Key-specific Shrinkage Techniques for Harmonic Models

Jeremy Pickens
Center for Intelligent Information Retrieval
Department of Computer Science
University of Massachusetts
Amherst, MA 01002 USA
jeremy@cs.umass.edu

1 Introduction

Statistical modeling of music is rapidly gaining acceptance as viable approach to a host of Music Information Retrieval related tasks, from transcription to ad hoc retrieval. As music may be viewed as an evolving pattern of notes over time, models which capture some statistical notion of sequence are preferred. The focus of this paper is on Markov models for ad hoc retrieval. In particular, we use the Harmonic Models created by [Pickens et al., 2002] as our baseline retrieval system and explain how they may be improved by better shrinkage procedures to improve parameter estimation.[1]

A quick review of harmonic model creation is necessary: First, principled heuristics are used to map simultaneities of notes onto distributions of chords. These distributions are further smoothed using windows of simultaneities at previous time steps to update the distribution at the current time step. Relative frequency counts of chord occurrences are then used to estimate parameters of a Markov model. The approach taken for retrieval is to estimate a model for every piece of music in the collection, to estimate a model of the query, and then rank the pieces of music by the KL divergence of query model to document model. However, if a document model has an estimate of zero for any probability the KL divergence score for the document goes to infinity. Shrinkage is the technique by which these estimates of zero probability are overcome.

2 Shrinkage

Shrinkage works by using data rich states of simpler models to better inform the estimates in data sparse states of more complex models [Freitag and McCallum, 1999]. In the initial work on Harmonic Models, shrinkage was done by creating a general model of music by collecting the counts for every piece in the collection into a single model, creating a global average. Whenever a zero probability estimate is encountered, the algorithm “backs off” to the estimate given by the global model.

[1]This work was supported in part by the Center for Intelligent Information Retrieval and in part by NSF grant #IIS-9905842. Any opinions, findings and conclusions or recommendations expressed in this material are the author(s) and do not necessarily reflect those of the sponsor.

In this work, we instead begin with the intuition that depending on the key in which a piece of music was written, there are going to be different estimates for different states. An analogy to text modeling may be helpful. Suppose you have models for two different topics: finance and ecology. Both topic models have some non-zero probability estimate for the word “bank”. However, the estimate of “bank” under the finance topic might be much higher, because banks and banking is a common feature of financial documents, while stream or river “bank” is an active, though less common, feature of ecological documents. Thus, if you determine that a particular document is about ecology, and not about finance, you would want to shrink that document’s estimate for “bank” toward the ecology topic model rather than toward the finance topic model.

By analogy, a key can be though of as the “topic” for a particular piece of music. No matter what the unique melodic or harmonic progression of that piece, the expected frequency of a C Major triad is going to be very different if that piece is written in C Major, versus if that piece is written in $F^\sharp$ Major. Yet if we smooth this piece with a global model, created by averaging every piece in the collection including those written in $F^\sharp$ Major, then we will be shrinking the estimates for this piece further away from what the true values might be.

So rather than create a global model, we create 24 “topic” models of music, one for each key (C Major, C minor$\ldots B$ Major, B minor). At indexing time, each piece in the collection is tagged by key. With no guarantee that accurate key information exists for any document in the collection, we instead use a simple heuristic key-finding algorithm: During harmonic modeling, after simultaneities are mapped onto distributions of major and minor triads, we sum the relative count for each triad over all simultaneities. The most frequently occuring triad is chosen; the root of this triad determines the tonic, while the third of this triad determines the modality, yielding the key “topic” label. The collection is partitioned by these labels, and each partition is averaged, to create 24 key-based models.

Finally, we are ready to propose two shrinkage techniques. First, the key of the document model to be shrunk is determined. Technique *one* is backoff. However, instead of backing off to a global model, we back off to that document’s corresponding key model. For technique *two*, all states in the document model are linearly interpolated with its corresponding key model, regardless of zero estimates or not. The parameters of the interpolation are given a maximum entropy value of 0.5 for the document model, 0.5 for the key-specific model.

Table 1: Mean Average Precision

	Twinkle (0)	Twinkle (1)	Twinkle (2)	Lachrimae (0)	Lachrimae (1)	Lachrimae (2)	Folia (0)	Folia (1)	Folia (2)
MM0, W1	0.145	0.143	0.172*	0.172	0.177*	0.179	0.333	0.376*	0.534*
MM0, W2	0.145	0.158	0.162	0.174	0.179	0.185*	0.337	0.383*	0.535*
MM0, W3	0.145	0.141	0.152	0.173	0.182*	0.187*	0.331	0.382*	0.539*
MM0, W4	0.130	0.132	0.140	0.172	0.181*	0.190*	0.328	0.378*	0.539*
MM1, W1	0.111	0.107	0.109	0.056	0.060*	0.180*	0.172	0.203*	0.151
MM1, W2	0.149	0.131*	0.123	0.136	0.167*	0.221*	0.315	0.349*	0.263*
MM1, W3	0.095	0.089	0.120*	0.177	0.177	0.221*	0.422	0.380*	0.291*
MM1, W4	0.104	0.093*	0.109	0.181	0.191*	0.224*	0.389	0.363*	0.303*
MM2, W1	0.150	0.169*	0.088*	0.030	0.039*	0.119*	0.105	0.162*	0.029*
MM2, W2	0.156	0.141*	0.096*	0.096	0.151*	0.221*	0.178	0.229*	0.108*
MM2, W3	0.117	0.102*	0.094	0.162	0.192*	0.232*	0.284	0.266*	0.144*
MM2, W4	0.083	0.081	0.091	0.195	0.203*	0.237*	0.329	0.274*	0.181*
Average	0.127	0.124 -2.4%	0.121 -4.7%	0.144	0.158* +9.7%	0.200* +38.9%	0.294	0.312* +6.1%	0.301* + 2.4%

Table 2: Precision at the top 5 retrieved pieces

	Twinkle (0)	Twinkle (1)	Twinkle (2)	Lachrimae (0)	Lachrimae (1)	Lachrimae (2)	Folia (0)	Folia (1)	Folia (2)
MM0, W1	0.485	0.469	0.539	0.461	0.507*	0.496	0.628	0.756*	0.800*
MM0, W2	0.515	0.577	0.546	0.440	0.517*	0.504	0.672	0.756*	0.788*
MM0, W3	0.531	0.539	0.523	0.427	0.528*	0.512*	0.628	0.744*	0.796*
MM0, W4	0.439	0.454	0.477	0.440	0.517*	0.525*	0.608	0.728*	0.796*
MM1, W1	0.285	0.315	0.431*	0.040	0.088*	0.584*	0.056	0.164*	0.512*
MM1, W2	0.539	0.515	0.462	0.216	0.469*	0.675*	0.404	0.620*	0.672*
MM1, W3	0.346	0.339	0.423	0.523	0.515	0.677*	0.788	0.712*	0.704*
MM1, W4	0.392	0.346	0.400	0.499	0.576*	0.685*	0.728	0.692	0.724
MM2, W1	0.408	0.462	0.377	0.032	0.035	0.507*	0.112	0.128	0.120
MM2, W2	0.431	0.546*	0.377	0.059	0.395*	0.736*	0.144	0.408*	0.352*
MM2, W3	0.446	0.400	0.346	0.419	0.643*	0.736*	0.480	0.596*	0.476
MM2, W4	0.331	0.331	0.346	0.619	0.672*	0.760*	0.732	0.616*	0.588*
Average	0.429	0.441 +2.8%	0.437 + 1.9%	0.348	0.455* +30.7%	0.616* +77.0%	0.498	0.577* +15.9%	0.611* +22.7%

3 Results

Tables 1 and 2 contain our results. We use the same 3000-piece collection as [Pickens et al., 2002], with transcribed audio versions of the polyphonic Twinkle, Lachrimae and Folia (TLF) sets as queries, and all score versions of these pieces tagged as relevant. Results are presented for the three model orders (MM0 through MM2) and four window smoothing sizes (W1 through W4). Column (0) is the original global backoff approach (the baseline), column (1) is key-specific backoff, and column (2) is key-specific linear interpolation. Asterisks indicate a statistically significant difference from the baseline (T-test).

The first observation we make is that, on average across all given parameter settings, key-specific shrinkage is either significantly better, or not significantly worse, than the original global technique. Improvements are greatest for the Lachrimae and Folia variations, with little significant difference on the Twinkle variations. While not all variations in the TLF sets are in the same key, there are more variations in non closely-related keys in the Twinkle set than in the others. Key-specific shrinkage serves to accentuate (without completely eliminating) differences between pieces in different keys, shrinking each document more toward others in its own and closely related keys. This polarization has the effect of improving retrieval when all relevant pieces are in closely related keys, but degrading performance when they are not. Even still, performance is no worse, on average, for the Twinkle set.

Our second observation is that precision at the top 5 retrieved pieces shows greater improvements than mean average precision. Indeed, even for the Twinkle set there is slight improvement on average (though still not significant), while for the Folia and Lachrimae sets there are huge improvements. Key-specific shrinkage can therefore be thought of as a “precision enhancing” technique, given that at least one variation sought by the user is in a similar key, which is true of many applications.

Our final observation is that for higher order models, backoff is better, while for lower order models, linear interpolation is better. While linear interpolation polarizes the relevant and non-relevant documents by key, higher order models do a better job finding relevant documents by “memorizing” the salient harmonic progressions. We conjecture that linear interpolation dilutes these patterns, sometimes forfeiting some of the earlier advantage gained through polarization. In the future these parameters could be better set using Expectation-Maximization.

In conclusion, key-specific shrinkage techniques have the potential to greatly improve retrieval accuracy.

References

[Freitag and McCallum, 1999] Freitag, D. and McCallum, A. K. (1999). Information extraction with hmms and shrinkage. In *Proceedings of the AAAI-99 Workshop on Machine Learning for Information Extraction*.

[Pickens et al., 2002] Pickens, J., Bello, J. P., Monti, G., Crawford, T., Dovey, M., Sandler, M., and Byrd, D. (2002). Polyphonic score retrieval using polyphonic audio queries: A harmonic modeling approach. In Fingerhut, M., editor, *Proceedings of the 3rd Annual International Symposium on Music Information Retrieval (ISMIR)*, pages 140–149, IRCAM - Centre Pompidou, Paris, France.