HCI is Different: We Need Something Else

James Fogarty
Computer Science & Engineering
DUB Group
University of Washington
Seattle, WA 98195 USA
jfogarty@cs.washington.edu

CHI 2009, April 4 – April 9, 2009, Boston, MA, USA
ACM 978-1-60558-247-4/08/04.

Abstract
The notion of a repository of shared home behavior datasets at first seems quite attractive. After all, the collection of even relatively small field datasets requires significant investment. But my experiences with five relatively interesting datasets have led me to believe that HCI is different, that we need something else. I suggest the possibility that what we need might be support for collaborating before and during the collection of datasets, as opposed to support for sharing datasets after they have been collected. The idea is fraught with problems, but I put it forward for the sake of discussion of a different approach.

Keywords
Shared datasets, human-computer interaction

Introduction
Interest in repositories of shared datasets stems in part from the success of such repositories in other fields. In statistical machine learning, for example, the UC Irvine Machine Learning Repository is an undeniable success. That repository has shaped how research is conducted, as the default approach to evaluating new algorithms has become to download several datasets from the repository and to compare the performance of the researcher's algorithm to the performance of other algorithms. If a researcher is claiming a general advance in statistical machine learning, it is

unacceptable to demonstrate that advance on only a single dataset, in part because so many datasets are readily available in the repository.

My experiences with five different datasets have led me to believe that HCI is different, and that we need something else. This position paper briefly reviews the datasets we have collected, then discusses the challenges in gaining new value from sharing them, then discusses the possibility that what we need might be support for collaborating before and during the collection of new datasets, as opposed to support for sharing datasets after they have been collected.

Our Datasets

For convenience, I will refer to the datasets we have collected as the *interruptibility video* dataset, the *interruptibility sensors* dataset, the *home pipes* dataset, the *programmer short video* dataset, and the *programmer long video* dataset.

Interruptibility Video

Collected as a part of our CHI 2003 and TOCHI 2005 work on sensor-based statistical models of human interruptibility [2, 5], this dataset contains a total of 602 hours of video recordings of four office workers in their normal work environments. Of that, 56 hours have been manually coded to indicate the occurrence of approximately 25 different events. The office workers were prompted at random intervals to self-report an estimate of their interruptibility, and our research focused on modeling these self-reports based on the manually-coded environmental cues.

Interruptibility Sensors

Collected in our CHI 2004 work on sensor-based statistical models of human interruptibility [3], this dataset contains approximately 100 interruptibility self-reports for each of 10 office workers in their normal work environments. We instrumented their environments with software monitoring their computer, a microphone to detect nearby speech, motion detectors, and contact switches to detect whether the door was open, cracked, or closed and whether their phone was off its hook.

Home Pipes

Collected in our UIST 2006 work on unobtrusive home activity sensing [1], this dataset contains six weeks of water usage data in a home shared by two adults. The data was collected using microphones pressed against the outside of water pipes, including pipes used to model activity (cold water, hot water, waste pipes) as well as ground truth sensors to directly capture the activities we were interested in modeling (toilets, showers, sinks, dishwashers, etc.). The dataset was manually inspected and annotated to indicate when each activity of interest occurred.

Programmer Short Video

Collected in our CHI 2005 work on models of programmer interruptibility [4], this dataset examines 20 programmers each working for 60 minutes to solve 5 tasks. I refer to it as the short dataset because each task lasted a relatively short period of time. At random intervals, participants were interrupted by a secondary task. We collected full screen captures of their desktops and synchronized event logs from the Eclipse integrated development environment.

Programmer Long Video

Collected in our CHI 2008 and AAAI 2008 work on the obstacles that software developers encounter in adopting statistical machine learning [6, 7], this dataset examines 10 programmers applying statistical machine learning algorithms and techniques to a task over the course of a 5 hour session. I refer to it as the long dataset because they worked on the same task for the entire session. We collected full screen captures of their desktops, audio/video recordings of their physical workspace, and frequent automatic snapshots of the files in their virtual workspace (allowing, for example, automated analyses measuring how well the programmer's system performed as it matured throughout the session).

The Challenge of Sharing

I believe the core challenge to gaining new value from sharing these types of datasets stems from a need to decide *what do we want to know?* In all of our interruptibility datasets, we wanted to know whether we could model a person's interruptibility using environmental cues. In our *home pipes* dataset, we wanted to know if we could model activity within the home using only the infrastructure sensors. In the *programmer long video* dataset, we wanted to examine the challenges that the programmers encountered.

Unlike a significant portion of machine learning research, where a dataset taken from the UC Irvine Machine Learning Repository can often be easily exchanged for another without greatly impacting the validity of the evaluation of a new machine learning algorithm, datasets in HCI are often carefully crafted to address specific questions. When attempting to gain new value by sharing an existing dataset, it is often the case that the careful crafting for the original intent of the data collection leaves the dataset of little or no value for answering a different question.

Part of this challenge is because of how we in the HCI research community define an interesting research contribution. If our work has already demonstrated the modeling of programmer interruptibility with an accuracy of 79.5% [4], the community will likely not be interested by another researcher using the same dataset to demonstrate an accuracy of 83.4%. Such an improvement would likely result from a relatively minor algorithmic difference of relatively little interest to the HCI research community. Instead, the HCI research community would likely be much more interested in follow-on work that takes a different perspective on the problem conception (a different *what do we want to know?*) and collects a new dataset addressing this different perspective.

Even when new research asks a question that might be answerable using data collected in prior work, there is often a need for some piece of information not collected in the prior work (because it was not part of *what do we want to know?* in that prior work). Here I have found the possibility of new value often hinges on video. In datasets without video (*interruptibility sensors*, *home pipes*), it may be impossible to obtain that new information. In datasets with large amounts of video (*interruptibility video*, *programmer long video*), the task of analyzing the video for the new information (with an uncertain likelihood of success) often does not compare well versus the option of designing a new data collection. Interestingly, the *programmer short video* dataset has been used in several publications beyond that for which it was originally collected. I think the video in this dataset managed to hit an interesting spot, where there is enough video to be interesting but not so much that it becomes unwieldy.

Collaborating Before and During Collection

Given the perspective that HCI research often requires deciding *what do we want to know?* before collecting an appropriate dataset, I suggest that what we need might be a venue for encouraging collaboration as people are preparing to collect new datasets.

Would you be willing to include additional prompts in an experience sampling study if you knew you were already collecting the sensor data needed to answer a question being asked by a different research group?

Would you be willing to add as a co-author, on your paper, the faculty member who oversaw the addition of your questions to their data collection? The graduate student who did the actual data collection? The undergraduate student who did the data transcription?

Would you be willing to advertise research questions that you want to study? Datasets you are planning to collect? Datasets you would like to know if others are considering collecting?

Given the often competitive nature of research, I think the answer to many of these questions is often *no*, for myself and for many other researchers. But I also think HCI as a field values research contributions that are difficult to convincingly answer using data originally collected for some other purpose. Supporting appropriate sharing of collected datasets has value, but we as a field also need to examine other strategies that may be better suited to the research questions we ask.

Acknowledgements

Thank you to all of my collaborators in the projects discussed here. This work was supported in part by the National Science Foundation under grant IIS-0812590.

References

1. Fogarty, J., Au, C. and Hudson, S.E. (2006). Sensing from the Basement: A Feasibility Study of Unobtrusive and Low-Cost Home Activity Recognition. *Proceedings of the ACM Symposium on User Interface Software and Technology* (UIST 2006), 91-100.
2. Fogarty, J., Hudson, S.E., Atkeson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C. and Yang, J. (2005). Predicting Human Interruptibility with Sensors. *ACM Transactions on Computer-Human Interaction* (TOCHI), **12**(1). 119-146.
3. Fogarty, J., Hudson, S.E. and Lai, J. (2004). Examining the Robustness of Sensor-Based Statistical Models of Human Interruptibility. *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI 2004), 207-214.
4. Fogarty, J., Ko, A.J., Aung, H.H., Golden, E., Tang, K.P. and Hudson, S.E. (2005). Examining Task Engagement in Sensor-Based Statistical Models of Human Interruptibility. *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI 2005), 331-340.
5. Hudson, S.E., Fogarty, J., Atkeson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C. and Yang, J. (2003). Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI 2003), 257-264.
6. Patel, K., Fogarty, J., Landay, J.A. and Harrison, B. (2008). Examining Difficulties Software Developers Encounter in the Adoption of Statistical Machine Learning. *Proceedings of the AAAI Conference on Artificial Intelligence* (AAAI 2008), Nectar Track, 1563-1566.
7. Patel, K., Fogarty, J., Landay, J.A. and Harrison, B. (2008). Investigating Statistical Machine Learning as a Tool for Software Development. *Proceedings of the ACM Conference on Human Factors in Computing Systems* (CHI 2008), 667-676.