A Generative Model for Rhythms

Jean-François Paiement, IDIAP Research Institute
Yves Grandvalet, CNRS
Samy Bengio, Google
Douglas Eck, Université de Montréal

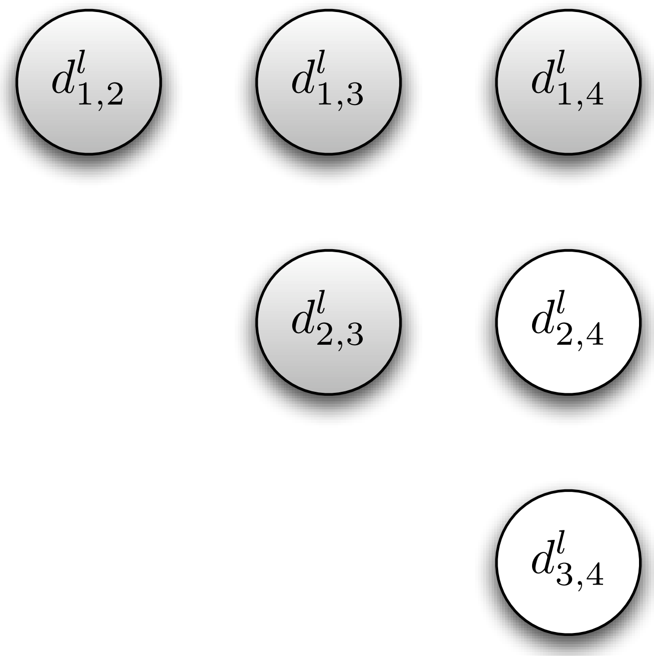
Motivation

- Music (and rhythm in particular) involves long term dependencies;
- Dependencies are characterized by hierarchical structure related to meter;
- We assume that distance patterns between subsequences are at least as important as the actual choice of notes in music structure.

Distance Patterns

- Repeated sequences of random notes sound like melody;
- Where can we put variability?
- Here, distance patterns refers to hierarchical distributions of distances between subsequences of equal lenght (*partition*).

Distance Model



$$\alpha^l_{i,j} = \min_{k \in \{1,\ldots,(i-1)\}} (d^l_{k,j} + d^l_{i,k})$$

$$\beta^l_{i,j} = \max_{k \in \{1,\ldots,(i-1)\}} (|d^l_{k,j} - d^l_{i,k}|)$$

$$\beta^l_{i,j} \leq d^l_{i,j} \leq \alpha^l_{i,j}$$

Rhythms

- We represent rhythms with 3 states on each position:
  1. Note onset;
  2. Note continuation;
  3. Silence.
- Thus, *d* can be chosen to be the Hamming distance.

Binomial Mixture Model

- We assume that the random variables
$$(d^l_{i,j} - \beta^l_{i,j})/(\alpha^l_{i,j} - \beta^l_{i,j})$$
are iid;
- We can model $d_{i,j} - \beta_{i,j}$ with a binomial distribution of parameters $(\alpha_{i,j} - \beta_{i,j}, p_{i,j})$;
- For more flexibility, we use a binomial mixture.

Hierarchy Learning

- The joint distribution of distances can be computed for many number of partitions;
- Parameters can be learned with the EM algorithm;
- We initialize the parameters with a variant of k-means.

Conditional Prediction

- We combine the distance model with a "local" HMM model;
- We solve the optimization problem

$$\begin{cases} \max\limits_{\tilde{x}_s,\dots,\tilde{x}_m} & p_{\mathrm{HMM}}(\tilde{x}_s,\dots,\tilde{x}_m|x_1,\dots,x_{s-1}) \\ \text{subject to} & \prod\limits_{r=1}^{h} p(D_{\rho_r}(\mathbf{x}^l)) \geq P_0 \end{cases},$$

- In practice, we solve the Lagrangian

$$\max_{\tilde{x}_s,\dots,\tilde{x}_m} \log p_{\mathrm{HMM}}(\tilde{x}_s,\dots,\tilde{x}_m|x_1,\dots,x_{s-1}) + \lambda \sum_{r=1}^{h} \log p(D_{\rho_r}(\mathbf{x}^l))$$

Experiments

- A jazz database of 47 standards and a subset of the "Nottingham" database (53 hornpipes tunes) were used for the experiments;
- We compared the proposed model with an HMM using conditional prediction accuracy computed with double cross-validation.

Conditional Accuracy

Jazz standards			
Observed	Predicted	HMM	Global
32	96	34.53%	54.61%
64	64	34.47%	55.55%
96	32	41.56%	47.21%

Hornpipes			
Observed	Predicted	HMM	Global
48	144	75.07%	83.02%
96	96	75.59%	82.11%
144	48	76.57%	80.07%

Dyadic structures

- Best results with deeper dyadic structures:

P	Accuracy
2	49.30%
2,4	49.27%
2,4,8	51.36%
2,4,8,16	55.55%

Demo