Leveraging Expert Domain Knowledge to Learn a Representation of Symbolic Music

Johanna Devaney
devaney.12@osu.edu
Ohio State University, United States of America

Music manuscripts offer as much potential as text manuscripts for data mining and, as with resources like Google Books, there is a wealth of data available online. Currently the largest resource, the International Music Score Library Project (IMSLP), has more than 370,000 scores in its database. While many of these scores exist only in image-based formats, ongoing improvements in the area of Optical Music Recognition (OMR) allow for automatic conversion from images to symbolic representations, which include information such as instrumentation, key signature, time signature, and the notes' pitches, metrical positions, and durations. In order to exploit the research potential of these symbolic music databases, a representation that captures temporal relationships within the music is needed that highlights the structurally significant parts of the musical surface, while ignoring ornamentation.

This paper describes a representation that emphasizes the more structurally significant parts of the musical surface and de-emphasizes less significant parts, such as ornamentation, by integrating human domain expertise and data-driven approaches within a temporal machine learning model. The representation contains less information than the musical surface but more than corresponding chord labels, which discard information about musical texture and are too generalized to use for detailed similarity and classification tasks. The weighting for the various components of the musical surface is determined from an initial harmonic analysis. This harmonic analysis will be performed by a hierarchical model of chord labels and phrases, which will function like a "language model" in speech recognition. In music theory, phrase models describe musical phrases in terms of the tonic, predominant, and dominant functions (Laitz 2012). The inclusion of the expert domain knowledge expressed in the phrase function model helps to resolve the ambiguity between the musical surface and appropriate chord labels in the harmonic analysis, namely whether a particular chord is likely to occur in a particular part of a phrase. Taken in combination with OMR, this representation could be used to render searchable all available scanned music. These searches would not be limited to melody, as is the current state-of-the-art, but would also allow for querying by chord progressions and/or formal structures. The representation can also facilitate automatic hierarchical analysis of musical structures and provides a basis from which to undertake classifications and similarity tasks. Classification tasks include harmonic analysis or assessing the likelihood of a particular composer having composed a piece of unknown provenance, while similarity tasks include longitudinal studies over a composer's career or across composers.

Much of the work on analyzing the growing wealth of music data has been heavily influenced by text retrieval methods through their use of N-grams, sequences of N contiguous symbols. N-grams work well in modeling monophonic sequences, such as when directly applied to the musical surface for monophonic melody retrieval (Pickens, 2001) and for chord retrieval when the chords occur as distinct vertical units (Scholz et al., 2009). This has been demonstrated effectively on peachnote.com (Viro, 2011) with an N-gram viewer similar to the one Google makes available for Google Books. One significant area where N-grams have problems, however, is for more complex textures where the notes of chords are not played simultaneously, which is true of a large proportion of western art music since 1750. One way to address this problem is to automatically segment the musical surface into beat-length frames and treat the contents of each frame as a "chord" (Radicioni and Espositio, 2006), which is well suited to chorale textures but is problematic for arpeggiations or other textures where the chords notes don't occur simultaneously. Another approach is to analyze chord labels rather than the musical surface, such as the system of de Haas et. al (2011), although these are often not available.

The representation described in this paper takes a different approach, using a conditional random fields (CRFs) model for developing both a data-driven model, where all of the feature functions and potentials are learned from the data, and a hybrid data-driven/rule-driven approach, where domain knowledge "rules" are used to design feature and potential functions. Data for the purely data-driven approach comes from a domain expert-labeled dataset of

Mozart and Beethoven piano works in theme and variation form (Devaney et al. 2015). The rule-driven approach incorporates the rules presented in textbooks used in undergraduate music theory curricula, primarily Laitz (2012). This hybrid data- and rule- driven approach is motivated by previous work that demonstrated that a combination of data- and rule-driven models performed better than either approach alone on music analysis tasks (Devaney and Shanahan 2013).

This paper will also discuss the implications of this use CRFs for analyzing other metrically structured cultural products, such as poetry or song lyrics, as well as how this approach could be generalized to other digital humanities projects, specifically for relatively "data- poor" problems where there is a large amount of domain expertise that can be modeled, such as the study of narrative in natural language. More broadly, this work presents a vision of the digital humanities, where large-scale data-driven approaches are balanced by deep domain knowledge and the types of humanistic questions being asked require the development of more sophisticated technology than is currently available.

Bibliography

de Haas, W.B., J.P. Magalhaes, R.C. Veltkamp, F. Wiering. (2011). HarmTrace: Improving Harmonic similarity estimation using functional harmony analysis. In *Proceedings of International Society of Music Information Retrieval conference (ISMIR)*. 67–72.

Devaney, J., C. Arthur, N. Condit-Schultz, and K. Nisula. (2015). Theme And Variation Encodings with Roman Numerals (TAVERN): A new data set for symbolic music analysis. In *Proceedings of the International Society for Music Information Retrieval (ISMIR) conference*, 728–34.

Devaney, J., and D. Shanahan. (2014). Evaluating Rule- and Exemplar-Based Computational Approaches to Modeling Harmonic Function in Music Theory Pedagogy. In *Proceedings of the 9th Conference on Interdisciplinary Musicology.*

Laitz, S. G. (2011). *The Complete Musician*. Oxford: Oxford University Press, 3rd edition. Pickens, J. 2001. A survey of feature selection techniques for music information retrieval.

Center for Intelligent Information Retrieval, Department of Computer Science, University of Massachusetts,

Technical Report.

Radicioni, D.P. and Esposito, R. (2006). Learning tonal harmony from Bach chorales. In *Proceedings of the International Conference on Cognitive Modelling.*

Scholz, R., Vincent, E., and F. Bimbot, F. (2009). Robust modeling of musical chord sequences using probabilistic N-grams. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP).* 53–6.

Viro, V. (2011). Peachnote: Music score search and analysis platform. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 359–62.