Playing Mozart by analogy: Learning Multi-level Timing and Dynamic Strategies

Gerhard Widmer and Asmir Tobudic
Presented By Arun P Chidambaram

---

Goal

- to investigate what extent a machine predict certain aspects of performance from real world data. E.g. Predictive models of Tempo, dynamics and timing

Hybrid Learning System that predicts at
- Note Level
- Phrase Level
- Combining the predictions into complex expressive curves for new pieces

---

Model

Stage 1
- Learning Algorithm 1
- System predicts tempo and dynamics shapes at different levels of phase structure
- Decompose curve into elementary patterns associated with individual phrases at phrase level

---

Model (cond..)

Stage 2
- Learning Algorithm 2
- Combines Stage 1 predictions with dynamics and local timing predictions by learned note level models

Stage 3
- Combining Expressive curves predicted at different levels into a final composite expression curve
Decomposition of Expression curves

Input to System:
- Scores of Musical pieces
- Tempo curves
- Dynamics curves

Represent as multiplicative factors example
tempo 1.5, loudness 1.5

Decomposition of expressive curves

- Phrase structures are done by hand
- Extract the training examples for phrase level and note level learning
- Complex curves are decomposed into basic shapes that represent the contribution of each phrase to overall expression curve.

Decomposition of expressive curves...contd

To approximate, we use the approximation function for shapes
Second degree polynomials
Ex: = ax^2 + bx + c

Decomposing a given Expression Curve
For each phrase at a given level
- Start with the highest level of phrasing and move to lowest
- Compute the polynomial that best suits the curve
- Subtract the tempo and dynamics
- Curve remaining after subtraction is used for next level of process
- Rudimentary expression curve left after all levels of phrase approximations have been subtracted is the Residual curve
Predicting Tempo and Dynamics

- Phrase level learning - nearest neighbor prediction
- Learning of residuals - PLCG
- Combining Phrase level and note level Predictions

Experiments and Results

Data:
- Mozart Piano Sonatas by a pianist on a computer controlled Grand piano
- Phrase Structure — Manually by a Musicologist

Measures:
- MSE
- MAE
Results by Sonata sections of Experiment

<table>
<thead>
<tr>
<th>Sonatas</th>
<th>Metre</th>
<th>MSe</th>
<th>MRE</th>
<th>MAE</th>
<th>MASE</th>
<th>Cmets</th>
<th>MSS</th>
<th>MSe</th>
<th>MRE</th>
<th>MAE</th>
<th>MASE</th>
<th>Cmets</th>
</tr>
</thead>
<tbody>
<tr>
<td>K.280</td>
<td>0.044</td>
<td>0.011</td>
<td>0.006</td>
<td>0.003</td>
<td>0.005</td>
<td>0.002</td>
<td>0.004</td>
<td>0.006</td>
<td>0.007</td>
<td>0.005</td>
<td>0.006</td>
<td>0.003</td>
</tr>
<tr>
<td>K.133</td>
<td>0.065</td>
<td>0.017</td>
<td>0.012</td>
<td>0.009</td>
<td>0.013</td>
<td>0.008</td>
<td>0.015</td>
<td>0.017</td>
<td>0.019</td>
<td>0.016</td>
<td>0.018</td>
<td>0.014</td>
</tr>
<tr>
<td>K.332</td>
<td>0.089</td>
<td>0.023</td>
<td>0.018</td>
<td>0.015</td>
<td>0.021</td>
<td>0.016</td>
<td>0.023</td>
<td>0.025</td>
<td>0.027</td>
<td>0.024</td>
<td>0.026</td>
<td>0.022</td>
</tr>
<tr>
<td>K.295</td>
<td>0.112</td>
<td>0.028</td>
<td>0.024</td>
<td>0.021</td>
<td>0.027</td>
<td>0.022</td>
<td>0.029</td>
<td>0.032</td>
<td>0.034</td>
<td>0.031</td>
<td>0.034</td>
<td>0.029</td>
</tr>
</tbody>
</table>

Result of Learning at Phrase-levels only

<table>
<thead>
<tr>
<th>Sonatas</th>
<th>Metre</th>
<th>MSe</th>
<th>MRE</th>
<th>MAE</th>
<th>MASE</th>
<th>Cmets</th>
<th>MSS</th>
<th>MSe</th>
<th>MRE</th>
<th>MAE</th>
<th>MASE</th>
<th>Cmets</th>
</tr>
</thead>
<tbody>
<tr>
<td>K.280</td>
<td>0.043</td>
<td>0.010</td>
<td>0.006</td>
<td>0.003</td>
<td>0.005</td>
<td>0.002</td>
<td>0.004</td>
<td>0.005</td>
<td>0.007</td>
<td>0.005</td>
<td>0.006</td>
<td>0.003</td>
</tr>
<tr>
<td>K.133</td>
<td>0.064</td>
<td>0.017</td>
<td>0.012</td>
<td>0.009</td>
<td>0.014</td>
<td>0.009</td>
<td>0.016</td>
<td>0.019</td>
<td>0.021</td>
<td>0.018</td>
<td>0.020</td>
<td>0.016</td>
</tr>
<tr>
<td>K.332</td>
<td>0.088</td>
<td>0.022</td>
<td>0.017</td>
<td>0.014</td>
<td>0.020</td>
<td>0.016</td>
<td>0.022</td>
<td>0.025</td>
<td>0.027</td>
<td>0.024</td>
<td>0.027</td>
<td>0.023</td>
</tr>
<tr>
<td>K.295</td>
<td>0.111</td>
<td>0.027</td>
<td>0.023</td>
<td>0.020</td>
<td>0.027</td>
<td>0.022</td>
<td>0.029</td>
<td>0.032</td>
<td>0.034</td>
<td>0.031</td>
<td>0.034</td>
<td>0.029</td>
</tr>
</tbody>
</table>

Summary of Wins vs. Losses between learning and no learning

<table>
<thead>
<tr>
<th>Wins</th>
<th>Losses</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.004</td>
<td>0.003</td>
</tr>
<tr>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>0.001</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Four level Polynomial Decomposition of Training data

<table>
<thead>
<tr>
<th>Metre</th>
<th>MSe</th>
<th>MRE</th>
<th>MAE</th>
<th>MASE</th>
<th>Cmets</th>
<th>MSS</th>
<th>MSe</th>
<th>MRE</th>
<th>MAE</th>
<th>MASE</th>
<th>Cmets</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.004</td>
<td>0.003</td>
<td>0.002</td>
<td>0.001</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.003</td>
<td>0.002</td>
<td>0.001</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.002</td>
<td>0.001</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.001</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Result of Learning at Phrase-levels only

Case of success

Major piano sonata K.280

Case of Failure

Failure for Mozart Sonata K.332
- Interpretation produced by Learning algorithm is terrible with respect to timing

Reasons
- Not similar to the pieces in training set
- Tempo curve not approximated properly
- Inappropriate phase structure analysis
Limitations

- Nearest Neighbor algorithm (learning algorithm) - doesn't produce interpretable models
- Attribute value Representation - doesn't allow user to refer to details of internal structure and content of phrases
- Individual Prediction of Phrasal shapes – too simple

Future Work

- More Expressive Representation Languages
- Better Learning Algorithms
- Predicting Interdependent concepts at different levels of resolution

Thank You