A Distance Model for Rhythms

Jean-François Paiement, IDIAP Research Institute
Yves Grandvalet, IDIAP and 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 [Handel, 1993];

• 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 length (partition).
Distance Model

\[ \alpha_{i,j}^l = \min_{k \in \{1, \ldots, (i-1)\}} \left( d_{k,j}^l + d_{i,k}^l \right) \]

\[ \beta_{i,j}^l = \max_{k \in \{1, \ldots, (i-1)\}} \left( |d_{k,j}^l - d_{i,k}^l| \right) \]

\[ \beta_{i,j}^l \leq d_{i,j}^l \leq \alpha_{i,j}^l \]
Rhythms

• We represent rhythms with 3 states on each position:

  1. Note onset;
  2. Note continuation;

• Thus, $d$ can be chosen to be the Hamming distance.
Distances Between Sub-sequences

• The similarity in a given position can be considered as an independent Bernouilli experiment with parameter $p_{i,j}$.

• The possible number of different positions is constrained by $\alpha_{i,j}$ and $\beta_{i,j}$.

\[
\begin{array}{ccccccc}
1 & 2 & 3 & 3 & 1 & 2 & 2 & 3 \\
1 & 2 & 2 & 3 & 1 & 2 & 2 & 2 \\
\end{array}
\]
Binomial Mixture Model

• We can model $d_{i,j} - \beta_{i,j}$ with a binomial distribution of parameters $(\alpha_{i,j} - \beta_{i,j}, p_{i,j})$;

• We model the sum of $\alpha_{i,j} - \beta_{i,j}$ Bernoulli experiments, each with probability $p_{i,j}$ of success.

• For more flexibility, we use a binomial mixture [McLachlan and Basford, 1988].
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 [Rabiner and Juang, 1993];

• We solve the optimization problem

\[
\begin{align*}
\max_{\tilde{x}_s, \ldots, \tilde{x}_m} & \quad p_{\text{HMM}}(\tilde{x}_s, \ldots, \tilde{x}_m | x_1, \ldots, x_{s-1}) \\
\text{subject to} & \quad \prod_{r=1}^{h} p(D_{\rho_r}(x^l)) \geq P_0,
\end{align*}
\]

• We solve the Lagrangian [Nocedal and Wright, 1999]

\[
\max_{\tilde{x}_s, \ldots, \tilde{x}_m} \log p_{\text{HMM}}(\tilde{x}_s, \ldots, \tilde{x}_m | x_1, \ldots, x_{s-1}) + \lambda \sum_{r=1}^{h} \log p(D_{\rho_r}(x^l))
\]
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

<table>
<thead>
<tr>
<th>Jazz standards</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>32</td>
<td>96</td>
<td>34.53%</td>
<td>54.61%</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>64</td>
<td>34.47%</td>
<td>55.55%</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>32</td>
<td>41.56%</td>
<td>47.21%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hornpipes</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>48</td>
<td>144</td>
<td>75.07%</td>
<td>83.02%</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>96</td>
<td>75.59%</td>
<td>82.11%</td>
</tr>
<tr>
<td></td>
<td>144</td>
<td>48</td>
<td>76.57%</td>
<td>80.07%</td>
</tr>
</tbody>
</table>
Dyadic structures

- Best results with deeper dyadic structures:

<table>
<thead>
<tr>
<th>P</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>49.30%</td>
</tr>
<tr>
<td>2,4</td>
<td>49.27%</td>
</tr>
<tr>
<td>2,4,8</td>
<td>51.36%</td>
</tr>
<tr>
<td>2,4,8,16</td>
<td>55.55%</td>
</tr>
</tbody>
</table>
Demo

The rhythms in the second halves are generated by the model:

- Four Brothers;
- Long Ago;
- But Beautiful.
Demo

The rhythms in the second halves are generated by the model:

- Four Brothers;
- Long Ago;
- But Beautiful.
The rhythms in the second halves are generated by the model:

- Four Brothers;
- Long Ago;
- But Beautiful.
Demo

The rhythms in the second halves are generated by the model:

- Four Brothers;
- Long Ago;
- But Beautiful.
Conclusion

• We introduced a generative model for distance patterns in temporal data;

• Leads to better prediction accuracy when combined with a “local” model;

• Can be sampled to generate rhythms similar to a training corpus.
Questions?