Differentiable Dynamic Programming for Time Series Alignment

Тимур Гарипов
Татьяна Шолохова
Павел Коваленко
Саня Щербаков

9 июня 2018

Постановка задачи

Рассматривается задача выравнивания временных рядов.

Дана нотная запись музыкальной композиции и аудиозапись этой композиции. Требуется каждому моменту времени в аудиозаписи сопоставить ноту, играемую в этот момент.

Описание датасета

В работе был использован датасет Bach 10, состоящий из 10 аудиозаписей фрагментов хоралов Баха, продолжительность фрагментов — от 25 до 40 секунд.

Каждая запись состоит из четырех дорожек, соответствующих четырем инструментам — скрипка, кларнет, саксофон и фагот. Есть как записи отдельных дорожек, так и сводная запись всех инструментов.

Для каждой дорожки дана ее идеальная нотная запись (18 различных нот), однако фактическая игра от нее немного отклоняется. Также для всех дорожек дано правильное выравнивание аудио– и нотной записи.

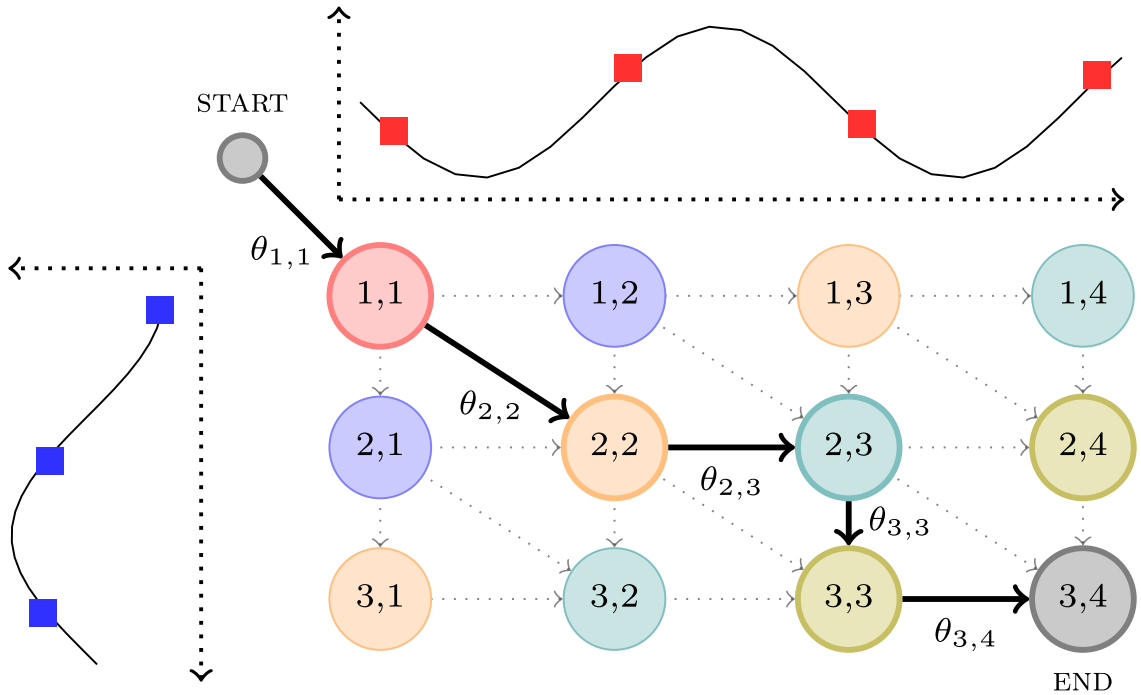
Выравнивание

Для обеих последовательностей — нотной и аудиозаписи — выделим с равными интервалами ключевые точки, для которых будем искать выравнивание.

Обозначим за N_A число нот в последовательности, а в аудиозаписи возьмем N_B точек с равными интервалами. Тогда выравнивание можно представить в виде бинарной матрицы $Y \in \{0,1\}^{N_A \times N_B}$. Единица в позиции (i,j) означает, что в j–й момент времени проигрывалась i–я нота.

Выравнивание

Предположим, что последовательность нот при игре не изменилась и ни одна из нот не была пропущена. Тогда выравнивание можно представить в виде пути в матрице Y из клетки $(1, 1)$ в клетку (N_A, N_B), при этом разрешены перемещения только вправо, вниз и вправо-вниз. Пример выравнивания — на рисунке ниже.



Метрика качества

Потребуем, чтобы N_B было больше N_A. Тогда можно потребовать, чтобы в каждый момент времени играла только одна нота, то есть для каждого момента времени требуется предсказать, какая нота сейчас играет.

Для матрицы это ограничение означает, что в каждом столбце может быть не больше одной единицы. Для пути в графе это ограничение равносильно запрету переходов вниз.

Метрика качества — *mean absolute deviation* — суммарное (по моментам времени) отклонение индекса предсказанной ноты от истинного индекса.

Простое решение

Разбиваем датасет на две части. На первой части обучаем классификатор на 18 классов — предсказываем вероятность того, что в момент времени t играет нота i.

Для второй части датасета построим матрицу $\theta \in \mathbb{R}^{N_A \times N_B}$.
θ_{ij} соответствует невероятности того, что в момент времени j играет нота номер i, то есть штрафу за предсказание ноты i для момента времени j. Этот штраф можно получить из классификатора.

Теперь требуется найти в матрице путь из клетки $(1, 1)$ в клетку (N_A, N_B) с наименьшим суммарным штрафом. Эту задачу можно решить за $N_A \times N_B$ операций при помощи динамического программирования.

Поиск минимального пути

Дана матрица $N_A \times N_B$ штрафов. Нужно найти путь из клетки $(1, 1)$ в клетку (N_A, N_B) с наименьшим суммарным штрафом. На пути из клетки можно перемещаться в ее соседа справа или справа-снизу.

Заведем матрицу D размера $N_A \times N_B$. D_{ij} равно минимальному штрафу, за который можно проложить путь из $(1, 1)$ в (i, j). Будем заполнять эту матрицу по столбцам.

База динамики

$D_{11} = \theta_{11}$, $D_{1j} = +\infty$

Шаг динамики

$D_{ij} = \min(D_{i-1,j}, D_{i-1,j-1}) + \theta_{ij}$

Аудио признаки

В статье предложено использовать следующие признаки для аудиодорожки:

- MFCC признаки — первые 5 коэффициентов.
- Root Mean Square Energy — энергия фрейма.
- Spectral Centroid — средняя частота спектра во фрейме.
- Spectral Bandwidth — разброс частот спектра во фрейме.

Были использованы реализации этих признаков из библиотеки librosa.

Аудио признаки

В качестве базовой модели для сравнения использовался описанный выше алгоритм. Базовый классификатор — логистическая регрессия (один против всех).

Первые 5 композиций использовались для обучения, остальные — для тестирования.

MAD для разных инструментов:

1. Скрипка:
2. Кларнет:
3. Саксофон:
4. Фагот: