OPTIMIZING DTW-BASED AUDIO-TO-MIDI ALIGNMENT

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

[1]

2. 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.