Modeling Musical Expression as a Score-Performance Hidden Markov Model With Invariant Risk Minimization

Quinn Ouyang
School of Music, College of Fine and Applied Arts
University of Illinois Urbana-Champaign
Champaign, Illinois, United States
qouyang3@illinois.edu

***Abstract*—We define score-performance analysis as a novel task of comparing a performance/relization of music to its symbolic representation (e.g. a recording vs. sheet music). One perspective of this task focuses on understanding *how* a musician performs rather than *what* they perform, i.e. obtaining a "profile" of their musical interpretation, style, etc. across their performances. We propose modeling this as a hidden Markov model where we can directly observe scores and performances, but not performance profiles. Applying invariant risk minimization yields on this chain yields an interpretable representation of musical expression that minimizes score information (in theory–we have not proved nor verified this).**

***Index Terms*—music score, musical expression, hidden Markov model, invariant risk minimzation, information theory**

I. Introduction

While most audio and music computation research centers around fairly well-defined tasks (e.g automatic music transcription or instrument classification), developments on more subjective problems remains scarce. One of the biggest gaps in the field is bridging between symbolic music representations and audio, especially as the academic community tends to seperate these domains into music information retrieval and signal processing, respectively. We take on the challenge of modeling the inherently human process of musical performance. (Cancino-Chacón et al. 2018)

Fundamentally, scores are manuscripts that represent a musical idea. They usually detail a sequence of pitches (notes) with broad performance directions (instruments, tempos, dynamics, etc.) to interpret and realize as a performance. As musicians inevitably read and play music differently, multiple performances of varying acoustic characteristics can come from a single score. Therefore, through a simplified perspective, a score can map to multiple performances but a performance can map to only one score.

We focus on quantifying musical interpretation, i.e. the transition between a music score and a performance of it.

II. Related Work

A. Quantifying Musical Expression

Very little computational research surrounding musical interpretation of scores exist. (Cancino-Chacón et al. 2018; Friberg, Bresin, and Sundberg 2006)

B. Hidden Markov Models

C. Invariant Risk Minimization

III. Formulation

Our goal is to learn a representation of musical expression Y from paired score-performance data $X := \{(s_i, p_i)\}_{i=1}^N$ where $s_i \in X_s$ and $p_i \in X_p$.

A. Minimizing Risk Across Environments

From the perspective of IRM, we inevitably sample scores from different environments, so X technically consists of sets of score-performance pairs where $X^e := \{(s_i^e, p_i^e)\}_{i=1}^N, \forall e \in E$ where $s_i^e \in X_s^e$ and $p_i^e \in X_p^e$ (Arjovsky et al. 2020; Huszár 2019).

To form a somewhat realistic structural equation model, we make several simplifying assumptions in the score-to-performer flow to relate these random variables.

- A score $s_i^e \in X_s^e$ only depends on its environment $e \in E$, i.e. $p(s_i^e, e) = p(s_i^e \mid e)p(e)$
- A performer's profile $y \in Y$ only depends on the score s_i^e it receives, i.e. $p(y_i, s^e) = p(y \mid s_i^e)$
- A performer's performance $p_i^e \in X_p^e$ depends on its score s_i^e and profile p_i^e, i.e. $p(p_i^e, s_i^e, y_e) = p(p_i^e \mid s_i^e, y_e)p(s_i^e, y_e)$

The chain rule of conditional probabilities yields: (Berdahl and Blandino 2020; Degirmenci 2014)

$$\begin{aligned} p(e, s_i, p_i, y) &= p(y \mid e, s_i, p_i)p(e, s_i, p_i) \\ &= p(y \mid e, s_i, p_i)p(p_i \mid e, s_i)p(e, s_i) \\ &= p(y \mid e, s_i, p_i)p(p_i \mid e, s_i)p(s_i \mid e)p(e) \\ &= p(y \mid s_i, p_i)p(p_i \mid s_i)p(s_i \mid e)p(e) \end{aligned}$$

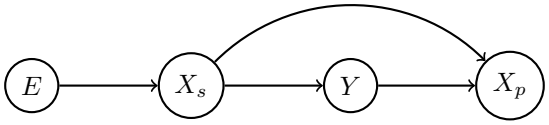
B. As a Communications Channel

We can interpret a performer's profile as a semi-deterministic encoder from a score s_i^e to a performance p_i^e which maximizes the mutual information between the two (though always performing music as closely to what a composer writes is debatable from a musicology standpoint). (Berdahl and Blandino 2020; Meyer 1957)

C. As a Hidden Markov Model

We can model this as a simplified feed-forward HMM:

- A set of score environments E directly influences the distribution of observed scores X_s
- A random variable Y describing a performance profile, dependent on the input score
- A set of performances X_p bijectively mapped their corresponding scores



REFERENCES

[1] Martin Arjovsky et al. *Invariant Risk Minimization*. 2020. arXiv: 1907.02893 [stat.ML].

[2] Edgar Berdahl and Michael Blandino. "Modeling a Musician Performing on a Digital Musical Instrument as a Communications Channel". In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. ¡conf-loc¿, ¡city¿Honolulu¡/city¿, ¡state¿HI¡/state¿, ¡country¿USA¡/country¿, ¡/conf-loc¿: Association for Computing Machinery, 2020, pp. 1–7. ISBN: 9781450368193. DOI: 10.1145/3334480.3382841. URL: https://doi.org/10.1145/3334480.3382841.

[3] Carlos E. Cancino-Chacón et al. "Computational Models of Expressive Music Performance: A Comprehensive and Critical Review". In: *Frontiers in Digital Humanities* 5 (2018). ISSN: 2297-2668. DOI: 10.3389/fdigh.2018.00025. URL: https://www.frontiersin.org/articles/10.3389/fdigh.2018.00025.

[4] Alperen Degirmenci. *Introduction to Hidden Markov Models*. 2014. URL: https://scholar.harvard.edu/files/adegirmenci/files/hmm_adegirmenci_2014.pdf.

[5] Anders Friberg, Roberto Bresin, and Johan Sundberg. "Overview of the KTH rule system for musical performance". In: *Advances in Cognitive Psychology* 2 (Jan. 2006). DOI: 10.2478/v10053-008-0052-x.

[6] Ferenc Huszár. *Invariant Risk Minimization: An Information Theoretic View*. 2019. URL: https://www.inference.vc/invariant-risk-minimization/.

[7] Leonard B. Meyer. "Meaning in Music and Information Theory". In: *The Journal of Aesthetics and Art Criticism* 15.4 (1957), pp. 412–424. ISSN: 00218529, 15406245. URL: http://www.jstor.org/stable/427154 (visited on 05/14/2024).