Modeling Users of Intelligent Systems

Stephanie Rosenthal
Carnegie Mellon University
Computer Science Department
Pittsburgh, PA 15213 USA
srosenth@cs.cmu.edu

Abstract
While many devices today increasingly have the ability to predict human activities, it is still difficult to build accurate personalized machine learning models. As users today will become responsible for helping to train their own models, we are interested in ways for applications to request labeled data from their users in a non-invasive way. This work focuses on opportunities for intelligent systems to ask their users for help through interactions over an extended period of time in order to improve their machine learning models. We focus on trading off the expected increase in accuracy with the potential interruptions that the questions may cause to improve the usability of such systems.

Keywords
experience sampling, asking for help, machine learning, long-term interactions

CHI 2011, May 7-12, 2011, Vancouver, British Columbia, USA.
ACM 978-1-4503-0268-5/11/05.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.
I.2.m Artificial Intelligence Miscellaneous

General Terms
Algorithms, Human Factors

Research Situation
I am a fourth year PhD student studying computer science at Carnegie Mellon University and am advised by Anind Dey (HCI) and Manuela Veloso (Artificial Intelligence/Robotics). I proposed my thesis in November 2010 and plan to graduate 18 months later. My thesis work focuses on the costs and benefits of interrupting humans during mixed-initiative interactions with intelligent systems that we use everyday such as mobile phones and robots. In particular, I am interested in opportunities for intelligent systems to learn to improve their machine learning models by asking for help from humans and, as a result, improve the usability of applications that use these systems.

Motivation and Related Work
Mobile phones and other devices we use today increasingly have the ability to predict human activities, autonomously keep track of a variety of statistics, and produce novel interaction techniques. HCI researchers are increasingly using these devices and machine

learning classifiers for a variety of purposes to improve usability. However, they often need to be calibrated to maximize the accuracy for each particular user to get the full benefit from the applications. While some machine learning techniques have been proposed to implicitly collect and learn from users [3], this is often hard to do in many domains without an understanding of all possible contexts of use. Instead, users must train their own machine learning classifiers by providing labeled data to learn from (*e.g.,* [8]).

Currently, there are two main ways to collect this data: user-initiated and device-initiated. As compliance can be difficult when users must enter their own data on their own initiative, we will focus on device-initiated experience sampling to collect data [7]. While much work in both machine learning and HCI has offered algorithms to determine when to ask for this training data, they have focused on the timeliness of questions to benefit the classifier accuracy without much regard to the user's context or interruptibility [1][2].

Horvitz suggests that it is unlikely that users will accept mixed-initiative systems including those that employ these collection techniques, if the systems interrupt the users inappropriately and ask possibly many questions [5]. Instead, they should infer the ideal times to ask users based on the uncertainties of the current machine learning models *and* the context-dependent costs of interruption to the user (*e.g.,* [6]).

Because users use a single mobile phone over a long period of time, there are opportunities to learn and crease personalized models of the interruptibility of users [4]. The model and its output can then be used by other applications on the phone (or elsewhere) to learn other classifiers in a usable way – allowing applications to reason about the interruptiblity of the user to determine when it is appropriate to ask for data labels while also maximizing the potential accuracy of the classifiers.

Statement of Thesis

My thesis contributes to the understanding of when it is timely to request this training data for machine learning as the user uses the device, based on both classifier needs for accuracy and the interruptibility of the user. In particular, this thesis will address the following research questions towards building these models:

1) Are users willing to help teach their mobile devices about their interruptibility and other classifiers while using these devices?
2) Can we learn accurate models of interruptibility on a mobile phone?
3) How can we combine user-dependent costs associated with interruption along with the expectations for accuracy to determine when to query a user for labeled data?
4) How can we extend these models to other domains – those that include multiple users or require other context-dependent costs?

Research Goals and Methods

This thesis will contribute models of humans for uses in intelligent systems, and algorithms to take into account these models when determining when to ask for help. Towards these goals, I will evaluate the work both theoretically and empirically.

In order to model humans, we must understand the theoretical constraints that humans have (here we consider interruptibility, accuracy, and patience) and how those affect human responses to questions. We know that availability and patience both limit the number of questions a person might answer, but that patience has to do with the number of times that the device asked, while availability is the probability of the device getting a response at all. Then, we can assign costs to asking questions to account for patience and probabilities for the availability of the user. This thesis will not only theoretically model these values but also calculate them empirically through user studies to ensure that the models are grounded in reality.

Then, using these models (both theoretical and empirical), I will introduce algorithms for the device to determine when to ask for help. These theoretical algorithms will trade off the costs of asking with the benefits of model accuracy or efficiency. Algorithms for a single user will determine whether to ask at a particular time, while algorithms for multiple users will not only determine whether to ask but also who to ask in order to minimize the expected costs. We will test the algorithms empirically against other experience sampling algorithms on real users participating in tasks to show that the new algorithms can maintain or increase accuracy or efficiency while increasing usability by decreasing annoyance with the device.

Dissertation Status

I have already accomplished several steps towards my overall thesis goal of devices that request help in a usable way. I have recently developed a mobile phone application that proactively learns users' interruption preferences to autonomously determine whether the phone should ring or be in silent mode. The underlying algorithm trades user-defined costs of answering questions with costs of misclassifying the phone state to determine when to ask for users' phone mode preferences over two weeks of training time. Then it uses the learned model to actually change the phone volume. We show that users are willing to answer the application’s questions over a prolonged period of time in order to receive the benefits of the automatic modification of the phone’s ring volume. Users were satisfied with our phone volume application in terms of its accuracy over two weeks of testing and were not as annoyed by the questions compared to other asking algorithms.

I have also defined symbiotic relationships to model the states, actions, and interactions of a mobile robot that performs services for humans but also requires help from them, and introduced the visitor companion task as an example of the relationship in which a mobile robot accompanies a single visitor to meetings. Here the robot benefits the visitor by performing these tasks and the visitor is asked for localization help so that the robot can perform more efficiently. The mobile robot is similar to the mobile phone in this task in that there is a single user that the robot can learn from over long-term interactions. I have shown that visitors are willing to help the robot in return for the robot’s services. While the robot did not need to learn the location classification model from scratch, I have shown that, with accurate localization help from users, the robot can localize faster and more accurately and navigate faster to its location compared to autonomous navigation and localization.

In the future, I will extend my model of humans to include multiple humans in three ways: modeling the availability and locations of the humans in the environment, including time so that the model can delay asking if no one is available, and taking into account patience of the users by keeping a history of questions asked. I will incorporate the new human model into my algorithm for balancing the costs of interruption with the benefits of accuracy that I demonstrated in the phone volume learning application. In the robot domain, this will include decisions about which person to ask for help and planning paths to proactively navigate towards those people during the robot's task. I will evaluate these new additions with a long-term study of our robot in the environment to show its effectiveness at employing the models and algorithms to maintain high performance and perceived usability. I expect this remaining work to take 14-18 months to complete.

Expected Contributions

This thesis will contribute principles for intelligent devices to proactively request and receive help from humans, while performing tasks of extended interaction over time. This thesis will show that users are willing to help intelligent systems in order to prove their classification models. We will model humans in the environment, in particular their interruptibility, time and patience. We will then introduce new autonomous algorithms that effectively use these models to balance the costs of asking for help with the benefits of improved accuracy. Future applications will be able to use these models and algorithms to learn new classifiers to improve task performance and maintain usability while asking for help over time. With these capabilities to learn accurate models in a usable way, applications can be deployed more quickly and with less training than is currently possible.

My thesis incorporates the knowledge of HCI in terms of interruption and human modeling with the machine learning needs to train accurate classifiers. At the Doctoral Consortium, I expect to receive feedback on my models of humans to ensure that they are grounded in reality and accurately capture difficulties that we have in creating accurate machine learning classifiers. I can contribute my knowledge of both machine learning and HCI to provide feedback about opportunities to incorporate machine learning into participants' research and about recent work that might be relevant to improve participants' current models.

References

[1] D. Cohn, L. Atlas and R. Ladner. Improving Generalization with Active Learning. *Machine Learning*. 0885-6125:15(2), 201-221, 1994.

[2] P. Donmez and J. G. Carbonell. Proactive learning: cost-sensitive active learning with multiple imperfect oracles. *Proc of CIKM '08*, 619-628, 2008.

[3] A. Faulring, *et al*. Successful User Interfaces for RADAR. *CHI 2008 Workshop on Usable Artificial Intelligence*, 2008.

[4] J. Fogarty, *et al*. Predicting human interruptibility with sensors. *ACM ToCHI* 12(1), pp. 119-146, 2005.

[5] E. Horvitz. Principles of mixed-initiative user interfaces. *Proc. of CHI '99,* 159-166, 1999.

[6] Kapoor, A. and Horvitz, E. Experience sampling for building predictive user models: a comparative study. In Proc. CHI 2008, 657–666, 2008.

[7] R. Larson and M. Csikszentmihalyi. The experience sampling method. *New Directions for Methodology of Social and Behavioral Science*, 15, 41-56, 1983.

[8] S. Stumpf, *et al*. Toward Harnessing User Feedback for Machine Learning. *Proc. IUI 2007*, 82 – 91, 2007.