Bayesian Model Selection for Harmonic Labelling

Christophe Rhodes

Goldsmiths, University of London

Friday 18th May

Overview

Harmonic Labelling

The task: identifying chords and assigning harmonic labels in popular music.

- currently to MIDI transcriptions of performances;
- could be applied to audio directly (given suitable processing).

Applications:

- generating fake books, guitar chords
- feeding into models of music cognition, melodic memory

Harmonic Labelling

The task: identifying chords and assigning harmonic labels in popular music.

- currently to MIDI transcriptions of performances;
- could be applied to audio directly (given suitable processing).

Applications:

- generating fake books, guitar chords
- feeding into models of music cognition, melodic memory

Harmonic Labelling

The task: identifying chords and assigning harmonic labels in popular music.

- currently to MIDI transcriptions of performances;
- could be applied to audio directly (given suitable processing).

Applications:

- generating fake books, guitar chords
- feeding into models of music cognition, melodic memory

Harmonic Labelling

The task: identifying chords and assigning harmonic labels in popular music.

- currently to MIDI transcriptions of performances;
- could be applied to audio directly (given suitable processing).

Applications:

- generating fake books, guitar chords
- feeding into models of music cognition, melodic memory

Building a better model

Previous work:

- preference rules / knowledge representation
- implicit models of harmony and key (e.g. profiles)
- embedding in suitable space (spiral, torus)

Common feature: label small section and then smooth.

We attempt to build a model with some desireable attributes:

- credible, at least at the descriptive level;
- quantitative enough to be usable in further inference;
- able to be extended incorporate new information quantitatively.

Building a better model

Previous work:

- preference rules / knowledge representation
- implicit models of harmony and key (e.g. profiles)
- embedding in suitable space (spiral, torus)

Common feature: label small section and then smooth.

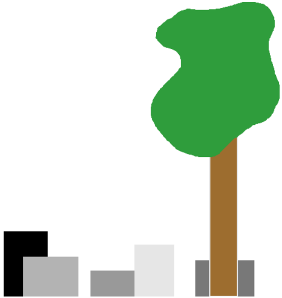
We attempt to build a model with some desireable attributes:

- credible, at least at the descriptive level;
- quantitative enough to be usable in further inference;
- able to be extended incorporate new information quantitatively.

The problem

What is the next number in the series -1, 3, 7, 11, ...?

How many boxes in this scene?



The solution

"accept the simplest explanation that fits the data". Why?

Build alternative classes of model which are capable of explaining the data, and compute and compare likelihoods of given data.

-1, 3, 7, 11, ...?

- $f(n) = x_0 + kn$: 15, 19
- $f'(n) = x_0 + dn^2 + cn^3$: -19.9, 1043.8

The chord model

For pitch-class vector $\mathbf{x}$, we express probability density given chord c as

$$p(\mathbf{x}|c;\Omega) = p_{\mathcal{D}}(t\bar{t}|c;\Omega)p_{\mathcal{D}}(rmd|c;\Omega)$$

where

$$p_{\mathcal{D}}(\mathbf{x}|\alpha_c) = \frac{1}{B(\alpha)}\prod_i x_i^{\alpha_i - 1} \qquad (\sum_i x_i = 1)$$

Then by Bayes' theorem, for a given pitch-class vector $\mathbf{x}$

$$p(c|\mathbf{x}\Omega) = \frac{p(\mathbf{x}|c\Omega)p(c\Omega)}{\sum_c p(\mathbf{x}|c\Omega)p(c\Omega)}$$

The chord model

For pitch-class vector $\mathbf{x}$, we express probability density given chord c as

$$p(\mathbf{x}|c;\Omega) = p_{\mathcal{D}}(t\bar{t}|c;\Omega)p_{\mathcal{D}}(rmd|c;\Omega)$$

where

$$p_{\mathcal{D}}(\mathbf{x}|\alpha_c) = \frac{1}{B(\alpha)}\prod_i x_i^{\alpha_i - 1} \qquad (\sum_i x_i = 1)$$

Then by Bayes' theorem, for a given pitch-class vector $\mathbf{x}$

$$p(c|\mathbf{x}\Omega) = \frac{p(\mathbf{x}|c\Omega)p(c\Omega)}{\sum_c p(\mathbf{x}|c\Omega)p(c\Omega)}$$

The chord model

For pitch-class vector **x**, we express probability density given chord c as

$$p(\mathbf{x}|c;\Omega) = p_{\mathcal{D}}(t\bar{t}|c;\Omega)p_{\mathcal{D}}(rmd|c;\Omega)$$

where

$$p_{\mathcal{D}}(\mathbf{x}|\alpha_c) = \frac{1}{B(\alpha)}\prod_i x_i^{\alpha_i - 1} \qquad (\sum_i x_i = 1)$$

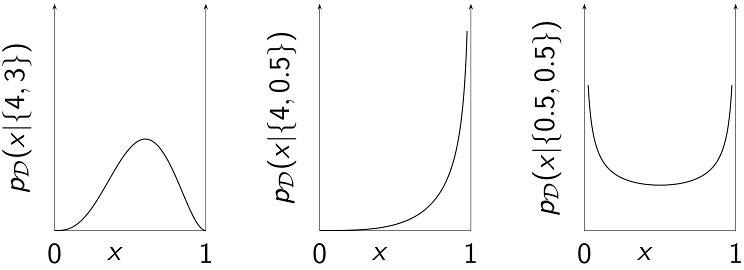
Then by Bayes' theorem, for a given pitch-class vector **x**

$$p(c|\mathbf{x}\Omega) = \frac{p(\mathbf{x}|c\Omega)p(c\Omega)}{\sum_c p(\mathbf{x}|c\Omega)p(c\Omega)}$$

Dirichlet distributions

$$p_{\mathcal{D}}(\mathbf{x}|\alpha_c) = \frac{1}{B(\alpha)} \prod_i x_i^{\alpha_i - 1} \qquad (\sum_i x_i = 1)$$

For two variables, choose one x (and the other is $1 - x$)



Parameter estimation

A Dirichlet distribution over k variables has k parameters. Our model has (in principle) two Dirichlet distributions per distinct chord.

Based on initial inspection of our corpus, we tie parameters such that there are only three distinct cases (instead of the $4 \times 12 \times 6$ that there are in principle for our chord repertoire):

- major or minor chord over a whole bar;
- major or minor chord over a sub-bar window;
- anything else (aug, dim, sus4, sus9).

Estimate parameters for these distributions

- Maximize likelihood of training set;
- Maximize posterior of training set given a suitable prior;
- Tune to maximize performance of labelling task on training set.

Parameter estimation

A Dirichlet distribution over k variables has k parameters.

Our model has (in principle) two Dirichlet distributions per distinct chord.

Based on initial inspection of our corpus, we tie parameters such that there are only three distinct cases (instead of the $4 \times 12 \times 6$ that there are in principle for our chord repertoire):

- major or minor chord over a whole bar;
- major or minor chord over a sub-bar window;
- anything else (aug, dim, sus4, sus9).

Estimate parameters for these distributions

- Maximize likelihood of training set;
- Maximize posterior of training set given a suitable prior;
- Tune to maximize performance of labelling task on training set.

Parameter estimation

A Dirichlet distribution over k variables has k parameters.

Our model has (in principle) two Dirichlet distributions per distinct chord.

Based on initial inspection of our corpus, we tie parameters such that there are only three distinct cases (instead of the $4 \times 12 \times 6$ that there are in principle for our chord repertoire):

- major or minor chord over a whole bar;
- major or minor chord over a sub-bar window;
- anything else (aug, dim, sus4, sus9).

Estimate parameters for these distributions

- Maximize likelihood of training set;
- Maximize posterior of training set given a suitable prior;
- Tune to maximize performance of labelling task on training set.

Limitations of this chord model

- no special treatment of bass note;
- more generally, no handling of register of individual notes;
- no modelling of transitions between chords.

Choosing a region

When does one chord end and another begin?
Assumptions:

- *barline* as fundamental division;
- new chords only on beats.

The first assumption is probably reasonable for our task; the second leads to problems in strongly-syncopated passages.

Choosing a region

When does one chord end and another begin?
Assumptions:

- *barline* as fundamental division;
- new chords only on beats.

The first assumption is probably reasonable for our task; the second leads to problems in strongly-syncopated passages.

Choosing a region

Models: all possible beatwise divisions of a bar. For example, for $\frac{4}{4}$,

- {4}
- {3,1}, {1,3}
- {2,2}
- {2,1,1}, {1,2,1}, {1,1,2}
- {1,1,1,1}

Choose between bar divisions ω using Bayesian model selection:

$$p(\omega|\mathbf{x}\Omega') \propto \sum_{c} p(\mathbf{x}|c\omega\Omega')p(c\omega\Omega')$$

Choosing a region

Models: all possible beatwise divisions of a bar. For example, for $\frac{4}{4}$,

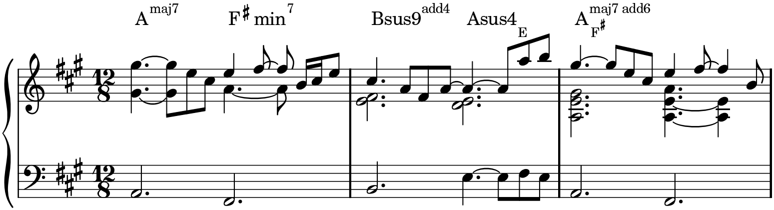
- $\{4\}$
- $\{3,1\}$, $\{1,3\}$
- $\{2,2\}$
- $\{2,1,1\}$, $\{1,2,1\}$, $\{1,1,2\}$
- $\{1,1,1,1\}$

Choose between bar divisions ω using Bayesian model selection:

$$p(\omega|\mathbf{x}\Omega') \propto \sum_{c} p(\mathbf{x}|c\omega\Omega')p(c\omega\Omega')$$

Example I

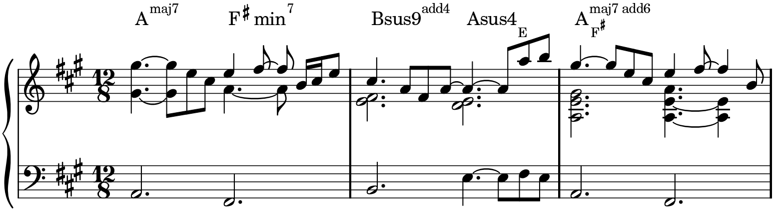
Saving all my love for you (Michael Masser)



Bass note assignments and extensions are heuristically derived *after* the harmonic labelling; future work would incorporate those judgments into the framework.

Example I

Saving all my love for you (Michael Masser)



Bass note assignments and extensions are heuristically derived *after* the harmonic labelling; future work would incorporate those judgments into the framework.

Evaluation: how good is our algorithm?

Maximum likelihood parameters estimated from training set (233 bars):

- 53% regions correctly bounded;
- 75% chords labelled correctly.

Parameters tuned to training set:

- 75% regions correctly bounded;
- 76% chords labelled correctly.

Example II

Lady Madonna (Lennon/McCartney)

Is there even a right answer?

Evaluation investigation

Current investigation: how much do experts' opinions on this task differ?

- send machine-generated labels to acknowledged experts for evaluation and corrections;
- four experts, forty excerpts;
- for each expert, score and audio provided for thirty and audio only for ten;
- lead sheet format – also include lead sheets from song books.

Watch this space...

Evaluation investigation

Current investigation: how much do experts' opinions on this task differ?

- send machine-generated labels to acknowledged experts for evaluation and corrections;
- four experts, forty excerpts;
- for each expert, score and audio provided for thirty and audio only for ten;
- lead sheet format – also include lead sheets from song books.

Watch this space...

Extensibility

If we have more domain knowledge, then we can use the same framework to incorporate that knowledge:

- bass note: $p(c|\mathbf{x}b\Omega') \propto p(\mathbf{x}|cb\Omega')p(c|b\Omega')$
- genre: $p(c|\mathbf{x}g\Omega'') \propto p(\mathbf{x}|cg\Omega'')p(c|g\Omega'')$

Bayesian inference doesn't give just one answer, but a probability distribution over labels and windows. We can quantify our uncertainty (e.g. distribution entropy).

Summary

- Can segment bars and generate harmonic labels with reasonable accuracy.
- Actual accuracy figures are indicative only: ongoing investigation into performance of human experts.
- Framework is extensible: can incorporate specific information (e.g. knowledge of bass note, alphabet of chord labels for a known genre) in a principled way.

Acknowledgments

- David Lewis, Daniel Müllensiefen
- Geerdes midimusic (`http://www.midimusic.de/`)
- EPSRC grants GR/S84750/01, EP/D038855/1