MODELING CHORD AND KEY STRUCTURE WITH MARKOV LOGIC

Hélène Papadopoulos and George Tzanetakis
Computer Science Department, University of Victoria
Victoria, B.C., V8P 5C2, Canada
helene.papadopoulos@lss.supelec.fr
gtzan@cs.uvic.ca

ABSTRACT

We propose the use of Markov Logic Networks (MLNs) as a highly flexible and expressive formalism for the harmonic analysis of audio signals. Using MLNs information about the physical and semantic content of the signal can be intuitively and compactly encoded and expert knowledge can be easily expressed and combined using a single unified formal model that combines probabilities and logic. In particular, we propose a new approach for joint estimation of chord and global key The proposed model is evaluated on a set of popular music songs. The results show that it can achieve similar performance to a state of the art Hidden Markov Model for chord estimation while at the same time estimating global key. In addition when prior information about global key is used it shows a small but statistically significant improvement in chord estimation performance. Our results demonstrate the potential of MLNs for music analysis as they can express both structured relational knowledge as well as uncertainty.

1. INTRODUCTION

Content-based music retrieval is an active and important field of research within the Music Information Retrieval (MIR) community, that deals with the extraction and processing of information from musical audio. Many applications, such as music classification or structural audio segmentation, are based on the use of musical descriptors, such as the key, the chord progression, the melody, or the instrumentation. Often regarded as an innate human ability, the automatic estimation of music content information proves to be a highly complex task, for at least two reasons. The first reason is the great variability of musical audio caused by the many modes of sound production and the wide range of possible combinations between the various acoustic events which make music signals extremely rich and complex from a physical point of view. The second reason is that the information of interest is generally very complex from a semantic point of view and many musical descriptors, that are strongly correlated, are necessary to characterize it. For instance, the chord progression is related to the metrical structure of a piece of music: chords change more often on strong beats than on other beat positions in the measure [9]. The chord progression is also related to the musical key: some chords are heard as more stable within an established tonal context [13]. Recent work has shown that the estimation of musical attributes would benefit from a unified musical analysis [4, 14, 15, 21]. However, most of existing MIR systems that estimate musical content from audio signals have relatively simple probabilistic structure and are constrained by limited hypotheses that do not model the underlying complexity of music. The idea of reinforcing the performance of object recognition by considering contextual information has been explored in other fields than MIR, such as computer vision [17].

As many real-world systems and signals, music signals exhibit both uncertainty and complex 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. However, alternative approaches towards a unification have been proposed within the emerging field of Statistical Relational Learning (SRL) [8]. Models in which statistical and relational knowledge are unified within a single representation formalism have emerged [6, 10, 18]. Among them, Markov Logic Networks (MLNs) [27], that 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. Moreover, 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 [2], collective classification [5], or entity resolution [32].

As far as we know, MLNs have not been used yet for music content processing. Chord recognition is one of the most popular MIR tasks as reflected by the number of related papers and the increasing number of contributions to the annual MIREX[3] evaluation. We propose MLNs as a highly flexible and expressive modeling language for es-

[1] http://alchemy.cs.washington.edu
[2] http://ias.cs.tum.edu/research/probcog
[3] http://www.music-ir.org/mirex/

timating the chord progression of a piece of music. The main contribution is to show how various types of information about the physics and the semantics of the signal can be intuitively and compactly encoded in a unified formalism. In addition, MLNs allow incorporating expert knowledge in the model in a flexible fashion. In particular, we show how prior information about the main key of an analyzed excerpt can be used to enhance the chord progression. We also propose a new approach for the estimation of harmonic structure and global key, in which the two attributes are estimated jointly and benefit from each other.

2. BACKGROUND

Previous approaches for chord estimation can be classified into two categories: approaches based on pattern-matching and probabilistic approaches. One of the advantages of probabilistic approaches is that they can model uncertainty and variability. Indeed, the realization of a chord produced in different conditions (instrumentation, dynamics, room acoustics, etc.) can result in significantly different signal observations. Moreover, probabilistic models allow incorporating context information to improve chord estimation. For example, chord transitions based on musical rules can be embedded in the model to improve estimation. A large number of existing algorithms are based on the use of Hidden Markov Models (HMM), see *e.g.* [29, 31]. One of the reasons is that chord transition rules may be incorporated into the state transition matrix of the HMM. In the framework of HMMs, additional context information, such as the key [4, 14], the meter [23] or the structure [16], can also be incorporated to improve the estimation.

Other statistical machine learning approaches for chord estimation include conditional random fields [3], which compared to HMMs do not require the observation vectors to be conditionally independent. The use of N-grams [30, 33] allows information about longer range chord dependencies to be considered. In contrast, HMMs make the Markovian assumption that each chord symbol only depends on the preceding one. In some of these approaches, context information is incorporated, such as in the graphical probabilistic model [20] where contextual information related to the meter is used, or in [15], where a 6-layered dynamic Bayesian network jointly modeling key, metric position, chord and bass pitch class is proposed.

Existing approaches for chord recognition, in particular HMMs, have been quite successful in modeling chord sequences. However, their limited probabilistic structure makes the incorporation of additional contextual information a complex task. More specifically, concerning chords and key interaction, state-of-the-art approaches may not fully exploit interrelationship between musical attributes, as in [24] and [19] where key estimation is based on the chord progression, but the chord estimation part does not benefit from key information. Other approaches [28] do not allow easily introducing expert knowledge (such as musical information about the key progression) that could help music content analysis. In this paper, we intend to show how such relational cues can be compactly modeled within the framework of Markov logic.

3. MARKOV LOGIC NETWORKS

A Markov Logic Network (MLN) is a set of weighted first-order logic formulas [27], 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 combination of Markov networks and first-order logic. A *Markov network* is a model for the joint distribution of a set of variables $X = (X_1, X_2, ..., X_n) \in \mathcal{X}$ [25], that is often represented as a log-linear model:

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

where Z is a normalization factor, and $f_j(x)$ are features of the 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. 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. Table 1 shows a simple KB. In a real world scheme, logic formulas are *generally* true, but not *always* true. The basic idea in Markov logic is to soften these constraints to handle uncertainty: when a world violates one formula in the KB, it is less probable than one that does not violate any formulas, 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.

Table 1. Example of a first-order KB and corresponding weights in the MLN.

Knowledge	*Logic formula*	*Weight*
A major chord implies an happy mood.	∀ x IsMajor(x) ⇒ IsHappyMood(x)	$w_1 = 0.5$
If two chords are neighbors, either the two are major chords or neither are.	∀ x ∀ y AreNeighbors(x, y) ⇒ (IsMajor(x) ⇔ IsMajor(y))	$w_2 = 1.1$

Formally, a *Markov logic network* L is defined [27] 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.

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 $\mathcal{X}$. The joint distribution of a possible world x is:

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

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).

Figure 1 shows the graph of the ground Markov network defined by the two formulas in Table 1 and the constants CMchord and GMchord. Each possible grounding of each predicate becomes a node in the corresponding Markov Network. There is an arc in the graph between each pair of atoms that appear together in some grounding of one of the formulas. The grounding process is illustrated in Figure 2.

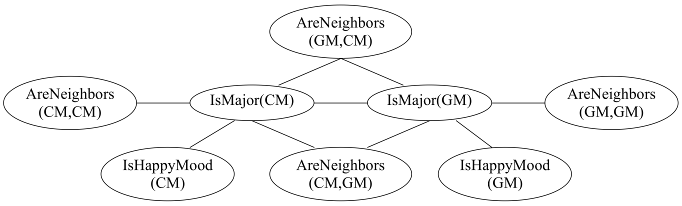

Figure 1. Ground Markov network obtained by applying the formulas in Table 1 to the constants CMchord (CM) and GMchord (GM).

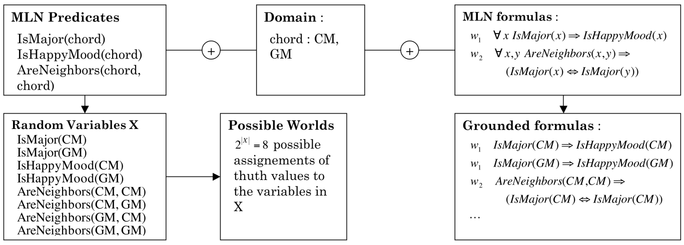
MLN Predicates
IsMajor(chord)
IsHappyMood(chord)
AreNeighbors(chord, chord)
Domain :
chord : CM, GM
MLN formulas :
w_1 $\forall x$ $IsMajor(x) \Rightarrow IsHappyMood(x)$
w_2 $\forall x,y$ $AreNeighbors(x,y) \Rightarrow (IsMajor(x) \Leftrightarrow IsMajor(y))$
Random Variables X
IsMajor(CM)
IsMajor(GM)
IsHappyMood(CM)
IsHappyMood(GM)
AreNeighbors(CM, CM)
AreNeighbors(CM, GM)
AreNeighbors(GM, CM)
AreNeighbors(GM, GM)
Possible Worlds
$2^{|X|} = 8$ possible assignements of thuth values to the variables in X
Grounded formulas :
w_1 $IsMajor(CM) \Rightarrow IsHappyMood(CM)$
w_1 $IsMajor(GM) \Rightarrow IsHappyMood(GM)$
w_2 $AreNeighbors(CM,CM) \Rightarrow (IsMajor(CM) \Leftrightarrow IsMajor(CM))$
...

Figure 2. Illustration of the grounding process of the Ground Markov network in Figure 1. Adapted from [12].

4. PROPOSED MODEL

In this section, we show how we can move from a standard HMM to a MLN, resulting in an elegant and concise representation with flexible modeling of context information.

4.1 Baseline HMM

We utilize a baseline model for chord estimation proposed in [22, 23] and briefly described here. The front-end of our model is based on the extraction of chroma feature vectors [7] that describe the signal. The chroma vectors are 12-dimensional vectors that represent the intensity of the twelve semitones of the Western tonal music scale, regardless of octave. We perform a *beat synchronous* analysis and compute one chroma vector per beat [4]. A chord lexicon composed of $I = 24$ major (M) and minor (m) triads is considered. The chord progression is then modeled as an ergodic 24-state HMM, each hidden state s_n (n denotes the time index) corresponding to a chord of the lexicon (CM, ..., BM, Cm, ..., Bm), and the observations being the chroma vectors o_n.

The HMM is specified using three probability distributions: the distribution $P(s_0)$ over initial states, the transition distribution $P(s_n|s_{n-1})$ and the observation distribution $P(o_n|s_n)$. The state-conditional observation 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 $I = 24$ major and minor triads. A state-transition matrix based on musical knowledge [19] is used to model the transition probabilities $P(s_n|s_{n-1})$, reflecting chord transition rules. The chord progression over time is estimated in a maximum likelihood sense by decoding the underlying sequence of hidden chords $S = (s_1, s_2, \dots, s_N)$ from the sequence of observed chroma vectors $O = (o_1, o_2 \dots, o_N)$ using the Viterbi decoding algorithm :

$$\hat{S} = \underset{S}{\operatorname{argmax}}(p(S, O)). \quad (2)$$

4.2 MLN for Chord Recognition

We now present a MLN for the problem of chord estimation, that is derived from the baseline HMM. MLNs are more general than HMMs, and we describe how the HMM structure can be expressed in a straightforward way using a MLN. Our MLN for chord recognition consists of a set of first-order formulas and their associated weights. It is described in Table 2. Given this set of rules with attached weights and a set of evidence literals, described in Table 3, Maximum A Posteriori (MAP) inference is used to infer the most likely state of the world.

Let $c_i, i \in [1, 24]$ denote the 24 chords of the dictionary, and $o_n, n \in [0, N-1]$ denote the succession of observed chroma vectors, with N being the total number of beat-synchronous frames of the analyzed song. The chord estimation problem can be formulated in Markov logic by defining formulas in the MLN using an unobserved predicate $Chord(c_i, t)$, meaning that chord c_i is played at frame t, and two observed ones, $Observation(o_n, t)$, meaning that we observe chroma o_n at frame t, and $Succ(t_1, t_2)$, meaning that t_1 and t_2 are successive frames. The constraints given by the prior, observation and transition probabilities of the baseline HMM form the abstract model. They are simply described by three MLN generic formulas. 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 2 using the symbol !. The evidence consists of a set of ground atoms that give the chroma observations corresponding to each frame, and the temporal succession of frames over time. The query is the chord progression.

[4] This is done by integrating a beat-tracker as a front-end of the system [26].

Table 2. Chord recognition MLN used for inference.

Predicate declarations	
$Observation(chroma!, time)$ $Chord(chord!, time)$ $Succ(time, time)$	
Weight	*Formula*
Prior observation probabilities:	
$log(P(CM(t=0)))$ $\cdots$ $log(P(Bm(t=0)))$	$Chord(CM, 0)$ $\cdots$ $Chord(Bm, 0)$
Probability that the observation (chroma) has been emitted by a chord:	
$log(P(o_0 \vert CM))$ $log(P(o_0 \vert C\#M))$ $\cdots$ $log(P(o_{N-1} \vert Bm))$	$Observation(o_0, t) \wedge Chord(CM, t)$ $Observation(o_0, t) \wedge Chord(C\#M, t)$ $\cdots$ $Observation(o_{N-1}, t) \wedge Chord(Bm, t)$
Probability to transit from one chord to another:	
$log(P(CM \vert CM))$ $log(P(C\#M \vert CM))$ $\cdots$ $log(P(Bm \vert Bm))$	$Chord(CM, t_1) \wedge Succ(t_2, t_1) \wedge Chord(CM, t_2)$ $Chord(CM, t_1) \wedge Succ(t_2, t_1) \wedge Chord(C\#M, t_2)$ $\cdots$ $Chord(Bm, t_1) \wedge Succ(t_2, t_1) \wedge Chord(Bm, t_2)$

Table 3. Evidence for MLN chord estimation.

// 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:
$Succ(1, 0)$
$\cdots$
$Succ(N-1, N-2)$

In many existing MLNs weights attached to formulas are obtained from training. However, we follow the baseline approach and use weights based on musical knowledge. They are directly obtained using the conditional prior, observation and transition probabilities of the baseline HMM.

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

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

for each combination of observation o_n and chord c_i. Conjunctions, by definition, have but one true grounding each. According to Eq.(2), the weight associated with each conjunction is set to $w = log(P(o_n|s_n = c_i))$, with $P(o_n|s_n)$ denoting the corresponding observation probability.

The transition probabilities are described using:

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

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

The prior observation probabilities are described using:

$$log(P(s_0 = c_i)) \quad Chord(c_i, 0) \qquad (5)$$

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

4.3 Including Prior Information on Key

In this section, we show how prior information about the key of the excerpt can be incorporated in the model. We assume that we know the key $k_i, i \in [1, 24]$ of the excerpt. Key is added as a functional predicate in Table 2 ($Key(key!, time)$) and given as evidence in the MLN by adding evidence predicates in Table 3 of the form:

$$Key(k_i, 0), Key(k_i, 1), \cdots, Key(k_i, N-1) \qquad (6)$$

Relying on the hypothesis that some chords are heard as more stable within an established tonal context [13], additional rules about key and chord relationship are incorporated in the model. Let $k_i, i \in [1, 24]$ denote the 24 major and minor keys and $c_j, j \in [1, 24]$ denote the 24 chords. For each pair of key and chords (k_i, c_j), we add the rule:

$$log(p_{ij}) \quad Key(k_i, t) \wedge Chord(c_j, t) \qquad (7)$$

where the values $p_{ij}, i, j \in [1, 24]$ define the prior distribution of chords $(c_1, \ldots, c_{24})$ given a key k_i. They are obtained from a set of key templates that represent the importance of each triad within a given key. The key templates are 24-dimensional vectors, each bin corresponding to one of the 24 major and minor triads. Two key templates, originally presented in [24], are considered. The first one, referred to as "weighted main chords relative" (WMCR) template, is derived from music knowledge, and attributes non-zero values to the bins corresponding to the most important triads in a given key (those built on the tonic, the subdominant and the dominant, plus the chord relative to the one built on the tonic) [13]. The second one, referred to as "cognitive-based" (CB) template, is built relying on a cognitive experiment conducted by Krumhansl [13], giving values corresponding to the rating of chords in harmonic-hierarchy experiments. Templates corresponding to C major (top) and C minor (bottom) keys are shown in Figure 3.

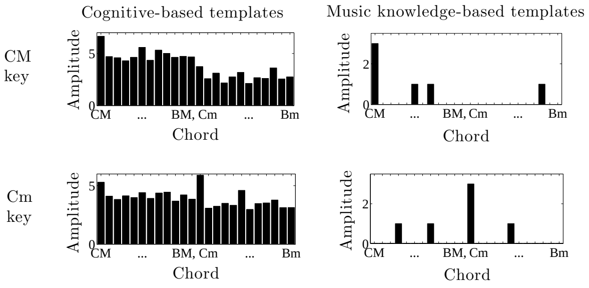

Figure 3. Key templates for chord and key modeling.

4.4 Joint Estimation of Chords and Key

The key can be estimated jointly with the chord progression by simply removing the evidence predicates about key listed in Eq. (6), that give prior information about the key context, and by considering Key as a query along with $Chord$. In addition, we add rules in Table 2 to model key modulations by using the set of formulas:

$$log(p_{ij}^{key}) \quad Key(k_i, t_1) \wedge Succ(t_2, t_1) \wedge Key(k_j, t_2)$$

for all pairs of keys $(k_i, k_j), i, j \in [1, 24]$. The values p_{ij}^{key}, that reflect probability to transit from one key to another, are derived from perceptual tests about proximity between the various musical keys [13]. However, because we focus on global key information in this paper, we manually give a high weight to the formulas corresponding to self-transitions (transition between two same keys) to favor constant key over the analyzed song.

4.5 Inference

The inference step consists of computing the answer to a query, here the chord progression and the key. Specifically, Maximum Probability Explanation (MPE), often denoted as Maximum A Posteriori (MAP) inference, finds the most probable state given the evidence. For inference, we used the toulbar2 branch & bound MPE inference [1], as implemented in the ProbCog toolbox. The graphic interface provided in ProbCog allows convenient editing of the MLN predicates and formulas, which are given as input to the algorithm. The answer to the query can then be directly computed. Although manageable on a standard laptop, the inference step has a high computational cost compared to the baseline algorithm (≈ 2 min (*chord only MLN*), ≈ 4 min (*key MLN*) against 6 sec (HMM, MATLAB) for processing 60s of audio on a MacBook Pro 2.4GHz Intel Core 2 Duo with 2GB RAM).

5. EVALUATION

The proposed model has been tested on a set of hand-labeled Beatles songs, a popular database used for the chord estimation task [11]. 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 180 Beatles songs but we reduced it to 141 songs, removing songs containing key modulations. The list of this subset can be found in [21].

Label accuracy (*LA*) is used to measure how the estimated chord/key is consistent with the ground truth. The *LA* chord estimation results correspond to the mean and standard deviation of correctly identified chords per song. The *LA* key estimation results indicate the percentage of songs for which the key has been correctly estimated. The results obtained with the various configurations of the proposed model are described in Tables 4 and 5. Paired sample t-tests at the 5% significance level are performed to determine whether there is statistical significance in the observed accuracy results between different configurations.

Table 4. Chords label accuracy (*LA*) results. *HMM*: baseline HMM, *Chord MLN*: chord-only MLN, *Prior key MLN*: MLN with prior key information, using the WMCR and CB key templates, *Joint chord/key MLN*: MLN for joint estimation of chords and key. *Stat. Sig.*: statistical significance between the model *Chord MLN* and others.

	Chord LA	*Stat. Sig.*
HMM	72.49 ± 14.68	**no**
Chord MLN	72.33 ± 14.78	
Prior key MLN, WMCR	$\mathbf{73.00 \pm 13.91}$	**yes**
Prior key MLN, CB	72.22 ± 14.48	no
Joint chord/key MLN	72.42 ± 14.46	no

Table 5. Key label accuracy (*LA*) results. *Joint chord/key MLN*: MLN for joint estimation of chords and key. *DTBM-chroma* and *DTBM-chord*: Direct Template-Based Method. Exact Estimation *EE*, Mirex Estimation *ME* and Exact + Neighbor *E+N* scores. *Stat. Sig.*: statistical significance between the model *Joint chord/key MLN* and others.

	EE	*EE*	*E+N*	*Stat. Sig.*
Joint chord/key MLN	82.27	88.09	94.32	
DTBM-chord	48.59	67.39	89.44	yes
DTBM-chroma	75.35	85.14	95.77	yes

The main interest of the proposed model lies in its simplicity and expressivity for compactly encoding physical content and semantic information in a unified formalism. Results show that the HMM structure can be concisely and elegantly embedded in a MLN. Although the inference algorithms used for each model are different, a song by song analysis shows that chord progressions estimated by the two models are extremely similar and the difference in the label accuracy results is not statistically significant.

To illustrate the flexibility of the MLN formalism, we also tested a scenario where some partial evidence about chords was added by adding evidence predicates of the form $Chord(c_i^{GT}, 0)$, $Chord(c_i^{GT}, 9)$, $Chord(c_i^{GT}, 19)$, $\cdots$, $Chord(c_i^{GT}, N-1)$, as prior information of 10% of the ground-truth chords $c_i^{GT}, i \in [1, 24]$. We tested this scenario on the song *A Taste of Honey*, for which the *chord only MLN* estimation results are poor. They were increased from 55.69% to 77.04%, which shows how additional evidence can be easily added and have a significant effect.

The MLN formalism incorporates prior information about key in a simple way. The CB key templates are not relevant for modeling chords given a key on our test-set, whereas the results are significantly better with the WMCR templates, that are more consistent with the harmonic/tonal content of our test-set by clearly favoring the main triads given a key. Incorporating prior information about key with minimal model changes improves the chord estimation results, and the difference is significant (Table 4).

In the *Prior key MLN*, coherent chords with the key context are favored, removing some errors obtained with the chord-only MLN. For instance, Figure 4 shows an excerpt of *Eleanor Rigby*, which is in E minor key. Between $24-30s$, the underlying Em harmony is disturbed by passing notes in the voice. The prior key information favors Em chords and reduces these errors. Prior key information can also reduce confusions due to ambiguous mapping. For instance, the song *The Word*, in DM key, contains several Ddom7 chords (D-F#-A-C), which are mapped to DM (D-F#-A) chords in our dictionary. Many of them are estimated as Dm chords with the *chord MLN*, whereas they are annotated as DM chords with the *Prior key MLN*. Introducing prior key information results in chord estimation that is more coherent with the tonal context.

By considering the key as a query, the proposed model can jointly estimate chords and key. Key estimation is based on the harmonic context, while the chords are estimated given a tonal context. Key information slightly improves the chord estimation results, but the difference is not statistically significant (see Table 4). Results in Table 5 show that the tonal context can be fairly inferred from the chords. Song by song analysis shows that harmonically close errors in the chord estimation (such as dominant or subdominant chords) do not affect the key estimation. Indeed, most of the keys are either correctly estimated or correspond to a neighboring key, as indicated by the MIREX 2007 key estimation score [5] (88.09%) and the *N+E* score (94.32%) that includes harmonically close keys [6] .

Following [24, 28], we compare our key estimation results to a *direct template-based method* (DTBM) that can be viewed as applying the Krumhansl-Schmuckler (K-S) key-finding algorithm [13] to the analyzed excerpt. We compute the correlation between a 12-dimensional vector that averages chroma vectors over time and the 24 key templates (*DTBM-chroma*) by Krumhansl. The estimated key is selected as the one that gives the highest value. To com-

[5] 1 for correct key, 0.5 for perfect fifth detection, 0.3 for relative major/minor, and 0.2 for parallel major/minor

[6] Parallel, relative, dominant or subdominant.

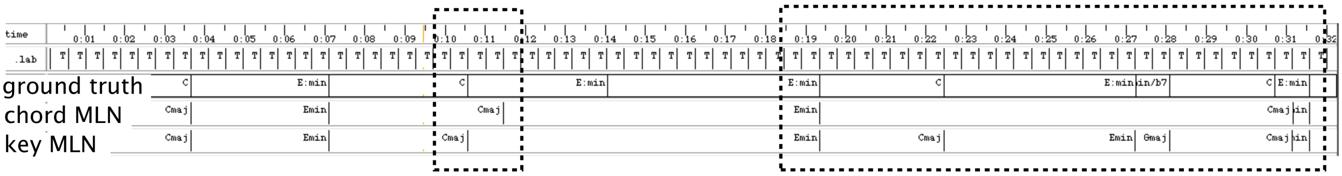

Figure 4. Chord estimation results for an excerpt of the song *Eleanor Rigby*.

pare the performances of the *Prior key MLN* with a baseline algorithm that estimates key from chords after they are predicted, we also report results obtained with a slightly modified version of the K-S algorithm that uses estimated chords instead of chroma: we compute the correlation between a 24-dimensional vector that accumulates the estimated chords over time (considering their duration) and the CB / WMCR templates (*DTBM-chord*)[7] . Results are presented in Table 5. In the *DTBM-chord* approach, errors in the estimation of the chord progression are propagated to the key estimation step, which explains the low *EE* results obtained. The results obtained with *DTBM-chroma* approach are higher, but in both cases, our model performs significantly better than the DTBM methods.

6. CONCLUSION AND FUTURE WORKS

In this article, we have introduced Markov logic networks as an expressive formalism to estimate music content from an audio signal. The results obtained with the *chord MLN* for the task of chord progression are equivalent to those obtained with the baseline *HMM*. Moreover, it allows introducing expert knowledge to enhance the estimation. We have focused on global key information. The model can be extended to local key estimation, which will be the purpose of future work. The proposed model has a great potential of improvement in the future. Context information (such as metrical structure, instrumentation, music knowledge, chord patterns, etc.) can be compactly and flexibly embedded in the model moving toward a unified analysis of music content. Training approaches will be considered. In particular, we will focus on the task of constructing new formulas by learning from the data and creating new predicates by composing base predicates, to compactly capture much more general regularities (predicate invention). As far as we know, Markov logic network have not been used for music content processing yet. We believe that this framework that combines ideas from logic and probabilities opens new interesting perspectives for our field.

7. ACKNOWLEDGMENT

The authors gratefully thank D. Jain for his help.

8. REFERENCES

[1] D. Allouche, S. de Givry, and T. Schiex. Toulbar2, an open source exact cost function network solver. Technical report, INRA, 2010.

[2] I.M. Bajwa. Context based meaning extraction by means of markov logic. *Int. J. Computer Theory and Engineering*, 2(1), 2010.

[3] J.A. Burgoyne, L. Pugin, C. Kereliuk, and I. Fujinaga. A Cross Validated Study Of Modelling Strategies For Auromatic Chord Recognition In Audio. In *ISMIR*, 2007.

[4] J.A. Burgoyne and L.K. Saul. Learning harmonic relationships in digital audio with Dirichlet-based hidden Markov models. In *ISMIR*, 2005.

[5] R. Crane and L.K. McDowell. Investigating markov logic networks for collective classification. In *ICAART*, 2012.

[6] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer. Learning probabilistic relational models. In *IJCAI*, 1999.

[7] T. Fujishima. Real-time chord recognition of musical sound: a system using common lisp music. In *ICMC*, 1999.

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

[9] M. Goto. An audio-based real-time beat tracking system for music with or without drum sounds. *J. New Music Res.*, 30(2), 2001.

[10] J.Y Halpern. An analysis of first-order logics of probability. In *IJCAI*, 1989.

[11] C. Harte, M. Sandler, S. Abdallah, and E. Gómez. Symbolic representation of musical chords: a proposed syntax for text annotations. In *ISMIR*, 2005.

[12] D. Jain. Knowledge engineering with markov logic networks: A review. In *KR*, 2011.

[13] C.L. Krumhansl. *Cognitive foundations of musical pitch*. Oxford University Press, New York, NY, USA, 1990.

[14] K. Lee and M. Slaney. Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio. *IEEE TASLP*, 16(2):291–301, 2008.

[15] M. Mauch and S. Dixon. Automatic chord transcription from audio using computational models of musical context. *IEEE TASLP*, 18(6), 2010.

[16] M. Mauch, K. Noland, and S. Dixon. Using musical structure to enhance automatic chord transcription. In *ISMIR*, 2009.

[17] Kevin Murphy, Antonio Torralba, and William T. Freeman. Graphical model for recognizing scenes and objects. In *Advances in Neural Information Processing Systems 16*. MIT Press, 2004.

[18] N.J. Nilsson. Probabilistic logic. *J. Artif. Intell*, 28:71–87, 1986.

[19] K. Noland and Sandler M. Key estimation using a hidden Markov model. In *ISMIR*, 2006.

[20] J.-F. Paiement, D. Eck, S. Bengio, and D. Barber. A graphical model for chord progressions embedded in a psychoacoustic space. In *ICML*, 2005.

[21] H. Papadopoulos. *Joint Estimation of Musical Content Information From an Audio Signal*. PhD thesis, Univ. Paris 6, France, 2010.

[22] H. Papadopoulos and G. Peeters. Large-Scale Study of Chord Estimation Algorithms Based on Chroma Representation and HMM. In *CBMI*, 2007.

[23] H. Papadopoulos and G. Peeters. Joint estimation of chords and downbeats. *IEEE TASLP*, 19(1), 2011.

[24] H. Papadopoulos and G. Peeters. Local Key Estimation from an Audio Signal Relying on Harmonic and Metrical Structures. *IEEE TASLP*, 2011.

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

[26] G. Peeters. Beat-marker location using a probabilistic framework and linear discriminant analysis. In *DAFx*, 2009.

[27] M. Richardson and P. Domingos. Markov logic networks. *J. Machine Learning*, 62, 2006.

[28] T. Rocher, M. Robine, P. Hanna, and L. Oudre. Concurrent Estimation of Chords and Keys From Audio. In *ISMIR*, 2010.

[29] M.P. Ryynänen and A.P. Klapuri. Automatic transcription of melody, bass line, and chords in polyphonic music. *Comp. Mus. J.*, 32(3), 2008.

[30] R. Scholz, E. Vincent, and F. Bimbot. Robust modeling of musical chord sequences using probabilistic N-grams. In *ICASSP*, 2008.

[31] A. Sheh and D.P.W. Ellis. Chord segmentation and recognition using EM-trained HMM. In *ISMIR*, 2003.

[32] P. Singla and P. P. Domingos. Memory-efficient inference in relational domains. In *AAAI*, 2006.

[33] K. Yoshii and M. Goto. A Vocabulary-Free Infinity-Gram Model for Nonparametric Bayesian Chord Progression Analysis. In *ISMIR*, 2011.

[7] We tested several segment durations and chord/key templates and report the results for the best configuration (segment length of 45s).