A Unified Probabilistic Model of Note Combinations and Chord Progressions

Kazuyoshi Yoshii **Matthias Mauch** **Masataka Goto**
National Institute of Advanced Industrial Science and Technology (AIST), Japan
{k.yoshii,m.goto}@aist.go.jp mail@matthiasmauch.net

Abstract

This paper presents a unified simultaneous and sequential model for note combinations and chord progressions. In chord progression analysis, n-gram models have often been used for modeling temporal sequences of *chord labels* (e.g., C major, D minor, and E# seventh). These models require us to specify the value of n and define a limited vocabulary of chord labels. On the other hand, our model is designed to directly modeling temporal sequences of *note combinations* without specifying the value of n, because we aim to use our model as a prior distribution on musical notes in polyphonic music transcription. To do this, we extend a nonparametric Bayesian n-gram model that was designed for modeling sequences of words in the field of computational linguistics. More specifically, our model can accept any combinations of notes as chords and allows each chord appearing in a sequence to have an unbounded and variable-length context. All possibilities of n are taken into account when predicting a next chord given precedent chords. Even when an unseen note combination (chord) emerges, we can estimate its n-gram probability by referring to its 0-gram probability, i.e., the combinatorial probability of note components. We tested our model by using the ground-truth chord annotations and automatic chord recognition results of the Beatles songs.

1 Introduction

Chord progression analysis forms a basis of music information research [1, 2]. Because the chord patterns used in musical pieces reflect the composer styles and musical genres, statistical models of chord patterns are considered to be useful for composer identification and genre classification. So far, n-gram models have often been used for chord progression analysis [2] because chord sequences exhibit strong short-term dependency. Note that an n-gram is defined as a subsequence of n chords in a chord sequence and n-gram models are based on $(n-1)$-order Markovian assumption that each chord in a chord sequence is assumed to depend on its $n-1$ precedent chords called a context. Using a limited amount of chord sequences, we aim to learn a probabilistic model that can compute the predictive probability (n-gram probability) of a next chord, given any context of length $n-1$. However, the observed n-grams are a limited subset of all kinds of n-grams because the number of all kinds of n-grams increases exponentially with increasing n. Therefore, the naive estimates of the probabilities of unobserved n-grams are zero. This means that a smoothing technique is required to interpolate sparse higher-order n-gram probabilities with reliable lower-order n-gram probabilities. In addition, we should tackle another problem that it is not appropriate to uniquely specify the value of n in advance because there are various chord patterns of different lengths.

To solve these problems, we focus on nonparametric Bayesian n-gram models that have been developed for modeling word sequences in the field of computational linguistics (CL). For example, the first probabilistic model of n-grams called a hierarchical Pitman-Yor language model (HPYLM) was proposed by Teh [3] to enable principled hierarchical smoothing. Mochihachi and Sumita [4] proposed a variable-order Pitman-Yor language model (VPYLM or infinity-gram model) that allows each word to depend on a different number of precedent words. More specifically, they extended the HPYLM by using the stick-breaking process (GEM distribution) as a prior distribution on n. This is a complete generative model of n-grams when a finite vocabulary of words is defined.

Table 1: Average perplexities for ground-truth data [6] via 10-fold cross validation

Training data	WB	GT	IKN	MKN	HPYLM	VPYLM	Proposed
Ground-truth data [6]	38.4	24.3	18.5	17.8	18.0	15.8	**14.5**
Recognition results [7]	39.3	28.7	33.1	29.0	27.1	22.5	**22.0**

In this paper we aim to formulate a nonparametric Bayesian n-gram model that could be used as a prior distribution on musical notes in music transcription based on probabilistic acoustic models (e.g., [5]). Such an n-gram model should be learned directly from musical notes, not from temporal sequences of conventional chord labels (e.g., C major, D minor, and E# seventh). To do this, the VPYLM is extended to allow any note combinations to form "musical units" (not limited to conventional chords). This enables us to estimate how likely multiple notes are to be simultaneously generated when precedent note combinations are given, without using a vocabulary of chord labels.

2 Nonparametric Bayesian Modeling

Our vocabulary-free infinity-gram model can be used for assessing *temporal sequences of note combinations*. We should therefore directly represent the generative process of musical notes in a probabilistic manner. We start with explaining the HPYLM [3], where a vocabulary of chord labels $\boldsymbol{W}$ and the value of n are given. The HPYLM is formulated by layering PYs in a hierarchical Bayesian manner. Suppose we have an n-gram distribution $G_{\boldsymbol{u}}$ over $\boldsymbol{W}$, where $\boldsymbol{u}$ is a context. An $n-1$-gram distribution $G_{\boldsymbol{u}^*}$ given the shortened context $\boldsymbol{u}^*$ is somewhat similar to $G_{\boldsymbol{u}}$. Here $G_{\boldsymbol{u}}$ is assumed to be drawn from a PY with base measure $G_{\boldsymbol{u}^*}$ as $G_{\boldsymbol{u}} \sim \mathrm{PY}(d_n, \theta_n, G_{\boldsymbol{u}^*})$, where d_n and θ_n are discount and strength parameters. Such a process can be recursively defined. Finally, the unigram distribution G_ϕ is given by $G_\phi \sim \mathrm{PY}(d_0, \theta_0, G_0)$ where G_0 is a global base measure over $\boldsymbol{W}$ (0-gram distribution), which is usually assumed to be an uniform distribution. However, this is not a reasonable assumption because different chord labels are not independent from each other.

To include all note combinations into $\boldsymbol{W}$ as musical units (simply called "chords" here), we propose a new global base measure based on note components of chords. This enables us to incorporate the dependency between different chords that share the same note degrees and expand the size of $\boldsymbol{W}$ to infinity in theory. In addition, the value of n is considered to diverge to infinity as in the VPYLM [4]. For Bayesian inference and parameter learning, we use the collapsed Gibbs sampling.

3 Evaluation and Future Direction

To conduct an experiment, we used chord-labels annotations on 137 major-scale songs of the Beatles [6] and transposed them to C major scale. Each label was converted to a combination of a root note and a 12-dimensional binary vector whose elements indicate the existences of the corresponding degrees, e.g., C major $\rightarrow$ C:100010010000 and D minor $\rightarrow$ D:100100010000 ("N" is a special symbol representing silence or untuned). The total number of all kinds of note combinations was 49153 (=$12 * 2^{12} + 1$). We compared our model with six existing methods: Good-Turing (GT), Witten-Bell (WB), interpolated Kneser-Ney (IKN), modified Kneser-Ney (MKN), HPYLM, and VPYLM. These models were trained by using the ground-truth annotations [6] or the chord-recognition results [7]. To evaluate the predictive performance, we measured perplexities via 10-fold cross validation. As listed in Table 1, the experimental results showed that our model significantly outperformed the other methods in the both cases. In the future, we plan to integrate our model with an acoustic model for improved multipitch analysis and chord recognition of polyphonic music audio signals.

Acknowledgments: This study was partially supported by the JSPS KAKENHI 23700184.

References

[1] J.-F. Paiement, D. Eck, S. Bengio, and D. Barber. A graphical model for chord progressions embedded in a psychoacoustic space. *ICML*, pp. 641–648, 2005.

[2] R. Scholz, E. Vincent, and F. Bimbot. Robust modeling of musical chord sequences using probabilistic N-grams. *ICASSP*, pp. 53–56, 2009.

[3] Y. W. Teh. A Bayesian interpretation of interpolated Kneser-Ney. Tech. Repo. TRA2/06, NUS, 2006.

[4] D. Mochihashi and E. Sumita. The infinite Markov model. *NIPS*, pp. 1017–1024, 2007.

[5] K. Yoshii and M. Goto. Infinite latent harmonic allocation: A nonparametric Bayesian approach to multipitch analysis. *ISMIR*, pp. 309–314, 2010.

[6] C. Harte, M. Sandler, S. Abdallah, and E. Gómez. Symbolic representation of musical chords: A proposed syntax for text annotations. *ISMIR*, pp.66–71, 2005.

[7] Matthias Mauch. *Automatic Chord Transcription Using Computational Models of Musical Context*. PhD thesis, Queen Mary, University of London, 2010.