




Protecting Student Privacy with Synthetic Data from Generative Adversarial Networks

Peter Bautista^(✉)  and Paul Salvador Inventado^(✉) 

California State University, Fullerton, CA 92831, USA
pinventado@fullerton.edu

Abstract. Educational data requires layers of protection that prohibit easy access to sensitive student data. However, the additional layers of security hinder research that relies on educational data to progress. In this paper, a Least Squares GAN (LSGAN) is proposed to create synthetic student performance datasets based on a master dataset without recreating samples. Synthetic data is less likely to be traced back to a student thereby reducing privacy issues. Two feature subsets were considered in the study: sequential, and all features. GANs trained on the sequential data produced new datasets that were representative of student performance from the training dataset, while the GAN trained on all features was not able to capture characteristics from the dataset. Based on the results, the synthetic dataset can provide an alternative unrestricted source of data without compromising student privacy.

1 Introduction

Student data is now easier to collect with the advent of learning platforms that make it easy to track learner behavior and performance [6]. Such data allows instructors and researchers to apply learning analytics and educational data mining techniques to analyze student learning that inform teaching [1]. However, as more data about students are collected, consumers of this data need to be more mindful about how it is used and shared to maintain student privacy [11].

Several policies exist to protect student data. In the United States, the Family Educational Rights and Privacy Act (FERPA) of 1974, 20 U.S.C. § 1232g (1974) requires federally funded institutions to get parental or student consent before disclosing personal information. The Children's Online Privacy Protection Act of 1998, 15 U.S.C. § 6501- 6506 (1998) requires web hosts and content providers to seek parental consent to store data about children under 13. The Student Digital Privacy and Parental Rights Act of 2015, H.R.2092, 114th Cong. (2015–2016) prohibits operators from selling personal information to third parties or collecting student information for purposes unrelated to educational activities. In academic settings, researchers are required to get approval from an institutional review board (IRB) in addition to the restrictions set by FERPA before collecting student information [15, 16].

This work aims to leverage generative adversarial networks (GANs) to produce synthetic data based on real student data. Generated data cannot be traced back to an individual thereby reducing privacy issues and satisfying privacy policy requirements [2].

2 Related Works

There are two approaches commonly considered in data privacy: privacy-preserving data publishing, and privacy-preserving data mining [8]. Privacy-preserving data publishing involves sharing sensitive information about individuals without violating their privacy. Privacy-preserving data mining involves the application of data mining without using sensitive information. This work focuses on privacy-preserving data publishing.

The simplest approach to protect students' privacy is to remove unique, identifying information such as names and student IDs. However, non-identifying student information can be checked against other data sources, like social media, and possibly uncover students' identities through implied relationships [10, 14].

Another approach to preserving privacy is k-anonymity and l-diversity. According to k-anonymity we can protect privacy by ensuring that each distinct pattern of key variables is possessed by at least a minimum of k records [13]. k-anonymity can be implemented by generalizing attributes with few observations into categories (e.g., age ranges instead of actual age), or injecting missing values (suppression) to data that is easy to distinguish. l-diversity measures the frequency of values in sensitive attributes (e.g., 2 instances of age 13). Records can be removed or further generalized and suppressed to maintain an l-diversity threshold [8]. A major drawback in this anonymization approach is the loss of data fidelity. Model performance may suffer as more data is lost or modified. There is a tradeoff between privacy and the utility of learning analytics and data mining [6].

Generative adversarial networks create completely synthetic data based on real data [5] and can be used to protect privacy. Baowaly et al., investigated different types of GANs to generate synthetic patient electronic health records (EHR), which are highly restricted data [2, 4]. They developed a medical boundary-seeking GAN (medBGAN) to produce patient data containing International Classification of Diseases (ICD) codes that are used for medical diagnoses. Their evaluation showed that the performance of logistic regression models trained on both real and synthetic data were comparable. GANs are also able to create similar characteristics to the real data such as distribution and mean [3, 9].

3 Data and Methodology

We developed a GAN and used it with student performance data collected from introductory computer science classes at a public Hispanic-serving institution to investigate how well it protected student privacy. The data set contained

information from 104 students who were enrolled in two sections of the same class taught by the same professor. The data consisted of 77 attributes describing grades for multiple assessments across the semester including quizzes, in-class activities, homework, projects, midterm exams, final exams, and their final grade.

Our GAN's discriminator is a recurrent neural network and its generator is a multi-layer densely connected neural network. Both networks utilized the least squares loss (LS Loss) for the loss function. We added regularizers to prevent the generator from producing the same outputs, also known as mode collapse. The output layer is the same shape as a row within the student data with ReLU activation function bounded from 0 to 100.

4 Results and Discussion

Two subsets of the original dataset were used for generation. The first included quizzes, midterms, and finals which were considered as sequential data where an earlier element of a row preceded those later in the same row. For example, students' performance in quiz 1 precedes the same student's performance for quiz 2 and quiz 3 in the same data point. There were 14 input features in total. The second subset contained all 77 initial graded attributes. The original dataset was split into 70 students for training and 34 for validation. The dataset generated by the GAN is the same size as the original with 104 students.

Figure 1 shows the average values of the 14 attributes from the synthetic data produced by our GAN and real student data. We measured the residuals to determine how well the generated data matched the real data. The average residual was 1.2% indicating that the synthetic data closely resembled real student data, but did not fit it exactly. We want to avoid residuals that are 0 as this might indicate overfitting. If there is overfitting, there is a possibility the GAN is recreating samples from the original dataset.

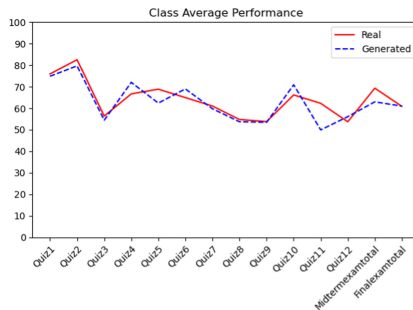


Fig. 1. Class average for features within sequential data

Looking closer at the individual rows, we find that in addition to creating a good fit for the data, the generator created unique rows. Figure 2 shows a

heatmap with attribute values from 14 randomly chosen real and synthetic students. The two heatmaps are essentially indistinguishable. Compared to real student data, the synthetic student data attribute values were close to their means, within the same range, followed similar patterns, but did not duplicate real student data.

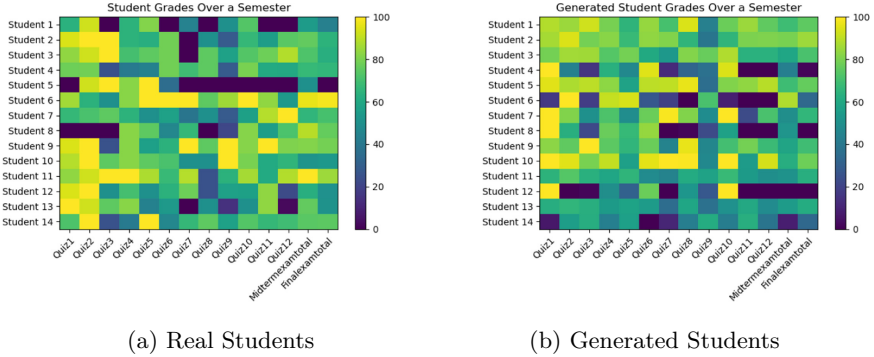


Fig. 2. Heatmap of student grades for quizzes, midterm, and final

Unfortunately, increasing the number of features caused the GAN to destabilize and resulted in the generator producing near zero outputs for a majority of features. Neither the generator nor discriminator were able to reach an acceptable level of performance even after increasing the number of training data.

5 Conclusion and Future Work

Overall, the GAN successfully generated synthetic student data that did not replicate real student data. It was able to recreate similar characteristics such as the averages and distributions of real data with only minor deviations. Therefore it can protect students' privacy.

Further work is necessary to generate synthetic data with more attributes without destabilization. Possible next steps include applying GANs on multiple attribute subsets or using autoencoders to reduce destabilization [2]. The data used in this work focused on continuous data, but we plan to explore its performance on categorical features such as ethnicity, gender, and major.

Other than protecting student privacy, GANs' generative nature increases data set sizes. As we saw in our experiment, the generator produced realistic data even with limited training data. Data generation can be useful for small scale-studies such as those conducted by instructors in their classes which often have limited data. Since model performance will likely depend on recent student performance data, data sources will contain information on students who are still in school. Therefore, these studies will benefit from GANs' privacy-preserving

nature. **Small-scale studies are important** because they inform researchers on how they might scale their research and it is also a common activity conducted by teachers to inform their teaching [7,12].

References

1. Baker, R.S., Inventado, P.S.: Educational data mining and learning analytics. In: Larusson, J.A., White, B. (eds.) *Learning Analytics*, pp. 61–75. Springer, New York (2014). https://doi.org/10.1007/978-1-4614-3305-7_4
2. Baowaly, M.K., Lin, C.C., Liu, C.L., Chen, K.T.: Synthesizing electronic health records using improved generative adversarial networks. *J. Am. Med. Inform. Assoc.* **26**(3), 228–241 (2019)
3. Behjati, R., Arisholm, E., Bedregal, M., Tan, C.: Synthetic test data generation using recurrent neural networks: a position paper. In: *2019 IEEE/ACM 7th International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE)*, pp. 22–27 (2019)
4. Choi, E., Biswal, S., Malin, B., Duke, J., Stewart, W.F., Sun, J.: Generating multi-label discrete patient records using generative adversarial networks (2018)
5. Goodfellow, I.J., et al.: *Generative adversarial networks* (2014)
6. Gursoy, M.E., Inan, A., Nergiz, M.E., Saygin, Y.: Privacy-preserving learning analytics: challenges and techniques. *IEEE Trans. Learn. Technol.* **10**(1), 68–81 (2016)
7. Kitchin, R., Lauriault, T.P.: Small data in the era of big data. *GeoJournal* **80**(4), 463–475 (2014). <https://doi.org/10.1007/s10708-014-9601-7>
8. Kyritsi, K.H., Zorkadis, V., Stavropoulos, E.C., Verykios, V.S.: The pursuit of patterns in educational data mining as a threat to student privacy. *J. Interact. Media Educ.* **2019**(1) (2019)
9. Lan, J., Guo, Q., Sun, H.: Demand side data generating based on conditional generative adversarial networks. *Energy Procedia* **152**, 1188–1193 (2018). <https://doi.org/10.1016/j.egypro.2018.09.157>. <http://www.sciencedirect.com/science/article/pii/S187661021830701X>. *Cleaner Energy for Cleaner Cities*
10. Narayanan, A., Shmatikov, V.: Robust de-anonymization of large sparse datasets. In: *Proceedings of the 2008 IEEE Symposium on Security and Privacy*. SP 2008, pp. 111–125. IEEE Computer Society, USA (2008). <https://doi.org/10.1109/SP.2008.33>
11. Pardo, A., Siemens, G.: Ethical and privacy principles for learning analytics. *Br. J. Edu. Technol.* **45**(3), 438–450 (2014)
12. Rodríguez-Triana, M.J., Martínez-Monés, A., Villagrà-Sobrino, S.: Learning analytics in small-scale teacher-led innovations: ethical and data privacy issues. *J. Learn. Anal.* **3**(1), 43–65 (2016)
13. Samarati, P., Sweeney, L.: Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression (1998)
14. Swenson, J.: Establishing an ethical literacy for learning analytics. In: *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge*, pp. 246–250 (2014)
15. Willis, J.E., Slade, S., Prinsloo, P.: Ethical oversight of student data in learning analytics: a typology derived from a cross-continental, cross-institutional perspective. *Educ. Tech. Res. Dev.* **64**(5), 881–901 (2016). <https://doi.org/10.1007/s11423-016-9463-4>
16. Zeide, E.: Student privacy principles for the age of big data: moving beyond FERPA and FIPPS. *Drexel L. Rev.* **8**, 339 (2015)