Interactive Active Learning for Self-Tracking in mHealth

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
Self-tracking algorithms for mobile health 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 discontinuation of use leaving those hoping to change their habits back where they started. 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 user through the use of multiple models and techniques that make use of the motivation, creativity, and intelligence of a dedicated community of users.

Author Keywords
Active Learning, Query By Committee, End-User Interactive Machine Learning, Self-tracking, Activity Recognition, Mobile Health

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous; See [http://acm.org/about/class/1998/]: for full list of ACM classifiers. This section is required.

Introduction

Self-tracking systems such as Strava, MapMyRun, Moves, and Fitbit have become a popular application of machine learning algorithms for activity recognition. They are increasingly being used by athletes looking to increase their performance, researchers studying human behavior in natural settings [3, 13], and as intervention technologies to better support mental health and well-being [8, 9]. However, for the average consumer of self-tracking devices, the user experience is considered opaque, making tracking inaccuracies difficult to understand [14, 6], a problem that ultimately leads to a breakdown in trust and discontinuation of use [12]. This is problematic for those hoping to apply the insight advertised by the developers of these systems as continued use is needed to affect long-term change in habits. The crux of this problem is a "black box" user experience in which the system passively collects data and presents the user with computational insight (e.g. information visualizations and recommendations) without notion of how it works or what the user can do to make their self-tracking experience more informative.

The design of current systems have several important user interaction challenges that need to be addressed. First, users need informative 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 based on this data, 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 a system errs can bias the user's process for making sensitive decisions concerning their health. Finally, these systems need to make the user feel confident in the accuracy of the metrics or suggestions that the system reports in order to mitigate discontinuation of use. One way in which accuracy is stifled is by using a universal model of activity which includes data from many users and has been proven to be less accurate than models calibrated by the individual end-user[7]. Additionally, there are invetable biases toward physical attributes, activities, and environments over-represented in the initial training data set of a universal model of activity that can lead to higher quality user experience for some and not others. Without feedback on the data used to train these models, it is difficult to understand how training data affects model performance.

To address these challenges, we must study novel ways of combining theories from human-computer interaction with those of machine learning. For example, there has been promising reseach in the application of *active learning*, a semi-supervised machine learning technique which makes judicious use of user interaction based on uncertainty, in web image search [5, 1]. While this research has allowed us to better understand how exposure and interaction with aspects of machine learners can aid the user in building functional mental models of the system, we need to conduct research that builds a better understanding of how machine learning should be applied so that users interacting with the system can effectively interact with and improve the performance of it. We believe that activity recognition and self-tracking are an appropriate domain for studying human-centered machine learning and propose the following :

1. studying how *feedback and interactivity can be designed for functional understanding of active learning-based mobile health systems by individual end-users*.
2. studying how a definition of uncertainty in active learning called ***query by committee*** *may enable individual end-users to explore multiple computational models in identifying system limitations and*

biases that occur at both the level of the individual end-user's personal model and the level of the global model. The *query by committee* technique can do this by presenting the user with estimations from models that vary in the samples used for the training data, the features used from the samples, or the algorithm which builds the model.

3. studying how *a human-centered machine learning approach can aid users in calibrating a self-tracking system to their lifestyle and environment as well as any changes to their lifestyle and environment that may occur.*

Background

This section provides the framing and background in the argument for active learning as an algorithm ripe for studying human-centered machine learning in the context of self-tracking and mobile health. The last section will build on these concepts to identify ways in which we might study them in the future.

Why Users Stop Tracking

When users buy or download physical or software products for self-tracking, they are typically under the impression that the activities they engage in will be tracked even when each user's expectation of a trackable activity varies wildly. During initial engagement this expectation is adjusted in attempt to make sense of the unexpected metrics displayed[12, 14, 6]. As the user continues to observe both expected and unexpected values in the information display of the product, they begin attempts to reproduce unexpected results, find the limitations of the system, and the nature of its errors. This is when the lack of transparency becomes most apparent and users begin losing trust in the system [14]. For those hoping to affect change in their own health behavior, this is particularly problematic. Changing health behavior is something that takes consistent attention over an extended period of time and is not easily affected by short periods of weak computational insight.

Universal, Personal, and Hybrid Models

When designing an activity recognition system that uses machine learning, it is important to balance the need for well-labeled training data with the immediate needs of the user. The most common approach is to use a *universal model* which combines data from many different users during an initial trial to train and evaluate a model of that activity. This allows the system to recognize an activity out-of-the-box without any need to calibrate it to an individual model of activity. Least common is the *personal model* which requires that the user actively label an activity before it can be recognized and tracked. Hybrid models leverage some form of model adaptation to strike a balance between the strengths of the universal model (robustness to variability, immediate use) and the strengths of the personal model (personalized calibration). Previous research has also shown that personal and hybrid models perform much better than universal models. If we can better understand how to design the process of calibrating and training a machine learning system to not feel interruptive, then it may become much easier to integrate these more powerful models into mHealth systems that use activity recognition. emphasize that better understanding of how to design for the training

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. Transparency through Active Learning
When a user decides to use wearable devices or self-tracking mobile applications to better understand their health and make informed decisions about their health, they need to be able to quickly identify the limits of such technology as it is often not apparent what is lost when adapting a medical technology for fashionability and 24/7 use. When an active learning system prompts the user for more information, the user has an opportunity to develop a mental model of scenarios and events which are perplexing to the learning system. This allows them to avoid making important health-sensitive decisions around information during the time in question.

Yang et al., suggest that self-tracking systems need to support testability, that is to support users' attempts at testing the system by providing test guidelines instead of leaving them to develop folk testing methods which yield little understanding of the system and its limitations. One approach that may help in developing design guidelines for such active learning prompts is to include known contextual information, such as time and place, about the uncertain event. This kind of transparency may allow the user to not only identify activities that are confusing, but also the places in which they commonly occur. For example, those who use repetitive hand gestures in their workplace such as carpenters using a hammer, may be inadvertently triggering a step detection algorithm. By identifying that the workplace is where this false positive attribution of a "step" to a physical activity occurs, the carpenter can begin ignoring "steps" during this time.

2. Query By Committee and Multiple Models
Previous research has shown how the use of multiple models on the same event can help users to identify the errors and limitations of a model as well as understand the relationship between the data, features, and algorithms used to analyze it [11]. To further understand how active learning can benefit users in making sense of the data, we propose that future research explore interaction designs which present the multiple models inherent in a committee and allow the user to explore these alternative views of the data used to infer the user's activity during a prompt for the ground truth label.

One way in which future research may allow us to explore ways of facilitating deeper engagement with self-tracking systems is to employ the concept of *cuts*, a subset of the data with some shared feature. Previous research explored the use of cuts in the domain of self-tracking as a way of reducing the complexity multi-faceted tracking systems such that the user can investigate correlations and identify relationships among varying aspects of the data collected [4]. For example, a user may explore a cut of the time in which they run by location or speed. The former cut may help them decide whether a different route may be more visually appealling during that time of day while the latter may help them understand the effect that time of day has on their performance.

Research in human-centered machine learning should focus on ways in which the cuts concept be applied to both the way in which we vary the models within an active learning committee and the design of the active learning prompt for labeled data when the committee disagrees. For example, if the models comprising the comittee are trained by different subsets of a sample stratified by a physical characteristic of the individuals included in the training set for a universal model (height, weight, age, gender, etc.), a disagreement within the committee could mean that some aspect of the selected attribute confounded the detection of an activity. Allowing the user to explore a cut of the attribute that was used to determine the committee by some attribute of the tracked data (time, location, heart rate, etc.) could expose biases and limitations inherent in the makeup of the universal training data. Additionally, by employing either a social computing technique such as collaborative filtering, or another end-user interactive machine learning technique such as the one used in ReGroup to create filters for different social groups[2], we might enable the users of the self-tracking system to correct biases in the universal training data that occur when certain characteristics are overrepresented, particularly when the overrepresentation results in an error-proned model and not one robust to the variations of the current population of end-users.

3. Holistic View of Self through Adaptable Self-Tracking
The kind of self-knowledge required for mobile health techniques based on serious behavioral intervention concepts like cognitive behavioral therapy must at least attempt to be holistic in order to be effective. By adopting a philosophy that developers should design self-tracking systems such that users can develop an accurate mental model of the underlying mechanics of the system allows us to begin designing for a more advanced or dedicated group of users who can learn to train a self-tracking system to detect new activities previously inconceivable by the development team. This kind of adaptable self-tracking has already been explored in a system called RecoFit which automatically detects and tracks repetitive activity without pre-conceived notions of which activities could be detected[10]. The popularity of products for self-tracking like FitBit, Jawbone Up, Moves, and Strava show that there is a dedicated community of users in the self-tracking domain. By researching novel ways in which the motivation, intelligence, and creativity of this community can be leveraged to train a wide variety of robust and accurate activity recognition models, we can find a path toward self-tracking systems that provide a holistic view of life instead of simply a step count or map of the places we have visited.

REFERENCES

1. Saleema Amershi, James Fogarty, Ashish Kapoor, and Desney Tan. 2011. Effective End-User Interaction with Machine Learning. (2011).
2. Saleema Amershi, James Fogarty, and Daniel Weld. 2012. Regroup: Interactive machine learning for on-demand group creation in social networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 21–30. http://dl.acm.org/citation.cfm?id=2207680
3. Nathan Eagle and Alex (Sandy) Pentland. 2005. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing* 10, 4 (Nov. 2005), 255–268. DOI: http://dx.doi.org/10.1007/s00779-005-0046-3
4. Daniel Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean Munson. 2014. Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In *Proceedings of the 2014 conference on Designing interactive systems*. ACM, 667–676. http://dl.acm.org/citation.cfm?id=2598558
5. 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

6. Daniel Harrison, Paul Marshall, Nadia Bianchi-Berthouze, and Jon Bird. 2015. Activity tracking: barriers, workarounds and customisation. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 617–621. http://dl.acm.org/citation.cfm?id=2805832
7. Jeffrey W. Lockhart and Gary M. Weiss. 2014. Limitations with activity recognition methodology & data sets. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp '14 Adjunct* (2014), 747–756. DOI: http://dx.doi.org/10.1145/2638728.2641306
8. Hong Lu, Andrew T Campbell, and Daniel Gatica-perez. 2012. StressSense : Detecting Stress in Unconstrained Acoustic Environments using Smartphones. *Ubicomp* (2012), 351–360.
9. David C Mohr, Stephen M Schueller, Enid Montague, Michelle Nicole Burns, and Parisa Rashidi. 2014. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. *Journal of medical Internet research* 16, 6 (Jan. 2014), e146–e146. DOI: http://dx.doi.org/10.2196/jmir.3077
10. Dan Morris, T. Scott Saponas, Andrew Guillory, and Ilya Kelner. 2014. RecoFit: using a wearable sensor to find, recognize, and count repetitive exercises. ACM Press, 3225–3234. DOI: http://dx.doi.org/10.1145/2556288.2557116
11. Kayur Patel, Steven M. Drucker, James Fogarty, Ashish Kapoor, and Desney S. Tan. 2011. Using multiple models to understand data. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, Vol. 22. Citeseer, 1723. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.208.3338&rep=rep1&type=pdf
12. Patrick C. Shih, Kyungsik Han, Erika Shehan Poole, Mary Beth Rosson, and John M. Carroll. 2015. Use and Adoption Challenges of Wearable Activity Trackers. 1 (2015), 1–12. https://www.ideals.illinois.edu/handle/2142/73649
13. Rui Wang, Gabriella Harari, Peilin Hao, Xia Zhou, and Andrew T. Campbell. 2015. SmartGPA : How Smartphones Can Assess and Predict Academic Performance of College Students. *ACM Conference on Ubiquitous Computing* UbiComp (2015), 7–11.
14. Rayoung Yang, Eunice Shin, Mark W. Newman, and Mark S. Ackerman. 2015. When fitness trackers don't'fit': end-user difficulties in the assessment of personal tracking device accuracy. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 623–634. http://dl.acm.org/citation.cfm?id=2804269