Beat-Synchronous Data-Driven Automatic Chord Labeling

Björn Schuller*, Florian Eyben, Gerhard Rigoll
Technische Universität München, Institute for Human-Machine Communication, 80333 München, Germany

Abstract

Automatic Chord Labeling becomes a challenge when dealing with original audio recordings, in particular of modern popular music. In this work we therefore suggest a data-driven approach applying Support Vector Machines (SVM) and Hidden-Markov-Models (HMM) as opposed to typical chord-template modeling. The feature basis is formed by pitch-tuned chromatic feature information. For synchronization with the rhythmic structure we use IIR comb-filter banks for tempo detection, meter recognition, and on-beat tracking. The chord base is built by all typical triads resulting in 76 classes. A musiological model is used to model the context of a chord. Extensive experimental results are reported on 11k chords of 7h of MP3 compressed popular music and demonstrate effectiveness over the traditional approach to Automatic Chord Labeling: 60% accuracy are reached for this challenging task when major and minor chords are tagged.

Introduction

The automatic recognition and transcription of musical chord progressions possesses a wide variety of applications : musicians can automatically transcribe their progression while jamming, or they can be offered a plug-in to media players to show them the current chord for play along. But chord knowledge can also be used as meta-information in many other Music Information Retrieval tasks, such as genre recognition (e.g. Jazz having many II-V-I successions, while e.g. Blues has many I-IV-V7s), or musical mood recognition (e.g. ratio of major/minor or 7/maj7 chords), or key recognition. Also, DJs can be provided with automatic synthesis of additional fitting notes as sub-basses or arpeggios, or tools that blend music at matched key/chord. One final application is music similarity analysis or finding of plagiarism (e.g. chord progression of *Johann Pachelbel's "Canon in D"* (*"Canon per 3 Violini e Basso"*), which is found in multiple contemporary popular pieces, such as *"Go West"*, *"Streets of London"*, *"All Together Now"*, *"Basket Case"*, *"Big City Life"* or *"Volverte a Ver"*). To save cost-intensive and partly not feasible manual labeling, we introduce a beat-synchronous and data-driven approach in the ongoing.

ChoRD Database

In order to have sufficient data for machine learning and testing, we annotated a total of 100 musical pieces that cover a good selection of typically aired pop and rock music with the tempo in bpm, the key, and each chord. As ground truth reference original scores were used. The alignment was carried out by three experienced musicians. 64 different artists are comprised. On average, 1.6 pieces per artist are used, however, only 18 artists are found more than once in the set: the highest number of songs per artist resembles 5 for *Delta Goodrem*, *James Blunt*, *Robbie Williams*, followed by *Celine Dion*, *Coldplay* and *Enya* with 4 songs, each, *Bon Jovi*, *Bryan Adams*, *Cher* with 3, each, and *All Saints*, *Backstreet Boys*, *Britney Spears*, *Keane*, *Phil Collins*, *Roxette*, and *The Corrs* with 2, each. These pieces all have constant tempo. The list of songs can be found at [1]. The original recordings are compressed to 128 kbit/s MP3 for the oncoming tests. The total playtime resembles 6h 58min 12sec, and 10,702 bars are contained. This set is referred to as *Chord Recognition Database*, respectively *ChoRD*.

The chords have been annotated in the 7 main classes: major (Maj), minor (Min), Suspended (Sus2, Sus4), Augmented (Aug), Diminished (Dim), and Power Chords (No3). Likewsie we cover all typical triads consisting of root, second/third/fourth, and fifth. Note that not 7x12, but only 6x12+4=76 final chord classes are obtained, as only 4 different augmented chords exist. This classes have been clustered and re-mapped for testing into the following sets: *76 classes* (all), *60 classes* (all w/o Aug, Dim), *48 classes* (all w/o Aug, Dim, No3), *36 MajMinO* (Maj, Min, anything else mapped onto others (O)), *24 MajMin* (Maj, Min), and *24 MajO* (Maj, O)). Note that the total of chords was kept constant by mapping chords that are not considered onto considered ones by their root and musical function (e.g. "C No3" is mapped onto "C Maj" if its function is accordingly).

In Table 1 the distribution of keys and chords within the ChoRD database is shown in detail for the classes major, minor, and other by root note.

Table 1: Distribution keys and chords in the ChoRD corpus.

Root	#Key	#Major	#Minor	#Other
A	7	511	459	57
A#	8	567	171	86
B	7	480	213	61
C	16	854	278	105
C#	5	312	315	61
D	3	557	349	94
D#	8	533	141	61
E	12	643	362	21
F	13	728	272	52
F#	4	407	209	44
G	12	719	287	103
G#	5	353	196	41
Sum	**100**	**6664**	**3252**	**786**

*Email:schuller@tum.de

Musiological Model

In order to model the context of a chord rather than recognize isolated chords, we employ a musiological model (MM), which resembles a typical language model (LM) as used in automatic speech recognizers. For training of the model we used the chord lead sheets of [2] after removal of doubles. These chord sheets are usually uploaded by users, which means that they are partly simplified, erroneous, or transposed into easily playable keys on guitar (e.g. G Major). However, for a statistical musiological model this is not too problematic, as we are only interested in typical chord successions. As the sheets often contain shortened progressions in a way that the chord succession is laid out only once, we use the following upsampling rule: assuming 60 to 100 bars for a typical rock and pop piece, we strictly repeat whenever a song has below 30 bars until 60-100 bars are reached. Chords were translated into the used target set by rule-based parsing (e.g. elimination of bass-notes, clustering of different spelling variants). Overall, 19,025 songs, resulting in a total of 1,573,803 chords are used for the MM. Table 2 shows the top-ranked uni- and bi-grams by frequency.

Table 2: Top-ranked chord uni- and bigrams by frequency.

Rank	1-gram	#	2-gram	#
1	G	244820	D-G	57500
2	D	227549	G-G	55106
3	A	198958	C-G	54702
4	C	188194	G-C	54040
5	E	130896	A-D	46162
6	F	87741	D-A	43534
7	B	72360	G-G	41090
8	Am	58929	A-A	40161
9	Em	57537	D-D	39710
10	A#	32583	E-A	36659

Automatic Chord Labeling

First, a musical piece is converted from MP3 to a monophonic, 44.1 kHz, 16 Bit wave. Next, the tempo, meter, and down-beat position are determined by IIR-comb filtering as described in [4]. According to the tempo, the song is partitioned into consecutive bars. Per bar a 12-dimensional CHROMA-based C.E.N.S. vector is computed [3]. In this process audio data passes a spectral transformation, dB(A)-correction, compensation of detuning and mapping to pitch classes. The result of this cascade is a 12-dimensional vector containing the intensities for each semitone, taking temporal development into account. Note that dB(A)-correction for adaption to human perception according to norm IEC/DIN 651 and pitch tuning are not standard operations in C.E.N.S. feature computation. For pitch tuning, we acquire the prominent frequency during a long-term analysis of the piece in the range between 130 Hz and 1 kHz. Next, the nearest reference frequency to the measured prominent frequency is detected and the semi-tone filter-bank is shifted, accordingly.

For classification we consider a data-free cross-correlation (CC) with a hard template ("1" for each note that is contained in the chord, "0" for any other note in the scale) as reference. For the proposed data-driven processing we compare SVM with HMM with and w/o the language model (MM). SVM performed best with linear Kernel, pairwise multi-class discrimination and SMO learning. In the case of HMM one continuous model with one emitting state per chord is trained by 20 Baum-Welch iterations. 1 mixture proved optimal. We use a context free grammar (word-loop) and Viterbi decoding to model the sequence of chords. If the MM is used (HMM+MM), Laplace smoothed class-based back-off-bigrams (Katz Back-Off, with cutoff 1) further proved optimal.

Results and Conclusion

For evaluation we use song-indepentent cyclic "leave-one-song-out" (LOSO) training and testing. In Table 3 mean accuracies are summarized. Note that only 36 and less classes provided sufficient statistics for HMM training.

Table 3: Accuracies ChoRD corpus, LOSO evaluation.

Accuracy [%]	CC	SVM	HMM	HMM +MM
76 Classes	28.06	37.31	-	-
60 Classes	29.39	37.43	-	-
48 Classes	29.39	37.96	-	-
36 MajMinO	28.37	36.71	45.39	48.84
24 MajO	33.21	43.70	44.54	43.33
24 MajMin	39.41	40.24	58.57	**60.13**

As can be seen, the data-driven approaches are superior, whereby HMM prevail. By language modeling a further gain is obtained, and the reduction to major and minor chords seems reasonable if appropriate. In future efforts we aim at chord enhancement by Non-Negative-Matrix-Factorization and use of stereophonic information.

Acknowledgment

This work highly benefits from the contributions of the student researchers Moritz Dausinger, Qianqian Xu, and Hermann Karl.

References

[1] http://www.mmk.ei.tum.de/~sch/chord.txt, 2008.

[2] http://www.olga.net, 2006.

[3] M. Müller, F. Kurth, and M. Clausen: Chroma-Based Statistical Audio Features for Audio Matching. In Proc. 2005 IEEEWorkshop on Applications of Signal Processing to Audio and Acoustics (WASPAA05), pp. 275—278, 2005.

[4] B. Schuller, F. Eyben, and G. Rigoll: Fast and robust meter and tempo recognition for the automatic discrimination of ballroom dance styles. In Proc. ICASSP 2007, vol. 1, pp. 217-–220, Honolulu, Hawaii, 2007.