A Distributed Model for Multiple Viewpoint Melodic Prediction

Srikanth Cherla$^{1,2}$, Tillman Weyde$^{1,2}$, Artur Garcez$^2$, Marcus Pearce$^3$

$^1$Music Informatics Research Group, City University London
$^2$Machine Learning Group, City University London
$^3$Centre for Digital Music, Queen Mary University of London

November 4, 2013
Outline

Introduction: Analysing sequences in symbolic music data

Background: Probabilistic modelling of melodic sequences

Approach: Modelling melodic sequences with RBMs

Results: Encouraging Prediction Performance
Introduction: Analysing sequences in symbolic music data

Background: Probabilistic modelling of melodic sequences

Approach: Modelling melodic sequences with RBMs

Results: Encouraging Prediction Performance
Sequential Information in Notated Music

- A wealth of information in notated music.
- Increasingly available
  - in different formats (MIDI, Kern, GP4, etc.).
  - for different kinds of music (classical, rock, pop, etc.)
- Analysis of sequences key to extracting information.
- Melody — Good starting point for a broader analysis.
Relevance

Scientific:
- Computational musicology
- Organizing music data
- Generating musical stimuli
- Aiding acoustic models
- Music education

Creative:
- Automatic music generation
- Compositional assistance
Introduction: Analysing sequences in symbolic music data

Background: Probabilistic modelling of melodic sequences

Approach: Modelling melodic sequences with RBMs

Results: Encouraging Prediction Performance
Information Dynamics of Music (IDyOM)

- Predictive models of musical structure using probabilistic learning (Pearce & Wiggins, 2004).
- Develop insights into the analysis of musical structure drawing on research in musicology (Whorley et al., 2013).
- Relate predictions to psychological and neural processing of music (Omigie et al., 2013).

Website: [www.idyom.org](http://www.idyom.org)
Multiple Viewpoint Systems for Music Prediction (Conklin & Witten, 1995)

- Framework for analysis of symbolic music data.
- *Viewpoint type* (feature) sequences extracted from score.
- One Markov model per *type*.
- Mixture/product-of-experts to combine multiple models.

![Musical notation](image)

<table>
<thead>
<tr>
<th>Viewpoint</th>
<th>Transformed sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>pitch</td>
<td>67  69  71  72  69  72  64  67  72  69</td>
</tr>
<tr>
<td>int</td>
<td>⊥  2   2   1  -3   3  -8   3   5   -3</td>
</tr>
<tr>
<td>onset</td>
<td>0   2   5   6   9  10  12  15  16  20</td>
</tr>
<tr>
<td>ioi</td>
<td>⊥  2   3   1   3   1   2   3   1   4</td>
</tr>
<tr>
<td>int ⊗ ioi</td>
<td>⊥  2,2  2,3  1,1 -3,3  3,1 -8,2  3,3  5,1 -3,4</td>
</tr>
</tbody>
</table>

(Image Courtesy: Darrell Conklin)
Motivating a Distributed Model

At present...

1. A more scalable way to *link* viewpoint types.
2. An alternative approach to one relying directly on occurrence statistics.

In the future...

- Interest in knowledge extraction from neural networks.
Introduction: Analysing sequences in symbolic music data

Background: Probabilistic modelling of melodic sequences

Approach: Modelling melodic sequences with RBMs

Results: Encouraging Prediction Performance
Goals

- Demonstrate the use of multiple-viewpoint systems with a distributed model - Restricted Boltzmann Machine.
- Compare the predictive performance of this model with the originally used Markov models on a melody corpus.
A bipartite network with binary stochastic units. 
Data in visible layer, features in hidden layer. 
Can model 
  - joint distribution $p(v_1, \ldots, v_r)$
  - conditional distribution $p(v_1, \ldots, v_c | v_{c+1} \ldots, v_r)$
Can be stacked into a deep network and trained efficiently.
A Distributed Melodic Prediction Model

- Viewpoint subsequence $s_{(t-n+1)}...t$ in visible layer.
- Models the conditional distribution $p(s_t|s_{(t-n+1)...(t-1)})$.
- Generalized softmax visible units.
- Viewpoint types linked by vector-concatenation.
- Trained generatively using Contrastive Divergence.
Introduction: Analysing sequences in symbolic music data

Background: Probabilistic modelling of melodic sequences

Approach: Modelling melodic sequences with RBMs

Results: Encouraging Prediction Performance
Evaluation Tasks

Predicting the next \textit{pitch} with

1. a model that uses context of type \textit{pitch}.
2. a model that uses context of type \textit{pitch} $\otimes$ \textit{dur}.
3. a simple mixture-of-experts combination of 1 and 2.
Evaluation Setup

Corpus

- As used in Pearce et al., 2004.
- Subset of the Essen Folk Song Collection.
- A collection of 8 datasets of chorale and folk melodies.
- A total of 54,308 musical events.

Evaluated models

- Context length $\in \{1, 2, 3, 4, 5, 6, 7, 8\}$
- Hidden units $\in \{100, 200, 400\}$
- Learning rate $\in \{0.01, 0.05\}$

Evaluation criterion — cross-entropy (to be minimized)

$$H_c(p_{mod}, D_{test}) = -\sum_{s_1^n \in D_{test}} \frac{\log_2 p_{mod}(s_n | s_1^{(n-1)})}{|D_{test}|}$$
Changing Context Length

- Dataset: Folk melodies of Nova-Scotia, Alsace, Yugoslavia, Switzerland, Austria, Germany; Chorale melodies
- Input: pitch, Target: pitch

Model Performance

![Graph showing model performance with varying context lengths. The graph compares IDyOM (bounded), IDyOM (unbounded), and RBM. The y-axis represents cross-entropy, and the x-axis represents context length (0 to 10). The graph shows the performance trends over different lengths.]
Combining “Multiple Viewpoints”

Dataset: 185 chorale melodies

- Input: *pitch*, Target: *pitch*

<table>
<thead>
<tr>
<th>context length</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>IDyOM</em></td>
<td>2.737</td>
<td>2.565</td>
<td>2.505</td>
<td>2.473</td>
</tr>
<tr>
<td><em>RBM</em></td>
<td>2.698</td>
<td>2.530</td>
<td>2.490</td>
<td>2.470</td>
</tr>
</tbody>
</table>

- Input: *pitch ⊗ duration*, Target: *pitch*

<table>
<thead>
<tr>
<th>context-length</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>IDyOM</em></td>
<td>2.761</td>
<td>2.562</td>
<td>2.522</td>
<td>2.502</td>
</tr>
<tr>
<td><em>RBM</em></td>
<td>2.660</td>
<td>2.512</td>
<td>2.481</td>
<td>2.519</td>
</tr>
</tbody>
</table>

- Input: *pitch ⊕ (pitch ⊗ duration)*, Target: *pitch*

<table>
<thead>
<tr>
<th>context length</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>RBM (combined)</em></td>
<td>2.663</td>
<td>2.486</td>
<td>2.462</td>
<td>2.413</td>
</tr>
</tbody>
</table>
Conclusions & Future Work

We presented the following

- A distributed model for multiple-viewpoint melodic prediction using Restricted Boltzmann Machines.
- Improved prediction results in comparison to previously evaluated Markov models.

Some interesting directions for future work

- Deeper networks.
- Musical interpretation of hidden layers.
- A distributed Short-Term Model.
- Polyphonic music.
- Interesting MIR applications.
We would like to thank

Darrell Conklin (Universidad del Pais Vasco)
Son Tran (City University London)
Thank you!

Questions?