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

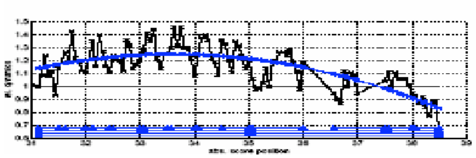
Represent as multiplicative factors example
tempo 1.5 ,loudness 1.5

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.

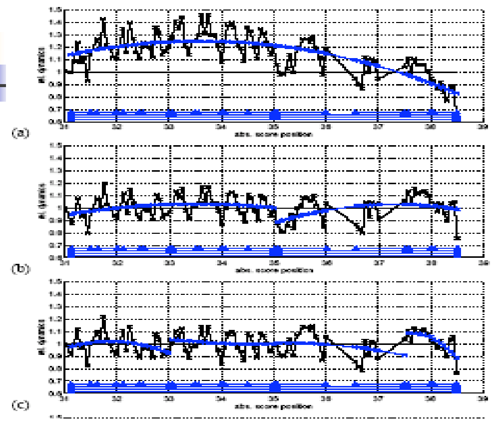
Decomposition of expressive curves…contd

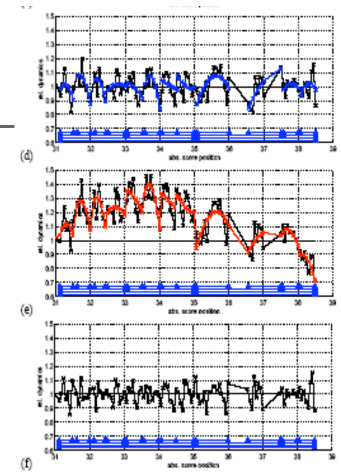
To approximate, we use the approximation function for shapes
Second degree polynomials
Ex - $ax^2 + bx + c$



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

	dynamics					tempo				
	MSE_D	MSE_L	MAE_D	MAE_L	$Corr_L$	MSE_D	MSE_L	MAE_D	MAE_L	$Corr_L$
kv279:1:1	.0383	.0411	.1643	.1544	.6212	.0348	.0406	.1220	.1479	.3550
kv279:1:2	.0318	.0737	.1479	.1975	.4204	.0244	.0335	.1004	.1327	.2984
kv280:1:1	.0313	.0266	.1432	.1226	.7080	.0254	.0192	.1053	.1032	.5821
kv280:1:2	.0281	.0491	.1365	.1642	.4711	.0250	.0304	.1074	.1232	.4010
kv280:2:1	.1558	.0831	.3498	.2002	.7168	.0343	.0187	.1189	.1079	.7518
kv280:2:2	.1424	.0879	.3178	.2235	.6980	.0406	.0431	.1349	.1400	.5128
kv280:3:1	.0334	.0134	.1539	.0916	.7765	.0343	.0244	.1218	.1136	.5813
kv280:3:2	.0226	.0728	.1231	.2089	.4590	.0454	.0418	.1365	.1327	.3953
kv282:1:1	.1126	.0465	.2792	.1721	.7667	.0295	.0315	.1212	.1160	.4222
kv282:1:2	.0920	.0521	.2537	.1782	.6976	.0227	.0421	.1096	.1477	.3460
kv282:1:3	.1230	.0613	.2595	.2105	.7200	.1011	.0583	.2354	.1815	.6676
kv283:1:1	.0283	.0234	.1423	.1194	.6007	.0183	.0274	.0918	.1193	.2441
kv283:1:2	.0371	.0520	.1611	.1629	.4406	.0178	.0275	.0932	.1208	.1948
kv283:3:1	.0404	.0320	.1633	.1323	.6030	.0225	.0214	.1024	.1085	.4460
kv283:3:2	.0417	.0402	.1676	.1466	.5336	.0238	.0254	.1069	.1150	.2948
kv332:2	.0919	.0844	.2554	.2370	.5475	.0286	.0416	.1110	.1520	.2787
Mean:	.0657	.0525	.2012	.1701	.6113	.0330	.0329	.1199	.1289	.4232

Summary of Wins vs. Losses between learning and no learning

	dynamics	tempo
Learning from all pieces:		
MSE	11+/5-	6+/10-
MAE	12+/4-	6+/10-
Learning from slow and fast pieces separately:		
MSE	14+/2-	8+/8-
MAE	14+/2-	8+/8-

Four level Polynomial Decomposition of Training data

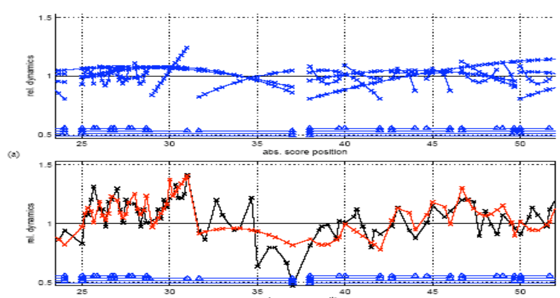
	MSE_D	MSE_P	MAE_D	MAE_P	$Corr_P$
dyn.	.0657	.0055	.2012	.0523	.9456
tempo	.0330	.0127	.1199	.0720	.7421

Result of Learning at Phrase-Levels only

	MSE_D	MSE_L	MAE_D	MAE_L	$Corr_L$
dyn.	.0657	.0533	.2012	.1718	.6027
tempo	.0330	.0339	.1199	.1308	.3877

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