International Society of the Learning Sciences

Modeling Unstructured Data: Teachers as Learners and Designers of Technology-enhanced Artificial Intelligence Curriculum

Cansu Tatar, North Carolina State University, ctatar@ncsu.edu
Michael Miller Yoder, Carnegie Mellon University, mamille3@andrew.cmu.edu
Madeline Coven, North Carolina State University, macoven@ncsu.edu
Kenia Wiedemann, Concord Consortium, kwiedemann@concord.org
Jie Chao, Concord Consortium, jchao@concord.org,
William Finzer, Concord Consortium, wfinzer@concord.org
Shiyan Jiang, North Carolina State University, sjiang24@ncsu.edu
Carolyn P. Rosé, Carnegie Mellon University, cp3a@andrew.cmu.edu

Abstract: In this paper, we present a co-design study with teachers to contribute towards development of a technology-enhanced Artificial Intelligence (AI) curriculum, focusing on modeling unstructured data. We created an initial design of a learning activity prototype and explored ways to incorporate the design into high school classes. Specifically, teachers explored text classification models with the prototype and reflected on the exploration as a user, learner, and teacher. They provided insights about learning opportunities in the activity and feedback for integrating it into their teaching. Findings from qualitative analysis demonstrate that exploring text classification models provided an accessible and comprehensive approach for integrated learning of mathematics, language arts, and computing with the potential of supporting the understanding of core AI concepts including identifying structure within unstructured data and reasoning about the roles of human insight in developing AI technologies.

Introduction

Artificial Intelligence (AI) has become a ubiquitous facet of our daily lives. Our young generation should be given high-quality academic training and environments in which they are empowered to participate in the public discourse about AI (AI4K12, 2019; Zimmermann-Niefield et al., 2019). It is a challenge to position youth to be thought leaders in this public discourse without fostering an in-depth understanding of what exactly AI is, how AI works, and why and where AI is used.

Data is the raw material from which artificial intelligence is imbued in artifacts through design. Learning from data is the process through which an AI program gains its "intelligence", and one that we argue is critical for students to devote more focus to. The growing attention to data literacy in pre-college education (e.g., Haldar, Chopra, Wong, Heller, & Konold, 2018) has partially responded to this learning goal. However, the data literacy community has primarily focused on structured data, which has already been quantified and distilled into variables arranged in tables, ready for modelers to manipulate and analyze (Haldar et al., 2018; Lee & Wilkerson, 2018). What has been neglected is the highly consequential and highly subjective process of operationalizing those variables from unstructured and semi-structured types of data such as text, image, audio, and video. The ISLS research community has acknowledged the great deal of creativity, insight, and variation that can be brought to bear in this process of operationalization in its investigation into Multivocality (Suthers et al., 2013). The message of the active and iterative role that data analysts must play when applying modeling technologies is one that we as researchers must also remind ourselves about (Rosé, 2018). Modelers need to extract features that define a representational space in which meaningful distinctions can be made between subsects of instances that take on different significance within an application or a conceptual model. This is what it means for AI developers to do the work of making a prediction problem learnable (Witten et al., 2016). Identifying the iterative process of structuring data, building models, and troubleshooting them in order to understand what their weaknesses tell us about how the structuring needs to be adjusted are the core activities that developers of AI need to be educated in, and which we attempt to make accessible to students. These foundational data practices are critical for students to understand the role of human insight in developing AI technology, but currently are missing in k-12 education.

Much is unknown about ways to advance students' understanding of the above-mentioned core AI concepts. Built on earlier work on student engagement in data wrangling practices (Jiang & Kahn, 2020), this study aimed to evaluate an initial design of a text classification learning activity and explore possible ways to incorporate the design into high school classes, especially non-computing classes. We report a co-design study of a three-year design-based research program (Brown, 1992) that aims to create a high school curriculum powered by StoryQ to engage students in practices of modeling unstructured data. In particular, we guided teachers to test-

drive (Penuel et al., 2007) a text classification learning activity using StoryQ prototype and conducted a semi-structured interview to elicit teachers' thoughts on the activity. Specifically, we examined:

1. How do teachers come to understand modeling unstructured data?
2. In what ways do teachers think modeling unstructured data can be integrated into their classes, especially non-computing classes?

Methods

Participants and context

Two teachers, Martin and Hector (pseudonyms), participated in co-designing the unit and technology. The research team met with them prior to the co-design session to discuss the project. Martin has taught ELA (i.e., English Language Arts) for 20 years and Hector had 21-year teaching experience in mathematics and computer science (CS). He was actively learning about teaching AI. We intentionally selected two teachers from different backgrounds as the curriculum will integrate the learning of mathematics, ELA, and computing. In addition, we envision that the curriculum will be co-taught by ELA and mathematics/CS teachers in future iterations.

The co-design session was conducted remotely via Zoom, a video conferencing tool, and included three sections: a five-minute introduction, one-hour clickbait activity, and one-hour semi-structured interview. We first described the goal of the co-design session and emphasized that teachers were expected to experience the learning activity as a user, learner, and teacher. Afterward, we walked teachers through our prototype and associated learning activities to understand how human and computer models determine whether a headline was a) clickbait for encouraging visitors to click on a link or b) news from professional news agencies. The clickbait activity was an exemplar topic for introducing text classification models. The prototype uses logistic regression over unigram features (i.e., single words) as the classification model, which allows predicting headlines into discrete labels by learning the relationship from a given set of headlines with actual labels. We chose it for our prototype as it is easy to interpret for novices and efficient to train (Witten et al., 2016). To solicit feedback, we conducted a semi-structured interview after the clickbait activity. In addition to the co-design session, teachers wrote reflections and contributed ideas about designing the curriculum and platform.

Data collection and analysis

We collected video recordings of the co-design session, including 1) teachers' interactions with the prototype when exploring text classification models to address RQs 1 and 2 and 2) their responses to interview questions to answer RQs 2 and 3. For analysis, we used interaction analysis (Jordan & Henderson, 1995), thematic coding (Braun & Clarke, 2006), and peer debrief. Three researchers developed analytics memos of video recordings, followed by peer-debriefing with four other researchers to ensure validity. Furthermore, we collected and coded teachers' written reflections and ideas throughout the project to gain more insights about their views of learning opportunities in modeling unstructured data and integrating the activity into high school classrooms.

Findings

RQ1: How do teachers come to understand modeling unstructured data?

We found that teachers' understanding of modeling unstructured data evolved after connecting human insight and computer decision making. First, Martin created a dot graph and then they discussed why a headline could be a clickbait or news based on their interpretation of the headline and prior knowledge. Since Martin had not interacted with the prototype and Hector used the prototype before, the research team asked Martin to share the screen. When reading and interpreting headlines, they wrote down rules (Martin typed based on discussions) for clickbait and news headlines and described the process of reading samples and generating rules as inductive reasoning.

Teachers then created a confusion matrix (Witten et al., 2016) to compare actual and predicted labels. Confusion matrix is a performance measurement for machine learning classification that compares predicted and actual labels in a two-dimensional array. As Martin described, "I'm immediately interested in all the ones that where the computer predicted incorrectly." In addition, he not only noticed errors from the computer model but also proposed how we could use the error as guidance to revise the model. Furthermore, Martin and Hector constantly drew on prior knowledge, such as equating the process of computer decision making as advanced search (e.g., headlines with superlative language were more likely to be clickbait).

After investigating the confusion matrix, Martin and Hector probed into computer decision making by creating a frequency and feature weight graph from the features table (Figure 1). They were prompted to

investigate the graphs of a) distribution and b) frequency and feature weight together to understand where the weights came from. Martin clicked feature "you" (the highest positive weight, representing that a headline containing "you" would more likely be predicted as clickbait) in the frequency and feature weight graph. Immediately, they noticed that only one headline containing "you" was not clickbait in the distribution graph. While they wondered why it had the highest positive weight, the research team explained that the high proportion of clickbait headlines to news headlines containing "you" contributed to the high positive weight. They came to understand that the feature weight was related to the proportion of clickbait headlines to news headlines containing the feature in the training data, which is the dataset that we used to train a machine learning model. The clickbait activity ended with teachers discussing sources of error from features and possible ways to improve the model.

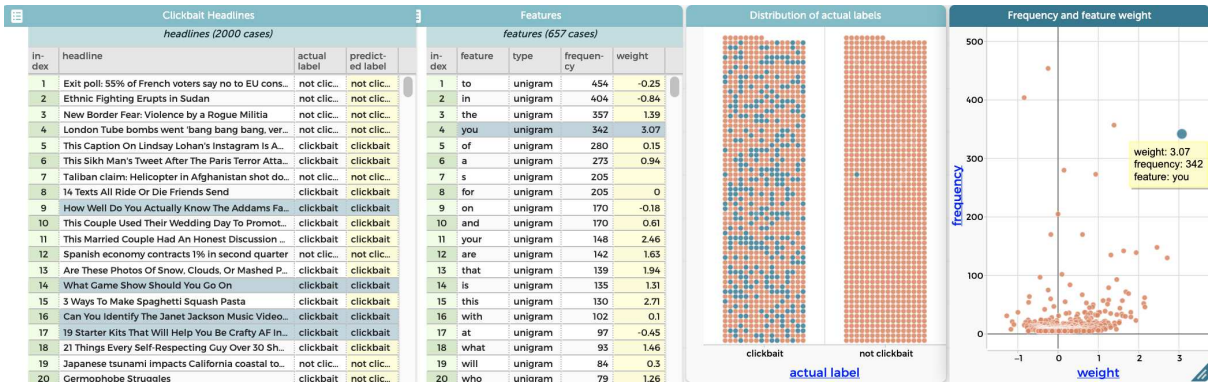

Figure 1. Features table (second) and graph (fourth). Unstructured texts needed to be converted to structured features for machine learning algorithms to exploit. One type of feature is unigram, or single word.

We can clearly see that teachers (in particular Martin, the ELA teacher) became highly engaged in reasoning about rules that computers could learn well, or not, such as the content. They developed an informal understanding of "feature" as "representation that the model can use." In general, while they were comfortable with connecting human insight and AI models, proposing ideas for programming rules into AI models for text classification, they (especially Martin) needed further guidance to interpret graphs with mathematical reasoning and understand how learning weights on features works from a machine learning perspective.

RQ 2: In what ways do teachers think modeling unstructured data can be integrated into their classes, especially non-computing classes?

Martin and Hector identified learning opportunities from different perspectives. Martin emphasized different layers of learning in ELA, from word choice and sentence structure to interpreting broader contexts, such as historical contexts surrounding text producers. He mentioned that models with basic features (e.g., unigram) might not be able to predict themes of stories without adding advanced features related to information about writers of stories. In addition, he suggested engaging students in analyzing misclassifications (i.e., predicted labels are different from actual labels) within and beyond texts, such as history surrounding the text. In terms of Math and CS learning, Hector discussed that the activity could introduce students to graph literacy, mathematical reasoning, error analysis in rich contexts, and the importance of human insight in developing AI technologies.

Both Martin and Hector suggested leveraging the exploration of text classification models as an interdisciplinary learning approach. They described the learning activity as a focused data analysis of language and a comprehensive approach to and unique context for integrating STEM and the humanities. Also, as described earlier, they stressed the need of helping students to interpret text by drawing on knowledge beyond text. Martin further framed the exploration of models as an innovative way of addressing the criticism in ELA, "most of the language arts for the last 20 years or so have been a form of new criticism, they call it new criticism, where you analyze a text using only the data within the text." From his perspective, when conducting error analysis, students could be guided to draw personal experience and knowledge about broader contexts to identify sources of error in computer models. In addition, he suggested integrating the activity into traditional writing lessons, such as having students write texts and using their writing as test data to measure model performance.

Discussion and implications

In this study, we examined how teachers began to understand models that were built from unstructured data and ways of integrating modeling unstructured data into their classes, especially non-computing classes. Specifically,

the activity sequence of HCC (Human decision making, Confusion matrix, Computer decision making) created an exciting opportunity for participants to understand the role of human insight in developing AI technology as well as reasoning about computer decision making in rich contexts. The confusion matrix, in particular, triggered in-depth error analysis such as analyzing the severity, number, and types of errors that models might produce and how models could be modified by applying human intelligence. While teachers had rich discussions about computer decision making by referring to rules that they identified, they were confused about the feature table, including how models learned feature weights. In the next round of co-design, we will extend the HCC into HC**P**C sequence. **P** stands for adjusting **perspectives** on models and modeling (i.e., practicing the iterative process of updating rules based on insights gained through probing into the behavior of models). We expect that this modified activity sequence might help participants to reason about model performance from a machine learning perspective.

Notably, teachers drew on prior knowledge and personal experience in the discussions about texts and contexts surrounding texts. Future studies could examine other text classification tasks and strategies to create broadly inclusive learning experiences that engage participants from diverse backgrounds in expressing their cultures and personalities. Also, teachers rarely questioned where the dataset came from and who created the actual label. Having participants manually label a small data set and create labels for classification might support them in questioning modeling processes instead of assuming the objectivity of data science and that labels come from nowhere (D'Ignazio & Klein, 2020). In summary, this work contributes to the research on creating a scalable learning environment for high school students to participate in text mining practices in an accessible, relatable, and empowering way, develop understandings of core concepts including machine learning and unstructured data, and envision their own future lives that are centered on or powered by artificial intelligence.

References

AI4K12.org (2019). *Five big ideas in AI* (Poster). Retrieved from https://bit.ly/ai4k12-five-big-ideas.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The journal of the learning sciences, 2*(2), 141-178.

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. Cambridge, MA: The MIT Press.

Haldar, L. C., Wong, N., Heller, J. I., & Konold, C. (2018). Students Making Sense of Multi-level Data. *Technology Innovations in Statistics Education*, *11*(1).

Jiang, S., & Kahn, J. (2020). Data wrangling practices and collaborative interactions with aggregated data. *International Journal of Computer-Supported Collaborative Learning*, *15*(3), 257-281.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences, 4*(1), 39–103.

Lee, V. R., & Wilkerson, M. (2018). Data use by middle and secondary students in the digital age: A status report and future prospects. Commissioned Paper for the National Academies of Sciences, Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6-12. Washington, D.C.

Penuel, W. R., Roschelle, J., & Shechtman, N. (2007). Designing formative assessment software with teachers: An analysis of the co-design process. *Research and practice in technology enhanced learning*, *2*(01), 51-74.

Rosé, C. P. (2018). Learning analytics in the Learning Sciences, invited chapter in F. Fischer, C. Hmelo-Silver, S. Goldman, & P. Reimann (Eds.) *International Handbook of the Learning Sciences*, Taylor & Francis.

Suthers, D. D., Lund, K., Rosé, C. P., & Teplovs, C. (2013). Achieving productive multivocality in the analysis of group interactions. In *Productive multivocality in the analysis of group interactions* (pp. 577-612). Springer, Boston, MA.

Witten, I. H., Frank, E., Hall, M., and Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*, fourth edition, Elsevier: San Francisco.

Zimmermann-Niefield, A., Turner, M., Murphy, B., Kane, S. K., & Shapiro, R. B. (2019, June). Youth Learning Machine Learning through Building Models of Athletic Moves. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (pp. 121-132). ACM.

Acknowledgments

This work was supported by the National Science Foundation (NSF) grants DRL-1949110. However, any opinions, findings, conclusions, or recommendations are our own and do not reflect the views of the NSF.