Identifying and Displaying EdTech Implementation Context Profiles

Brian Wright[1], Christine Tomasik[2], Alexandra Rivera[1], Caroline Bush[1], Brennan Danek[1]

[1] School of Data Science, University of Virginia, Charlottesville, Virginia, USA
[2] EdTech Evidence Exchange, Washington, DC, USA
brianwright@virginia.edu

Abstract. Education technology (edtech) is increasingly prevalent in classrooms, yet 85% of the technologies currently implemented are a bad fit for the school, or are poorly implemented. This is especially problematic for poorly funded schools typically seen in minority communities. To address this limitation, the EdTech Evidence Exchange has collected survey responses to characterize the contexts in which technologies are being implemented. Variable selection via penalized regression and unsupervised methods extracts a subset of the factors that are the most informative in characterizing schools' contexts. This subset is then used to fit a Gaussian Mixture Model to create soft clusters of schools with similar contexts. A new feature, "Schools Like Mine," combines soft classification and Euclidean distance to identify and rank schools by similarity. This research will hopefully reduce the likelihood of failed software investments.

Keywords: Education Technology, Machine Learning, Variable Selection, Gaussian Mixture Modeling, Survey Responses

1 Introduction

Identifying a subset of variables that captures important sources of variation in a data set enhances the interpretability of subsequent analyses. The results of cluster analyses informed by a variable subset are easier to understand and communicate, for example, and allow the knowledge gained to yield a broader impact. The variable subset identified as part of this work is used to fit a Gaussian Mixture Model (GMM) and underlies a matching algorithm that quantifies the similarity between entities and returns a list of the most similar entities to a point of interest.

2 Background

The EdTech Evidence Exchange (i.e., the Exchange) is a non-profit that leverages educator insights to provide education technology (edtech) decision-makers with context-sensitive evidence to inform decisions about edtech selection and implementation.

After activating more than 10 working groups and 300 stakeholders, the initiative hosted the first-ever EdTech Efficacy Academic Symposium to understand the problem more fully. The output from the 2017 Symposium was overwhelming agreement on three fronts: context matters in edtech implementation; the current chaotic state of edtech is no single actor's fault; and somebody needs to lead a broad research effort to understand and share what works where, and why. As of August 2020, Jefferson Education Exchange changed its name to the EdTech Evidence Exchange.

A first step in filling this gap was developing the language to describe the context in which these technologies are implemented. The EdTech Evidence Exchange launched the EdTech Genome Project [1] and led, coordinated, and convened a diverse technical working network of more than 140 researchers, practitioners, experts, system leaders, and industry representatives and sought feedback from thousands of educators at multiple stages in the three-year research initiative. The outcome: a common language for describing, defining, and measuring education contexts so that we can understand what makes any two districts or schools "similar" in the ways that matter most to the implementation of education technology. Ultimately, this endeavor identified and defined ten context variables, as well as ways to measure them. Each variable was then broken down into a small set of subscales via exploratory factor analyses and subsequent confirmatory factor analyses [4]; this process yielded 47 subscales total, with three to six subscales per context variable.

Currently, the Exchange Platform offers a "Districts Like Mine" feature that uses the Context Inventory data to quantify the degree of similarity between school districts.

This method of identifying similar school districts works well enough, but has several limitations. First, while context factors within each of the ten variables have been analyzed to prevent redundancy, they have not been examined in aggregate. The Platform also hopes to provide more detailed information to accompany the similarity scores, but this task is too daunting when all 47 factors are still at play. Finally, there is substantial variation among schools within a single district, so matching at the district level is likely inappropriate. Thus, the goal of this work is to improve the current EdTech Evidence Exchange Platform's matching algorithm to produce more meaningful and interpretable results. The methods developed below identify the most informative subset of the 47 factors that describe a school's context and use this knowledge to produce an improved "Schools Like Mine" algorithm that yields school-level matches for Platform users.

3 Data Description

The data on the Exchange Platform consists of basic identifying information (district ID, school ID, etc.) for each user and their question responses, averaged at the factor level such that values range continuously between one and five across 47 factors.

At the time this research was completed, the data set included responses from 1574 educators (teachers and administrators) from more than ten states around the country. However, 113 responses were removed from the data set on account of a large number

(> 10, or 21.3%) of missing values. The final data set, therefore, contained factor-level scores from 1461 educators with a limited number of missing values per entry. (For a full list of the factors and survey validation refer to E. A. Kohler, L. M. Elreda, and K. Tindle, "EdTech Context Inventory: Factor analyses for ten instruments to measure edtech implementation context features," Computers & Education, vol. 195, p. 104709, 2023)

4 Methodology

4.1 Subset Selection

The complete data set on the Exchange Platform includes 47 factors that comprise ten context variables. High correlations between various variables indicate that some factors may be redundant when analyzed across all context variables [4]. Variable selection methods were employed in order to identify a subset of the available factors to be used in subsequent analyses. Reducing the number of factors also improves the overall interpretability of the Exchange Platform's analyses.

Supervised and unsupervised methods of variable selection were employed for the selection of a factor subset. Both methods are discussed in greater detail below.

Ridge Regression. Ridge regression does not directly eliminate any variables but rather shrinks all coefficients asymptotically toward zero. Thus, a threshold of 0.05 is used to remove factors with little contribution.

The EdTech Context Inventory does not include a true response variable to use in ridge regression. However, the Exchange Platform also contains data from an Implementation Survey, which includes two variables that measure the perceived overall success of edtech implementation as it relates to instruction and student outcomes, respectively. These two variables serve as pseudo-response variables, allowing the use of ridge regression for factor selection. Two-factor subsets are thus generated using ridge regression, one for each pseudo-response variable (sklearn, Ridge_CV; [5]).

Sparse PCA. Rather than relying on a response variable to identify informative variables, unsupervised methods identify those variables that explain the greatest amount of the variance present in the data set. One such method is Sparse Principal Component Analysis (sparse PCA), which applies the L1 norm penalty such that each principal component includes loadings of only a subset of the variables in the data set. Sparse PCA supplements factor selection via ridge regression because it focuses on explaining the inherent variance in contexts among the schools on the Exchange Platform and avoids the potential biases that come from relying on implementation data.

Ten relatively sparse components were identified and the factors identified in the first five components represented the factor subset found by sparse PCA. Principal components in regular PCA are required to be orthogonal, and so each component represents a different axis in the variation in the data. While sparse principal components are not strictly orthogonal, they are nearly orthogonal. Thus, the first five components are used to identify factors that represent multiple axes of variance. Four λ values

(6.5,7,7.5,8) were used in this method, as they each produce sufficiently sparse models from the first five components while still accounting for roughly half of the total variance.

4.2 Clustering

Following subset variable selection, the identified factors were used to fit a Gaussian Mixture Model (GMM) to help classify and describe the schools on the Exchange Platform. The optimal number of clusters and covariance type were identified using BIC.[5].

Evaluation of Model Stability. Cluster stability of the GMM is evaluated via bootstrap resampling and calculation of the Jaccard coefficient [3]; this is an important step as GMMs are non-deterministic. The cluster stability algorithm first takes hard cluster classifications from the original dataset and model. Then, for each bootstrapped sample, the data is sampled with replacement and used to fit a new GMM using the same parameters as the original. Hard cluster classifications are then calculated from the bootstrapped sample and model. Classifications from the original GMM model are then compared with classifications from the bootstrapped model by using the Jaccard coefficient. The Jaccard coefficient measures the similarity between two sample sets (A, B) as

$$\frac{|A \cap B|}{|A \cup B|}$$

or the size of the intersection divided by the size of the union. The bootstrapped cluster with the highest Jaccard coefficient for each original cluster is considered the bootstrapped equivalent cluster, and its Jaccard coefficient is stored. Coefficients are averaged across multiple bootstraps. Average Jaccard coefficients greater than 0.5 are considered to be defining the same cluster, and scores ranging from 0.6-0.7 are considered "good". This evaluation of cluster stability will be used periodically to ensure model quality.

4.3 Matching Algorithm

The Exchange Platform currently offers its users a "Districts Like Mine" service. This work improves upon this service by producing a "Schools Like Mine" feature from which a Platform user receives a ranked list of up to 20 schools most similar to their institution. A combination of soft clustering from the GMM fit in the previous step and the calculation of the Euclidean distances between schools on the Exchange Platform is used to generate this ranked list.

Euclidean distances between all schools on the Exchange platform are calculated using pairwise distance methods [5]. Soft clustering from the GMM is then used to narrow down the schools considered for the ranked list. Clusters of interest include those to which the probability of membership to that cluster is at least 5% for the caller's school. Similarly, other schools on the Platform are considered for the ranked list if they belong, with a probability of at least 5%, to one of these identified clusters of interest. Rankings are determined by the Euclidean distance to the caller's school (D), with a shorter distance conferring a higher rank. Finally, the Euclidean distances are converted to a score that is scaled to the maximum distance from the user's school in the Exchange Platform data set (D_{max}). Scores are calculated using the following equation

$$Score = 1 - \frac{D}{D_{max}}$$

such that the most distant school returns a score of zero and a school with factor values identical to the user's returns a score of one. This scale was chosen to assist with interpretability, as educators may view these scores as being analogous to percent matches.

This method returns a ranked list of up to 20 schools with the highest similarity scores (the shortest distance) that are identified by their district ID number and school ID number. In the event that fewer than 20 schools are in the clusters of interest, all schools within those clusters are returned.

5 Case Study

The following is an example of how these methods will be applied on the EdTech Evidence Exchange Platform.

5.1 Subset Selection

Subset selection using data from the Exchange Platform identified 22 of the 47 factors, with representation from all ten context variables as seen in Figure 1.

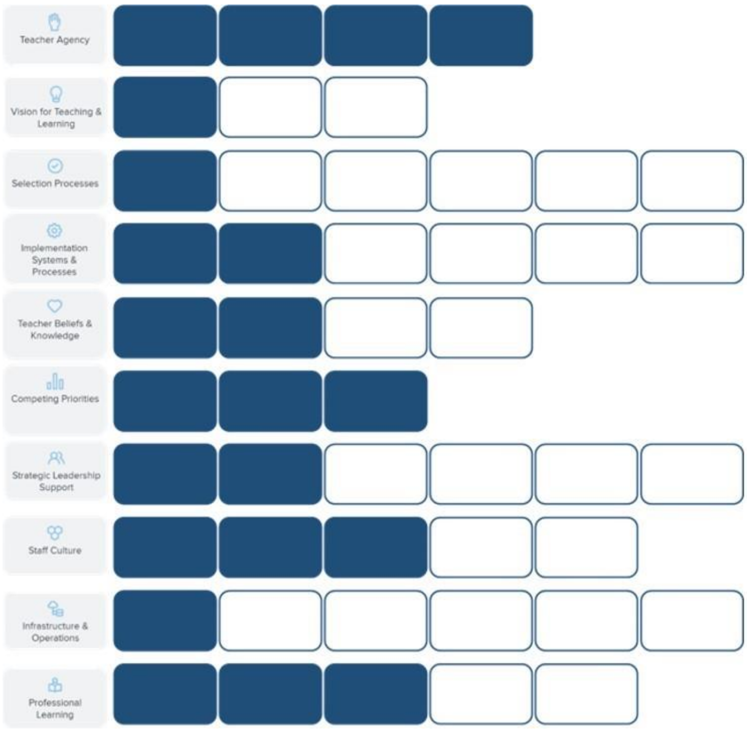

Fig. 1. Subset selection identified 22 factors spanning the ten context features

5.2 Matching Algorithm

Existing Schools. To simulate current users, two schools were randomly sampled from the data set and posed at the callers of the "Schools Like Mine" function. The first school received a list of 15 schools from nine districts with scores ranging from 0.8238 to 0.6411. Combined, these 16 schools most likely belong to a small, distinct cluster. The second school received a list of 20 schools from 12 districts (including its own) with scores ranging from 0.8883 to 0.8167.

A "New" School. The "Schools Like Mine" feature is available to all Exchange Platform users, regardless of whether their school was on the platform when the GMM was last trained. Indeed, this "new" school with randomly generated data received a list of 11 schools, each from a unique district, and the scores ranged from 0.5197 to 0.2612.

6 Conclusion

The methods described in this paper improve upon the current comparison system, leveraging supervised and unsupervised methods to select the factors that most effectively describe a school's context. This subset is then used in the new "Schools Like Mine" feature, which identifies similar schools in terms of the most informative context factors. The methods provide more robust outcomes with significantly more discrete information on variables contributing to differences between districts. If used by future districts looking to purchase mathematic software it will provide excellent points of reference for what works and does not. This could reduce the risk of failed software purchases which is critical, especially for lesser-funded schools and/or districts.

References

1. E. Barton et al., “The EdTech Genome Project Report,” EdTech Evidence Exchange. Accessed: October 2022. [Online]. Available: https://edtechevidence.org/AboutUs/TheGenomeProject/.
2. J. Camacho, A. Smilde, E. Saccenti, J. Westerhuis, and R. Bro, “All sparse PCA models are wrong, but some are useful. Part II: Limitations and problems of deflation,” *Chemometrics and Intelligent Laboratory Systems*, vol. 208, p. 104212, 2021.
3. (n.d.). 3.3.2.10. Jaccard similarity coefficient score. [Online]. Available: https://scikit-learn.org/stable/modules/model evaluation.html
4. E. A. Kohler, L. M. Elreda, and K. Tindle, “EdTech Context Inventory: Factor analyses for ten instruments to measure edtech implementation context features,” *Computers & Education*, vol. 195, p. 104709, 2023.
5. F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
6. H. Zou, T. Hastie, and R. Tibshirani, “Sparse principal component analysis,” Journal of Computational and Graphical Statistics, vol. 15, no. 2, pp. 265–286, 2006. Available: https://hastie.su.domains/Papers/ spc jcgs.pdf