Online Symbolic Music Alignment with Offline Reinforcement Learning

Silvan David Peter

Institute of Computational Perception, Johannes Kepler University Linz, Austria

JOHANNES KEPLER UNIVERSITY LINZ

Institute of Computational Perception

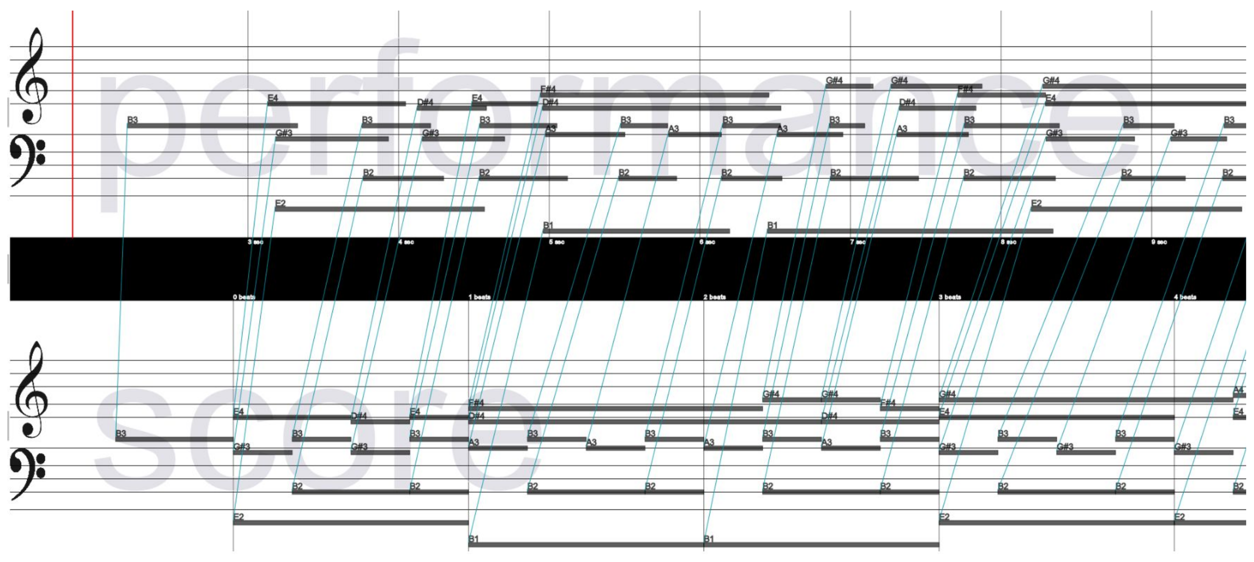
Task and Models

In this article, we present two models tackling note-wise music alignment, i.e., alignment of a MIDI performance with a corresponding musicXML score by means of matched pairs of notes. This type of alignment can happen in two ways: offline (with access to the full recording of the performance) or online (following the performance as it happens, possibly in realtime).

The two models are:

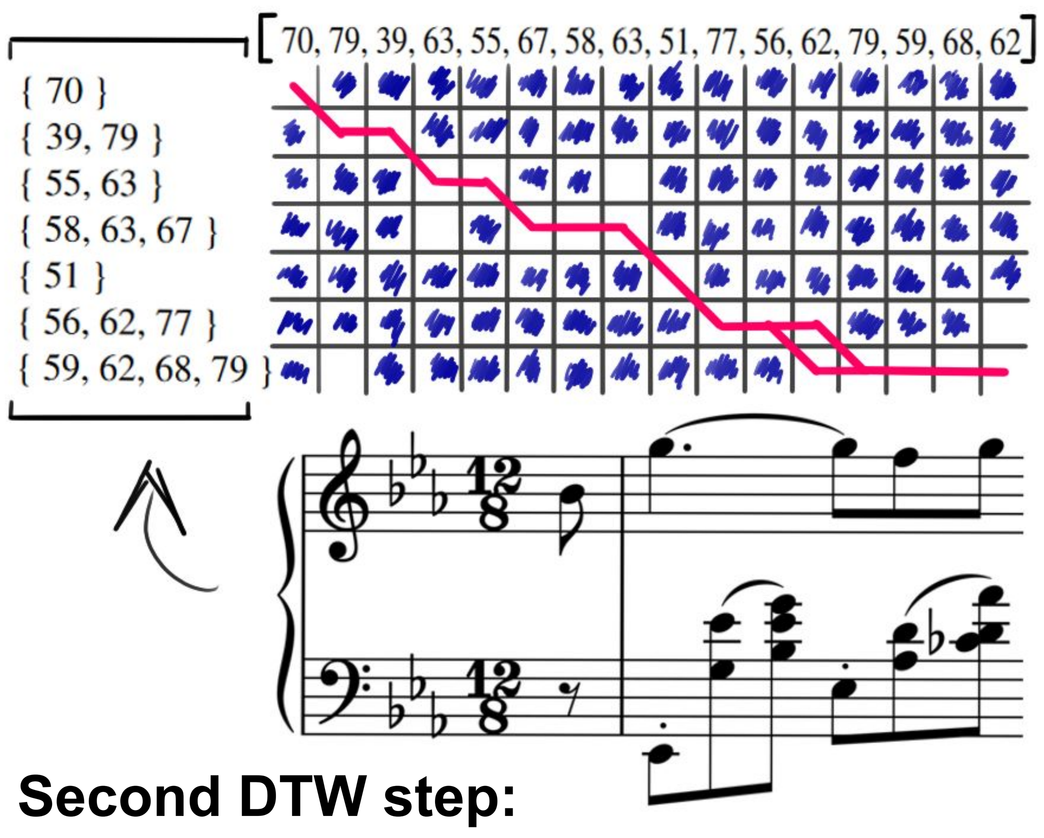
- a two-step dynamic time warping (**DTW**)-based **offline** model
- a reinforcement learning (**RL**)-based **online** model

The models share the design approach of completely separating the handling of pitch and timing information.



Left: a piano roll representation of the type of alignments investigated. Each note in both score and performance is either matched, inserted, or deleted.

Offline DTW model



First DTW step:
Pitch and pitch-set based features to create a coarse mapping from performance to score.

Using this mapping, each score onset is projected to an approximate performance time.

Second DTW step:
All notes are separated into pitch-wise sequences. The score notes of each pitch are mapped to performance time using the approximation of the first DTW path. Then, a DTW is computed between the score onsets (mapped to p. time) and the performance time, for each pitch and using the time delta between onsets as local metric. This DTW path minimizes the global distance between the sequences. For each note which is connected to multiple reference notes (horizontal or vertical steps in DTW matrix), we match the notes with minimal distance, marking the other insertions and deletions, respectively.

Dataset	DTW Offline	Nakamura
Magaloff	98.4 ± 0.9 %	97.8 ± 1.4 %
Zeilinger	99.3 ± 0.9 %	98.8 ± 1.2 %
Batik	99.4 ± 0.7 %	98.5 ± 2.1 %
Vienna 4x22	99.8 ± 0.4 %	99.5 ± 0.5 %
Combined	99.0 ± 1.0 %	98.5 ± 1.5 %

Table 1. Dataset-wise averaged F-scores and standard deviations of each model.

Offline Evaluation

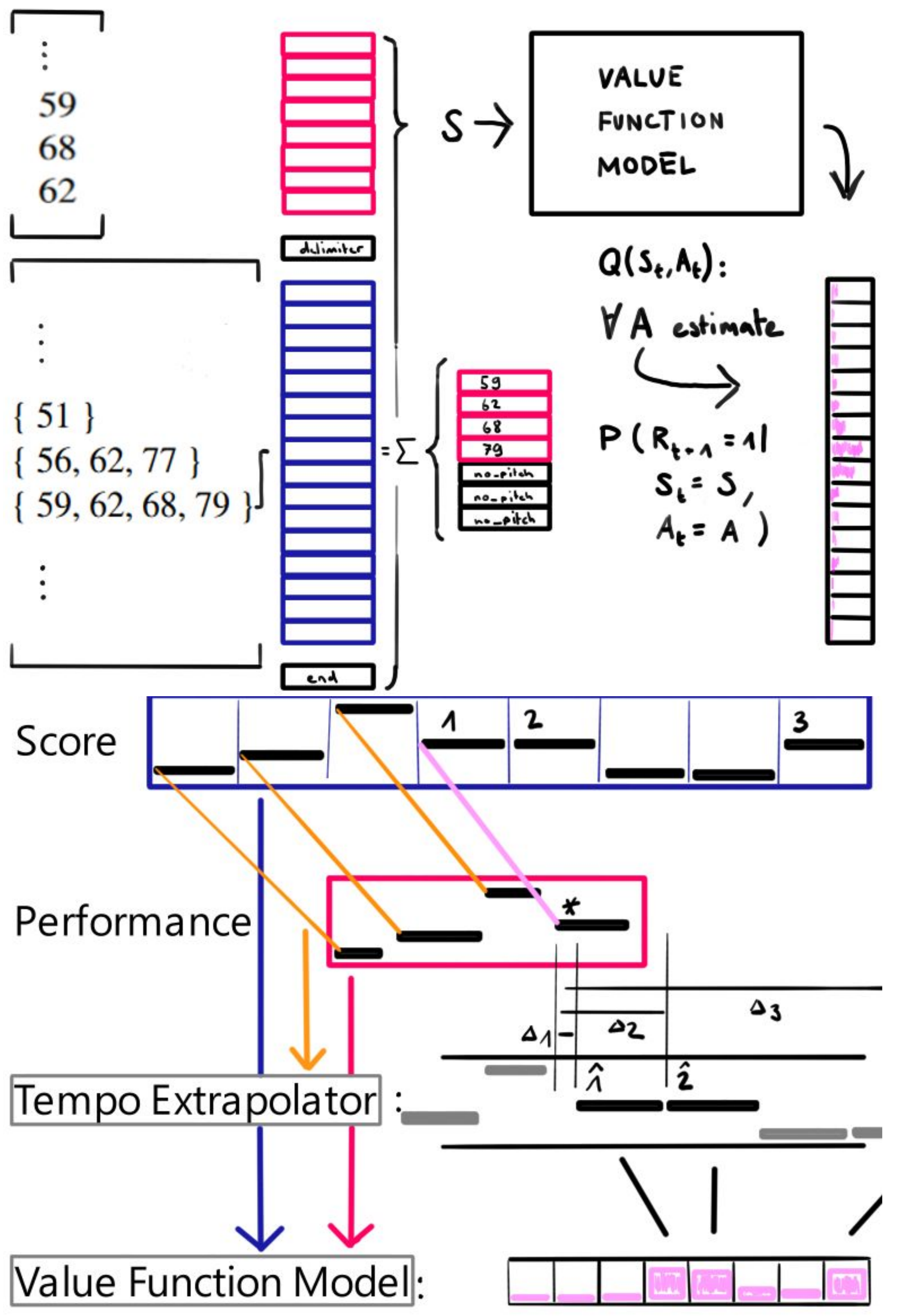
The offline model is not tuned and hence is evaluated on the whole of four high-quality datasets of note-aligned solo piano music.

The proposed model outperforms the previous state of the art significantly on all datasets except for Vienna4x22, where the state of the art already reaches a perfect F-score for many performances (see Table 1).

Online RL model

We (partially) formalize the online note alignment task as reinforcement learning problem. Specifically, we define an agent as moving on the score. The local score context, i.e., 8 score onsets before and after the agents, as well as the current performance context, i.e., the last 8 performed onsets, are given to the agent as current state (S) of its environment. The agent can pick from 16 actions (A), i.e. move to one of the score onsets within the context window. Only the choice of the score onset that corresponds to the most recent performed note receives a positive reward, all other actions are not rewarded.

We learn the agent's state-action value function Q(S,A) using a small attention-based neural network. We design the agent to be completely myopic, i.e., the value associated with each state-action tuple corresponds only to the immediate reward, not any delayed information. We further sample possible contexts from a dataset which turns the reinforcement learning problem into an offline one with supervised value function training.



Left top: the performance (red) and score (blue) context are passed to the network as sequence of tokens.The network faces a token classification problem where each score onset token is classified by its expected reward.

There are several ways to derive a policy from this value function. In our case, we add timing information to the purely pitch-based value estimator via local tempo estimation (left bottom). For the highest ranked actions, the agent picks the one with the lowest extrapolated onset time estimation error.

Online Evaluation

Model	Async	≤ 25ms	≤ 50ms	≤ 100ms
OLTW	60.6 ms	38.0 %	63.3 %	86.7 %
GAM	36.0 ms	89.0 %	91.4 %	94.6 %
OAM	15.7 ms	91.4 %	93.8 %	96.6 %

Table 4. Asynchrony of the models in score follower setting. Column "Async" presents the median asynchrony. Columns 3, 4, 5 present the percentage of onset estimates with lower asynchrony than 25ms, 50ms and 100ms, respectively.

Piece	OAM	DTW Offline	Nakamura
B. Op. 53 3rd. m.	99.0 %	99.4 %	98.2 %
C. Op. 9 No. 1	97.6 %	98.4 %	98.8 %
C. Op. 9 No. 2	97.4 %	99.1 %	97.6 %
C. Op. 10 No. 11	90.3 %	96.3 %	94.3 %
C. Op. 60	95.1 %	97.9 %	94.7 %

Table 3. Piece-wise F-scores of each model. OAM = Online Alignment Model, DTW Offline = model of section 3.3, Nakamura = reference SOTA model [11].

The online model is evaluated on an unseen test set of five pieces in two ways:

- as a score follower predicting the current score onset (Table 4), measured in asynchrony
- as a note alignment model using the same F-score used in the offline case

Implementations:

https://github.com/sildater/parangonar

pip install parangonar

References: [1] [2] [3]