OPTIMIZING DTW-BASED AUDIO-TO-MIDI ALIGNMENT AND MATCHING

Colin Raffel and Daniel P. W. Ellis

LabROSA
Columbia University
New York, NY

ABSTRACT

Dynamic Time Warping (DTW) has proven to be an extremely effective method for both aligning and matching recordings of songs to corresponding MIDI transcriptions. The performance of DTW-based approaches in this domain is heavily effected by system design choices, such as the representation used for the audio and MIDI data and DTW's adjustable hyperparameters. We propose a method for optimizing the design of DTW-based alignment and matching systems. Our technique uses Bayesian optimization to tune system design and hyperparameters over a synthetically created dataset of audio and MIDI pairs. We then perform an exhaustive search over DTW score normalization techniques in order to determine an optimal method for reporting a reliable alignment confidence score, which is necessary for matching tasks. Using our approach, we are able to create a DTW-based system which is conceptually simple and highly accurate at both alignment and matching. We also verified that our system achieves high performance in a large-scale qualitative evaluation of results on real-world data.

Index Terms**—** Dynamic Time Warping, Audio to MIDI Alignment, Sequence Retrieval, Bayesian Optimization, Hyperparameter Optimization

1. INTRODUCTION

Why is MIDI to audio alignment important? Matching?

Systems which do alignment

Systems which do matching [1] plus KDD literature

2. DTW-BASED ALIGNMENT

Formal definition of DTW-based alignment, with parameter and representation discussion

Discussion of extracting a confidence score, normalization methods

3. CREATING A SYNTHETIC ALIGNMENT DATASET

Collection of MIDI files

"Easy" corruption, for alignment accuracy

"Hard" corruption, for matching accuracy

4. OPTIMIZING DTW-BASED ALIGNMENT

Short overview of Bayesian optimization

Parameter space (including multiplicative penalty)

Random trials

Discussion of best-performing aligners; also best aligner with beats

5. OPTIMIZING CONFIDENCE REPORTING

Grid search

Statistical tests used

Choosing the best alignment scheme (with algorithm box?)

6. QUALITATIVE EVALUATION

Data preparation

Evaluation criteria

Results

7. AVENUES FOR IMPROVEMENT

Augmentation with MUDA, partial alignments, robustness to missing instruments, re-training specifically on subsequences

8. REFERENCES

[1] Ning Hu, Roger B. Dannenberg, and George Tzanetakis, "Polyphonic audio matching and alignment for music retrieval," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2003, pp. 185–188.