A Principled Approach to Using Machine Learning in Qualitative Education Research

Alejandra J. Magana
Department of Computer and Information Technology
Purdue University
West Lafayette, IN USA
admagana@purdue.edu

Mireille Boutin
School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN USA
mboutin@purdue.edu

***Abstract*— This Full Paper in the Research Category presents a principled approach to integrate machine learning within qualitative education research. More specifically, we show how to build on an existing theory or conceptual framework using machine learning applied to qualitative data in order to make valid conclusions. Our model is guided by the assessment triangle. One case study is presented. The study focuses on habits of mind and their relationship to course outcomes. Patterns among students are identified using the n-TARP clustering method and validated statistically. Students are represented by a profile representing the patterns they follow and their individual course outcomes. We subsequently test for the existence of a relationship between the patterns of habits of mind and the course outcomes using a statistical approach in order to meaningfully interpret the profiles.**

Keywords—Assessment Trianglet; Learning Analytics; Clustering; Habits of Mind

I. Introduction

The National Research Council [1] has called for "scientifically-based" education research approaches needed to improve policy and practice. Such call has focused on identifying common ground principles about what entails quality research in education. In their endeavor, the NRC (2002) identified six guiding principles of education research. These principles focused on (a) significant research questions posed that can be investigated empirically; (b) building on and refining existing theories or conceptual frameworks; (c) selection of appropriate and effective methods for addressing research questions; (d) clear explanation of procedures and valid conclusions that conform scientific inferences via explicit chain of reasoning; (e) verification and validation of findings and results generalizable to broader populations and settings; and (f) disclosure and dissemination of research findings subjected to professional scrutiny by peers. In tandem, education research fields have also called for the creation or extension of formal models and methodologies for education research and evaluation [2].

One area that has significant potential to strengthen or supplement traditional methods of research in education (i.e., qualitative and quantitative approaches) are technology-based research methods such as educational data mining and learning analytics. Technology-based research methods integrate learning sciences with computational and quantitative methods and visualizations [3]. Technology-based research methods and assessment can allow the integration of multiple sources into a single analysis that can improve the generation of learner profiles, which in turn can result in the adaptation of the content or the learning environment [3]. Specifically, in the case of machine learning, advances in computing have resulted in automatic clustering methods. Among other things, these clustering methods can help infer learner patterns and profiles, which summarize the observations at a high level of granularity that is still amenable to human understanding.

This paper presents a principle-based method that combines qualitative and quantitative research methods enhanced with machine learning procedures. Specifically, we present an approach where we first start by following a theoretical approach to identify students' habits of mind patterns. The student profiles are then built by listing the patterns which apply to the student, along with the level of performance (grade) of the student. Finally, we use statistical testing and existing theory to interpret the newly identified patterns in the data and measure their influence on the levels of performance.

II. Background

The following two sections provide an overview of the use of machine learning methods in general, and clustering methods specifically, for supporting education research. Our background literature first starts by describing the foundations of clustering approaches for pattern analysis. It will then be followed by a brief review of the literature identifying ways in which machine learning in general, and clustering approaches specifically, have been used as means to support education research and assessment.

A. Clustering Approaches to Pattern Discovery

Discovering patterns is a core topic within the larger field of data analysis. One way to discover patterns is to look for data points that form groups or "clusters". The task of finding these groups is known as "clustering." In high level terms, clustering is accomplished by classifying data points into groups such that points in the same group are more similar than points in different groups (i.e., large intra-class similarity and low inter-class similarity). In machine learning, clustering is viewed as a special case of pattern recognition called "unsupervised pattern recognition" where the labels of the patterns need to be discovered from the data itself [4]. In statistical machine learning [5], the data points given are viewed as observations of

978-1-5386-1174-6/18/$31.00 ©2018 IEEE

random variables whose probability density function depend on the class to which they belong; the class specific probability density functions (marginal densities) are estimated from the data. In this context, a good pattern is one where the marginal densities for the different classes have little overlap.

The use of clustering in education research dates as far back as the 1970s [6]. Todays' large data storage capacities and fast computers support many automatic clustering methods (e.g., k-means [7, 8], kernel k-means [9], expectation-maximization algorithms [10, 11] and DBSCAN [12]. Still, using these methods in education applications is not straightforward, as these algorithms typically require the data to take the form of points in some vector spaces, as opposed to text, attributes or other types of qualitative data. Another challenge is that education data is often too complex to be accurately represented by low-dimensional feature vectors; clustering high-dimensional data requires specialized algorithm that typically perform best with a large number of training points, which makes them unsuitable for analyzing the data of studies with few participants.

It has recently been observed [13, 14] that real data can be quite easy to cluster, as the data tends to contain several good linear separations. There are so many linear separations that a random pick often yields a good quality clustering. This observation is the basis of the RP1D clustering approach [14], a fast clustering algorithm constructed using a tree structure. Each node of the tree divides the data into two groups by looking for a good linear separation at random. More specifically, a random vector is generated and the data at the node is projected onto the line spanned by that random vector. The best threshold is then selected so to separate the projected data in such a way that the variance of each 1D cluster is minimized. This process is repeated n times and the best clustering among the n clustering obtained is selected to cluster the data at that node. The procedure to cluster the node data is called "n-TARP", where the acronym TARP stands for "Thresholding After Random Projection" [15].

In experiments [14], RP1D has been shown to outperform state-of-the-art high-dimensional clustering methods. For small datasets, a single n-TARP, with n properly large (e.g., n=200) can be used to obtain a good binary clustering of the data. As the clustering is performed based on the separation of a single feature (i.e., the projection of the data onto the line spanned by a random vector), the clustering criteria is one-dimensional, which makes it possible to cluster even small data sets (about 15 subjects). If the data set contains more points (subjects) the remaining points can be used to test the statistical significance of the clustering found [15].

B. *Automatic Clustering Approaches in Education*

The field of educational data mining and learning analytics has emerged as a response to advances in machine learning, natural language processing, and data mining in general. It has also emerged based on the need for integrating effective analytical and assessment approaches into education and consequently promote data-driven decision-making processes in teaching and learning. For instance, a review of the literature that described advances in learning analytics and educational data mining identified that the majority of the uses of these approaches centered on modeling student behavior, prediction of performance, retention and dropout, promotion of reflection and awareness behaviors, and support assessment and feedback [3]. Specifically, clustering methods in education have been primarily used in the identification of learner profiles based on their interactions with learning materials [16, 17], as well as for modeling students' behaviors [18, 19], and strategies they use to approach the learning process [20]. By doing so, researchers aim to provide adaptive assessment and feedback services [e.g., 21] or provide learners with a set of recommendations [e.g., 22, 23]. Clustering approaches have also been used to predict student performance based on multiple indicators including demographic characteristics, grades, student-generated artifacts, and student level of engagement, among others [24, 25].

The previously explored body of literature suggest that the main use of clustering has been to try to capture student performance, actions or evidence that have a determined value assigned to it (e.g., a grade, counts of access to a specific material, time on task, etc.) by using large quantitative datasets. The use of clustering approaches on qualitative data, on the other hand, has been rarely explored. However, the analysis of open-ended (qualitative) behaviors can provide insights into student learning not previously afforded with traditional analytical approaches [26]. For instance, Worsley and Blikstein (2014) presented a human-computer analysis approach that started with hand-coded video data that was then further analyzed following machine learning procedures.

Such studies focused on quantitative large datasets have primarily followed a data-driven approach to learning. However, the fields of learning analytics and educational data mining have made a call for the consideration of learning theory for the analysis and interpretation of educational data [27]. Three exemplary qualitative studies that have grounded their analytical procedures as well as data interpretation in theory were identified. The first study was conducted by Worsley and Blikstein [26], who used theories of expertise to ground their investigation. Specifically, students in their study were first classified based on their level of experience in the domain of engineering design considering the amount of formal instruction and teacher ratings. Worsley and Blikstein [26], also followed analytical approaches proposed by Atman et al. [28] to generate a coding scheme. Their initial coding scheme used grounded theory approaches and then was supplemented with Design Stages for higher-level object manipulation (i.e., realize, plan, evaluate, modify, revert). The data coded was used as the input to a two-step machine learning algorithm. In the first step, clustering is used to quantize participant actions. In the second step, the sequences of (quantized) actions for all participants are clustered into groups and a chi--squared test is performed to test the validity of the grouping. In the discussion section, Worsley and Blikstein [26], went back to the literature in design and expertise, and used it to interpret the clusters and the dimensions they identified for each of them.

A second study from Colthrope and colleagues [29], used self-regulated learning as a lens to design intra-semester assessment tasks, which then were associated with academic achievement. Specifically, Colthrope et al., [29], used Zimmerman's [30] self-regulatory phases to design four meta-learning assessment tasks during the semester. The tasks consisted of six questions aimed at helping students articulate their own learning and engage in learning strategies. Students' responses to the tasks along with their academic performance were the input for an exploratory cluster analysis. More specifically, different clusterings of the student tasks responses were obtained using two methods, and the difference in grade between the students in different clusters was used to judge the quality of these different task clusterings. This analysis method was used to determine if there were differences in academic performance between groups measured with course grade, student regulation of their learning as reported via the six tasks, and extent and timing of lecture recording use along with timing of assessment submission. Results of the study suggested that prompting students to engage in meta-learning is beneficial in increasing student self-regulatory skills and consequently deepening their understanding of their learning [31].

The third study identified was from Andrade, Delandshere and Danish [32], who grounded their study in literature about epistemological frames [33]. Specifically, Andrade et al., (2016), used epistemological frame analysis to analyze behaviors during a post-interview after an inquiry activity. A machine learning approach based on Hidden Markov Models was then used to identify what patterns in students' frames became visible without relying on their speech. More specifically, the data was fitted using 2- to 7-state Hidden Markov Models, and the 3-state model was found to be the best fit. The three corresponding profiles identified were further interpreted and discussed grounded in literature of mechanistic reasoning following Russ et al.'s [34] guidelines.

The three latter studies demonstrate how clustering approaches can be used following a theory-driven approach primarily starting from qualitative data. These three studies also demonstrate how findings from the cluster analysis were then further interpreted under the lens of theories of learning. Our goal in this article is to generalize this approach by grounding it on principles of assessment and demonstrate with a case study how the model can be applied. We further strengthen our model by emphasizing the potentially large number of possible patterns present in the data and the necessity to validate the chosen patterns and findings using statistical tests performed on different data than the data used to find the patterns.

III. A Principled Approach for Detecting Learner Profiles Model (PALP Model)

To guide educational researchers in this endeavor we present a framework herein called a Principled Approach for Detecting Learner Profiles (PALP) model (see Figure 1 for an overview), which was guided by the assessment triangle [35]. The assessment triangle has been identified as particularly useful in design of valid assessments [36]. For instance, Streveler and colleagues used the assessment triangle to design and validate a concept inventory [37]. This paper extends the use of the assessment triangle to the process generation, analysis and interpretation of educational data. Figure 1 presents our proposed approach that considers the three components of the assessment triangle including the cognition (Fig 1a), observation (Fig 1b), and interpretation (Fig 1c) corners. Detailed procedures are described in a case study in the next sections.

Our model details the procedures that logically connect the theoretical groundings on how students develop scientific behaviors and domain knowledge (the cognition vertex), the procedures followed to tease out the evidence of student learning and performance (the observation vertex), and the ways in which the assessment results were analyzed, interpreted and validated (the interpretation vertex). A critical characteristic of the

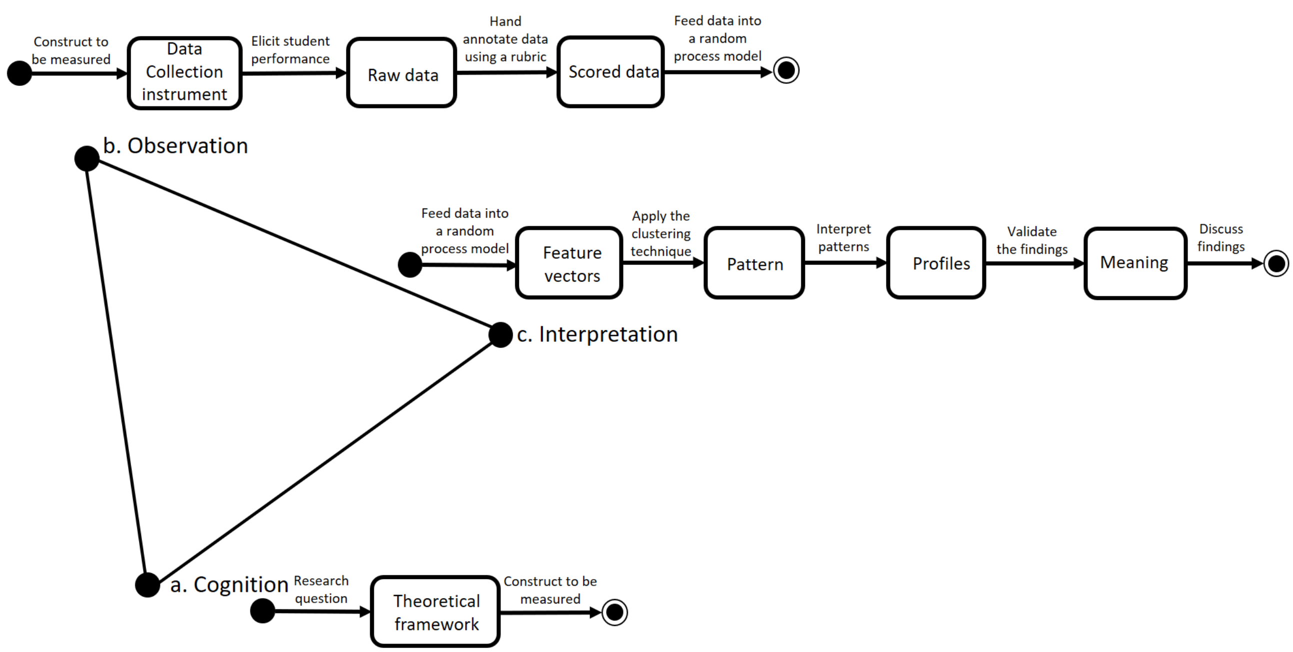

Fig 1. Procedures of the PALP model following guidelines of the assessment triangle.

assessment triangle framework is the alignment of its three corners. That is, theories about how students learn must be consistent with the kinds of evidence of the learning identified, along with the methods used to analyze and interpret the results of the evidence identified through assessment or data collection instruments.

We present our approach with a specific case study that demonstrates how the PALP model can be applied. Our study aims to identify learner profiles based on the different modes of thinking 28 students exhibited when explaining their understanding of digital signal processing concepts through the process of creating and peer-reviewing online content [38]. The following sections describe how the assessment triangle was used to approach our analyses utilizing the PALP model. As we describe our approach we will also provide details of how our proposed model combines traditional research and assessment methods such as qualitative and quantitative procedures to data analyses, and combines them with clustering techniques that supported the identification of patterns in the data and the analysis of their impact on course outcomes.

A. The Cognition Vertex

The cognition vertex (Fig 1a), deals with the theoretical foundations of the phenomenon to be measured. That is, it refers to a theory or conceptual framework that would describe or hypothesize about how students develop knowledge or behaviors conductive to learning. For our first case study aiming to identify profiles of students' ways of thinking when explaining phenomena in the domain of signal processing concepts, we grounded our analysis on the identification of students habits of mind [39]. Habits of mind refer to modes of thinking required for STEM students to become effective problem solvers [39]. These include individual's responses to problems where the answers are not immediately known, Specifically, we centered our investigation on scientific habits of mind including mathematical, logical and attitudinal modes of thinking required for science, mathematics, technology, and engineering students to become problem solvers capable of transferring such skills to new contexts [40]. Such habits of mind include a combination of certain knowledge, skills and attitudes with science literacy skills including quantitative, communication, manual and critical-response skills [39].

Once constructs are grounded in the proper learning theory or theoretical framework, the next step consists on identifying the type of overt activities or artifacts that will demonstrate evidence of the learning, performance, or behavior. For the case study used to exemplify our approach, we used students' explanations of a specific course topic produced through an online learning activity called ``slectures" [41].

B. The Observation Vertex

The observation corner (Fig 1b), refers to the procedures followed to specify the evidence of student learning and performance. That is, the observation vertex deals with the operationalization and characterization of students' performances, behaviors or evidence of learning. The chosen mechanism of evidence should be aligned with the cognition vertex. Figure 1b depicts the proposed steps that describe the operationalization of the constructs under study along with the collection and interpretation of the data. Note that the data can take any form, including text, video, or interview transcripts.

Once the data has been collected using the proper mechanism to capture the evidence of learning or performance to be investigated, the next step is to characterize that evidence in terms of the construct to be investigated. In our approach, this step is usually achieved via a rubric. The design of the rubric should be guided by the conceptual or theoretical framework identified as part of the cognition corner. In general, the data can be scored by hand or with the assistance of Natural Language Processing (NLP) techniques. However, cases where only a small number of subjects are considered are better scored by hand, as NLP techniques require extensive amounts of training data in order to get reliable results. The scored data is not in a form that can be handled by most existing clustering methods, which take as input a vector of real-valued features called a "feature vector." Thus, the data is transformed into a feature vector through modeling. More specifically, the sequence of annotations in the data is viewed as the observation of a parametric random process; the parameters of the random process are specific to each test subject, and are inferred from the scored data of that test subject. The estimated values of the parameters are then used as the entries of feature vector representing the test subject. In our case, the parameters are taken to be the probabilities of each rubric item/value being tagged to a part of the text produced by the student. That probability is estimated by the ratio of the number of times the specific rubric item/value was actually tagged, divided by the total number of tags in the student's text.

The collection of all the feature vectors (all test subjects) is then divided randomly into two sets: the first set, called the *observation data*, is used to observe the patterns in the data. The second set, called the *validation data*, will be used later in the interpretation vertex. The feature vectors of the first set of test subjects is used as the input of a random clustering algorithm in order to find groupings. We use the n-TARP clustering method [15] several times to obtain several different binary clusterings of the data. Each binary clustering divides the data into two groups according to a randomly selected criterion. The criterion corresponds to a one-dimensional feature whose value is thresholded in order to divide the dataset into two groups. An example of grouping is illustrated in [42]. In that example, two groups of students are first found using n-TARP clustering on data representing habits of mind of students. The statistics of the data for each group are used to compare the two groups. Inspection reveals that the groups statistics differ in ways that can be interpreted ("habits developed" students, and "habits developing" students).

Each binary grouping corresponds to an observed pattern in the *observation data*, which is used to assign a category to each learner: Category 1, if the student belongs to the first group, and Category 2, if the student belongs to the second group. Outcome variables (e.g., grades) are also recorded. Finally, the student profile is written as the sequence of the group index for each grouping, concatenated together with the outcome variables. For example, [1,2,1,B] would represent a student that has achieved a grade of B in the class and belongs to group 1 according to the first and third binary clusterings, and group 2 according to the second binary clustering.

C. The Interpretation Vertex

The interpretation vertex (Fig 1c), considers the methods used to interpret the observations. That is the interpretation corner deals with the way in which results derived from the observation corner are analyzed, validated and interpreted. More specifically, the patterns (clusters) are statistically validated, the dependency of the outcome variable on the patterns is measured, and the extent to which this measure is meaningful is statistically validated. Subsequently, the findings are interpreted.

The validity of each clustering is determined using a statistical test performed on the *validation data*. Because the clustering is only present in the one-dimensional projected data, the statistical test is also performed in one-dimension (i.e., on the projection of the *validation data*). We use a permutation test in order to accommodate the small data size. The clusterings that are found to be non-statistically significant can be dropped from the student profile representation at this point. If no statistically significant clustering is found (e.g., if the data is too small), then the interpretation phase is terminated and another round along the assessment triangle can be undertaken with this finding in mind. If at least one significant cluster is found, then the extent to which the outcome depends on the (valid) patterns is analyzed by looking at the difference between the outcome variable for test subjects belonging to different groups. In our experiments, the outcome variable is the course grade, and we compare the grade distribution of the course grades between two different groups (clusters) by looking at the absolute value of the difference between the mean grade of the two groups (Fig 2).

Since there can be many different (valid) groupings, we measure the extent of the effect of the outcome (grade) on the patterns by plotting the empirical cumulative distribution function (CDF) of the grade difference for all the random groupings observed. Following the approach of [15,38], the statistical significance of this CDF is tested by comparing the resulting curve to a null-hypothesis CDF curve. The null-hypothesis is obtained by dividing the grades at random into two several times, computing the mean difference each time so to generate another (random) CDF curve. Repeating this process, several curves are generated and the expected CDF curve is estimated, along with that within a confidence threshold from the expected CDF. Examples of these curves, along with more details about their construction, can be found in [15,38]. Regions of the CDF curves that lie below the hypothesis testing curve correspond to significant grade differences for the corresponding fraction of patterns considered; these statistically different grade differences are highlighted in pink in the Figure 2.

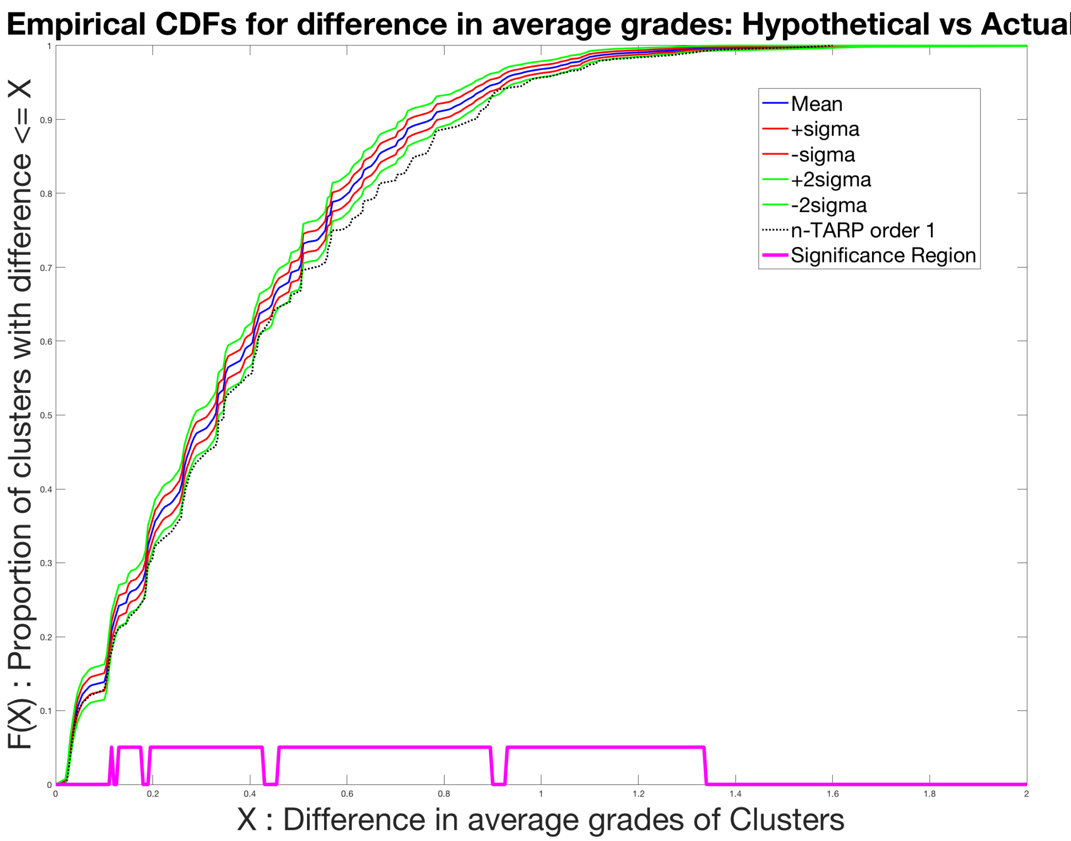

Fig 2. Statistical test of dependency between course outcomes and habits of mind using CDF curves of grade difference between groups.

The presence of these pink regions in Figure 2 allows us to conclude that some patterns of habits of mind are associated with significant grade differences. For example, some patterns of habits of mind (about 40%) were associated with an average grade difference of at least 0.5 (since the curve value at a grade difference of 0.5 is about 0.6); the difference is statistically significant at the chosen significance level (two standard deviations from the mean). More details about this method can be found in [15].

Each student is associated with a profile consisting of the list of clusters to which they belong, among all the valid binary clusterings found. Such a profile can be represented by a list of 0's and 1's, where a zero corresponds to the first cluster (below threshold), and one corresponds to the second cluster (above threshold). These profiles summarize the qualitative data for each student, as interpreted through the lens of the rubric, in a concise, numerical fashion. Our results (Fig.2) indicate that these student profiles (and thus the items of our theory) are meaningfully related to the course outcome.

IV. Discussion, Conclusion and Future Work

A weaknesses identified by Papamitsiou and Economides [3] in their review of the literature was the lack of evidence in the effectiveness of the use of educational data mining and learning analytics approaches with qualitative data. They concluded that qualitative methods have not yet provided significant results. Our study presents a model that is highly suitable for qualitative data. Our approach also presents a theory-driven approach to the use of machine learning; the previously explored body of research on the use of clustering techniques in education suggest that the most common method is a data-driven approach. Furthermore, theory-driven approaches that utilize clustering approaches have been also scarcely reported in the literature. There is a body of work that uses clustering for primarily analyzing quantitative data. Most such work is based on finding and analyzing a single clustering structure. Our work emphasizes the likely existence of many different clustering structures corresponding to different patterns in the data, and the importance of testing their statistical significance before putting any meaning into these patterns. We also build on the many clusterings found to test for dependencies of outcome variables (grade) on the many items of a theory (e.g., habits of mind). Moreover, our model also may help researchers strengthen their findings through validation via statistical approaches.

Acknowledgment

Special thanks to Tarun Yellamraju for help with data analysis and figure creation. This material is based upon work supported by the National Science Foundation under Grant No. EEC-1544244.

References

[1] [NRC], "Scientific research in education," in *Committee on Scientific Principles for Education Research*, R. J. Shavelson and L. Towne, Eds., ed Washington, D.C.: National Academies Press, 2002.

[2] [NSF], "Research on Methodologies for STEM Education," National Science Foundation, Washington, D.C.2017.

[3] Z. Papamitsiou and A. A. Economides, "Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence," *Journal of Educational Technology & Society,* vol. 17, p. 49, 2014.

[4] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*: Wiley, New York, 1973.

[5] K. Fukunaga, *Introduction to statistical pattern recognition*: Academic press, 2013.

[6] R. J. Shavelson, "Applications of cluster analysis in educational research: looking for a needle in a haystack," *British Educational Research Journal,* vol. 5, pp. 45-53, 1979.

[7] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics),* vol. 28, pp. 100-108, 1979.

[8] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern recognition letters,* vol. 31, pp. 651-666, 2010.

[9] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation,* vol. 10, pp. 1299-1319, 1998.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society. Series B (methodological),* pp. 1-38, 1977.

[11] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: an efficient data clustering method for very large databases," in *ACM Sigmod Record*, 1996, pp. 103-114.

[12] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Kdd*, 1996, pp. 226-231.

[13] S. Han and M. Boutin, "The hidden structure of image datasets," in *Image Processing (ICIP), 2015 IEEE International Conference on*, 2015, pp. 1095-1099.

[14] T. Yellamraju and M. Boutin, "Clusterability and Clustering of Images and Other "Real" High-Dimensional Data," *IEEE Transactions on Image Processing,* 2018.

[15] T. Yellamraju and M. Boutin, "Pattern Dependence Detection using n-TARP Clustering," *arXiv preprint arXiv:1806.05297,* 2018.

[16] R. F. Kizilcec, C. Piech, and E. Schneider, "Deconstructing disengagement: analyzing learner subpopulations in massive open online courses," in *Proceedings of the third international conference on learning analytics and knowledge*, 2013, pp. 170-179.

[17] T. Hecking, S. Ziebarth, and H. U. Hoppe, "Analysis of dynamic resource access patterns in online courses," *Journal of Learning Analytics,* vol. 1, pp. 34-60, 2014.

[18] C. Romero, S. Ventura, A. Zafra, and P. De Bra, "Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems," *Computers & Education,* vol. 53, pp. 828-840, 2009.

[19] M. h. Abdous, H. Wu, and C.-J. Yen, "Using data mining for predicting relationships between online question theme and final grade," *Journal of Educational Technology & Society,* vol. 15, p. 77, 2012.

[20] D. Gasevic, J. Jovanovic, A. Pardo, and S. Dawson, "Detecting learning strategies with analytics: Links with self-reported measures and academic performance," *Journal of Learning Analytics,* vol. 4, pp. 113-128, 2017.

[21] C.-M. Chen and M.-C. Chen, "Mobile formative assessment tool based on data mining techniques for supporting web-based learning," *Computers & Education,* vol. 52, pp. 256-273, 2009.

[22] A. Klašnja-Milićević, B. Vesin, M. Ivanović, and Z. Budimac, "E-Learning personalization based on hybrid recommendation strategy and learning style identification," *Computers & Education,* vol. 56, pp. 885-899, 2011.

[23] M. K. Khribi, M. Jemni, and O. Nasraoui, "Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval," in *Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on*, 2008, pp. 241-245.

[24] I. Lykourentzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis, and V. Loumos, "Dropout prediction in e-learning courses through the combination of machine learning techniques," *Computers & Education,* vol. 53, pp. 950-965, 2009.

[25] J. Luan, "Data mining as driven by knowledge management in higher education-persistence clustering and prediction. Keynote for SPSS Public Conference," University of California San Francisco, San Francisco, CA2001.

[26] M. Worsley and P. Blikstein, "Analyzing Engineering Design through the Lens of Computation," *Journal of Learning Analytics,* vol. 1, pp. 151-186, 2014.

[27] A. F. Wise and D. W. Shaffer, "Why theory matters more than ever in the age of big data," *Journal of Learning Analytics,* vol. 2, pp. 5-13, 2015.

[28] C. J. Atman, M. E. Cardella, J. Turns, and R. S. Adams, "Comparing freshman and senior engineering design processes: an in-depth follow-up study," *Design Studies,* vol. 26, pp. 325-357, 2005.

[29] K. Colthorpe, K. Zimbardi, L. Ainscough, and S. Anderson, "Know thy student! Combining learning analytics and critical reflections to develop a targeted intervention for promoting self-regulated learning," *Journal of Learning Analytics,* vol. 2, pp. 134-155, 2015.

[30] B. J. Zimmerman, *Attaining self-regulation: A social cognitive perspective.* Burlington, MA: Elsevier Academic Press, 2000.

[31] J. Hattie, *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*: Routledge, 2008.

[32] A. Andrade, G. Delandshere, and J. A. Danish, "Using Multimodal Learning Analytics to Model Student Behavior: A Systematic Analysis of Epistemological Framing," *Journal of Learning Analytics,* vol. 3, pp. 282-306, 2016.

[33] R. S. Russ, V. R. Lee, and B. L. Sherin, "Framing in cognitive clinical interviews about intuitive science knowledge: Dynamic student understandings of the discourse interaction," *Science Education,* vol. 96, pp. 573-599, 2012.

[34] R. S. Russ, R. E. Scherr, D. Hammer, and J. Mikeska, "Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science," *Science Education,* vol. 92, pp. 499-525, 2008.

[35] J. Pellegrino, N. Chudowsky, and R. Glaser, *Knowing what students know*: National Academy Press Washington, DC, 2003.

[36] S. F. Marion and J. W. Pellegrino, "A validity framework for evaluating the technical quality of alternate assessments," *Educational Measurement: Issues and Practice,* vol. 25, pp. 47-57, 2006.

[37] R. A. Streveler, R. L. Miller, A. I. Santiago-Román, M. A. Nelson, M. R. Geist, and B. M. Olds, "Rigorous methodology for concept inventory development: Using the'assessment triangle' to develop and test the thermal and transport science concept inventory (TTCI)," *International Journal of Engineering Education,* vol. 27, p. 968, 2011.

[38] T. Yellamraju, A. J. Magana, and M. Boutin, "Investigating students' habits of mind in a course on digital signal processing," *Manuscript submitted for publication,* 2017.

[39] [AAAS], "Project 2061. Benchmarks for Science Literacy A Tool for Curriculum Reform," A. A. f. t. A. o. Science, Ed., ed. Washington, DC: ERIC Clearinghouse, 1995.

[40] A. L. Costa and B. Kallick, "Habits of mind," ed, 2005.

[41] M. Boutin and J. Lax, "Engaging graduate students through online lecture creation," in *Frontiers in Education Conference (FIE), 2015 IEEE*, 2015, pp. 1-4.

[42] A. J. Magana, C. Vieira, and M. Boutin, "Characterizing Engineering Learners' Preferences for Active and Passive Learning Methods," *IEEE Transactions on Education,* vol. 61, pp. 46-54, 2017.