Raphael and Stoddard's Harmonic Analysis Model: A Summary

Mark Granroth-Wilding

June 8, 2011

1 Introduction

This document presents a brief summary of the sequence labelling model proposed by Christopher Raphael and Joshua Stoddard in *Functional Harmonic Analysis Using Probabilistic Models*, 2004. It summarizes the key details of the model's parameters and how they are trained. It does not describe the model in full or motivate it in any way. Refer to their paper for that.

2 Basic Model

- Fixed minimal time division, q.
- y_i : pitches with onset time in period $[iq, i(q+1))$.
- $y_i = \{y_i^1, \ldots, y_i^K\}$.
- Pitches represent pitch-classes in $T = \{0, \ldots, 11\}$. $0 \Rightarrow C$, etc.
- Labels are chosen from L,

$$L = T \times M \times C = \{0, \ldots, 11\} \times \{major, minor\} \times \{I, II, \ldots, VII\}$$

- All secondary functionality handled as modulation.
- HMM, where states are analysis labels chosen from L, and emissions are pitch-class sets from $\mathcal{P}(T)$.

3 Parameter Tying: Transition Distribution

1. $$\begin{aligned} p(x'|x) = & \; p(t', m', c'|t, m, c) \\ = & \; p(t', m'|t, m) \cdot p(c'|t', m', t, m, c) \end{aligned}$$

   Assumes choice of t' and m' is not dependent on c.

   (First term is $p(t', m'|t, m)$, not $p(t'm'|t, m, c)$.)

2. $p(x'|x) = \; p(t' - t, m'|m) \cdot p(c'|t' - t, m', m, c)$

   Assumes we choose a key interval, rather than absolute key conditioned on previous key. Simply absolute pitch invariance of key roots.

3. $p(x'|x) = \; p(t' - t, m'|m) \cdot p(c'|c)$

   Assumes choice of chord doesn't depend on absolute key: another aspect of absolute pitch invariance. It also doesn't depend on mode (major/minor). Not a very strong assumption.

4. $$p(x'|x) = \; p(t' - t, m'|m) \cdot \begin{cases} p(c'|c) & t' = t, m' = m, \\ p(c) & otherwise \end{cases}$$

   Assumes that when key or mode changes, choice of chord doesn't depend on previous chord, or anything else, but is drawn from a random distribution of chords. Quite strong, but probably doesn't make a huge difference and massively reduces parameters.

Final transition distribution:

$$p(x'|x) = \quad p(t'-t, m'|m) \cdot \begin{cases} p(c'|c) & t'=t, m'=m, \\ p(c) & \textit{otherwise} \end{cases}$$

4 Parameter Tying: Emission Distribution

Vector r contains certain observations of the onset times of notes in y.

$$r^k = \begin{cases} 0 & y^k \textit{ occurs at start of bar} \\ 1 & y^k \textit{ occurs in middles of bar} \\ 2 & y^k \textit{ occurs on offbeat (2 or 4)} \\ 3 & \textit{otherwise} \end{cases}$$

$$\begin{aligned} p(y|x,r) = & \quad p(y^1, \dots, y^K|x, r^1, \dots, r^K) \\ = & \quad \prod_{k=1}^{K} p(y^k|x, r^k) \end{aligned}$$

Assumes note onsets are drawn from independent rhythm distributions. They claim this is the most problematic assumption and suggest (but don't try) an alternative.

$$p(y|x,r) = \prod_{k=1}^{K} \frac{p(d(y^k,x)|r^k)}{V(d(y^k,x))}$$

d projects pitch classes onto an abstraction of their relation to the chord (from the label x).

$$d(y^k,x) = d(y^k,t,m,c) = \begin{cases} 1 & y^k \textit{ is root of chord } t,m,c \\ 2 & y^k \textit{ is 3rd of chord } t,m,c \\ 3 & y^k \textit{ is 5th of chord } t,m,c \\ 4 & y^k \textit{ is in scale } t,m \\ 5 & \textit{otherwise} \end{cases}$$

V is simply the number of pitch classes equated by each value of d. This means the probability for a category is spread over the pitch classes in the category.

$$V(1) = V(2) = V(3) = 1 \qquad V(4) = 4 \qquad V(5) = 5$$

Now only 30 parameters in emission distribution.

Final emission distribution:

$$p(y|x,r) = \prod_{k=1}^{K} \frac{p(d(y^k,x)|r^k)}{V(d(y^k,x))}$$

5 Training

Train using Baum-Welch.

Initialize with chord type model. No dependence on rhythmic position. Hand-set initial emission probabilities, though they don't say exactly how.

$$\begin{aligned} p(d(y^k,x)|r^k) = & \quad p(d(y^k,x)) \\ = & \quad \begin{cases} \textit{high} & d \in \{1,2,3\} \\ \textit{medium} & d = 4 \\ \textit{low} & d = 5 \end{cases} \end{aligned}$$

They claim transition parameter initialization doesn't make any difference. I suppose we can just set all transitions equiprobable:

$$
\begin{aligned}
p(x'|x) = &\quad p(t'-t, m'|m) \cdot \begin{cases} p(c'|c) & t'=t, m'=m \\ p(c') & \textit{otherwise} \end{cases} \\
= &\quad \tfrac{1}{12\times 2\times 7} \\
= &\quad \tfrac{1}{168}
\end{aligned}
$$

6 Voice Leading Model

They vaguely suggest a way of incorporating some notion of voice-leading into the model, dispensing with the first assumption of the emission distribution.

They don't try it, so nor will we for now.