Computational Modelling of Harmony

Simon Dixon

Centre for Digital Music,
Queen Mary University of London,
Mile End Rd, London E1 4NS, UK
`simon.dixon@elec.qmul.ac.uk`
`http://www.elec.qmul.ac.uk/people/simond`

Abstract. Many computational models for processing music fail to capture essential aspects of the high-level musical structure and context, and this limits their usefulness, particularly for musically informed users. In this talk I describe two recent approaches to modelling musical harmony which attempt to reduce the gap between computational models and human understanding of music. The first is a chord transcription system which uses a high-level model of musical context in which chord, key, metric position, bass note, chroma features and repetition structure are integrated in a Bayesian framework, achieving state-of-the-art performance. The second approach uses inductive logic programming to learn logical descriptions of harmonic sequences which characterise particular styles or genres. Each approach brings us one step closer to modelling music in the way it is conceptualised by humans.

Key words: Chord transcription, inductive logic programming, musical harmony

1 Introduction

Music is a complex phenomenon. Human understanding of music is at best incomplete, and computational models used in our research community fail to capture much of what is understood about music. Nevertheless, in the last decade we have seen remarkable progress in Music Information Retrieval research. This progress is particularly remarkable considering the naivete of the musical models used. Two examples are the bag-of-frames approach to music similarity [Aucouturier et al., 2007], and the periodicity pattern approach to rhythm analysis [Dixon et al., 2003], which are both independent of the order of musical notes, whereas temporal order is an essential feature of melody, rhythm and harmonic progression. This talk will present recent work on modelling musical harmony in order to come closer to modelling music as a musician might conceptualise it.

2 Chord Transcription

When a musician transcribes the chords of a piece of music, the chord labels are not assigned solely on the basis of local pitch content of the signal. Musical

context such as the key, metrical position and even the large-scale structure of the music play an important role in the interpretation of harmony. The goal of our recent work on chord transcription [Mauch and Dixon, 2010b,Mauch, 2010] is to propose computational models that integrate musical context into the automatic chord estimation process. We employ a dynamic Bayesian network (DBN) to combine models of metric position, key, chord, bass note and beat-synchronous bass and treble chroma into a single high-level musical context model. The most probable sequence of metric positions, keys, chords and bass notes is estimated via Viterbi inference.

A DBN is a graphical model representing a succession of simple Bayesian networks in time. These are assumed to be Markovian and time-invariant, so the model can be expressed recursively in two time slices: the initial slice and the recursive slice. Each node in the network represents a random variable, which might be an observed node (in our case the bass and treble chroma) or a hidden node (the key, metrical position, chord and bass pitch class nodes). Edges in the graph denote dependencies between variables. In the recursive slice, the bass chroma class is dependent on the bass pitch class, the treble chroma is dependent on the chord, the bass pitch class is dependent on the chord and the previous chord, while the chord is dependent on the previous chord, the key and the metric position. Finally, the key and metric position are only dependent on their previous values. The dependencies between nodes are expressed as conditional probability distributions, which assign high probabilities to the following "normal" situations: the metrical position advances one beat at a time, the key does not change, the chord does not contain non-key pitch classes or change on a weak metric position, and the bass note is the chord bass (particularly on the first beat of the chord) or otherwise a chord note. For more details see [Mauch, 2010].

Using a standard test set of 210 songs used in the MIREX chord detection task, our model achieved an accuracy of 73%, with each component of the model contributing significantly to the result. This improves on the best result at MIREX 2009 for pre-trained systems. Further improvements have been made via two extensions of this model: taking advantage of repeated structural segments (e.g. verses or choruses), and refining the front-end audio processing.

Most musical pieces have segments which occur more than once in the piece, and there are two reasons for wishing to identify these repetitions. First, multiple sets of data provide us with extra information which can be shared between the repeated segments to improve detection performance. Second, in the interest of consistency, we can ensure that the repeated sections are labelled with the same set of chord symbols. We developed an algorithm that automatically extracts the repetition structure from a beat-synchronous chroma representation [Mauch et al., 2009], which ranked first in the 2009 MIREX Structural Segmentation task. Using this algorithm, we merged the chroma representations of matching segments and found a significant performance increase (to 75% on the MIREX score).

A further improvement was achieved by modifying the front end audio processing. We found that by learning chord profiles as Gaussian mixtures, the recognition rate of some chords can be improved. However this did not result in an overall improvement, as the performance on the most common chords reduced. Instead, an approximate pitch transcription method using non-negative least squares was employed to reduce the effect of upper harmonics in the chroma representations [Mauch and Dixon, 2010a]. This results in both a qualitative (reduction of specific errors) and quantitative (a substantial overall increase in accuracy) improvement in results, with a MIREX score of 79% (without using segmentation), which again is significantly better than the state of the art. By combining both of the above enhancements we reach an accuracy of 81%, a statistically significant improvement over the best result (74%) in the 2009 MIREX Chord Detection tasks and over our own previously mentioned results.

3 Logic-Based Modelling of Harmony

First order logic (FOL) is a natural formalism for representing harmony, as it is sufficiently general for describing combinations and sequences of notes of arbitrary complexity, and there are well-studied approaches for performing inference, pattern matching and pattern discovery using subsets of FOL. Logic-based representations can also be presented in an intuitive way to non-expert users. Inductive logic programming (ILP) has been used for various musical tasks, including inference of harmony [Ramirez, 2003] and counterpoint [Morales, 1997] rules from musical examples, as well as rules for expressive performance [Widmer, 2003].

In our work, we use ILP to learn sequences of chords that might be characteristic of a musical style [Anglade and Dixon, 2008], and test the models on classification tasks [Anglade and Dixon, 2009,Anglade et al., 2009]. To allow for human-readable classification models we represent pieces of music as lists of chords and induce characterisations of musical genres using subsequences of these chord lists expressed as context-free definite clause grammars.

As test data we used a collection of 856 pieces covering 3 genres, each of which was divided into a further 3 subgenres: academic music (Baroque, Classical, Romantic), popular music (Pop, Blues, Celtic) and jazz (Pre-bop, Bop, Bossa Nova). The data is represented in the Band in a Box format, containing a symbolic encoding of the chords, which were extracted and encoded in logic. The Band in a Box software is designed to produce an accompaniment based on the chord symbols, using a MIDI synthesiser. In further experiments we tested the classification method using an automatic transcription of chords from this synthesised audio data, in order to test the robustness of the system to errors in the chord symbols.

The experiments were performed with the first-order logic decision tree induction algorithm, Tilde, which learns a classification model based on a vocabulary of predicates supplied by the user. In our case, we described the chords in terms of their root note, scale degree, chord category (e.g. major, minor, dominant

seventh), and intervals between successive root notes, and we constrained the learning algorithm to generate rules containing subsequences of length at least two chords.

The results for various classification tasks are shown in Table 1. All results are significantly above the baseline, but performance clearly decreases for more difficult tasks. Perfect classification is not to be expected from harmony data, since other aspects of music such as instrumentation (timbre), rhythm and melody are also involved in defining and recognising musical styles.

Classification Task	Baseline	Symbolic	Audio
Academic – Jazz	0.55	0.947	0.912
Academic – Popular	0.55	0.826	0.728
Jazz – Popular	0.61	0.891	0.807
Academic – Popular – Jazz	0.40	0.805	0.696
All 9 subgenres	0.21	0.525	0.415

Table 1. Classification results.

Analysis of the most common rules extracted from the decision tree models built during these experiments reveals some interesting and/or well-known jazz, academic and popular music harmony patterns. For example, while a perfect cadence is common to both academic and jazz styles, the chord categories distinguish the styles very well, with academic music using triads and jazz using seventh chords:

genre(**academic**,A,B,Key) :- **gap(A,C),**
                     degreeAndCategory(5,maj,C,D,Key),
                     degreeAndCategory(1,maj,D,E,Key),
                     gap(E,B).

[Coverage: academic=133/235; jazz=10/338]

genre(**jazz**,A,B,Key) :- **gap(A,C),**
                 degreeAndCategory(5,7,C,D,Key),
                 degreeAndCategory(1,maj7,D,E,Key),
                 gap(E,B).

[Coverage: jazz=146/338; academic=0/235]

In recent work we have combined the classifier with a state of the art timbre-based classifier and shown that a small but significant improvement in classification performance can be observed on some data sets.

References

[Anglade and Dixon, 2008] Anglade, A. and Dixon, S. (2008). Characterisation of harmony with inductive logic programming. In *9th International Conference on Music Information Retrieval*, pages 63–68.

[Anglade and Dixon, 2009] Anglade, A. and Dixon, S. (2009). First-order logic classification models of musical genres based on harmony. In *6th Sound and Music Computing Conference*, pages 309–314.

[Anglade et al., 2009] Anglade, A., Ramirez, R., and Dixon, S. (2009). Genre classification using harmony rules induced from automatic chord transcriptions. In *10th International Society for Music Information Retrieval Conference*.

[Aucouturier et al., 2007] Aucouturier, J.-J., Defréville, B., and Pachet, F. (2007). The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music. *Journal of the Acoustical Society of America*, 122(2).

[Dixon et al., 2003] Dixon, S., Pampalk, E., and Widmer, G. (2003). Classification of dance music by periodicity patterns. In *4th International Conference on Music Information Retrieval*, pages 159–165.

[Mauch, 2010] Mauch, M. (2010). *Automatic Chord Transcription from Audio Using Computational Models of Musical Context*. PhD thesis, Queen Mary University of London, Centre for Digital Music.

[Mauch and Dixon, 2010a] Mauch, M. and Dixon, S. (2010a). Approximate note transcription for the improved identification of difficult chords. In *11th International Society for Music Information Retrieval Conference*.

[Mauch and Dixon, 2010b] Mauch, M. and Dixon, S. (2010b). Simultaneous estimation of chords and musical context from audio. *IEEE Transactions on Audio, Speech and Language Processing*, 18. Accepted for publication.

[Mauch et al., 2009] Mauch, M., Noland, K., and Dixon, S. (2009). Using musical structure to enhance automatic chord transcription. In *10th International Society for Music Information Retrieval Conference*, pages 231–236.

[Morales, 1997] Morales, E. (1997). PAL: A pattern-based first-order inductive system. *Machine Learning*, 26(2–3):227–252.

[Ramirez, 2003] Ramirez, R. (2003). Inducing musical rules with ILP. In *Proceedings of the International Conference on Logic Programming*, pages 502–504.

[Widmer, 2003] Widmer, G. (2003). Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence*, 146(2):129–148.