End-User Machine Learning in Music Composition and Performance

Rebecca Fiebrink
Department of Computer Science (also Music)
Princeton University
fiebrink@princeton.edu

Dan Trueman
Department of Music
Princeton University
dtrueman@princeton.edu

ABSTRACT
We discuss our work creating the Wekinator software for end-user interactive machine learning, and we outline five key findings pertaining to our observations of its use in music composition and performance.

Keywords
Interactive machine learning, music, creativity, embodied design.

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces—Interaction Styles; I.2.6 Artificial Intelligence: Learning.

INTRODUCTION

Over the last several years, we have been investigating the use of interactive machine learning for semantic audio analysis [8] and new music composition and performance [3][6][7][14]. A central goal of this work is to enable students, composers, artists, and hobbyists to apply machine learning to their work—i.e., to enable end-users to work more effectively with intelligent systems. We begin this position paper by briefly describing the Wekinator, our software toolkit for end-user interactive machine learning in music and other real-time domains. We then highlight five key findings from our observations of students and composers applying the Wekinator to their work. These findings provide exciting examples of how end-user machine learning can be useful in the real world, and they underscore the fact that that research pertaining to end-user interactions with intelligent systems must consider a broad scope of potential applications and human-computer interactions. We conclude by outlining our current research questions and summarizing our goals for the workshop.

THE WEKINATOR

In 2008, we created the first version of the Wekinator software[1] [3][7], which allows end users to apply supervised learning to create custom, interactive, real-time systems. For example, the Wekinator can be used to build models that recognize a musician's instrumental playing techniques from video or sensor data (e.g., [14]), to create new musical instruments in which a performer's gestures are mapped in real-time to the control of sound synthesis parameters (e.g., [19]) or to construct other novel interactive systems in which people control audio synthesis, animations, video games, or other real-time processes through movement, sound, or other actions.

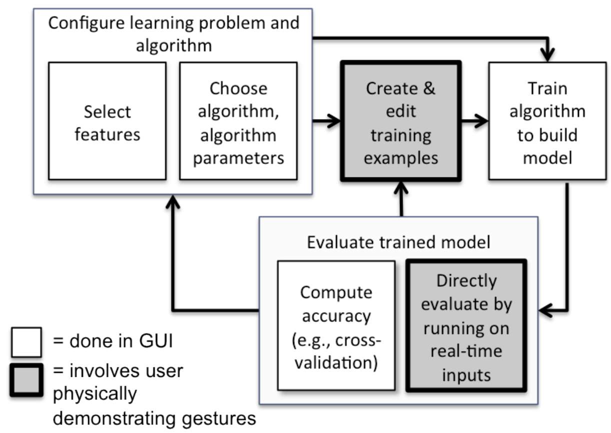

Figure 1. The interactive machine learning workflow supported by the Wekinator. (From [5].)

The Wekinator provides a set of graphical interfaces for users to iteratively create training examples through real-time demonstration, apply supervised learning algorithms to create trained models from those examples, evaluate those models by cross-validation or by directly applying them to real-time inputs (e.g., new gestures), and modify models through changes to the features, learning algorithm, and training data. This interactive workflow is illustrated in Figure 1.

Since its creation, the Wekinator has been used by numerous composers and artists—both students and professionals—to create publically-performed musical compositions and interactive installations,[2] such as the one shown in Figure 2. We have also used the Wekinator as a research platform to study composers' relationships to

[1] Downloadable at http://code.google.com/p/wekinator/

[2] See http://www.cs.princeton.edu/~fiebrink/thesis/resources.html for links to audio and video and [3] for discussion of these works.

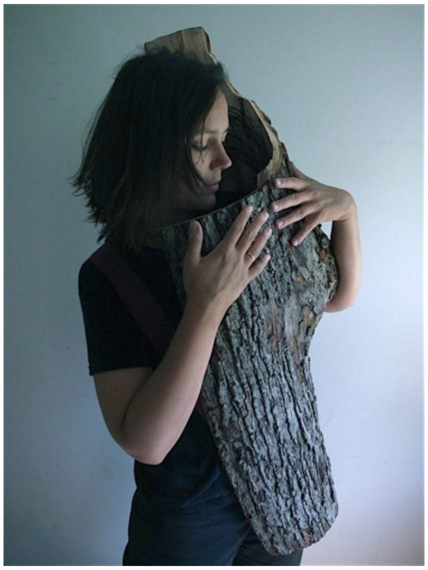

Figure 2. Michelle Nagai with her instrument, the MARtLET. Light sensors embedded in the tree bark drive sound synthesis parameters through the use of Wekinator-created models.

technology in the instrument-building process [6], to investigate new types of interactions for end-user machine learning [4], and to learn about the opportunities and challenges that interactive machine learning presents to end-users [3][5][6][14][15].

One of the questions that sparked our interest at the inception of the Wekinator project—and a question that still interests us today—regards how to effectively insert human interaction into machine learning algorithms. This question has also received considerable attention in recent projects such as CueFlik [1][9], EnsembleMatrix [18], and ManiMatrix [10], not to mention the foundational Crayons system [2]. Additionally, we are fundamentally interested in how to make the machine learning process understandable, predictable, debuggable, and usable by end users who may not possess expertise in machine learning or programming. These questions have begun to be rigorously addressed in other work in the community, such as that by Stumpf et al. [17]; these questions are also closely related to certain research on end-user programming (see, e.g., [12]).

HIGHLIGHTED FINDINGS

As outlined above, the aims of our work share significant commonalities with other research on end-user machine learning and end-user interaction with intelligent systems. At the same time, our experience deploying end-user machine learning software in artistic and creative domains and our observations of users applying the Wekinator to build real systems for professional use have revealed some new ways of understanding the potential applications and challenges of end-user machine learning. Here, we highlight several key findings that have arisen out of these more unique aspects of our work.

Model Evaluation

In [5], we discussed a subset of these findings pertaining to human model evaluation in interactive machine learning. Among these findings:

1. **When building models (e.g., gesture classifiers), users have a diverse array of goals for model behavior that extends far beyond model accuracy.** For example, users may care about misclassification costs or decision boundary shapes and locations. The relative importance of such criteria may vary appreciably across users and applications. In any case, the "conventional" goal of supervised learning—to build a model that generalizes most accurately from the training data—may be somewhat mismatched with users' real goals for their models.

2. **Application of interactive machine learning does not necessarily just entail the training of a learning algorithm; it can also entail the "training" of the user in unexpected—and possibly advantageous—ways**. For example, in under an hour using the Wekinator, some undergraduate students reported discovering that minimizing the noise in their training examples and balancing the number of training examples in each class led to better models. A cellist using the Wekinator for bow gesture classification reported that the Wekinator spurred her to improve her articulation technique: when she failed to create a classifier capable of distinguishing certain articulations, she realized that she was failing to execute those articulations correctly. In general, iterative model-building with the Wekinator results in many (if not all) users learning something about what may be accurately and/or quickly learnable by the learning algorithms, given their chosen gestural inputs or audio features.

3. **Users can employ their knowledge about what is accurately or quickly learnable in changing the learning concept they choose to teach the system.** Sometimes we observed these changes come about when a user who had unrealistic expectations for what machine learning could accomplish became convinced to temper his goals; other times, a user ended up making learning problems even harder once she discovered that a harder (and more useful) concept might be learnable. In any case, interactive, end-user machine learning can put the **user** in charge of managing inevitable, critical tradeoffs between concept complexity, model accuracy or quality, and model usefulness in the intended application context.

Creativity and Embodiment

Our observations of professional composers and students also suggest that interactive machine learning can play a critical role supporting creative, embodied work. We discuss these two findings below:

1. **Interactive machine learning can support creativity**. Researchers such as Shneiderman [16] and Resnick et al. [13] have proposed general principles for designing technology for use in creative work. For example according to Resnick et al. [13], creativity support tools should support exploration, discovery, and sketching; support diverse users and applications; and operate seamlessly with users' other tools. In light of such guidelines, interactive machine learning has marvelous potential as a creativity support tool, and the composers we have worked with have indeed valued the Wekinator for precisely these reasons. For example, composers can "sketch out the boundaries" of a new instrument by demonstrating a few gestures of interest paired with a few sounds of interest. Then, they can rely on the Wekinator's neural networks to fill in the details of this sketch (i.e., create a model that will produce a sound for any new gesture) and see if they like the results. They can create dozens of alternative sketches within minutes and efficiently explore their design space, or they can add training examples to sketches they like and begin to refine a model into a carefully-crafted instrument. They can do all this without prior machine learning expertise and without programming, and they can "patch" trained models directly into their existing gesture sensor systems (to provide features) and preferred sound synthesis software (to translate model outputs into sound).

2. **Interactive machine learning can support an embodied approach to design.** Embodied experience and expertise are critical considerations in the design of interactive systems, as noted by Klemmer et al. [11], and we have observed that interactive machine learning can naturally incorporate the body into the design process. Composers who worked with the Wekinator deeply valued the ability to create new instruments that "felt right" to play. The Wekinator allowed them a better way to create such instruments compared to their usual approaches (such as writing code to specify the relationship between gesture and sound): much of their time with the Wekinator was spent creating training examples through physical demonstration of gestures and evaluating trained models by gesturing in real-time and listening to the sounds that were produced. Their attention was focused on crafting and evaluating the relationships between gesture and sound, not on writing code or designing mathematical functions translating sensor values into synthesis parameters. Some composers have especially valued the "playalong" interface for training data creation [4], which captures a stream of training examples as a user gestures along with a dynamic sonic prompt or "score" in real-time. This interface allows composers to practice gesturing to the prompt and then encode directly into the training data their embodied understanding of which gestures best "match" the score.

CURRENT CHALLENGES

In order to expand the set of problems for which end-user machine learning is useful, and to expand the set of users who are able to use it effectively, our current work is focused on addressing several challenges, including:

1. Supporting end-user feature design and selection for complex input domains such as audio and video
2. Supporting learning of overlapping, hierarchically-structured, or otherwise "messy" concepts (e.g., teaching the system to respond to both a gesture's identity and its dynamics)
3. Building high-accuracy systems for domains such as health, where it is critical both that the system has learned the correct concept accurately and that users have the ability to verify and trust the model.

GOALS FOR THE WORKSHOP

We seek to engage with other researchers and practitioners investigating the types of questions we believe to be fundamentally important to this burgeoning research area: How might HCI, machine learning, and other disciplines forge more effective collaborations to develop algorithms and systems in which human and machine intelligence more effectively complement each other? How can intelligent systems be made more understandable, flexible, and debuggable? What might end users accomplish with intelligent systems that has not previously been possible?

REFERENCES

[1] Amershi, S., J. Fogarty, A. Kapoor, and D. Tan. 2010. "Examining multiple potential models in end-user interactive concept learning." *Proc. CHI'10,* 1357–1360.

[2] J. Fails and D. Olsen, Jr. 2003. "Interactive machine learning." *Proc. IUI'03*, 39–45.

[3] Fiebrink, R. 2011. Real-time Human Interaction with Supervised Learning Algorithms for Music Composition and Performance. PhD thesis, Princeton University.

[4] Fiebrink, R., P. R. Cook, and D. Trueman. 2009. "Play-along mapping of musical controllers." *Proc. International Computer Music Conference.*

[5] Fiebrink, R., P. R. Cook, and D. Trueman. 2010. "Human model evaluation in interactive supervised learning." *Proc. CHI'10.*

[6] Fiebrink, R., D. Trueman, C. Britt, M. Nagai, K. Kaczmarek, M. Early, M.R. Daniel, A. Hege, and P. R. Cook. 2010. "Toward understanding human-computer interactions in composing the instrument." *Proc. International Computer Music Conference.*

[7] Fiebrink, R., D. Trueman, and P. R. Cook. 2009. "A meta-instrument for interactive, on-the-fly machine learning." *Proc. New Interfaces for Musical Expression*.

[8] Fiebrink, R., G. Wang, and P. R. Cook. 2008. "Foundations for on-the-fly learning in the ChucK programming language." *Proc. International Computer Music Conference*.

[9] Fogarty, J., D. Tan, A. Kapoor, and S. Winder. 2008. "CueFlik: Interactive concept learning in image search." *Proc. CHI'08*, 29–38.

[10] Kapoor, A., B. Lee, D. Tan, and E. Horvitz. 2010. "Interactive optimization for steering machine classification." *Proc. CHI'10*, 1343–1352.

[11] Klemmer, S.R., B. Hartmann, and L. Takayama. 2006. "How bodies matter: Five themes for interaction design." *Proc. Designing Interactive Systems*, 140–149.

[12] Kulesza, T., W.K. Wong, S. Stumpf, S. Perona, R. White, M.M. Burnett, I. Oberst, and A.J. Ko. 2009. "Fixing the program my computer learned: Barriers for end users, challenges for the machine." *Proc. IUI'09*, 187–196.

[13] Resnick, M., B. Myers, K. Nakakoji, B. Shneiderman, R. Pausch, T. Selker, and M. Eisenberg. 2005. "Design principles for tools to support creative thinking." *Report of Workshop on Creativity Support Tools*, Washington, DC, USA, 13-14 June.

[14] Schedel, M., and R. Fiebrink. 2011. "A demonstration of bow articulation recognition with Wekinator and K-Bow." *Proc. International Computer Music Conference*.

[15] Schedel, M., P. Perry, and R. Fiebrink. 2011. "Wekinating 000000Swan: Using machine learning to create and control complex artistic systems." *Proc. New Interfaces for Musical Expression*.

[16] Shneiderman, B. 2000. "Creating creativity: User interfaces for supporting innovation." *ACM TOCHI* 7:114–138.

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

[18] Talbot, J., B. Lee, A. Kapoor, and D. Tan. 2009. "EnsembleMatrix: Interactive visualization to support machine learning with multiple classifiers." *Proc. CHI'09*, 1283–1292.

[19] Trueman, D. 2010. "Clapping Machine Music Variations: A composition for acoustic/laptop ensemble." *Proc. International Computer Music Conference*.