EDITORIAL

Guest editorial: Machine learning in and for music

Gerhard Widmer

Published online: 29 August 2006 Springer Science + Business Media, LLC 2006

Music has always held a special fascination for researchers in mathematics and computer science. It is both an abstract, formal system and a physical event; it relates to areas of human endeavour as diverse as the arts, aesthetics, philosophy, mathematics, acoustics, physiology, or history; and it poses deep questions regarding human nature, intellect, and emotions. Artificial Intelligence, in particular, has dreamed of devising intelligent musical machines almost from its inception.

Music has also played its role in the history of machine learning research. Initially, musical tasks were mostly used as toy domains to demonstrate the properties of new learning algorithms, and the musical problems addressed were usually greatly simplified (the guest editor's own early research is no exception). But recent years have seen a dramatic change in this regard. Machine learning researchers have begun to address musical tasks of real-world complexity—from the analysis of the artistic style of famous musicians to systems that learn to accompany, or improvise with, human musicians, and from programs that classify audio recordings into musical genres to computers that can recognise artists from the way they play.

This trend towards 'real-world' applications has been accelerated by the dramatic changes that are currently going on in the digital music market. With the Internet becoming the central medium for the distribution of music and Web-based music stores offering literally millions of digital music tracks, there is a pressing need for things like intelligent and adaptive music search and recommendation systems, automatic style recognition and music classification, etc. The new field of *Music Information Retrieval* (MIR) is a direct response to this need, and the research published at the International Conference on Music Information Retrieval (ISMIR) and similar events is closely watched by major players in the music industry. The potential of this kind of research is also being recognised by the large international research funding organisations. At the European level, for instance, several large-scale research efforts in the area of intelligent computing and learning in music and audio have recently been funded—e.g., the projects SIMAC (www.semanticaudio.org), Semantic HiFi (http://shf.ircam.fr), and S2S² (http://www.s2s2.org).

G. Widmer (🖂)

Department of Computational Perception, Johannes Kepler University Linz, Austria and Austrian Research Institute for Artificial Intelligence, Vienna, Austria www.cp.jku.at/people/widmer

But machine learning is also beginning to make substantial contributions to music as an art form, and to music research. For instance, learning algorithms have recently made a number of discoveries about the art of expressive music performance (Widmer, 2003; Widmer et al., 2003) that have met with great interest in the world of musicology.

These and other developments suggested that now is the right time for a special issue on ML & Music, in order to give the machine learning community an impression of the breadth and depth of research in this field, and maybe also motivate machine learning researchers to consider musical problems for inspiration. In response to a corresponding call for submissions, 28 papers were submitted, 7 of which were finally selected for presentation in this special issue. They demonstrate the whole breadth of work that is currently being carried out—from audio processing to high-level symbolic representations, from classical through jazz to popular music, from audio transcription to analysis and classification, from music analysis to music performance. Perhaps naturally, the focus in these articles is more on the peculiarities of the musical problems addressed, and on the musical significance of the results, than on general theoretical contributions to machine learning. They demonstrate how state-of-the-art machine learning technology can make useful contributions to music.

The article by *Darrell Conklin* focuses on music analysis, and on the level of symbolic representation. A particular representation for musical abstractions in melodies (or generally musical lines) is presented, and it is shown how this representation can facilitate various analytical tasks via machine learning—the discovery of sequential patterns and rules at the level of melodic segments, and the statistical modelling and recognition of musical styles. Structural abstractions is one of the key concepts in music, and principled ways of representing such abstractions have been the focus of much research in AI and music. The concept of modular 'viewpoints' as presented in this paper will probably prove useful beyond music analysis tasks proper—for instance, in music query and retrieval in large databases.

The next three articles relate to performed music and the aspect of expressive performance, which is of central importance at least in classical music and in jazz. Still at the level of symbolic representations (i.e., musical scores and MIDI files), *Graham Grindlay* and *David Helmbold* demonstrate the use of structured probabilistic models for music performance research. They use Hierarchical Hidden Markov Models (HHMMs) to model expressive tempo variations (expressive timing) in classical piano performances, and their relation to the melodic context of the piece. Training data are manually measured timing profiles from a substantial number of piano performances by both world-class artists and piano students. The power of the resulting models is demonstrated in several experimental settings, including the automatic 'expressive timing profiles. The models also give interesting insights into the stylistic consistency of piano students and famous pianists.

The connection between a discrete, symbolic representation of music (the musical score) and the level of actual audio recordings is made in the paper by *Christopher Raphael*, where a method is presented for matching complex audio recordings onto the printed score of the music. Score-performance matching is an important prerequisite for many interesting applications, such as automated accompaniment systems or intelligent music synchronisation. It is a hard problem, especially with music that is performed expressively, that is, with strongly varying tempo (among other things). The approach proposed in this paper is based on a graphical model that explicitly includes a model of the time-varying tempo process. The performance of the model, demonstrated on complex orchestra recordings, is quite impressive, and the readers are invited to listen to some of the sound examples provided by the author.

The paper by *Maarten Grachten et al.* takes us into the world of jazz and jazz performance. It is concerned with a problem that may seem rather specific, but is nevertheless important O Springer

in music content processing, namely, the musically 'valid' transformation of expressively performed music from one tempo to another. The expressive aspects of the performance (which in jazz include not just details of timing and dynamics, but also ornamentations, insertions of notes, etc.) cannot simply be transformed linearly when the tempo of the playback is changed; that would result in recordings that would, in many cases, immediately be recognised as unmusical or 'odd' by listeners. The authors present a case-based approach that heavily relies on an abstract representation of melodies (similar in spirit to Darrell Conklin's 'viewpoints', though based on a different music theory), and on case adaptation. Using a music similarity metric that was derived automatically from human similarity judgments, it is shown that the resulting system, TempoExpress, performs tempo transformations on expressively played jazz saxophone phrases in a musically better way than uniform time stretching would.

The three remaining articles pertain more directly to the application fields *Content-based Music Processing* and *Music Information Retrieval*. *Daniel Ellis* and *Graham Poliner* show that the identification of the melody in polyphonic audio recordings can be cast (and solved well) as a classification problem, in a surprisingly simple and assumption-free way. Support vector machines trained directly on audio signal frames achieve a surprisingly high accuracy in identifying the correct melody pitch in each frame. The authors also show that smoothing of the results via HMMs can improve the results considerably. Altogether, this paper is a good example of how current machine learning technology, when applied in a thoughtful way, can solve musical problems of practical relevance. The extraction of melodic content from audio recordings is useful for a wide range of applications in MIR.

Another audio labeling problem of practical relevance is addressed by *Ning Hu* and *Roger Dannenberg*. They present a relatively simple approach to bootstrapping a model for onset detection (the precise location of time points in an audio recording where an individual note starts). Starting from an initial, imprecise alignment between an audio recording and the corresponding written score of the music, a cycle of repeated training and onset estimation leads to a precise onset detector. The important aspect about this process is that it does not require hand-labelled training examples (which would be costly to obtain). Experiments show very good results both on monophonic and polyphonic audio recordings. Again, onset detection is of great importance in many content-based music processing tasks such as transcription or beat and rhythm identification and synchronisation.

Finally, the paper by *James Bergstra et al.* briefly reports on experiments with automatic genre classification, i.e., the recognition of musical styles from audio recordings. This is an important benchmark test in Music Information Retrieval, but also potentially useful in a large variety of commercial applications. Using a standard boosting algorithm and a set of cleverly selected 'mid-level' features extracted from the audio recordings, the authors demonstrate very good classification performance on several collections of digital music tracks. Systematic experiments show that it is the careful selection of the temporal extent of the audio features that contributes much to the success of the classifier. The method described here won the 2005 MIREX genre classification contest (an annual series of contests to compare algorithms in music information retrieval) and thus demonstrates the current state of the art in automatic genre recognition.

I do hope that the mix of articles collected in this special issue will meet with interest and curiosity on the part of the readers, and succeed in conveying the promise and some of the fascination of music as an application and experimentation field for machine learning.

Acknowledgments The editor of the special issue greatly appreciates support for his and his group's research in machine learning and music by a number of institutions, in particular: the *European Commission* (projects "Semantic Interaction with Music Audio Contents" (SIMAC; FP6-507142), "Sound to Sense, Sense to Sound" (S2S²; IST-2004-03773), and "Knowledge Exploration in Science and Technology" (COST Action 282)); the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung* (FWF; projects "Artificial Intelligence Models of Musical Expression" (Y99-START) and "Operational Models of Music Similarity for Music Information Retrieval" (L112-N04)); and the *Vienna Science and Technology Fund* (WWTF; project "Interfaces to Music" (CI010)). The Austrian Research Institute for Artificial Intelligence acknowledges basic financial support from the Austrian Federal Ministries BMBWK and BMVIT.

References

- Widmer, G. (2003). Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. Artificial Intelligence, 146(2), 129–148.
- Widmer, G., Dixon, S., Goebl, W., Pampalk, E., & Tobudic, A. (2003). In search of the Horowitz factor. AI Magazine, 24(3), 111–130.