In Proceedings of the 5th AIAA/ACM/NASA/IEEE Computers in Aerospace Conference, Long Beach, CA, 125--135, October 21-23.
6. Pardos, Z. A., Baker, R. S., San Pedro, M., Gowda, S. M., & Gowda, S. M. 2014. Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics*, 1(1), 107-128.
7. Prieto-Díaz, R. 1990. Domain analysis: An introduction. *ACM SIGSOFT Software Engineering Notes*, 15(2), 47-54.
8. San Pedro, M. O. Z., Baker, R. S., Bowers, A. J., & Heffernan, N. T. 2013. Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In Proceedings of the 6th international conference on educational data mining (pp. 177-184).
9. Shute, V. J. 2011. Stealth assessment in computer-based games to support learning. In S. Tobias, & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503-524). Charlotte, NC: Information Age.
10. Stenerson, M. E., Salmon, A., Berland, M., & Squire, K. 2014, October. Adage: an open API for data collection in educational games. In Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play (pp. 437-438). ACM.
11. Werner, L., McDowell, C., & Denner, J. 2013. A first step in learning analytics: pre-processing low-level Alice logging data of middle school students. *JEDM-Journal of Educational Data Mining*, 5(2), 11-37.