Friedrich-Alexander-Universität Erlangen-Nürnberg

Friedrich-Alexander-Universität
Erlangen-Nürnberg

Internship Report

Report Title
Training with Unaligned Dataset: Soft Dynamic Time Warping

submitted by

Quang Hoang Nguyen Vo

submitted

September 2, 2025

Supervisors

Msc. Johannes Zeitler
Prof. Dr. Meinard Müller

AUDIO LABS

International Audio Laboratories Erlangen
A joint institution of the
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and
the Fraunhofer-Institut für Integrierte Schaltungen IIS.

Fraunhofer
IIS

Abstract

The evolution of Deep Neural Networks (DNNs) has shifted the paradigm of music information retrieval (MIR) from heuristic and mathematical models to data-driven approaches, which rely on large amounts of labelled training data. However, it introduces challenges when training with weakly aligned datasets. In this project, we investigate the characteristics of differential dynamic time warping (dDTW) through the soft-DTW (sDTW) algorithm when training with weakly aligned data. The main objective is to integrate soft-DTW as a loss function in the training process of a template-based chord recognition model. The dataset will have its chord label timestamps distorted or removed to simulate weakly or unaligned data. The dDTW loss function will then be used to train the model with the distorted dataset. The results will be compared with those obtained using the original dataset and the Connectionist Temporal Classification (CTC) loss function. Additional tasks may include experimenting and evaluating the performance of sDTW with different stablizing strategies.

Contents

Chapter 1

Introduction

Generally speaking, the introduction chapter should introduce the topic of the thesis and motivate the importance of it. Moreover, the introduction should give an outline of the thesis and point out the contributions of this work.

You can logically group the chapters of the thesis by using so-called parts. An example of how to insert a part-page containing a teaser-image is included in this template. Typically, you will have an introductory chapter that gives a broad overview. Then, the first part of the thesis might start after the introduction.

Chapter 2

Soft Dynamic Time Warping Algorithm

2.1 Forward Pass

2.2 Backward Pass

2.3 Tables

Non-English or Math	Frequency	Comments
Ø	1 in 1,000	For Swedish names
π	1 in 5	Common in math
\$	4 in 5	Used in business
Ψ_1^2	1 in 40,000	Unexplained usage

Table 2.1. Frequency of Special Characters. Note that this table does not contain any vertical lines which makes the table look more tidy.

Chapter 3

Experimental Setup

3.1 Data Preparation

For the experiments, we will use "the Beatles" dataset taken from Isophonics [?], which contains audio recordings and their corresponding chord annotations. This dataset is denoted as "strongly-aligned" since the chord labels are precisely aligned with the audio. Based on this dataset, we create a "soft" version by removing the repetitions in the chord annotations, resulting in a shorter chord label sequence, as illustrated in Figure **??**. This simulates a weakly-aligned dataset where the exact timing of chord changes is unknown.

3.2 Model Architecture

Given the aim of this experiment is to evaluate the performance of the proposed SDTW loss function, the network architecture plays a minor role and are kept simple. Therefore, we used a basic chord recognition model (dChord) consisting of a single convolutional layer followed by a softmax layer. Preceding this layer is a log-compression layer to reduce the dynamic range of the input features and a feature normalization layer to standardize the input features. The model takes as input a sequence of chroma features extracted from the audio signal and outputs a hot encoded vector representing the predicted chord.

3.3 Training Procedure

The training procedure will involve optimizing the model parameters using the dDTW loss function on the distorted dataset. The model will be evaluated on both the distorted and original datasets to assess its robustness to label noise. Hyperparameter tuning and regularization techniques may be applied to improve generalization.

3.4 Results

3.5 Discussion

In this section, we will discuss the implications of the results obtained from the experiments. We will analyze the performance of the proposed method in comparison to existing approaches and highlight its strengths and weaknesses. Additionally, we will explore potential avenues for future research and improvements.

Chapter 4

Conclusions

4.1 Limitation

4.2 Future Work

Draw the conclusions in the big picture of the thesis! Then, indicate future work.

Bibliography