End-user Interactions with Intelligent and Autonomous Systems

Simone Stumpf
Centre for HCI Design
School of Informatics
City University London
Northampton Square
London EC1 0HB
Simone.Stumpf.1@city.ac.uk

Weng-Keen Wong
School of Electrical Engineering
and Computer Science
Oregon State University
Corvallis, OR 97331
wong@eecs.oregonstate.edu

Margaret Burnett
School of Electrical Engineering
and Computer Science
Oregon State University
Corvallis, OR 97331
burnett@eecs.oregonstate.edu

Volkmar Pipek
Institut für Wirtschaftsinformatik
Universität Siegen
Hölderlinstr. 3
57068 Siegen
volkmar.pipek@uni-siegen.de

CHI'12, May 5–10, 2012, Austin, Texas, USA.
ACM 978-1-4503-1016-1/12/05.

Abstract
Systems that learn from or personalize themselves to users are quickly becoming mainstream yet interaction with these systems is limited and often uninformative for the end user. This workshop focuses on approaches and challenges to explore making these systems transparent, controllable and ultimately trustworthy to end users. The aims of the workshop are to help establish connections among researchers and industrial practitioners using real-world problems as catalysts to facilitate the exchange of approaches, solutions, and ideas about how to better support end users.

Keywords
End users; machine learning; personalization; intelligent assistants; autonomous systems

ACM Classification Keywords
H.5.m [Information interfaces and presentation]: Miscellaneous

General Terms
Human Factors

Introduction
Systems that learn from or personalize themselves to end users, such as email inbox filters, gesture

recognition systems, network device alarms, smart home applications, music recommender, healthcare decision or fraud detection systems, are quickly becoming mainstream. Consumers and business specialists now often interact with systems on a daily basis in a form of "human-in-the loop" learning (e.g., [8, 1]). Yet, interacting with even well-designed systems is limited and often uninformative for the end user because of the internal complexity and current "black box" nature of most intelligent systems.

This workshop focuses on approaches and challenges in this area. The aims of the workshop are to help establish connections among researchers and industrial practitioners bringing real-world problems to the table, and to facilitate the exchange of approaches, solutions, and ideas about how to better support end users of intelligent systems.

Interaction Challenges for End Users

Traditionally, researchers have investigated end-user interactions with intelligent and autonomous systems to inform Artificial Intelligence (AI) or Intelligent User Interfaces (IUI). Only recently has a HCI stance emerged with end users as the main focus. Three example areas of research are:

Transparency

Making the system's decisions and behavior understandable by end users is a first step in successful interactions, and we look forward to exploring new approaches during the workshop. Recent attempts have moved beyond explaining rule-based systems [25] toward more complex algorithms [20, 26]. Examples of explanations for specific decisions include *why…* and *why not…* descriptions [14, 16], visual depictions of the assistant's known correct predictions versus its known failures [29], confidence of the system in making predictions [13, 19], and electronic "door tags" displaying predictions of worker interruptibility with the reasons (e.g., "talking detected" [31]). Recent work by Lim and Dey has resulted in a toolkit for applications to generate explanations for popular machine learning systems [17]. However, some complex systems' understandability by end users, for example those employing neural networks [3], has received scant attention. As collaborative approaches (e.g. Collective Intelligence) receive more attention, transparency of systems based on interactions of multiple users may be necessary [21].

Control

Recent research has made inroads into supporting end user interactions to directly shape a system's decisions or reasoning via *programming by demonstration* [5, 15] or via input of more training examples [2, 7]. A different way is to incorporate user interactions that can cause a deeper change in the underlying algorithms. These approaches include imposing user constraints [4, 23, 12], similarity metric learning from rich user feedback [9], clustering documents using a set of user-selected representative words for each class [18], and allowing the user to directly build a decision tree for the data set with the help of visualization techniques [32]. Several methods have also allowed end users to modify features that the algorithm uses, for example by reweighting them [12, 27, 28] or labeling them [22, 33, 6]. The workshop will explore the range of interaction approaches for end users to shape learning and personalization.

User Experience

Traditionally, system "accuracy" underpins evaluation methods for intelligent and autonomous systems. However, systems that personalize themselves suggest the need for additional measures, such as trust and satisfaction, to comprehensively assess the user experience of interacting with these systems. There has been some initial work to investigate what makes these kinds of system trustworthy [10]. It has also been found that making the reasoning transparent can improve perceptions of satisfaction and reliability toward music recommendations [24], as well as other types of recommender systems [11, 30]. However, experienced users' satisfaction with the system may actually *decrease* [19]. The workshop will discuss measures for evaluating user interactions with intelligent and autonomous systems.

Workshop Goals

This workshop has three primary goals: (1) to map out the space of current work and to exchange information about current approaches among researchers already working in this area; (2) to generally share information among researchers already working in this area with HCI researchers new to the area, and with practitioners interested in current and future techniques that can be embodied in their systems; and (3) to match real-world end-user problems and applications with potential solutions or approaches drawn from new and emerging research findings.

References

[1] Amershi, S., Fogarty, J., Kapoor, A. and Tan, D. Examining Multiple Potential Models in End-User Interactive Concept Learning. *Proc. CHI*, ACM (2010), 1357-1360.

[2] Cohn, D.A., Ghahramani, Z. and Jordan, M. I. Active learning with statistical models. *J. Artificial Intelligence Research* 4 (1996), 129-145.

[3] Craven, M.W. Extracting comprehensible models from trained neural networks. Ph.D. thesis. School of Computer Science, University of Wisconsin, Madison, WI (1996).

[4] Culotta, A., Kristjansson, T., McCallum, A. and Viola, P. Corrective feedback and persistent learning for information extraction, *Artificial Intelligence* 170 (2006), 1101-1122.

[5] Cypher, A. (ed.) *Watch What I Do: Programming by Demonstration*. MIT Press, Cambridge, MA, 1993.

[6] Druck, G., Mann, G. and McCallum A. Learning from labeled features using generalized expectation criteria. *Proc. SIGIR*, ACM (2008).

[7] Fails, J.A. and Olsen, D.R. Interactive machine learning. *Proc. IUI*, ACM (2003), 39-45.

[8] Fiebrink, R., Cook, P. and Trueman, D. Human model evaluation in interactive supervised learning. *Proc. CHI*, ACM (2011), 147-156.

[9] Fogarty, J., Tan, D., Kapoor, A. and Winder, S. CueFlik: Interactive concept learning in image search. *Proc. CHI* ACM, (2008), 29-38.

[10] Glass, A., McGuinness, D. and Wolverton, M. Toward establishing trust in adaptive agents. *Proc. IUI*, ACM (2008), 227-236.

[11] Herlocker, J., Konstan, J. and Riedl, J. Explaining collaborative filtering recommendations. *Proc. CSCW*, ACM (2000), 241-250.

[12] Kapoor, A., Lee, B., Tan, D. and Horvitz, E. Interactive optimization for steering machine classification. *Proc. CHI*, ACM (2010) 1343-1352.

[13] Kulesza, T., Stumpf, S., Burnett, M., Wong, W.-K., Riche, Y., Moore, T., Oberst, I., Shinsel, A. and McIntosh, K. Explanatory debugging: Supporting end-user debugging of machine-learned programs, *Proc. VL/HCC*, IEEE (2010), 41-48.

[14] Kulesza, T., Wong, W.-K., Stumpf, S., Perona, S., White, R., Burnett, M., Oberst, I. and Ko, A. J. Fixing the program my computer learned: Barriers for end users, barriers for the machine. *Proc. IUI*, ACM (2009), 187-196.

[15] Lieberman, H., (ed.) *Your Wish is My Command: Programming By Example*. Morgan Kaufmann, 2001.

[16] Lim, B. Y., Dey, A. K. and Avrahami, D. Why and why not explanations improve the intelligibility of context-aware intelligent systems. *Proc. CHI*, ACM (2009), 2119-2128.

[17] Lim, B. Y. and Dey, A. K. Toolkit to support intelligibility in context-aware applications. *Proc. UbiComp*, ACM (2010), 13-22.

[18] Liu, B., Li, X., Lee, W. and Yu, P. Text classification by labeling words. *Proc. AAAI*, AAAI (2004).

[19] McNee, S. M., Lam, S. K., Guetzlaff, C., Konstan, J. A. and Riedl, J. Confidence displays and training in recommender systems. *Proc. INTERACT*, IFIP (2003), 176-183.

[20] Pazzani MJ. Representation of electronic mail filtering profiles: a user study. *Proc. IUI,* ACM (2000). 202–206.

[21] Pipek, V., and Wulf, V. Infrastructuring: Towards an Integrated Perspective on the Design and Use of Information Technology. *JAIS 10,* 5 (2009), 447-473.

[22] Raghavan, H. and Allan, J. An interactive algorithm for asking and incorporating feature feedback into support vector machines. *Proc. SIGIR*, ACM (2007), 79-86.

[23] Shilman, M., Tan, D. and Simard, P. CueTIP: A mixed-initiative interface for correcting handwriting errors. *Proc. UIST*, ACM (2006), 323-332.

[24] Sinha, R. R. and Swearingen, K. The role of transparency in recommender systems. *Ext. Abstracts CHI*, ACM (2002), 830-831.

[25] Sotos JG. MYCIN and NEOMYCIN: Two approaches to generating explanations in rule-based expert systems. *Aviation, space, and environmental medicine* 61, 10 (1990), 950–954.

[26] Stumpf, S., Rajaram, V., Li, L., Burnett, M., Dietterich, T., Sullivan, E., Drummond, R. and Herlocker, J. Toward harnessing user feedback for machine learning. *Proc. IUI*, ACM (2007), 82-91.

[27] Stumpf, S., Sullivan, E., Fitzhenry, E., Oberst, I., Wong, W.-K. and Burnett, M. Integrating rich user feedback into intelligent user interfaces. *Proc. IUI*, ACM (2008), 50-59.

[28] Stumpf, S., Rajaram, V., Li, L., Wong, W.-K., Burnett, M., Dietterich, T., Sullivan, E. and Herlocker J. Interacting meaningfully with machine learning systems: Three experiments. *Int. J. Human-Computer Studies* 67, 8 (2009), 639-662.

[29] Talbot, J., Lee, B., Tan, D. and Kapoor, A. EnsembleMatrix: Interactive visualization to support machine learning with multiple classifiers. *Proc. CHI*, ACM (2009), 1283-1292.

[30] Tintarev, N. and Masthoff, J. Effective explanations of recommendations: user-centered design. *Proc. Recommender Systems*, ACM (2007), 153-156.

[31] Tullio, J., Dey, A., Chalecki, J. and Fogarty, J. How it works: a field study of non-technical users interacting with an intelligent system. *Proc. CHI*, ACM (2007), 31-40.

[32] Ware, M., Frank, E., Holmes, G., Hall, M. and Witten, I. H. Interactive machine learning: Letting users build classifiers. *Int. J. Human-Computer Studies* 55, (2001), 281-292.

[33] Wong, W.-K., Oberst, I., Das, S., Moore, T., Stumpf, S., McIntosh, K., and Burnett, M. End-user feature labeling: A locally-weighted regression approach. *Proc. IUI*, ACM (2011).