Principles and Applications of Collaborative Interactive Machine Teaching

Behnoosh Mohammadzadeh

Principles and Applications of Collaborative Interactive Machine Teaching
Entraînement interactif et collaboratif d'algorithmes d'apprentissage : principes et applications

BEHNOOSH MOHAMMADZADEH, Université Paris-Saclay, CNRS, LISN, France

While human-centered approaches to Machine Learning investigate diverse human roles within the interaction loop, the concept of Interactive Machine Teaching (IMT) has proposed with an emphasis on leveraging the teaching abilities of humans to build machine learning systems. Yet, most systems and studies in this field are devoted to single users. This doctoral thesis focuses on collaborative interactive machine teaching, exploring how people can collectively structure the teaching process and understand their experience. The research introduces TeachTOK, a collaborative interactive machine learning application that enables groups of users to curate data and train a model together incrementally. The thesis presents an exploratory study in collaborative interactive machine teaching, highlighting different collaboration patterns in different teaching activities and how collaboration helps revise teaching strategies. Based on these findings, we expand the work with novel research questions for the subsequent user study to investigate the potential of collaboration in other forms of involving humans in the loop for improving the quality of ML models, specifically regarding bias, harmful behaviours, and fairness.

Les approches centrées sur l'humain en apprentissage automatique explorent les différents rôles humains dans la boucle d'interaction. Le concept d'Enseignement Automatique Interactif (Interactive Machine Teaching) met l'accent sur l'utilisation des compétences pédagogiques des humains pour developer des systèmes d'apprentissage automatique. La plupart des systèmes et des études dans ce domaine sont néanmoins conçus pour des utilisateurs individuels. Cette thèse de doctorat se concentre sur l'enseignement automatique interactif collaboratif, cherchant à comprendre comment les utilisateurs peuvent structurer collectivement le processus d'enseignement. La recherche présente TeachTOK, une application collaborative interactive d'apprentissage automatique permettant à des groupes d'utilisateurs de curater des données et d'entraîner progressivement un modèle ensemble. La thèse explore différents schémas de collaboration dans différentes activités d'enseignement et comment la collaboration aide à réviser les stratégies d'enseignement. Basé sur ces résultats, le travail est étendu avec de nouvelles questions de recherche pour une étude ultérieure, visant à explorer le potentiel de la collaboration pour améliorer la qualité des modèles d'apprentissage automatique, en particulier en ce qui concerne les biais, les comportements nuisibles et l'équité.

CCS Concepts: • **Interactive Machine Learning, Machine Teaching, Collaborative Interaction**;

Mots Clés et Phrases Supplémentaires: Apprentissage automatique interactif, enseignement automatique, interaction collaborative, thesis

Reference:
Behnoosh Mohammadzadeh. 2024. Principles and Applications of Collaborative Interactive Machine Teaching. *IHM '24 : Rencontres Doctorales de la 35^{e} conférence Francophone sur l'Interaction Humain-Machine, March 25–29, 2024, Paris, France*

1 INTRODUCTION

Machine Learning (ML) has experienced significant growth in recent years, impacting diverse users. However, despite its widespread use, there is limited public scrutiny, which leads to potential biases and harmful consequences. In current ML practices, end users lack control over crucial aspects like training data and performance evaluation. Recent research suggests involving end-users in human-in-the-loop approaches to mitigate social risks in ML systems [17]. Including

diverse stakeholders in training and evaluation could enhance performance, transparency, and fairness. However, making ML accessible to a broader audience remains challenging due to the technical expertise usually required in the development process.

To overcome this, research in Human-Computer Interaction (HCI) explores involving end users in designing machine learning (ML) models through Interactive Machine Learning (IML) [1, 6]. This interdisciplinary field merges HCI and ML, allowing ML novices to participate in various stages of model development, such as dataset creation, labelling, and advanced control over features. The process enables end users to apply their domain knowledge iteratively, creating customized and transparent models for their specific applications [12, 13]. As part of this endeavour, recent studies emphasize Machine Teaching, focusing on the human teacher's ability to convey concepts to learning machines efficiently [18, 21]. This shift in perspective offers insights into how novices in ML can contribute to designing more accessible and democratic technologies [19]. Ramos et al. [18] further proposed the notion of Interactive Machine Teaching (IMT), where the focus is on leveraging the teaching skills of humans and implicit and explicit forms of their knowledge (labels, features, rules, etc.) in the design of IML systems. The authors emphasize that IMT is distinct from IML by a focus on model-building, and the specific role of the human-in-the-loop as a teacher.

Most studies in machine teaching have primarily focused on engaging individual users, whether they are novices or domain experts. Current machine learning practices supporting multiple users are typically confined to crowdsourcing approaches to ensure data quality(Bonnet et al., Chang et al., Ferrario et al., Heimerl et al., Kellenberger et al., Kulesza et al.). While some work leverages crowd annotation to overcome biases and lack of diversity [9], users and domain experts in these scenarios are mainly involved in data annotation, lacking significant influence over model design and training. In contrast, Machine Teaching and Interactive Machine Learning integrate users more closely into building and evaluating ML models. The collaborative involvement of users throughout the ML training process in Interactive Machine Learning presents potential challenges and mutual benefits for both ML practitioners and end-users. Yet, there needs to be more knowledge regarding the extent of these possibilities and challenges.

Therefore, in this doctoral thesis, "Principles and Applications of Collaborative Machine Teaching", we developed a collaborative IML application to study how a group of users can be involved in a collaborative scenario to train an ML model. We are interested in developing a human-centered approach by designing interactions to foster collaboration in IMT and involve end-users in various tasks with diverse roles to enhance the sociotechnical impacts of ML models.

2 FIRST EXPLORATORY STUDY

2.1 Overview

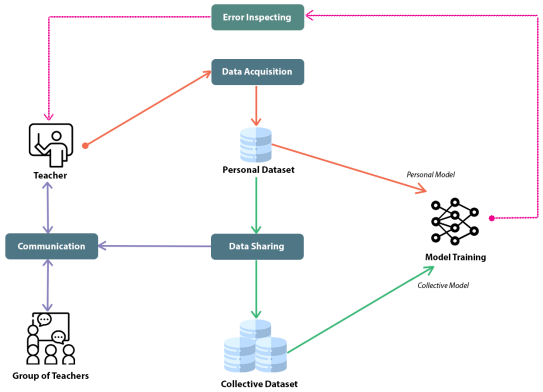
We designed and developed a web-based collaborative machine teaching application with Marcelle[1], an open-source toolkit for IML systems [8]. This collaborative prototype enables a group of users to build a model collectively through dataset curation, interactive model training and performance inspection, and real-time communication. Second, we conducted a user study with 12 novices in machine learning, forming three teams that competed for nine days to teach an image classifier to recognize 10 dance styles. The results describe emerging collaboration in machine teaching task as well as opportunities for reflection and discussion through interacting with our application. From these results, we derived a set of implications for the design of future collaborative MT systems. Finally, we submitted the work to the 2024 ACM IUI Conference, which was accepted and will be published soon [16].

[1]https://marcelle.dev/

2.2 TeachTOK: A Collaborative Machine Teaching System

We designed TeachTOK to study how a group of people, as an ML system end-users, would collaborate to solve a machine teaching task. Our main requirements for the design of this application were that (1) people should be able to train the model using their own images; (2) they should be able to assess performance in real-time and inspect errors; (3) the teaching should be collective, meaning that people are organized in a group where they share a common dataset and model, and they should be able to communicate with each other; (4) the collaboration process should be asynchronous and distributed, meaning that people can contribute to the task from anywhere at different moments.

A screeshot of the application and its workflow is illustrated in Figure 2, and 1. TeachTOK draws upon interactive machine teaching systems previously described in the literature [6, 18, 19] but considers that data and models are shared among a group of teachers.



schema.png schema.png

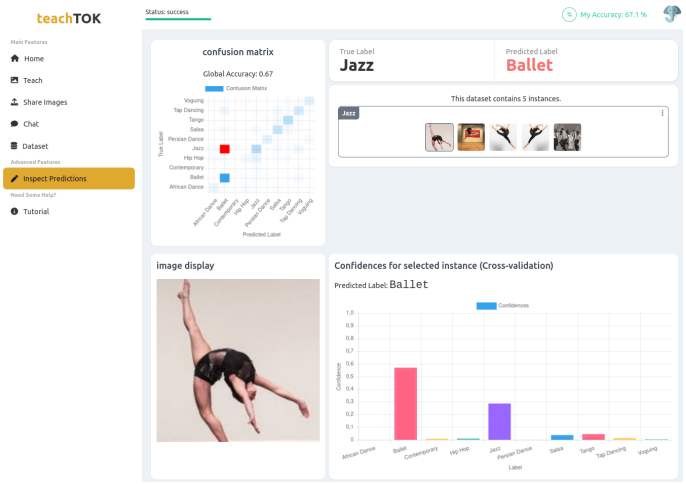
Fig. 1. Workflow of a collaborative interactive machine learning application

In TeachTOK, a group of users constitutes a collective dataset of images together to train an image classifier. Users receive feedback on the classifier's performance, and they can inspect errors. To contribute, users can upload images to the platform to build a personal dataset. This personal dataset is used to retrain the classifier locally, with both the collective data and the new contributions, updating performance measures to let the user assess the effect of their contributions. They can then share their new data publicly with the group. The data is added to the collective dataset and synchronized with all group members.

2.3 Preliminary results

Through both a quantitative and qualitative analysis (thematic analysis[3]), we ended up with the following results.

(1) According to Ramos et al.[18], the machine teacher is engaged in three activities: *Planning*, to identify diverse, challenging examples to teach, reflect on their strategy, and adjust their approach as they assess the evolution of concepts.; *Explaining*, to provide the necessary knowledge to the learning algorithm such as labelling data for classification; and *Reviewing*, to evaluate the confusions, debug errors, and correct labels to gain a comprehensive understanding of the model performance. We found that regardless of the different approaches for coordinating the teaching task, all teams proceeded to collaborate on the Planning activity.



matrix (2).png matrix (2).png

Fig. 2. TeachTOK enables users to inspect errors using an interactive confusion matrix (top left). Clicking on the confusion matrix displays the corresponding data (top right). Clicking on an image displays its associated prediction with the latest model, in particular through a bar chart of class confidences (bottom right).

(2) Although participants were tasked with building a classifier within a time constraint, focusing on accuracy as the primary performance metric, we observed that differences in teaching strategies emerged, with participants realizing the impact of diverse image selection methods. The study revealed that collaboration facilitated the identification of biases, fostering communication among team members to refine teaching strategies and integrate considerations such as diversity into the collective approach.

(3) TeachTOK's design facilitated individual reflection on the teaching task through a tight loop integrating reviewing, planning and explaining. Participants also used the communication tools to deliberate with other team members about the task and share strategies and advice.

(4) Participants required more discussion and effective collaboration on all teaching activities (Planning, Explaining, and Reviewing) to converge a collective knowledge of the model and the teaching task.

3 FUTURE DIRECTION

The user study, focusing on accuracy, revealed that participants considered additional criteria like diversity in their teaching strategies, prompting the need for a systematic study on the impact of diversity and participant reflections on bias and fairness.

Moreover, we observed that participants employed various techniques to assess model quality, including inspecting datasets and analyzing confusion and predictions to understand errors and failure causes. However, we showed that participants in all teams mostly performed the reviewing tasks individually. This prompts us to explore what form of collaboration might emerge in the context of these activities. Participants demonstrated the need for a shared understanding of the errors. Achieving this collective agreement requires continuing study group discussions and collaboration within the Reviewing activity.

Building upon our first exploratory study on Collaborative Interactive Machine Teaching, we got motivated to go deeper into the potential of end-user collaboration on the Reviewing activity to discover errors and biases and enhance

the fairness of ML models. This leads us to review related research fields to first grasp a comprehensive understanding of the human perception of Fairness [10, 22] and, second, identify collective practices in scaffolding end-users with tools to audit algorithms [5, 20] they frequently use.

REFERENCES

[1] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the people: The role of humans in interactive machine learning. *Ai Magazine* 35, 4 (2014), 105–120.

[2] Pierre Bonnet, Alexis Joly, Jean-Michel Faton, Susan Brown, David Kimiti, Benjamin Deneu, Maximilien Servajean, Antoine Affouard, Jean-Christophe Lombardo, Laura Mary, et al. 2020. How citizen scientists contribute to monitor protected areas thanks to automatic plant identification tools. *Ecological Solutions and Evidence* 1, 2 (2020), e12023.

[3] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[4] Joseph Chee Chang, Saleema Amershi, and Ece Kamar. 2017. Revolt: Collaborative crowdsourcing for labeling machine learning datasets. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2334–2346.

[5] Alicia DeVos, Aditi Dhabalia, Hong Shen, Kenneth Holstein, and Motahhare Eslami. 2022. Toward User-Driven Algorithm Auditing: Investigating users' strategies for uncovering harmful algorithmic behavior. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.

[6] John J. Dudley and Per Ola Kristensson. 2018. A Review of User Interface Design for Interactive Machine Learning. *ACM Trans. Interact. Intell. Syst.* 8, 2, Article 8 (jun 2018), 37 pages. https://doi.org/10.1145/3185517

[7] Andrea Ferrario, Raphael Weibel, and Stefan Feuerriegel. 2020. ALEEDSA: Augmented reality for interactive machine learning. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[8] Jules Françoise, Baptiste Caramiaux, and Téo Sanchez. 2021. Marcelle: Composing Interactive Machine Learning Workflows and Interfaces. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 39–53.

[9] Mitchell L Gordon, Michelle S Lam, Joon Sung Park, Kayur Patel, Jeff Hancock, Tatsunori Hashimoto, and Michael S Bernstein. 2022. Jury learning: Integrating dissenting voices into machine learning models. In *CHI Conference on Human Factors in Computing Systems*. 1–19.

[10] Galen Harrison, Julia Hanson, Christine Jacinto, Julio Ramirez, and Blase Ur. 2020. An empirical study on the perceived fairness of realistic, imperfect machine learning models. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 392–402.

[11] Alexander Heimerl, Tobias Baur, Florian Lingenfelser, Johannes Wagner, and Elisabeth André. 2019. NOVA-a tool for eXplainable Cooperative Machine Learning. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 109–115.

[12] Fred Hohman, Kanit Wongsuphasawat, Mary Beth Kery, and Kayur Patel. 2020. Understanding and visualizing data iteration in machine learning. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[13] Andreas Holzinger. 2016. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Informatics* 3, 2 (2016), 119–131.

[14] Benjamin Kellenberger, Devis Tuia, and Dan Morris. 2020. AIDE: Accelerating image-based ecological surveys with interactive machine learning. *Methods in Ecology and Evolution* 11, 12 (2020), 1716–1727.

[15] Todd Kulesza, Saleema Amershi, Rich Caruana, Danyel Fisher, and Denis Charles. 2014. Structured labeling for facilitating concept evolution in machine learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3075–3084.

[16] Behnoosh Mohammadzadeh, Jules Françoise, Michèle Gouiffès, and Baptiste Caramiaux. 2024. Studying Collaborative Interactive Machine Teaching in Image Classification. In *Proceedings of the 29th International Conference on Intelligent User Interfaces [TO APPEAR]*. –.

[17] Yuri Nakao, Simone Stumpf, Subeida Ahmed, Aisha Naseer, and Lorenzo Strappelli. 2022. Toward involving end-users in interactive human-in-the-loop AI fairness. *ACM Trans. Interact. Intell. Syst.* 12, 3 (Sept. 2022), 1–30.

[18] Gonzalo Ramos, Christopher Meek, Patrice Simard, Jina Suh, and Soroush Ghorashi. 2020. Interactive machine teaching: a human-centered approach to building machine-learned models. *Human–Computer Interaction* 35, 5-6 (2020), 413–451.

[19] Téo Sanchez, Baptiste Caramiaux, Jules Françoise, Frédéric Bevilacqua, and Wendy E Mackay. 2021. How do people train a machine? Strategies and (Mis) Understandings. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–26.

[20] Hong Shen, Alicia DeVos, Motahhare Eslami, and Kenneth Holstein. 2021. Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–29.

[21] Patrice Y Simard, Saleema Amershi, David M Chickering, Alicia Edelman Pelton, Soroush Ghorashi, Christopher Meek, Gonzalo Ramos, Jina Suh, Johan Verwey, Mo Wang, et al. 2017. Machine teaching: A new paradigm for building machine learning systems. *arXiv preprint arXiv:1707.06742* (2017).

[22] Megha Srivastava, Hoda Heidari, and Andreas Krause. 2019. Mathematical notions vs. human perception of fairness: A descriptive approach to fairness for machine learning. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2459–2468.