Annual Progress Seminar - II Report

on

"Machine Learning Algorithms for Music Composition"

Large

Submitted

in partial fulfillment of

the requirements of the degree of

Bachealor of Technology (Information Technology)

by

Mr. Prathmesh Bendal

(171080076)

Under Supervision of

Mr.Pranav .A. Nerurkar

Department of Computer Engineering & Information Technology

Veermata Jijabai Technological Institute, Mumbai - 400019

(Autonomous Institute Affiliated to University of Mumbai)

2018 - 2019

CERTIFICATE

This is to certify that, **Mr Prathmesh Bendal (ID- 171080076)** , a student of **Btech (Information Technology)** has completed an Annual Progress Seminar-II held in November 2018 for research work on **“Machine Learning Algorithms ”** to our satisfaction.

Mr Pranav A.Nerurkar	Dr. M. M. Chandane	Dr. V. B. Nikam
Supervisor	Co-Supervisor	Head, Dept of CE & IT

dateFebruary 28, 2019

Declaration of the Student

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources.

I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea / data / fact / source in my submission.

I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Mr. Prathmesh. D. Bendal
(Reg no : 171080076) DateFebruary
28, 2019

Abstract

This thesis examines machine learning through the lens of human-computer interaction in order to address fundamental questions surrounding the application of machine learning to real-life problems, including: Can we make machine learning algorithms more usable and useful? Can we better understand the real-world consequences of algorithm choices and user interface designs for end-user machine learning? How can human interaction play a role in enabling users to efficiently create useful machine learning systems, in enabling successful application of algorithms by machine learning novices, and in ultimately making it possible in practice to apply machine learning to new problems?

The scope of the research presented here is the application of supervised learning algorithms to contemporary computer music composition and performance. Computer music is a domain rich with computational problems requiring the modeling of complex phenomena, the construction of real-time interactive systems, and the support of human creativity. Though varied, many of these problems may be addressed using machine learning techniques, including supervised learning in particular. This work endeavors to gain a deeper knowledge of the human factors surrounding the application of supervised learning to these types of problems, to make supervised learning algorithms more usable by musicians, and to study how supervised learning can function as a creative tool.

This thesis presents a general-purpose software system for applying standard supervised learning algorithms in music and other real-time problem domains. This system, called the Wekinator, supports human interaction throughout the entire supervised learning process, including the generation of training examples and the application of trained models to real-time inputs. The Wekinator is published as a freely-available, open source software project, and several composers have already employed it in the creation of new musical instruments and compositions. This thesis also presents work utilizing the Wekinator to study human-computer interaction with supervised learning in computer music. Research is presented which includes a participatory design process with practicing composers, pedagogical use with non-expert users in an undergraduate classroom, a study of the design of a gesture recognition system for a sensor-augmented cello bow, and case studies with

Acknowledgements

I am indebted to the support, guidance, and inspiration of many people, without whom my research and this thesis would not be possible. I am tremendously grateful to have had the opportunity to work with my advisor, Perry Cook, who has tirelessly provided insight, encouragement, and inspiration throughout my graduate study at Princeton. Perry's guidance and the creative incubation of the Sound Lab have been critical in shaping the way I think about research, and I've had an immense amount of fun as one of his students.

I am incredibly appreciative of everyone else who has played a part in making the Wekinator software what it is and in making this thesis possible. To all the composers who participated in the "Music and Machine Learning" seminar in 2009—Anne Hege, Cameron Britt, Dan Trueman, Konrad Kaczmarek, Michael Early, Michelle Nagai, and MR Daniel—thank you for partaking in the adventure; for sharing your many ideas, questions, and complaints; and for the chance to create music together. To all the COS/MUS 314 (PLOrk) 2010 students: thank you for bringing your enthusiasm and creativity to your work with the Wekinator and throughout the course; it was a pleasure to teach you, and I learned a lot from working with you. To Meg Schedel: thanks for your vision, your Max/MSP expertise, your many ideas for improving the

Contents

1 Introduction

Supervised learning and Computer Music

Computer music is a domain rich with computational problems requiring the modeling of complex phenomena, including the relationships between low-level audio and sensor signals and higher-level analyses of pitch, harmony, instrumentation, human gesture and intention, and even emotion. The ability to accurately model these complex relationships has far-reaching implications for transforming not only live performance, but also music composition, studio production, sound design, the design of new musical instruments, and content-based music search and recommendation. Supervised learning is an attractive tool for dealing with complex musical phenomena across all these problems and application domains, and it has been applied to many such problems in music. For example, to build a pitch classier capable of transcribing the melody from an audio le, one might train a learning algorithm on a dataset of recorded songs that have each been explicitly annotated with their melodic pitch content over time, as in the work of Ellis and Poliner (2006). Or, to build a system in which a human performer wearing a sensor glove can "play" a computer synthesis algorithm in real-time, one could train an algorithm on the glove sensor outputs paired with corresponding synthesis parameter values for dierent hand gestures, as in work by Modler et al. (1998).

Many computer music applications possess additional characteristics, beyond the requirement for dealing with complexity, that make them suitable to the application of supervised learning algorithms. As in our gesture glove example, computer music performances often incorporate signicant real-time interactivity, wherein a computer's actions are driven by real-time analysis of live, human-generated signals. Also, music is a creative domain, and the creativity of composers or performers can be enhanced or inhibited by the hardware and software tools they use to do their work. As we will show, supervised learning can be a particularly eective tool for building real-time interactive systems and for supporting creative work.

2 Goals and Contribution

Many computer music applications possess additional characteristics, beyond the requirement for dealing with complexity, that make them suitable to the application of supervised learning algorithms. As in our gesture glove example, computer music performances often incorporate signicant real-time interactivity, wherein a computer's actions are driven by real-time analysis of live, human-generated signals. Also, music is a creative domain, and the creativity of composers or performers can be enhanced or inhibited by the hardware and software tools they use to do their work. As we will show, supervised learning can be a particularly eective tool for building real-time interactive systems and for supporting creative work.

We have also used the Wekinator to study human-computer interaction with supervised learning in several musical applications with dierent types of users. These users have been composers writing new music and designing new interactive musical instruments, students studying interactive systems-building in an undergraduate computer music course, and a composer/cellist developing a gesture recognition system for a sensor-augmented cello bow. Through this work, we have demonstrated the breadth and importance of the roles that interaction—encompassing both humancomputer control and computer-human feedback—can play in the development of supervised learning systems, and we have achieved a greater understanding of the dierences between interactive and conventional machine learning contexts. We have also gained a better understanding of the requirements and challenges in the analysis and design of algorithms and interfaces for interactive supervised learning in real-time and creative problem domains

3 Brief History

Computer music has grown dramatically in the past sixty years as an area of academic inquiry and artistic pursuit. With the advent of minicomputers and then personal computers, many more musicians and researchers gained access to the means to develop their own musical software and computer music compositions. A host of programming languages and environments such as Max (Puckette 1991, now Max/MSP) and CSOUND (Boulanger 2000) evolved to meet the needs of these users. Realtime control over digital sound became possible through the use of specialized DSP workstations such as the IRCAM Signal Processing Workstation (Lindemann et al. 1990), as well as digital synthesizers such as the Yamaha DX-7 (Reid 2001); digital music thus became something that one could perform live. Composers and technologists began to envision, build, and use new mechanisms for controlling the computer in real-time, ranging from manipulating GUIs (such as those in Max), to interfaces like the MIDI keyboard which closely resembled conventional instruments, to acoustic instruments augmented with sensors (such as the hypercello of Paradiso and Gershenfeld 1997), to entirely new sensor-based gestural interfaces (such as the The Hands by Waisvisz 1985). Increasing processing power has enabled the real-time execution of complex software programs for accompanying and improvising alongside human musicians (e.g., see work by Raphael 2010 and Kapur 2007), while the shrinking size of computing systems has enabled increasingly portable and populous computer music performance ensembles like the Princeton Laptop Orchestra (Trueman et al. 2006) and the Stanford Mobile Phone Orchestra (Wang et al. 2008).

Some computer music performance technologies developed in academic contexts also end up being adopted by mainstream artists. For example, John Chowning discovered the FM synthesis algorithm while working at Stanford University in 1967 (Chowning 1973). Packaged into the Yamaha DX7 synthesizer in 1983, FM synthesis enabled countless artists of the 1980's to eciently create a wide range of articially synthesized sounds (Reid 2001). More recently, Max/MSP has been used by the English alternative rock band Radiohead (Pareles 2007), and the Reactable tabletop interface (developed at the Universitat Pompeu Fabra by Jorda‘ et al. 2007) was used by the Icelandic singer-songwriter Bjork on her 2007 world tour (Andrews 2007).

4 Scope

Computer music involves a wide range of live performance practices. One signicant dimension of variability among these practices is the extent to which human performers exercise control or inuence over the live actions of the computer. At one extreme of this spectrum of human involvement lie practices such as tape music and algorithmic computer music, in which the sounds produced by the computer are uninuenced by human performers, or in which there are no human performers at all. At the other extreme lie practices in which the computer is highly responsive to human actions. For example, a human performer might control the pitch, articulation, volume, and timbre of a computer synthesis algorithm through her gestures using a hardware controller. In this case, the controller and computer software together function as an expressive musical instrument. Work mentioned above by Waisvisz (1985) falls into this category, as do notable works by Laetitia Sonami (Bongers 2000), Trueman and Cook (2000), and many others. Alternatively, the computer might listen to the sound of a human musician performing on an acoustic instrument and respond by producing its own musically appropriate output. In this case, the computer may play a role more akin to a human accompanist or collaborator.

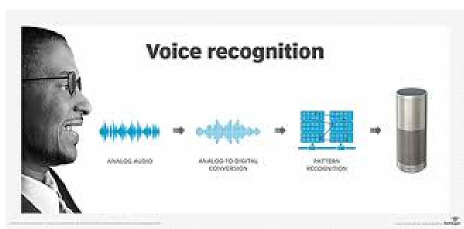

Figure 1: In interactive computer music, actions of a human performer are sensed in real-time by a microphone, controller, and/or other sensing mechanism, and communicated to the computer. The computer interprets these actions, and this interpretation is used to control or inuence its future actions. The outcome of the computer action—for example, changes in the timbre of its sound or in the type melodic material it generates, provides real-time feedback to the human performer

5 conclusion

In this thesis, we have examined applied machine learning through the lens of humancomputer interaction. In doing so, we have created, studied, and improved systems for users to interactively apply supervised learning algorithms to their work in computer music composition, performance, and instrument design. We have created a useful software tool that has aided students and professional composers in creating new musical works, and in doing so, we have demonstrated the feasibility and ecacy of interactive machine learning in this application domain. Our work with users has led to a clearer characterization of the requirements and goals of interactive machine learning users and of the dierent roles that interaction may play in allowing them to design and evaluate systems, to learn to become more eective users of machine learning, and to work creatively. As a result, this work has both empowered musicians to create new forms of art and contributed to a broader HCI perspective on machine learning practice. In this section, we present a summary of our work and highlight the contributions that are most signicant to future research in HCI and machine learning, as well as to research and creative work in computer music. these contribution include:

1. A new software tool allowing real-time human interaction with supervised learning algorithms and, within it, a new "playalong" interaction for training data creation

2. A demonstration of the important roles that interaction—encompassing both human-computer control and computer-human feedback—can play in the development of supervised learning systems, and a greater understanding of the dierences between interactive and conventional machine learning contexts

3. A better understanding of the requirements and challenges in the analysis and design of algorithms and interfaces for interactive supervised learning in realtime and creative problem domains.

4. A clearer characterization of composers' goals and priorities for interacting with computers in music composition and instrument design, and a demonstration that interactive supervised learning is useful in supporting composers in their work.

5. A demonstration of the usefulness of interactive supervised learning as a creativity support tool.

Our work has emphasized the importance of considering the human context of machine learning practice. By providing users with the ability to engage in a wider variety of interactions with supervised learning algorithms, the software we have created has enabled people to put standard learning algorithms to use to do their work more eciently and eectively, and to imagine and accomplish goals that were not possible with the tools previously available to them. Our work has shown that these interactions can empower end users to build interactive systems for themselves and for others while employing an approach to design that leverages their domain knowledge, embodied expertise, and understanding of the context in which the trained models will be used. Furthermore, these interactions can inspire users to consider new design possibilities and engage in self-reection and growth as machine learning users and creative beings

References

Bishop, C. M. 2007. Pattern Recognition and Machine Learning, 2nd ed. Springer

Burges, C. 1998. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2, 2, 121–167

Modler, P., Hofmann, F., and Zannos, I. 1998. Gesture recognition by neural networks and the expression of emotions. IEEE Transactions on Systems, Man, and Cybernetics 2, 1072–1075

Quinlan, J. R. 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann

Turnbull, D., Barrington, L., Torres, D., and Lanckriet, G. R. G. 2008. Semantic annotation and retrieval of music and sound eects. IEEE Transactions on Audio, Speech, and Language Processing 16, 2, 467–476.

Bencina, R. 2005. The Metasurface: Applying natural neighbor interpolation to two-to-many mapping. In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME). Vancouver, BC, Canada, 101–104.

Schapire, R. E. 2003. The boosting approach to machine learning: An overview. Lecture Notes in Statistics, 149–172.