Empowering Teachers to Integrate Machine Learning into K-12 Scientific Discovery

Xiaofei Zhou*, Jingwan Tang*, Sufian Mushtaq, Xiaoyu Wan, and Zhen Bai

The University of Rochester, Rochester NY 14623, USA
xiaofei.zhou@rochester.edu
jingwan.tang@rochester.edu
sufianmushtaq@rochester.edu
xwan3@ur.rochester.edu
zhen.bai@rochester.edu

Abstract. With the recent demand to cultivate scientific mindsets in the era of Artificial Intelligence (AI), this paper investigates opportunities of integrating Machine Learning (ML) with K-12 scientific discovery. We co-designed a learning environment called SmileyDiscovery with science teachers. Preliminary findings indicate that SmileyDiscovery efficiently supports teachers to obtain ML skills and create ML-empowered scientific discovery activities.

Keywords: ML education · K-12 scientific discovery · Co-design with teachers .

1 Introduction

In today's science community, Machine Learning (ML) is considered as the new engine to accelerate knowledge discovery for major STEM domains [1]. There is a natural connection between ML and K-12 scientific practices of curriculum standards [2] through pattern identification and prediction. Although recent research sheds light on connecting ML and K-12 STEM activities [3, 4], there remains a research gap of preparing STEM teachers to utilize ML as a new learning tool to engage students with scientific discovery. This study explores opportunities and challenges of integrating ML and K-12 scientific discovery. We developed a ML-empowered scientific discovery learning environment named "SmileyDiscovery", through the co-design research method [5] in collaboration with a science educator and eighteen K-12 teachers in a teacher training course.

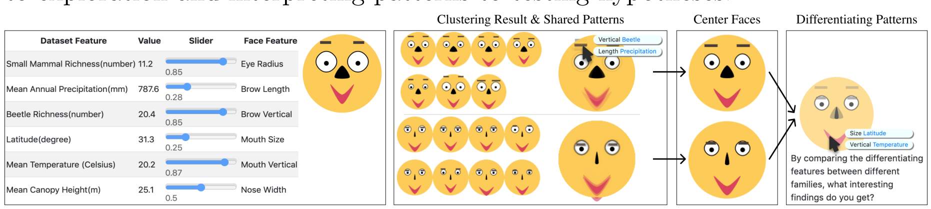
2 SmileyDiscovery System Design

SmileyDiscovery extends an existing system named SmileyCluster [3] with a variety of ML-empowered scientific discovery activities. SmileyCluster aims to introduce accessible ML to K-12 students by simplifying k-means clustering through visual observation of emoji faces. It maps data to facial features and enables interpretation of patterns by arranging and overlaying emoji faces to compare similarity/dissimilarity of data within/between clusters (**Fig. 1**). To investigate

* Equal contribution.

meaningful learning activities aligned with curriculum standards, and specific connections between similarity-based ML and key learning processes involved in scientific discovery in K-12 classrooms, we conducted a two-phase co-design study.

Phase 1 To explore initial ML-empowered learning activities, we collaborated with a science educator with extensive research and teaching experience of K-12 science education. We developed three scientific discovery activities in SmileyDiscovery. The topics include dynamic ecosystems, wine chemistry and breast cancer diagnosis with corresponding datasets [6–8]. SmileyDiscovery connects k-means clustering components with key scientific discovery components suggested by the How Science Works framework [9], such as linking pairwise comparison to exploration and interpreting patterns to testing hypotheses.



(a) A SmileyDiscovery feature to introduce the **mapping mechanism** between dataset features and facial features.

(b) SmileyDiscovery features for **pattern interpretation** by comparing shared/differentiating features of/between clusters.

Fig. 1. Highlights from SmileyDiscovery (Dynamic Ecosystems learning activity).

Phase 2 To gather feedback and develop a generalized instructional design for ML-empowered scientific discovery activities, we conducted a co-design workshop with 18 STEM teachers (3 elementary, 7 middle and 8 high school teachers) in a remote K-12 teacher training course provided by the Warner School of Education at the University of Rochester. We assigned participants to four groups based on their grade levels and conducted this study in two sessions, one week apart. In session 1, the teachers were introduced to SmileyDiscovery, and asked to complete one of the three pre-designed SmileyDiscovery activities individually, with a focus group interview at the end. In session 2, teachers worked in groups to collaboratively design a learning activity with a teaching topic of their own by customizing SmileyDiscovery features, and discussed system limitations and potential improvements to better facilitate learning and teaching.

3 Preliminary Findings

The pre- and post-study of learning gains show that SmileyDiscovery helps teachers obtain essential k-means clustering concepts and methods. In addition, during the co-design workshop, all four groups of teachers successfully designed ML-empowered scientific discovery activities in different STEM contexts that reflect the main scientific inquiry processes of the 5E model of scientific instruction - engage, explore, explain, elaborate and evaluate [10]. We are currently conducting an in-depth analysis of the learning activities designed by each teacher group, and feedback about limitations and potential improvements for SmileyDiscovery. These findings will inform the next iteration of SmileyDiscovery, and future evaluation for authentic ML-empowered learning in K-12 STEM classrooms.

References

1. Gil, Y., Greaves, M., Hendler, J., Hirsh, H.: Amplify scientific discovery with artificial intelligence. Science 346(6206), 171–172 (2014)
2. NGSS Lead States: Appendix F: Science and Engineering Practices in the NGSS. Next Generation Science Standards: For States, By States. (2013)
3. Wan. X., Zhou, X., Ye, Z., Mortensen, C., Bai, Z.: SmileyCluster: Supporting accessible machine learning in K-12 scientific discovery. In: Proceedings of the 19th ACM International Conference on Interaction Design and Children (in-press article) (2020)
4. Zhang, Y., Wang, J., Bolduc, F., Murray, W. G., Staffen, W.: A preliminary report of integrating science and computing teaching using logic programming. In: 3rd AAAI Conference on Artificial Intelligence Proceedings, vol. 33, pp. 9737–9744. AAAI Press (2019)
5. Roschelle, J., Penuel, W. R.: Co-design of innovations with teachers: Definition and dynamics. In: Proceedings of the 7th International Conference on Learning Sciences, pp. 606-612. International Society of Learning Sciences (2006)
6. Cortez, P., Cerdeira, A., Almeida, F., Matos, T., Reis, J.: Modeling wine preferences by data mining from physicochemical properties. Decision Support Systems 47(4), 547–553 (2009)
7. UCI Machine Learning Repository http://archive.ics.uci.edu/ml. Last accessed 7 June 2020
8. National Ecological Observatory Network, http://data.neonscience.org. Provisional data downloaded from on 25 February 2020
9. Understanding Science, http://www.understandingscience.org. Last accessed 7 June 2020
10. Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., Landes, N.: The BSCS 5E instructional model: Origins and effectiveness. Colorado Springs, Co: BSCS (1999)