Распознавание нот из музыки

Дмитрий Протасов

МФТИ

Научный руководитель: к.т.н. И. А. Матвеев

18 мая 2024

Цель исследования

Мотивация

Создание генеративных музыкальных моделей удобно в пространстве MIDI. Однако, MIDI-датасетов недостаточно, большинство песен доступны только в аудиоформате. Решение — алгоритм преобразования аудио в MIDI.

Цель

Улучшить качество транскрипции существующих моделей автоматической музыкальной транскрипции.

Метод решения

Предлагается использовать ранее известные методы для разделеления вокальных и инструментальных частей на отдельные аудиозаписи, а также применять квантизацию по bpm и фильтрацию по музыкальной тональности для улучшения качества транскрипции.

Литература

- Automatic Musical Instrument Recognition, 2001 (Eronen)
- Jointist: A multi-faceted approach to music-source-separation, instrument-recognition, and transcription, 2023.
- Basic-pitch: lightweight instrument-agnostic model for polyphonic note transcription and multipitch estimation, 2022.

Формальная постановка задачи

Рассматривается аудиосигнал $A(t) : [0, T] \to \mathbb{R}$, целью является транскрибировать его в последовательность событий MIDI $S = \{(n_k, i_k, t_{on_k}, t_{off_k}) | k = 1, \ldots, N\}$, где n_k — номер MIDI ноты, i_k — тип инструмента, t_{on_k} и t_{off_k} — время начала и окончания ноты, N — число нот в MIDI последовательности.

Автоматическая музыкальная транскрипция включает следующие подзадачи:

- Восстановление отдельных музыкальных источников $A_j(t)$ из исходной аудиозаписи $A(t) = \sum_j A_j(t)$
- Классификация каждого $A_j(t)$ в $i_k \in I$, где $I = \{i_1, i_2, \ldots, i_M\}$.
- Транскрипция $A_m(t) \to S_m = \{(n_k, i_k, t_{on_k}, t_{off_k}) | k = 1, \ldots, N\}$, где $A_m(t)$ — аудиосигнал (как правило, моноинструментальный)

Цель задачи – максимизировать метрику F_{no}, описанную на следующем слайде

Метрика F_{no}

Метрика F_{no} (F-measure-no-offset) используется для оценки качества транскрибированных музыкальных записей, учитывая точность, полноту и перекрытие временных интервалов нот. Нота считается правильно транскрибированной, если выполняются следующие условия:

- Начало ноты находится в пределах ± 50 мс.
- Высота тона в пределах ± 50 центов.
- Если offset_ratio задан, окончание ноты должно быть в пределах этого значения от длительности эталонной ноты. Если не задан, окончание ноты не учитывается.

Предложенный метод

Предлагается гибридное решение, интегрирующее элементы AMT, такие как определение тональности и оценка BPM.

Квантизация BPM:

$$t_{on}^{quant} = \left\lfloor \frac{t_{on} \cdot BPM}{60} \right\rceil \cdot \frac{60}{BPM}, \tag{1}$$

$$t_{off}^{quant} = \left\lfloor \frac{t_{off} \cdot BPM}{60} \right\rceil \cdot \frac{60}{BPM}, \tag{2}$$

Фильтрация по тональности:

$$F_{key}(n) = \begin{cases} 1, & \text{если } n \in \text{Тональность} \\ 0, & \text{иначе} \end{cases} \tag{3}$$

Это позволяет повысить точность транскрипции, сокращая количество потенциальных ошибок.

Теоретическое обоснование

Теорема (Протасов, 2024)

Пусть $A(t)$ — аудиосигнал, обладающий нулевой фазой и находящийся в одной тональности, и $S = \{(n_i, t_{on_i}, t_{off_i})\}$ — последовательность событий MIDI, полученная из некоторой AMT-модели при трансформации $A(t)$. Пусть $S' = \{(n'_i, t'_{on_i}, t'_{off_i})\}$ — последовательность, полученная после применения квантизации BPM и фильтрации по тональности к S. Тогда метрики Accuracy и F_{no} не уменьшатся, то есть: $F(S) \leq F(S')$ для $F \in \{\text{Accuracy}, F_{no}\}$.

Вычислительный эксперимент

Эксперименты проводились на датасете BabySlakh и синтетическом MIDI датасете с целью проверки гипотезы об улучшении точности транскрипции через анализ тональности и BPM.

Использовались метрики F_{no} с разными параметрами «offset» для оценки эффективности подходов, включая базовую модель AMT, модель с анализом тональности, модель с квантизацией BPM, и комбинированную модель.

Результаты

Модель	F_{no} (без ‘offset‘)	F_{no} (с ‘offset=0.2‘)
Обычная модель	0.4189	0.1669
+ Key Estimation	0.4129	0.1639
+ BPM Quantization	0.4149	0.1647
+ Key + BPM	0.4089	0.1620

Таблица: Сравнение эффективности различных методов по улучшении транскрипции модели на датасете BabySlakh.

Заключение и планы на будущее

Заключение

- Предложен подход по улучшению качества
- Гипотеза протестирована на датасете babyslakh и синтетическом датасете

Планы на будущее

- Рассмотреть задачу Audio Style Transfer для улучшения транскрипции
- Применить и протестировать подход по улучшению распознавания вокальных партий с помощью Timestamped-ASR

Проект про поиск оптимальных покрытий

- Новая статья (в процессе написания)
  - Для всех $n \geqslant 6$ на плоском торе доказана нижняя оценка методом Ленца
  - Получен ряд верхних оценок на интервальное хроматическое число тора при $3 \leqslant n \leqslant 8$ на интервальное хроматическое число тора
- D.S. Protasov, A.D. Tolmachev, V.A. Voronov “Optimal partitions of the flat torus into parts of smaller diameter”
  - Подана на конференцию IEEE