Convolutional Composer Classification

Harsh Verma, John Thickstun

Probabilistic Classification

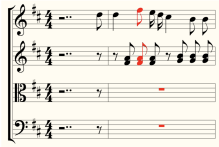
Learn to attribute a score $\mathbf{x}$ to a composer y_i :

$$p(y_i|\mathbf{x}) = \frac{\exp(f_\theta(\mathbf{x})_{y_i})}{\sum_{k=1}^{C} \exp(f_\theta(\mathbf{x})_{y_k})}.$$

- **How do we featurize a score $\mathbf{x}$?**
- **How do we parameterize the classifier $f_\theta(\mathbf{x})$?**

Digital Scores

Visual Representation



Digital (Kern) Representation

23:	2..r	2..r	2..r	2..r
24:	8r	8r	8r	8dd
25:	=1	=1	=1	=1
26:	1r	1r	8r	4dd
27:	.	.	8f# 8a	.
28:	.	.	8a 8f#	8ff#
29:	.	.	8a 8f#	16ee
30:	.	.	.	16dd

Featurizing a Score

Tensor Representation

We encode a score as a 3-axis binary tensor:

$$\mathbf{x} \in \mathcal{S} = \{0,1\}^{T\times P\times(N+D+1)}.$$

$\mathbf{x}_{t,p,n} = 1$ iff pitch n occurs at time t in spine p,
$\mathbf{x}_{t,p,N+d} = 1$ iff note-value d occurs at time t in spine p,
$\mathbf{x}_{t,p,N+D} = 1$ iff pitch n continues at time t in spine p,

- T - The number of rows of pitch/note-value data in the score S.
- P - The maximum number of concurrent **kern columns (spines).
- N - The number of pitches, ranging from C1 to F#7.
- D - The number of distinct note values (i.e. durations).

f and d denote pitch & duration components of the score:

$$\mathbf{x} = \mathbf{f} \oplus \mathbf{d}, \quad \mathbf{f} \in \{0,1\}^{T\times P\times N}, \quad \mathbf{d} \in \{0,1\}^{T\times P\times(D+1)}.$$

Classification Models

Model #1: Voice Convolution

Convolve over each spine (part); pool the part representations:

$$h_{t,p}(\mathbf{x};\theta) = \text{relu}\left((W^1_\theta)^\top \mathbf{x}_{t:t+n,p}\right), \qquad W^1_\theta \in \mathbb{R}^{n(N+D+1)\times k},$$

$$h^2_{t,p}(\mathbf{x};\theta) = \text{relu}\left((W^2_\theta)^\top h_{t:t+n,p}(\mathbf{x};\theta)\right), \qquad W^2_\theta \in \mathbb{R}^{nk_1\times k_2},$$

$$h_{\text{conv}}(\mathbf{x};\theta) = \frac{1}{TP}\sum_{t=1}^{T}\sum_{p=1}^{P} h^2_{t,p}(\mathbf{x};\theta),$$

$$f_\theta(x) = (W_\theta)^\top h_{\text{conv}}(\mathbf{x};\theta), \qquad W_\theta \in \mathbb{R}^{(N+D+1)\times C}.$$

Model #2: Harmonic Convolution

Convolve over frequencies across parts; pool temporally:

$$h_{t,u}(\mathbf{f};\theta) = \text{relu}\left((W^1_\theta)^\top \mathbf{f}_{t,:,u:u+j}\right), \qquad W^1_\theta \in \mathbb{R}^{jP\times k},$$

$$h_t(\mathbf{f};\theta) = \frac{1}{N}\sum_{u=1}^{N} h_{t,u}(\mathbf{f},\theta),$$

$$h^2_t(\mathbf{x};\theta) = \text{relu}\left((W^2_\theta)^\top h_t(\mathbf{f};\theta) + (W^3_\theta)^\top \mathbf{d}_t\right), \qquad W^2_\theta \in \mathbb{R}^{k\times k_2}, W^3_\theta \in \mathbb{R}^{(D+1)\times k2},$$

$$h_{\text{harmonic}}(\mathbf{x};\theta) = \frac{1}{T}\sum_{t=1}^{T} h^2_t(\mathbf{x};\theta),$$

$$f_\theta(\mathbf{x}) = (W_\theta)^\top h_{\text{harmonic}}(\mathbf{x};\theta), \qquad W_\theta \in \mathbb{R}^{(N+D+1)\times C}.$$

Model #3: Voice/Harmonic Hybrid

Combine the representations of Models 1 and 2; train jointly:

$$f_\theta(\mathbf{x}) = (W^c_\theta)^\top h_{\text{conv}}(\mathbf{x};\theta) + (W^h_\theta)^\top h_{\text{harmonic}}(\mathbf{x};\theta), \qquad W^c_\theta \in \mathbb{R}^{k_2\times C}, W^h_\theta \in \mathbb{R}^{k_2\times C}.$$

Hybrid Model Results

Overall Results

Composer	Dates	Sub-Collection	Scores	Accuracy
Du Fay	1397-1474	Choral	35	74.3
Ockeghem	1410-1497	Choral	98	72.4
Busnois	1430-1492	Choral	68	60.3
Martini	1440-1497	Choral	122	73.8
Compere	1445-1518	Choral	27	37.0
Josquin	1450-1521	Choral	423	82.3
de la Rue	1452-1518	Choral	178	79.2
Orto	1460-1529	Choral	43	48.8
Japart	1474-1507	Choral	22	13.6
Corelli	1653-1713	Chamber	188	95.2
Vivaldi	1678-1741	Chamber	33	54.5
Bach	1685-1750	Choral/Keyboard	466	97.6
D. Scarlatti	1685-1757	Keyboard	59	72.9
Haydn	1732-1809	Chamber	209	82.3
Mozart	1756-1791	Chamber/Keyboard	151	67.5
Beethoven	1770-1827	Keyboard	169	89.3
Hummel	1778-1837	Keyboard	24	91.7
Chopin	1810-1849	Keyboard	76	68.3
Joplin	1868-1917	Keyboard	47	91.5
Overall	1397-1917	-	2,438	81.7

3 Composer Confusion Matrix

	Bach	Haydn	Beethoven
Bach	99.8	0.2	0.0
Haydn	3.4	93.3	3.3
Beethoven	3.0	10.6	86.4

http://homes.cs.washington.edu/~thickstn/ismir2019classification/