Interactive Active Learning for Understanding Workplace Systems And The Potential Problems They Invite

Scott Allen Cambo
Northwestern University
Evanston, IL, 60202, USA
scottallencambo@u.northwestern.edu
Darren Gergle
Northwestern University
Evanston, IL, 60202, USA
dgergle@northwestern.edu

Abstract
Machine Learning Algorithms in the workplace often employ "black box" user experience design making it difficult or impossible to understand how they work, when they do not work, and how to fix it when they do not work. Additionally, these problems can lead to bigger issues, such as algorithmic discrimination, that go unnoticed and leave little understanding of how to rectify the problem once it is identified. We propose research on interactive machine learning system design by studying how active learning approaches can allow for functional understanding of machine learning processes, identification of errors by the worker through the use of multiple models, and the use of collaborative filtering as a way of mitigating algorithmic discrimination in the workplace.

Author Keywords
Active Learning, Query By Committee, End-User Interactive Machine Learning, Algorithmic Discrimination

ACM Classification Keywords
H.5.2 [User Interfaces: Input devices and strategies]: Miscellaneous

Introduction
The presence of machine learning algorithms in the workplace has increased as data-driven software products for

business solutions have become more popular. However, many of these applications employ "black box" user experiences in which the system passively collects data and presents the user with computational insight in the form of predictions and recommendations. These black box user experience design patterns present a few important challenges in the workplace. First, users need feedback and interaction to develop a functional understanding of computational systems. When a system passively collects data in the background and provides only a prediction or recommendation, it can be difficult to act on this insight effectively. Second, machine learning systems are imperfect and without feedback or interactivity it can be difficult or impossible to understand what the system's limitations are and what might be causing them. The way in which these errors occur can bias a worker's actions. Additionally, it can cause accidental discrimination in the workplace or result in discriminatory advertising or customer service [5, 6]. Finally, when the computational system does not allow for a work group to impose their own values upon the system's decision making process, such as racial equality, we invite the risk of operating on predictions and recommendations which violate those values.

To address these challenges, We propose research on interactive machine learning interfaces by studying an approach called *Active learning*. Active learning systems make judicious use of user interactivity to label data for training a machine learner and evaluating its performance. These systems have the potential to serve as a framework for human-computer work processes. One approach to active learning that I find particularly compelling is a technique called *Query By Committee*, which maintains a committee of varying models simultaneously to determine the uncertainty of a prediction or recommendation. To understand the potential benefits and challenges of using workplace learning systems that employ an active learning approach, we propose:

1. studying how *feedback and interactivity can be designed for functional understanding of active learning systems by the workers who use them*.
2. studying how a *query by committee system enables a worker to utilize multiple computational models in identifying the kinds of system limitations and biases that can lead to unintended algorithmic descrimination*.
3. studying how *user interactions and workflow processes can be designed in tandem to provide affordances which enable workers to rectify algorithmic discrimination and biases and minimize the limitations inherent in learning systems.*

Background

Workplace Machine Learning Systems

While workplace algorithms include those which replace managerial tasks such as delegation (e.g. Uber's algorithm for assigning drivers to passengers), the proposed work focuses on algorithms which use passively collected data from workers to both track task completion rates and provide recommendations for optimizing task completion. An example of this can be seen in UPS drivers who have recently begun using a system called Orion. Orion tracks each driver's workflow and uses the data to derive a set of rules for optimizing the process of delivering packages. While using Orion and following its recommendations have proven to increase the number of packages delivered in a day, UPS drivers are left to exercise their own judgement and expertise in deciding when to follow its recommendations [4, 7]. Each driver is required to partake in nearly a week of personalized training and yet many still don't like, understand, or trust the system.

The experiences of these UPS drivers mirror those of Uber drivers studied by Lee et al. seeking to understand the algorithm which assigns drivers (workers) to passengers (customers) [3]. In the study, Lee et al. found that Uber drivers who had more knowledge of the algorithm's inner workings had advantages over their fellow Uber drivers because they found ways to work around the work assignment process to better fit into their work life while others were forced to simply reject passengers and decrease their acceptance rate. Lee et al. suggest that if we design such that understanding of the algorithm's decisions is central to the worker's user experience, then drivers will feel more comfortable with the passenger assignments that it makes. More generally, designing for understanding the algorithm may result in better worker cooperation with algorithmic work suggestions and assignments.

End-User Interactive Machine Learning

Expanding on the importance of user-centered design, we should specifically consider how feedback from the algorithm is displayed and how interacting with that feedback can benefit the learning algorithm's performance as well as the overall work task performance. Previous research by Fogarty et al. began studying this through CueFlik, an experimental web image search program in which users enter a text search query and the system returns images which match the text semantics. Forgarty et al. discovered that when users are presented with both feedback on the system's best and worst guesses and the ability to interact with the guesses by reording their importance, removing them, or replacing them with better examples, they were able to produce better results, in a shorter amount of time, with fewer training examples [2]. Further studying of the CueFlik system showed that displaying multiple potential models of the target concept and providing undo/redo interactions lead to quicker completion of the web image search task[1]. Displaying multiple models and undo/redo affordances are well known as important design features in human-computer interaction and yet they are often missing from workplace software systems employing work assignment or machine learning algorithms.

Active Learning and Query By Committee

The main advantage of active learning is that it only requests that an example of data be labeled for use in training a learning algorithm when uncertainty of that example's label is high. One approach to determining uncertainty is *query by comittee* in which a *comittee* of models are trained using previously collected data and polling for labels of new data is only done when there are large disagreements within the committee. There are many ways that the models within the comittee can be differentiated. The system can use different learning algorithms (e.g. SVM, Random Forest decision trees, K-Nearest Neighbors, etc.) or a single type of learning algorithm can use either various random samples or various sampling strategies to create different models.

An Approach to User Interaction Design For Active Learning Systems

1. Designing for Trust and Functional Understanding

When it is the workers' responsibility to determine when to follow the recommendations of a learning algorithm and when not to follow those recommendations, it is important that the worker can understand what kinds of errors the system is prone to. For example, if a UPS driver consistantly receives a recommendation for a route that seems counterintuitive to his own experience, he will likely avoid following any of the system's recommendations. However, if the system expresses uncertainty about its recommendation and suggests that the user tries the route and report the number of packages delivered so that it can learn, then the driver can learn what factors confound the system and what they can do to help. Additionally, trust in the system can be gained when the driver observes that the system's recommendations are based on the driver's own actions.

2. Using Multiple Models in Identifying Limitations, Biases, and Discrimination

By using multiple approaches to build a committee of models, the system can form a notion of uncertainty when the committee disagrees. However, if we design the system so that the driver is not only made aware that route recommendation is uncertain, but they also receive what each of the committee members' (models') predictions are, then the driver can begin to learn what some of the inherent biases are in each model's approach. For the committee which varies in learning algorithms, the driver learns how these algorithms err and when they should or should not trust them.

For the comittee which varies in sampling strategies, the driver may want to explore the differences in these samples. Doing so can bring to light some incredibly important biases in learning algorithms that often go unrecognized. For example, unintened algorithmic racial discrimination can occur when the learning algorithm is trained on many examples from one ethnicity and few from another. A system, like Orion, can use a query by committee approach in which one model is initially trained using samples randomly from the full data set which is subject to the same biases present in the total population and another is initially trained using examples that evenly represent each race. When the worker or driver is prompted with a discrepancy in predictions from this committee, they might observe that such discrepancy is indicative of racial bias when comparing the predictions of the two models as well as the makeup of the samples which produced each of them. Understanding how to design machine learning systems such that there are opportunities for humans to identify algorithmic racial discrimination is very important to mitigating the problem.

3. Enabling System Correction and Optimization through User Interaction and Collaborative Filtering

Utilizing systems with learning algorithms in the workplace require that the strengths of human cognition and expertise are combined with the strengths of the computational system. Committee disagreements like the algorithmic discrimination example in the previous section may bring problems to light which merit interaction beyond getting the ground truth label. Using techniques like *collaborative filtering*, a work group can collectively identify problematic modeling strategies by flagging a set of model parameters, such as those which resulted in a racially discriminating model, as inappropriate. If enough workers flag the set of model parameters as in appropriate, the model can be dismissed. It is important that researchers in CSCW, HCI, AI, and Machine Learning focus their efforts on designing systems that allow us to rectify problematic biases of learning algorithms like algorithmic discrimination so that recommendations from these learning systems reflect the values of the workers who must use them.

REFERENCES

1. Saleema Amershi, James Fogarty, Ashish Kapoor, and Desney Tan. 2010. Examining multiple potential models in end-user interactive concept learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1357–1360. http://dl.acm.org/citation.cfm?id=1753531

2. James Fogarty, Desney Tan, Ashish Kapoor, and Simon Winder. 2008. CueFlik: interactive concept learning in image search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 29–38.

http://dl.acm.org/citation.cfm?id=1357061

3. Min Kyung Lee, Daniel Kusbit, Evan Metsky, and Laura Dabbish. 2015. Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers. ACM Press, 1603–1612. DOI: http://dx.doi.org/10.1145/2702123.2702548
4. Steven Rosenbush and Laura Stevens. 2015. At UPS, the Algorithm Is the Driver. *Wall Street Journal* (Feb. 2015). http://www.wsj.com/articles/at-ups-the-algorithm-is-the-driver-1424136536
5. Latanya Sweeney. 2013. *Discrimination in Online Ad Delivery*. SSRN Scholarly Paper ID 2208240. Social Science Research Network, Rochester, NY. http://papers.ssrn.com/abstract=2208240
6. Zeynep Tufekci. 2015. Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency. *J. on Telecomm. & High Tech. L.* 13 (2015), 203. http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/jtelhtel13&section=18
7. Jacob Goldstein Twitter. To Increase Productivity, UPS Monitors Drivers' Every Move. (????). http://www.npr.org/sections/money/2014/04/17/303770907/to-increase-productivity-ups-monitors-drivers-every-move