Exploiting Structural Relationships in Audio Music Signals Using Markov Logic Networks

Hélène Papadopoulos, George Tzanetakis

EXPLOITING STRUCTURAL RELATIONSHIPS IN AUDIO MUSIC SIGNALS USING MARKOV LOGIC NETWORKS

Hélène Papadopoulos

Laboratoire des Signaux et Systèmes
UMR 8506, CNRS-SUPELEC-Univ. Paris-Sud, France
helene.papadopoulos[at]lss.supelec.fr

George Tzanetakis

Computer Science Department
University of Victoria, Canada
gtzan@cs.uvic.ca

ABSTRACT

We propose an innovative approach for music description at several time-scales in a single unified formalism. More specifically, chord information at the analysis-frame level and global semantic structure are integrated in an elegant and flexible model. Using Markov Logic Networks (MLNs) low-level signal features are encoded with high-level information expressed by logical rules, without the need of a transcription step. Our results demonstrate the potential of MLNs for music analysis as they can express both structured relational knowledge through logic as well as uncertainty through probabilities.

Index Terms— Music Information Retrieval, Markov Logic Networks, Chord Detection, Structure Analysis

1. INTRODUCTION

Music audio signals are very complex, both because of the intrinsic nature of audio, and because of the information they convey. Signal observations are generally incomplete and noisy. Besides, the great variability of audio signals, due to the many modes of sound production and the wide range of possible combinations between the various acoustic events, make music signals extremely rich and complex from a physical point of view. Music audio signals are also complex from a semantic point of view and convey multi-faceted and strongly interrelated information (e.g. harmony, metric, structure, etc.).

The extraction of relevant content information from audio signals of music is one of the most important aspects of *Music Information Retrieval* (MIR). Although there is a number of existing approaches that take into account interrelations between several dimensions in music (e.g. [1, 2]), most existing computational models extracting content information tend to focus on a single music attribute, which is contrary to the human understanding and perception of music that processes holistically the global musical context [3]. Dealing with real audio recordings thus requires the ability to handle complex relational and rich probabilistic structure at multiple levels of representation. Existing approaches for musical retrieval tasks fail to capture both of these aspects.

Probabilistic graphical models are popular for music retrieval tasks. In particular Hidden Markov models (HMM) have been quite successful in modeling various tasks where objects can be represented as sequential phenomena, such as in the case of chord estimation [4] or beat tracking [5]. However, an important limitation of HMMs is that it is hard to express dependencies in the data. HMMs make the Markovian assumption that each frame only depends on the preceding one. Other formalisms that allow considering more complex dependencies between data in the model have been explored, such as conditional random fields [6], N-grams [7, 8] or tree structures [2]. Although probabilistic models can handle the inherent uncertainty of audio, most of them fail to capture important aspects of higher-level musical relational structure and context. This aspect has been more specifically explored within the framework of logic.

A major advantage of the logic framework is that its expressiveness allows modeling music rules in a compact and human-readable way, thus providing an intuitive description of music. knowledge, such as music theory, can be introduced to construct rules that reflect the human understanding of music [9]. Another advantage is that logical inference of rules allows taking into account all events including those which are rare [10]. Inductive Logic Programming (ILP) [11] refers to logical inference techniques that are subset of First-Order Logic (FOL). These approaches combine logic programming with machine learning. They have been widely used to model and learn music rules, especially in the context of harmony characterization and in the context of expressive music performance. Approaches based on logic have focused on symbolic representations such as the MIDI file format, rather than on audio.

In the context of harmony characterization, pattern-based first-order inductive systems capable of learning new concepts from examples and background knowledge [12], or counterpoint rules for two-voice musical pieces in symbolic format [13] have been proposed. An inductive approach for learning generic rules from a set of popular music harmonization examples to capture common chord patterns is described in [14]. Some ILP-based approaches for the automatic characterization of harmony in symbolic representations [15] and classification of musical genres [16] have been extended to audio [17]. However, they require a transcription step, the harmony characterization being induced from the output of an audio chord transcription algorithm and not directly from audio. In the context of expressive music performance, algorithms for discovering general rules that can describe fundamental principles of expressive music performance [18, 19, 20, 9] have also been proposed. The inductive logic programming approaches are not directly applied to audio, but on symbolic representations. This generally requires a transcription step, such as melody transcription [9].

Real data such as music signals exhibit both uncertainty and rich relational structure. Until recent years, these two aspects have been generally treated separately, probability being the standard way to represent uncertainty in knowledge, while logical representation being used to represent complex relational information. Music retrieval tasks would benefit from a unification of logical and probabilistic knowledge representations. As reflected by previous works, both aspects are important in music, and should be fully considered. However, traditional machine learning approaches are not able to cope with rich relational structure, while logic-based approaches are not able to cope with the uncertainty of audio and need a transcription step to apply logical inference on a symbolic representation. Approaches towards a unification have been proposed within the emerging field of Statistical Relational Learning (SRL) [21]. They combine first order logic, relational representations and logical inference, with concepts of probability theory and machine learning [22].

Many models in which statistical and relational knowledge

are unified within a single representation formalism have emerged [23, 24, 25]. Among them, Markov Logic Networks (MLNs) [26], which combine first-order logic and probabilistic graphical models (Markov networks) have received considerable attention in recent years. Their popularity is due to their expressiveness and simplicity for compactly representing a wide variety of knowledge and reasoning about data with complex dependencies. Multiple learning and inference algorithms for MLNs have been proposed, for which open-source implementations are available, for example the *Alchemy*[1] and *ProbCog* [2] software packages. MLNs have thus been used for many tasks in artificial intelligence (AI), such as meaning extraction [27], collective classification [28], or entity resolution [29].

We are interested in providing a multi-level description of music structure, at the analysis frame, phrase and global structure scale, in which information specific to the various strata interact. This paper presents some steps towards this direction. In traditional computational models, it is not easy to express dependencies between various semantic levels. In [30], we have introduced MLNs as a highly flexible and expressive formalism for the analysis of music audio signals, showing that chord and key information can be jointly modeled into a single unified MLN model. In this work we show that the MLNs framework can be further explored to integrate information at different time scales within a single formalism.

More specifically, we consider here the problem of modeling the harmonic progression of a music signal at the analysis-frame level, taking into account a more global semantic level. A number of works focus on the task of automatic analysis of the musical structure from audio signals, see e.g. [31, 32, 33, 34]. Music pieces are structured at several time scales, from musical phrases to longer sections that generally have multiple occurrences (with possible variations) within the same musical piece. Each segment type can be categorized and distinguished from the others according to several parameters such as the timbre, the musical key, the chord progression, the tempo progression etc. Here, we focus on popular music where pieces can be segmented into specific repetitive segments with labels such as *chorus*, *verse*, or *refrain*. Segments are considered as similar if they represent the same musical content, regardless of their instrumentation. In particular, two same sections are likely to have similar harmonic structures. In this work, we use this structural information to obtain mid-level representations of music in terms of chord progression that has a "structural consistency" [35].

Previous works have already used the structure as a cue to obtain a "structurally consistent" mid-level representation of music. In the work of Dannenberg [35], music structure is used to constrain a beat tracking program based on the idea that similar segments of music should have corresponding beats and tempo variation. A work more closely related to this article is [36] in which the repetitive structure of songs is used to enhance chord extraction. A chromagram is extracted from the signal, and segments corresponding to a given type of section are replaced by the average of the chromagram over all the instance of the same segment type over the whole song, so that similar structural segments are labelled with the exact same chord progression. A limitation of this work is that it relies on the hypothesis that the chord sequence is the same in all sections of the same type. However, repeated segments are often transformed up to a certain extent and present variations between several occurrences [37]. Moreover, in the case that one segment of the chromagram is blurred (e.g. because of noise or percussive sounds), this will automatically affect all same segments, and thus degrade the chord estimation.

[1] `http://alchemy.cs.washington.edu`

[2] `http://ias.cs.tum.edu/research/probcog`

Here, we show that prior structural information can be used to enhance chord estimation in a more elegant and flexible way within the framework of Markov Logic Networks. We do not constrain the model to have the exact same chord progression in all sections of the same type, but we only *favor* same chord progressions for all instances of the same segment type, so that variations between similar segments can be taken into account. Moreover, the proposed formalism has a good potential of improvement in the future by incorporating more context information and discovering new predicates.

Although our final goal is to develop a fully automatic model where an automatic segmentation is used, in this article, the segmentation of the song in beats and in structure is given as prior information. As in [36], structure information within a given song is incorporated relying on segment types whose instances are harmonically similar and also have the same length in beats [3].

2. MARKOV LOGIC NETWORKS

A Markov Logic Network (MLN) is a set of weighted first-order logic formulas [26], that can be seen as a template for the construction of probabilistic graphical models. We present a short overview of the underlying concepts with specific examples from the modeling of chord structure. A MLN is a hybrid of Markov networks and first-order logic. A *Markov network* [38] is a model for the joint distribution of a set of variables $X = (X_1, X_2, ..., X_n) \in \mathcal{X}$, that is often represented as a log-linear model:

$$P(X = x) = \frac{1}{Z}\exp(\sum_j w_j f_j(x)) \quad (1)$$

where Z is a normalization factor, and the value $f_j(x)$ are features associated with state x (x is an assignment to the random variables X). Here, we will focus on binary features, $f_j(x) \in 0, 1$.

A first-order domain is defined by a set of *constants* (that is assumed finite) representing objects in the domain (e.g. CMchord, GMchord) and a set of *predicates* representing properties of those objects (e.g. IsMajor(x), IsHappyMood(x)) and relations between them (e.g. AreNeighbors(x, y)). A predicate can be *grounded* by replacing its variables with constants (e.g. IsMajor(CMchord), IsHappyMood(CMchord), AreNeighbors(CMchord, GMchord)). A *world* is an assignment of a truth value to each possible ground predicate (or atom). A *first-order knowledge base* (KB) is a set of formulas in first-order logic, constructed from predicates using logical connectives and quantifiers. For instance, the knowledge "*Major chords imply happy mood*" can be described using the formula $\forall x, IsMajor(x) \Rightarrow IsHappyMood(x)$. A first-order KB can be seen as a set of hard constraints on the set of possible worlds: if a world violates even one formula, it has zero probability. In real world schemes, logic formulas are *generally* true, but not *always*. The basic idea in Markov logic is to soften these constraints to handle uncertainty: a world that violates one formula in the KB is less probable than one that does not violate any formula but not impossible. The weight associated with each formula reflects how strong a constraint is, i.e. how unlikely a world is in which that formula is violated.

Formally, a *Markov logic network* L is defined [26] as a set of pairs (F_i, w_i), where F_i is a formula in first-order logic and w_i is a real number associated with the formula. Together with a finite set of constants C (to which the predicates appearing in the formulas can be applied), it defines a ground Markov network $M_{L,C}$, as follows:

1. $M_{L,C}$ contains one binary node for each possible grounding of each predicate appearing in L. The node value is 1 if the ground predicate is true, and 0 otherwise.

[3] Instances of a segment type may differ in length within the song. In such a case, following [36], to fulfill the requirement of equal length instances, only the part of the segment type that is similar in all instances is considered. The remaining parts are labeled as additional one instance-segments.

2. $M_{L,C}$ contains one feature for each possible grounding of each formula F_i in L. The feature value is 1 if the ground formula is true, and 0 otherwise. The feature weight is the w_i associated with F_i in L.

A ground Markov logic network specifies a probability distribution over the set of possible worlds $\mathcal{X}$, i.e. the set of possible assignments of truth values to each of the ground atoms in X. The joint distribution of a possible world x is:

$$P(X = x) = \frac{1}{Z} exp(\sum_i w_i n_i(x)) = \frac{exp(\sum_i w_i n_i(x))}{\sum_{x' \in \mathcal{X}} exp(\sum_i w_i n_i(x'))}$$

where the sum is over indices of MLN formulas and $n_i(x)$ is the number of true groundings of formula F_i in x (i.e. $n_i(x)$ is the number of times the i^{th} formula is satisfied by possible world x).

3. MODEL

We now present a MLN for modeling the chord progression incorporating *a priori* structural information. The front-end of our model is based on the extraction from the signal of chroma features [39] that are 12-dimensional vectors representing the intensity of the 12 semitones of the Western tonal music scale, regardless of octave. We perform a *beat synchronous* analysis and compute one chroma vector per beat. A chord lexicon composed of 24 major M and minor m triads is considered (CM, ..., BM, Cm, ..., Bm).

The structure of the domain is represented by a set of weighted logical formulas that are described in Table 1. Given this set of rules with attached weights and a set of evidence literals, described in Table 2, Maximum A Posteriori (MAP) inference is used to infer the most likely state of the world. Structural information both at the beat-synchronous and at global semantic level are added using two time predicates at multiple time-scale, $Succ$ and $SuccStr$.

Table 1. MLN for joint chord and structure description.

Predicate declarations	
$Observation(chroma!, time)$	$Succ(time, time)$
$Chord(chord!, time)$	$SuccStr(time, time)$
Weight	*Formula*
Prior observation chord probabilities:	
$log(P(CM(t=0)))$	$Chord(CM, 0)$
...	...
$log(P(Bm(t=0)))$	$Chord(Bm, 0)$
Probability that the observation (chroma) has been emitted by a chord:	
$log(P(o_0\|CM))$	$Observation(o_0, t) \wedge Chord(CM, t)$
$log(P(o_0\|C\#M))$	$Observation(o_0, t) \wedge Chord(C\#M, t)$
...	...
$log(P(o_{N-1}\|Bm))$	$Observation(o_{N-1}, t) \wedge Chord(Bm, t)$
Probability of transition between two successive chords:	
$log(P(CM\|CM))$	$Chord(CM, t_1) \wedge Succ(t_2, t_1) \wedge Chord(CM, t_2)$
$log(P(C\#M\|CM))$	$Chord(CM, t_1) \wedge Succ(t_2, t_1) \wedge Chord(C\#M, t_2)$
...	...
$log(P(Bm\|Bm))$	$Chord(Bm, t_1) \wedge Succ(t_2, t_1) \wedge Chord(Bm, t_2)$
Probability that similar segments have the same chord progression:	
w_{struct}	$Chord(CM, t_1) \wedge SuccStr(t_2, t_1) \wedge Chord(CM, t_2)$
w_{struct}	$Chord(C\#M, t_1) \wedge SuccStr(t_2, t_1) \wedge Chord(C\#M, t_2)$
...	...
w_{struct}	$Chord(Bm, t_1) \wedge SuccStr(t_2, t_1) \wedge Chord(Bm, t_2)$

Table 2. Evidence for chord and structure description.

// We observe a chroma at each time frame:
$Observation(o_0, 0) \cdots$
$Observation(o_{N-1}, N-1)$
// We know the temporal order of the frames:
$Succ(1, 0) \cdots$
$Succ(N-1, N-2)$
//Prior information about similar segments in the structure:
$SuccStr(1, 10)$
$SuccStr(2, 11) \cdots$

3.1. Beat-Synchronous Time-Scale

The chord progression at the beat-synchronous frame level can be modeled by a classic ergodic 24-state HMM such as the one presented in [4][4], each hidden state corresponding to a chord of the lexicon, and the observations being the chroma vectors. The HMM is specified by the prior, observation and transition probabilities distributions. As we show in [30], the chord progression can be equivalently modeled in the MLN framework considering three generic formulas, described in Eqs. (2, 4, 6), that reflect the constraints given by the three distributions given by the HMM. This model does not consider high-level structural relationships and will be referred to as *MLN_chord* in what follows. It is briefly described below.

Let $c_i, i \in [1, 24]$ denote the 24 chords of the dictionary, $o_n, n \in [0, N-1]$ denote the succession of observed chroma vectors, n being the time index, and N being the total number of beat-synchronous frames of the analyzed song, and $s_n, n \in [0, N-1]$ denotes the succession of hidden states.

To model the chord progression at the beat-synchronous frame level, we use an unobservable predicate $Chord(c_i, t)$, meaning that chord c_i is played at frame t, and two observable ones, the predicate $Observation(o_n, t)$, meaning that we observe chroma o_n at frame t, and the temporal predicate$Succ(t_1, t_2)$, meaning that t_1 and t_2 are successive frames.

The prior observation probabilities are described using:

$$log(P(s_0 = c_i)) \quad Chord(c_i, 0) \tag{2}$$

for each chord $c_i, i \in [1, 24]$, and with $P(s_0)$ denoting the prior distribution of states.

The conditional observation probabilities are described using a set of conjunctions of the form:

$$\forall t \in [0, N-1] \quad log(P(o_n|s_n = c_i)) \quad Observation(o_n, t) \wedge Chord(c_i, t) \tag{3}$$

for each combination of observation o_n and chord c_i, and with $P(o_n|s_n)$ denoting the corresponding observation probability. Note that conjunctions, by definition, have but one true grounding each.

The transition probabilities are described using:

$$\forall t_1, t_2 \in [0, N-1] \quad log(P(s_n = c_i|s_{n-1} = c_j)) \quad Chord(c_i, t_1) \wedge Succ(t_2, t_1) \wedge Chord(c_j, t_2) \tag{4}$$

for all pairs of chords $(c_i, c_j), i, j \in [1, 24]$, and with $P(s_n|s_{n-1})$ denoting the corresponding transition probability.

The weights attached to formulas can be obtained from training. However, in this work, following [30, 4] weights are based on musical knowledge. The distribution $P(s_0)$ over initial states is chosen as uniform. The observation distribution probabilities $P(o_n|s_n)$ are obtained by computing the correlation between the observation vectors (the chroma vectors) and a set of chord templates which are the theoretical chroma vectors corresponding to the 24 major and minor triads. A state-transition matrix based on musical knowledge [40] is used to model the transition probabilities $P(s_n|s_{n-1})$, reflecting chord transition rules. More details can be found in [30, 4].

Note that for each conditional distribution, only mutually exclusive and exhaustive sets of formulas are used, *i.e.* exactly one of them is true. For instance, there is one and only one possible chord per frame. This is indicated in Table 1 using the symbol !.

Evidence consists of a set of ground atoms that give chroma observations corresponding to each frame, and the temporal succession of frames over time using the beat-level temporal predicate $Succ$.

3.2. Global Semantic Structure Time-Scale

Prior structural information at the global semantic level, based on the idea that segments of the same type have a similar chord progression, is incorporated using the time predicate $SuccStr$. This predicate allows considering wider temporal windows, as opposed to consecutive frames via the $Succ$ predicate.

The position of segments of same type in the song is given as evidence. Let K denote the number of distinct segments. Each segment $s_k, k \in [1, K]$ may be characterized by its beginning position (in frames) $b_k \in [1, N]$, and its length in beats l_k. For each pair of same

[4]Model evaluated during the MIREX 2009 contest.

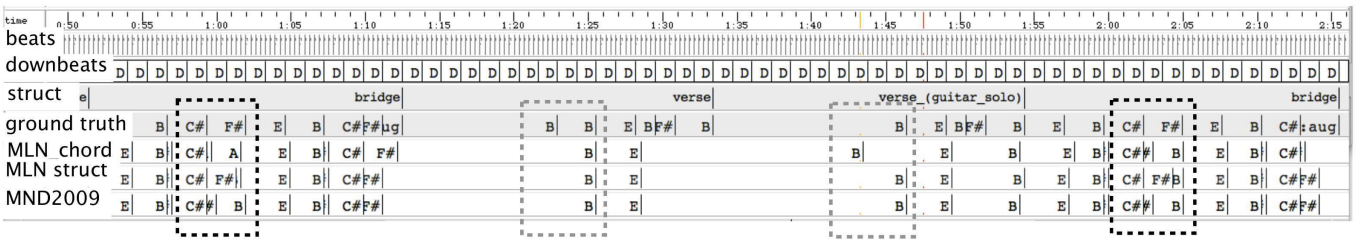

Fig. 1. Chord estimation results for an excerpt of the song *One After 909*.

segment type $(s_k, s_{k'})$, the position of matching beat-synchronous frames (likely to be the same chord type) is given as evidence[5]:

$$\begin{matrix} SuccStr(s_k(b_k), s'_k(b_{k'})) \\ \cdots \\ SuccStr(s_k(b_k + l_k - 1), s'_{k'}(b_{k'} + l_{k'} - 1)) \end{matrix} \quad (5)$$

The following set of formulas is added to the Markov logic network to express how strong the constraint that two same segments have a similar chord progression is:

$$\begin{matrix} \forall t_1, t_2 \in [0, N-1] \quad w_{struct} \\ Chord(c_i, t_1) \wedge SuccStr(t_2, t_1) \wedge Chord(c_i, t_2) \end{matrix} \quad (6)$$

for all chord $c_i, i \in [1, 24]$, and with weight w_{struct}, reflecting how strong the constraint is, manually set. In practice, w_{struct} will be a small positive value (in our experiments, $p_{struct} = -log(0.95)$) to favor similar chord progressions in same segment types.

This model that incorporates prior information on global semantic structure will be referred to as *MLN_struct* in what follows.

3.3. Inference

The inference task consists of computing the answer to a query (here the chord progression), and finding the most probable state of the world y given some evidence x. Specifically, Maximum A Posteriori (MAP) inference, finds the most probable state given the evidence. For inference, we used the exact solver toulbar2 branch & bound MPE inference [41] with the ProbCog toolbox, which graphic interface allows convenient editing of the MLN predicates and formulas given as input to the algorithm.

4. EVALUATION

The proposed model has been tested on a set of hand-labeled Beatles songs, a popular database used for the chord estimation task [42]. All the recordings are polyphonic, multi-instrumental songs containing drums and vocal parts. We map the complex chords in the annotation (such as major and minor 6^{th}, 7^{th}, 9^{th}) to their root triads. The original set comprises of Beatles songs but we reduced it to 143 songs, removing songs for which the structure was ambiguous (perceptually ambiguous metrical structure or segment repetitions)[6].

We compare the results of the model *MLN_struct* with the baseline method *MLN_chord*, and with the baseline method modified to account for the structure in a similar way to [36], by replacing chromagram portions of same segments types by their average. Note that the basis signal features (chroma) are the same for all three methods.

The results obtained with the various configurations of the proposed model are described in Table 3. The label accuracy *LA* chord estimation results correspond to the mean and standard deviation of correctly identified chords per song. Paired sample t-tests at the 5% significance level were performed to determine whether there is statistical significance in the results between different configurations.

The proposed approach compactly encodes physical signal content and higher-level semantic information in a unified formalism.

[5]Note that the values $s_k(b_k), \ldots, s'_{k'}(b_{k'} + l_{k'} - 1)$ in Eq.(5) correspond to beat time-instants. Note also that $l_{k'} = l_k$.

[6]The list of this subset can be found in `http://opihi.cs.uvic.ca/icassp2013mln.html`.

Table 3. Chord results obtained with various methods. *Stat. Sig.*: statistical significance between the model *MLN_struct* and others.

	Chord LA results	*Stat. Sig.*
MLN_chord	72.57 ± 13.51	}*yes*
MLN_struct	74.03 ± 13.90	
[36]	73.90 ± 13.79	}*no*

Results show that global semantic information can be concisely and elegantly combined with information at the analysis frame scale so that chord estimation results are significantly improved, and more consistent with the global structure, as illustrated in Figure 1 (see the gray dashed rectangles, *MLN_chord* and *MLN_struct*).

The results obtained with the proposed model fairly compare with the previously proposed approach [36] that uses global structure information to enhance chord estimation. Moreover, the proposed model allows for taking into account variations between segments by favoring instead of exactly constraining the chord progression to be the same for segments of the same type, as illustrated in Figure 1. In the bridge sections, in the black dashed rectangles, the underlying harmony is $F\#$ major. In the first instance of the bridge section, the harmony is disturbed by a descending chromatic scale in the bass, which is not the case for the second instance. Averaging the chromagram of the two instances (as in [36]) results into errors in the chord estimation, whereas in the case of *MLN_struct*, the first instance benefits from the signal content of the second instance and the harmonic content is better estimated.

There is no significant difference between the [36] and *MLN_struct* models, but we expect that other music styles such as jazz music where repetitions of segments result in more complex variations due to improvisation would further benefit from the flexibility of the proposed model. This is left for future work.

5. CONCLUSION AND FUTURE WORK

In this article, we have proposed Markov logic as a formalism that enables intuitive, effective, and expressive reasoning about complex relational structure and uncertainty of music data. Chord and structure are integrated in a single unified formalism, resulting in a more elegant and flexible model, compared to existing more ad-hoc approaches. This work is a new step towards a unified multi-scale description of audio in which information specific to various semantic levels (analysis frame, phrase and global structure) interact.

Future work will focus on extending this approach to a fully automatic one, by incorporating estimated beats and structure location, possibly using penalties according to the degree of reliability of their estimation. The proposed model has great potential for improvement. It allows for incorporation of other context information by adding new logical rules, and future work will in particular consider combination with the model described in [30]. Relational structure has been derived from background musical knowledge. A major objective is now to explore the use of learning algorithms in the framework of Markov logic to automatically discover and model new structural rules, and to take advantage of the flexibility of the MLN framework to combine this information from training with background music knowledge.

6. REFERENCES

[1] J.A. Burgoyne and L.K. Saul, "Learning harmonic relationships in digital audio with dirichlet-based hidden markov models," in *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, 2005.

[2] J.-F. Paiement, D. Eck, S. Bengio, and D. Barber, "A graphical model for chord progressions embedded in a psychoacoustic space," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2005, pp. 641–648.

[3] J.B. Prince, Schmuckler M.A., and W.F. Thompson, "The effect of task and pitch structure on pitch-time interactions in music," *Memory & Cognition*, vol. 37, pp. 368–381, 2009.

[4] H. Papadopoulos and G. Peeters, "Joint estimation of chords and downbeats," *IEEE Trans. Aud., Sp. and Lang. Proc.*, vol. 19, no. 1, pp. 138–152, 2011.

[5] A.P. Klapuri, A. Eronen, and J. Astola, "Analysis of the meter of acoustic musical signals," *IEEE Trans. Aud., Sp. and Lang. Proc.*, vol. 14, no. 1, pp. 342–355, 2006.

[6] J.A. Burgoyne, L. Pugin, C. Kereliuk, and I. Fujinaga, "A cross validated study of modeling strategies for automatic chord recognition in audio," in *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, 2007, pp. 251–254.

[7] R. Scholz, E. Vincent, and F. Bimbot, "Robust modeling of musical chord sequences using probabilistic n-grams," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processings (ICASSP)*, 2008, pp. 53–56.

[8] K. Yoshii and M. Goto, "A vocabulary-free infinity-gram model for nonparametric bayesian chord progression analysis," in *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, 2011, pp. 645–650.

[9] R.l Ramirez, A. Hazan, E. Maestre, X. Serra, V.A. Petrushin, and L. Khan, *A Data Mining Approach to Expressive Music Performance Modeling*, pp. 362–380, Springer London, 2007.

[10] A. Anglade and S. Dixon, "Towards logic-based representations of musical harmony for classification, retrieval and knowledge discovery," in *Proceedings of the International Workshop on Machine Learning and Music (MML)*, 2008.

[11] S. Muggleton, "Inductive logic programming," *New Generation Computing*, vol. 8, pp. 295–318, 1991.

[12] E. Morales and R. Morales, "Learning musical rules," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 1995, pp. 81–85.

[13] E.F. Morales, "Pal: A pattern-based first-order inductive system," *Machine Learning*, vol. 26, no. 2, pp. 227–252, 1997-02-01.

[14] R.l Ramirez and C. Palamidessi, *Inducing Musical Rules with ILP*, vol. 2916, pp. 502–504, Springer Berlin / Heidelberg, 2003.

[15] A. Anglade and S. Dixon, "Characterisation of harmony with inductive logic programming," in *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, 2008, pp. 63–68.

[16] A. Anglade, R. Ramirez, and S. Dixon, "Genre classification using harmony rules induced from automatic chord transcriptions," in *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, 2009.

[17] A. Anglade, E. Benetos, M. Mauch, and S. Dixon, "Improving music genre classification using automatically induced harmony rules," *Journal of New Music Research*, vol. 39, no. 4, pp. 349–361, 2010.

[18] G. Widmer, "Discovering simple rules in complex data: A meta-learning algorithm and some surprising musical discoveries," *Artificial Intelligence*, vol. 146, no. 2, pp. 129–148, 2003.

[19] M. Dovey, N. Lavrac, and S. Wrobel, *Analysis of Rachmaninoff's piano performances using inductive logic programming (Extended abstract)*, vol. 912, pp. 279–282, Springer Berlin / Heidelberg, 1995.

[20] E. Van Baelen, L. de Raedt, and S. Muggleton, *Analysis and prediction of piano performances using inductive logic programming*, vol. 1314, pp. 55–71, Springer Berlin / Heidelberg, 1997.

[21] L. Getoor and B. Taskar, *Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)*, The MIT Press, 608 p., 2007.

[22] L. de Raedt and K. Kersting, "Probabilistic inductive logic programming," in *Probabilistic Inductive Logic Programming*, L. De Raedt, P. Frasconi, K. Kersting, and S. Muggleton, Eds., vol. 4911 of *Lecture Notes in Computer Science*, pp. 1–27. Springer Berlin / Heidelberg, 2008.

[23] N.J. Nilsson, "Probabilistic logic," *J. Artif. Intell*, vol. 28, pp. 71–87, 1986.

[24] J.Y Halpern, "An analysis of first-order logics of probability," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 1989, vol. 46, pp. 311–350.

[25] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer, "Learning probabilistic relational models," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 1999, pp. 1300–1309.

[26] M. Richardson and P. Domingos, "Markov logic networks," *J. Machine Learning*, vol. 62, pp. 107–136, 2006.

[27] I.M. Bajwa, "Context based meaning extraction by means of markov logic," *Int. J. Computer Theory and Engineering*, vol. 2, no. 1, pp. 35–38, 2010.

[28] R. Crane and L.K. McDowell, "Investigating markov logic networks for collective classification," in *Proceedings of the International Conference on Agents and Artificial Intelligence (ICAART)*, 2012.

[29] P. Singla and P. Domingos, "Memory-efficient inference in relational domains," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2006, pp. 488–493.

[30] H. Papadopoulos and G. Tzanetakis, "Modeling chord and key structure with markov logic," in *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, 2012.

[31] R.B. Dannenberg and N. Hu, "Pattern discovery techniques for music audio," *Journal of New Music Research*, vol. 32, no. 2, pp. 153–163, 2003.

[32] M. Müller and F. Kurth, "Towards structural analysis of audio recordings in the presence of musical variations," *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, pp. 163–163, 2007.

[33] J. Paulus and A. Klapuri, "Music structure analysis using a probabilistic fitness measure and a greedy search algorithm," *IEEE Trans. Aud., Sp. and Lang. Proc.*, vol. 17, no. 6, pp. 1159–1170, 2009.

[34] R.B. Dannenberg, M. Goto, D. Havelock, S. Kuwano, and M. Vorländer, *Music Structure Analysis from Acoustic Signals*, pp. 305–331, Springer New York, 2009.

[35] R. Dannenberg, "Toward automated holistic beat tracking, music analysis, and understanding," in *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, 2005.

[36] M. Mauch, K. Noland, and S. Dixon, "Using musical structure to enhance automatic chord transcription," in *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, 2009.

[37] A. Volk, W. Bas de Haas, and P. van Kranenburg, "Towards modelling variation in music as foundation for similarity," in *Proceedings of the International Conference on Music Perception & Cognition (ICMPC)*, 2012, pp. 1085–1094.

[38] J. Pearl, *Probabilistic reasoning in intelligent systems: Networks of plausible inference*, San Francisco, CA: Morgan Kaufmann, 552 p., 1988.

[39] T. Fujishima, "Real-time chord recognition of musical sound: a system using common lisp music," in *Proceedings of the International Computer Music Conference (ICMC)*, 1999, pp. 464–467.

[40] K. Noland and Sandler M., "Key estimation using a hidden markov model," in *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, 2006.

[41] D. Allouche, S. de Givry, and T. Schiex, "Toulbar2, an open source exact cost function network solver," Tech. Rep., INRA, 2010.

[42] C. Harte, M. Sandler, S. Abdallah, and E. Gómez, "Symbolic representation of musical chords: a proposed syntax for text annotations," in *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, 2005.