

Using Computational Methods to Analyze Educational Data

Camilo Vieira

Department of Education Universidad del
Norte
Barranquilla, Colombia
cvieira@uninorte.edu.co

Alejandra J. Magana

Computer and Information Technology
Purdue University
West Lafayette, United States
admagana@purdue.edu

Mireille Boutin

School of Electrical and Computer
Engineering
Purdue University
West Lafayette, United States
mboutin@purdue.edu

Abstract—This paper proposes a special session on the use of computational methods for analyzing educational data. Computation has permeated all disciplines because it provides unique opportunities to represent knowledge and understand complex phenomena. In education, disciplines such as learning analytics and educational data mining have emerged to better understand educational phenomena. This special session will discuss three different approaches to use computational methods to analyze qualitative educational data. After the discussion, the participants will be able to implement these methods using R programming, while reflecting on how they can use these methods in their own context.

Keywords— *Pattern recognition, clustering, computation, research methods, educational research, information visualization*

I. GOAL

This special session will discuss and put in practice computational methods that are increasingly used in educational research to understand complex data. The research team conducted a special session at FIE 2017 [1] to discuss three innovative pattern recognition methods, providing an opportunity for the participants to apply these methods using R programming in a hands-on activity. The special session for 2019 will include additional methods and visualization techniques that can be used to understand complex educational data. Some of the topics of this session include interactive visualizations, finding meaningful groups using clustering methods, and validation methods such as the permutation test. The participants in this session will have the opportunity to critically evaluate when and how to use these methods and to practice creating visualizations and implementing clustering and validation methods using R programming language.

II. JUSTIFICATION

Computation has permeated all disciplines and is denominated the third-pillar of science along with the theoretical and experimental approaches. The use of computation enables researchers to process and visualize large amounts of data, create models and simulations to understand complex phenomena, and to make predictions using computational devices [2]. In education, computation can capture data about the learning process while students are engaged in a learning task providing rich data to understand how learning occurs. Researchers can also take advantage of

interactive visualizations that provide valuable insights into complex datasets [3, 4]. Computational methods of data analysis allow researchers to better understand educational phenomena compared to traditional statistical analysis techniques that tend to aggregate all students as a uniform group. This session will provide firsthand experience for educational researchers interested in using these methods in their own projects.

III. DESCRIPTION AND INTERACTION

The facilitators will divide the 80-minute session into three parts as follows: (1) introduction; (2) application; and (3) reflection.

A. Introduction

During the first part of the session (35 – minutes), the facilitators will introduce different computational methods such as visualization, clustering, and the permutation test. This introduction will include open discussions about what these methods are, how they work, why they are useful, and under what conditions they can be useful. The facilitators will provide sample data sets to the participants to discuss how and when to use these methods. However, the participants will also be encouraged to discuss how they could use these methods in their own projects. For instance, the research team has used visualization and clustering techniques to validate the groups identified in a qualitative phenomenographic analysis for a think-aloud protocol (under review). Figure 1 shows the number of instances of the different types of knowledge (columns) students (rows) used in the different steps (grouping columns) of the modeling and simulation process [5]. The size of the symbol represents the number of instances, while the different shapes represent differences among groups of students. These groups were automatically identified using K-means clustering method.

Another example of visualizations to provide insights is a literature review the first author of this paper conducted on visual learning analytics [6]. The authors used visualizations both to summarize their dataset (Figure 2a), and to find gaps in existing literature (Figure 2b).

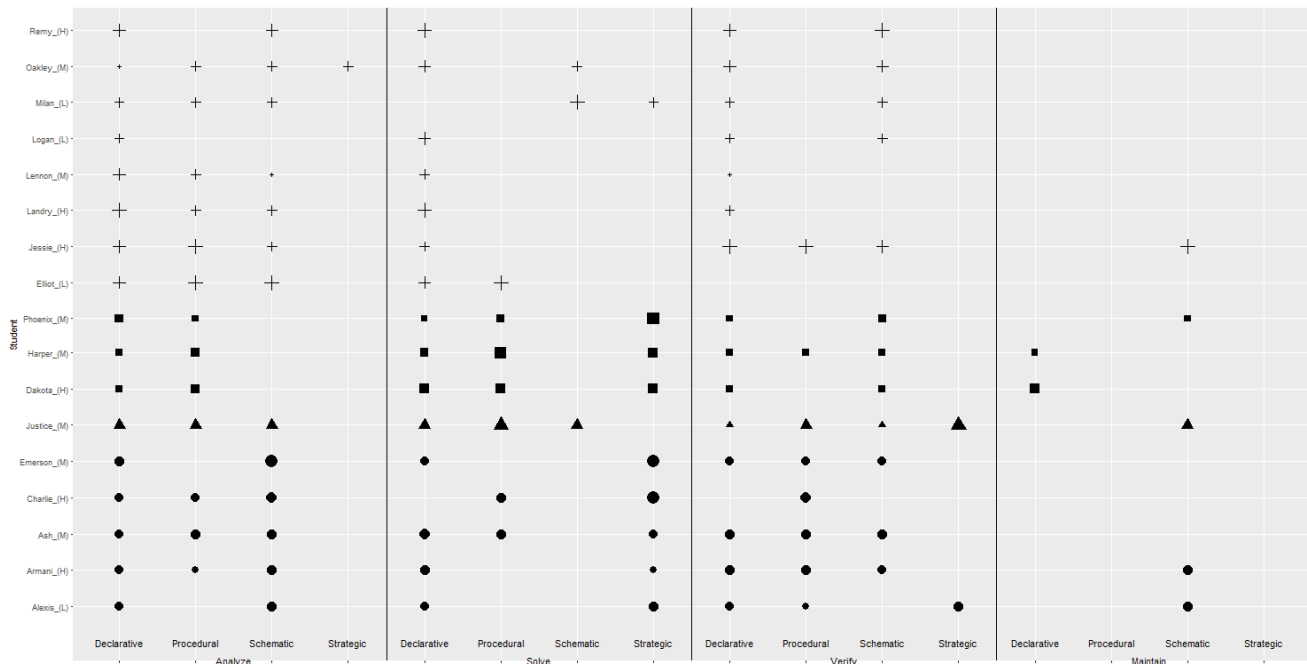


Fig. 1. Number of instances of types of knowledge students used during a modeling and simulation process

Figure 2a shows two heatmaps that represent the number of studies that used a specific data source in a specific context (left), and the number of studies that used a specific type of visualization for a specific purpose (right). Figure 2b depicts a scatter plot of scores the researchers computed based on how sophisticated the visualization was, and how much connection the research process did to an educational theory. This visualization allowed the researchers to demonstrate that there was a gap in the reviewed research that would achieve high levels in both dimensions.

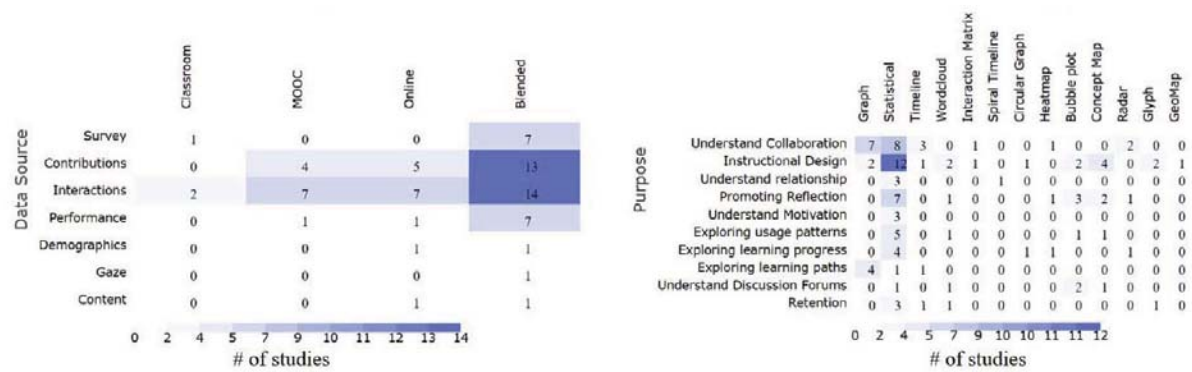
B. Application

The second part of the special session (35 – minutes) will include an introduction to R programming language and to the RStudio programming environment. The research team will bring eight laptops to the special session with the required software already installed. If participants decide to use their own laptops, the software will be available for installation from a USB drive. All the software is free and open-source, so the participants should not have any problem to install it. The instructional materials for this session comprise a tutorial in the form of worked examples plus practice activities, which participants will follow working on dyads or groups. The special session held in 2017 showed us this was an effective

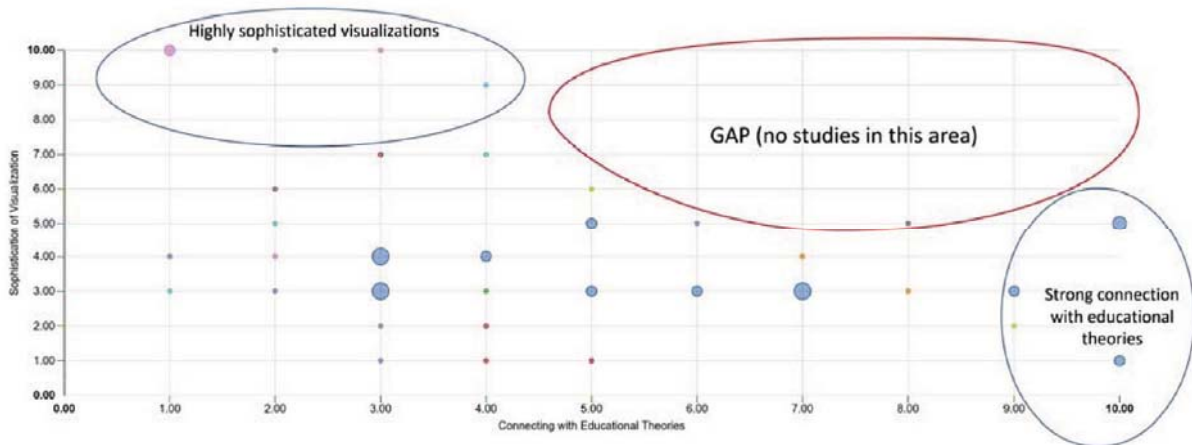
approach for engaging participants in R programming in such a short time. Some participants actually asked for permission to use the tutorial within their own research groups, which was obviously granted. Given the restrictions on time, participants will be asked to choose one of the methods discussed during the first part to implement it using R. However, the implementation of the three methods will be freely available to all participants for future exploration.

C. Reflection

To conclude the special session, the participants will turn back into group discussion led by the research team. Some of the questions to be discussed include: (1) what are the affordances of these methods for educational research compared to traditional methods? (2) how can you use these methods in your own research? And (3) how was your experience implementing these methods in R?. To wrap up presenters will emphasize role of theory and other theoretical frameworks to pursue either theory-driven or data-driven studies when applying learning analytics and educational data mining approaches (e.g., [7]).



(a)



(b)

Fig. 2. Sample visualizations for a literature review in learning analytics – (a) heatmaps to summarize the dataset; (b) scatter plot to identify possible gaps in existing literature

IV. FUTURE WORK

The future work in the use of computational methods for educational research includes the interdisciplinary collaboration of educational researchers and computer scientists to better integrate these methods. Different fields such as educational data mining, learning analytics, and machine learning for education have recently positioned, including specialized conferences and journals, and special issues in important journals in education. However, a recent review of literature [6] found that computing specialists sometimes focus on sophisticated methods and visualization strategies with a limited connection to educational literature. Likewise, educational researchers use theories from education to inform their studies, but have a limited use of specialized literature in computational methods and visualization. Hence, future work should promote this integration of both words to make an informed used of computational methods for educational phenomena.

ACKNOWLEDGMENT

This work was supported in part by the U.S. National Science foundation under the award No. EEC #1826099. The content is solely the responsibility of the authors and does not necessarily represent the official views of NSF.

REFERENCES

- [1] C. Vieira, A. J. Magana & M. Boutin, "Using pattern recognition techniques to analyze educational data," in *Proceedings of the IEEE-ERM 47th Annual Frontiers in Education (FIE) Conference*, Indianapolis, Indiana, 2017.
- [2] NSF, "Advisory Committee for Cyberinfrastructure Task Force on Grand Challenges - Final Report," 2011.
- [3] G. Siemens & D. Gasevic, "Guest editorial-learning and knowledge analytics," *Educational Technology & Society*, vol. 15, n° 3, pp. 1-3, 2012.
- [4] P. D. Ritsos & J. C. Roberts, "Towards more visual analytics in learning analytics," de *Proceedings of the 5th EuroVis Workshop on Visual Analytics*, 2014.
- [5] A. J., Magana, H. W., Fennell, C. Vieira, & M. L. Falk, "Characterizing the interplay of cognitive and metacognitive knowledge in computational modeling and simulation practices". *Journal of Engineering Education*. 1-28

- [6] A. J., Magana & M. Boutin, "A principled approach to using machine learning in qualitative education research," In *Proceedings of the IEEE-ERM 48th Annual Frontiers in Education (FIE) Conference*. San Jose, California. October 3-6, 2018.
- [7] A. J., Magana & M. Boutin, "A principled approach to using machine learning in qualitative education research," In *Proceedings of the IEEE-ERM 48th Annual Frontiers in Education (FIE) Conference*. San Jose, California. October 3-6, 2018C. Vieira, P. Parsons & V. Byrd, "Visual learning analytics of educational data: A systematic literature review and research agenda," *Computers & Education*, vol. 122, pp. 119-135, 2018.