TOPIC MODELS REVEAL SCALE SYSTEMS IN POPULAR MUSIC

David R. W. Sears
Texas Tech University
david.sears@ttu.edu

Justin R. Glosson
Texas Tech University
justin.glosson@ttu.edu

ABSTRACT

This late-breaking demo explores the potential for topic models to discover scale systems in triadic corpora representing both the common-practice and popular music traditions.

1. INTRODUCTION

Triadic corpora associated with common-practice music typically include annotations for both key and mode, but those associated with popular music generally eschew modal or other scalar annotations due to the difficulty of the task [1, 2]. As a result, recent attempts to model various aspects of harmonic organization in popular music necessarily conflate several scale systems [3, 4]. Nevertheless, several music-analytic studies over the last few decades have explored recurrent modal and pentatonic progressions in popular music using close-reading methodologies (e.g., [5, 6]). Studies examining the aeolian cadence, ♭VI-♭VII-i, and the double-plagal cadence, ♭VII-IV-I, provide two obvious examples (e.g., [5, 7]). How, then, might we identify the scale systems reflected in these harmonic progressions using computational methods?

This late-breaking demo explores the potential for topic models from natural language processing to discover the scale systems reflected in harmonic annotations from common-practice and popular music corpora. To that end, Section 2 describes the Latent Dirichlet allocation (LDA) architecture. Section 3 evaluates the performance of LDA as an unsupervised classifier for major- and minor-mode excerpts in common-practice music, and then explores the topics identified in a corpus of contemporary popular music using the same model pipeline. Finally, Section 4 offers avenues for future research.

2. TOPIC MODELS

Topic models attempt to derive thematic topics from a collection of text documents, of which LDA is perhaps the most well-known example [8]. Space limitations preclude a formal description of the method, but in short, LDA is a generative probabilistic model that represents each document D as a mixture of latent topics $\theta_1, ..., \theta_D$, and each topic K as a mixture of words $\varphi_1, ..., \varphi_K$. LDA then models each topic-word distribution using a sparse Dirichlet prior based on the assumption that only a small set of words have high probability. For our purposes, the appeal of LDA is thus that it might analogously model a collection of harmonic annotations as a mixture of underlying scale systems, and each scale system as a mixture of chords. In this way, a piece could reflect a mixture of scale systems resulting from tonicizations, modulations, modal mixture, and the like.

Data set	N_{pieces}	N_{excerpts}	Style
ABC	70	473	CP
BCMH	100	408	CP
TAVERN	27	119	CP
McGill Billboard	738	NA	popular
RollingStone-200	200	NA	popular

Note. CP = common-practice.

Table 1. Data sets and descriptive statistics for the corpus.

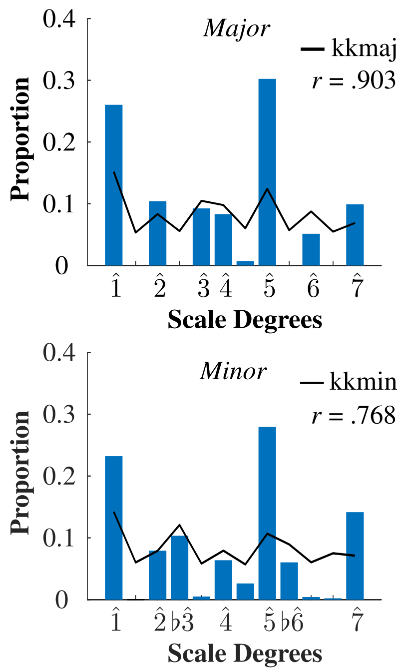

Figure 1. Bar plots of the scale-degree content reflected in the top 20 chords from each topic in the common-practice corpus.

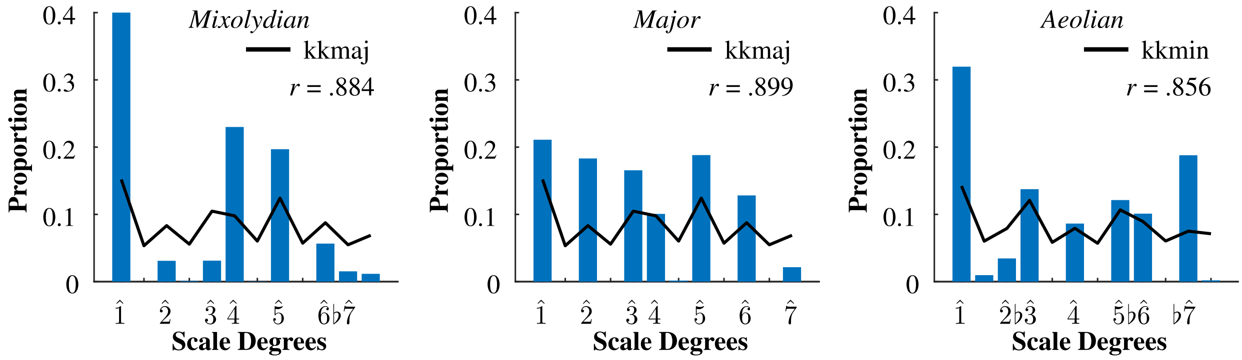

Figure 2. Bar plots of the scale-degree content reflected in the top 20 chords from each topic in the popular corpus.

3. EXPERIMENTS

3.1 Corpora & Evaluation

Table 1 provides descriptive statistics for the data sets in this study. The *common-practice* corpus (N=197) consists of three data sets of Roman numeral annotations: the Annotated Beethoven Corpus (ABC) [9], the Bach Melody-Harmony Corpus (BCMH) [10], and the Theme and Variation Encodings with Roman Numerals (TAVERN) data set [11]. To effectively model chord symbols reflecting the major and minor modes, each piece was additionally divided into shorter excerpts based on annotated changes in mode (N=1000). The *popular* music corpus (N=938) consists of two data sets of chord annotations: the Rolling Stone Corpus (RS-200) [2], and the McGill Billboard data set [1]. [1]

To assess model performance, each corpus was divided into a 50/50 train/test split stratified by data set, with the train split used to identify the number of topics K based on estimates of model perplexity [12]. [2] Classification performance for the common-practice corpus was then estimated using the weighted F_1-measure. [3] Finally, to identify the underlying scale system, topic model scores were assigned to the chord members from the top 20 chords in each topic. These scores were then weighted based on each chord member's role within the chord (*root*, *bass*, *other*) in order to maximize the resulting scale-degree distribution's correlation with the corresponding goodness-of-fit key profile [13]. [4]

CP		Popular		
Major	*Minor*	*Mixolydian*	*Major*	*Aeolian*
I	i	I	I	i
V	V	IV	ii^7	$\flat$VII
V^7	V^7	V	vi^7	$\flat$VI
I^6	i^6	I^{d7}	ii^7	$\flat$III
IV	iv	vi	iii	i^7
$\mathrm{V}(^6_4)$	$\mathrm{V}(^6_4)$	V^7	V^7	I_p
V^6_5	V^6	IV^{d7}	vi^7	iv
V^6_5	iv^6	VI^6_4	iii^7	v
V^4_2	$\flat$VI	ii	I^{M7}	iv^7
vi	V^6_5	$\flat$VII	I^6	v^7

Note. CP = common-practice. ^{d7} = dominant seventh; ^{M7} = major seventh; $_\mathrm{p}$ = power chord (i.e., with missing fifth).

Table 2. Top 10 chords for topics from each model.

3.2 Results

LDA produced two topics that significantly reduced model perplexity for the common-practice corpus (p = .016). Classification based on topic assignment also produced a weighted F_1 estimate of .929. Shown in Figure 1, the optimized weights for the top 20 chords from both topics were significantly correlated to the corresponding goodness-of-fit key profiles (optimized weights: major – *root* = .02; *bass* = .92, *other* = .06; minor – *root* = .02; *bass* = .84, *other* = .14). For both modes, privileging the bass voice in the final scale-degree distribution maximized the correlation coefficient.

LDA produced three topics that significantly reduced model perplexity for the popular corpus (p = .047). Shown in Figure 2 and Table 2, the top 20 chords from topic 1 are primarily triadic and reflect scale-degree content from the mixolydian mode (e.g., I^{d7} and $\flat$VII). By comparison, topic 2 features diatonic seventh chords and scale-degree content from the major (or ionian) mode. Finally, topic 3 clearly represents the aeolian mode, but may also reflect harmonies associated with (minor) pentatonic scales [5].

4. CONCLUSION

Future research could refine the current approach by exploring coherence measures for the estimated LDA topics (e.g., UMass), or by examining other hyperparameter settings (i.e., α and β). Nevertheless, given the success of the model architecture employed here, topic modeling may be an essential first step in future studies of triadic harmony.

[1] Due to differences in the encoding schemes across data sets, we converted the chord symbols from each data set into a single, standard Roman numeral representation scheme (for further details, see [3]).

[2] Perplexity was measured for between 1 and 10 topics using 10-fold cross-validation within the train split, again stratified by data set. The highest K_i was selected that significantly differed from K_{i-1} using a two-sample t-test with Bonferroni correction.

[3] Since LDA outputs a probability distribution over topics for each excerpt, excerpts in the test split were assigned the topic label that received the highest probability estimate in the distribution.

[4] Code is available for download at `https://github.com/PeARL-laboratory/ScalesinPop`

5. REFERENCES

[1] J. A. Burgoyne, J. Wild, and I. Fujinaga, "An expert ground-truth set for audio chord recognition and music analysis," in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, A. Klapuri and C. Leider, Eds., Miami, FL, 2011, Book Section, pp. 423–428.

[2] T. Declercq and D. Temperley, "A corpus analysis of rock harmony," *Popular Music*, vol. 30, pp. 47–70, 2011.

[3] D. R. W. Sears and D. Forrest, "Triadic patterns across classical and popular music corpora: Stylistic conventions, or characteristic idioms?" *Journal of Mathematics and Music*, 2021.

[4] C. W. White and I. Quinn, "Chord context and harmonic function in tonal music," *Music Theory Spectrum*, vol. 40, no. 2, pp. 314–350, 2018.

[5] N. Biamonte, "Triadic modal and pentatonic patterns in rock music," *Music Theory Spectrum*, vol. 32, no. 2, pp. 95–110, 2010.

[6] C. Doll, *Hearing Harmony: Toward a Tonal Theory for the Rock Era*. Ann Arbor, MI: University of Michigan Press, 2017.

[7] W. Everett, "Pitch down the middle," in *Expression in pop-rock music: Critical and analytical essays*, 2nd ed., W. Everett, Ed. New York: Routledge Press, 2008, pp. 111–174.

[8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[9] M. Neuwirth, D. Harasim, F. C. Moss, and M. Rohrmeier, "The Annotated Beethoven Corpus (ABC): A dataset of harmonic analyses of all Beethoven string quartets," *Frontiers in Digital Humanities*, vol. 5, no. 16, 2018.

[10] J. Verbeten and D. R. W. Sears, "Pattern discovery using melodic-harmonic reductions of bach chorales," in *The Biennial meeting of the Society for Music Perception and Cognition (SMPC)*, New York, NY, 2019.

[11] J. Devaney, C. Arthur, N. Condit-Schultz, and K. Nisula, "Theme and variation encodings with Roman numerals (TAVERN): A new data set for symbolic music analysis," in *Proceedings of the 16th International Society for Music Information Retrieval (ISMIR)*, Málaga, Spain, 2015, Book Section, pp. 728–734.

[12] H. M. Wallach, I. Murray, R. Salakhutdinov, and D. Mimno, "Evaluation methods for topic models," in *Proceedings of the 26th International Conference on Machine Learning (ICML)*, Montreal, Canada, pp. 1105–1112.

[13] C. L. Krumhansl and E. Kessler, "Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys," *Psychological Review*, vol. 89, no. 4, pp. 334–368, 1982.