

ПРОБЛЕМА АВТОМАТИЧЕСКОЙ МУЗЫКАЛЬНОЙ ТРАНСКРИПЦИИ

Е. Р. Связова

sviyazovaer@gmail.com

ФГБОУ ВО «Уфимский государственный авиационный технический университет» (УГАТУ)

Аннотация. Рассматривается проблема автоматической музыкальной транскрипции, выделяются основные этапы музыкальной транскрипции, производится обзор существующих решений. Предлагается разработать собственное приложение для решения задачи. Описываются архитектура приложения, процесс работы приложения и место нейронной сети в нем.

Ключевые слова: музыкальная транскрипция; цифровая аудиозапись; нотный текст; спектрограмма; сверточная нейронная сеть; дискретное преобразование Фурье; piano roll; MIDI.

1. АКТУАЛЬНОСТЬ

Музыка занимает важное место в жизни многих людей. Люди вовлекаются в музыку как слушатели, исполнители, создатели. Начиная с роли слушателя, часто впоследствии хочется попробовать себя в роли исполнителя. И здесь недостаточно одних навыков игры на музыкальном инструменте – нужна «инструкция», которая позволит исполнить конкретную композицию. В качестве такой «инструкции» выступает нотный текст.

Нотный текст содержит информацию о высоте, длительности и порядке извлечения звуков, темпе и инструментальном составе композиции. Он записывается с помощью особого языка – современной музыкальной нотации. Нотация отражает некоторые ограничения, связанные с историческими и популярными музыкальными традициями: соотношения между частотами звуков, составляющих композицию, дискретность длительностей звуков и др.

На данный момент наиболее распространенный способ хранения и передачи композиции – ее цифровая аудиозапись. Процесс получения нотного текста на основе аудиозаписи композиции называется музыкальной транскрипцией. Этот процесс можно осуществить вручную, однако он предъявляет высокие требования к его ис-

полнителю – музыкальный слух, опыт игры на различных инструментах – и занимает много времени. В связи с чем появляется проблема автоматической музыкальной транскрипции (АМТ).

В то же время акцент при анализе мультимедийной информации сместился от технических к содержательным аспектам. Для музыки это информация о мелодии, ритме, жанре, настроении композиции и т.п. Эта информация может быть полезна во множестве областей, в частности:

1) информационный поиск в музыке – поиск композиций на основе сходства мелодических, гармонических и структурных характеристик;

2) обработка аудиозаписей – например, разложение записи на дорожки с отдельными инструментами, их раздельная обработка;

3) системы интерактивного взаимодействия человека и ЭВМ – например, автоматическое создание аккомпанемента в процессе живого исполнения;

4) программное обеспечение (ПО) для связанного с музыкой оборудования, например, для создания световых эффектов;

5) расширение выборки материалов для музыковедческих исследований путем анализа концертных записей и импровизаций [1].

Описание композиции с помощью музыкальной нотации может быть той проме-

жуточной информацией, которая позволит извлечь некоторые содержательные характеристики композиции. Таким образом, результаты, полученные в ходе решения проблемы АМТ, могут быть использованы для решения проблемы извлечения содержательной информации из музыкальной аудиозаписи.

2. ЭТАПЫ ЗАДАЧИ АМТ

При решении задачи АМТ аудиозапись рассматривается как сумма отдельных специфических звуков. Каждому звуку ставится в соответствие нота – атомарный звук, извлекаемый с помощью музыкального инструмента, и основная единица музыкальной нотации. Эти два фактора определяют специфику звуков. Например, присутствие в звуке основной частоты с наибольшей амплитудой и ее гармоник с меньшими амплитудами, затухающий характер звука, набор основных частот и отношения между ними, особая организация звуков во времени и др. В связи с этим в решении задачи АМТ можно выделить несколько этапов.

1. Получение частотно-временного представления аудиозаписи – обычно спектрограммы.

2. Выделение на спектрограмме отдельных звуков – нот. При этом определяются их характеристики с точки зрения звука: в частности, основная частота, момент начала и длительность звука в абсолютных временных единицах (напр. секундах).

3. Создание нотного текста. При этом определяются характеристики нот и записи в целом с точки зрения музыкальной нотации: в частности, высота ноты, длительность ноты в относительных временных единицах, положение ноты в такте.

3. ОБЗОР СУЩЕСТВУЮЩИХ РЕШЕНИЙ

Задача АМТ нетривиальна, поэтому существуют решения, позволяющие выполнить программно лишь некоторые этапы, предлагая выполнить остаток работы вручную или с помощью других инструментов.

Распространены программные решения, не выполняющие автоматически ни один из этапов, но предоставляющие инструменты, многократно упрощающие их выполнение.

Основные возможности таких программ – проигрывание аудиозаписи с разной скоростью без искажений по высоте, проигрывание отдельных отрезков записи и их зацикливание. Также возможны наличие виртуальной клавиатуры музыкального инструмента, создание текстовых меток и комментариев, привязанных к отрезкам записи. Примеры программ – Tune Transcriber, Anytune.

Подзадачи выделения нот и определения их высоты наиболее требовательны к пользователю. Поэтому одной из лидирующих программ для помощи в транскрипции музыки является Transcribe! Она предоставляет множество инструментов для упрощения транскрипции, выполняет автоматически первые два этапа задачи АМТ и определяет высоту обнаруженных нот. Результаты представляются в piano roll – подобном виде. Задача записи нотного текста решается пользователем самостоятельно.

Третья группа программ позволяет решить задачу АМТ полностью. Как правило, такие программы содержат инструменты для корректировки промежуточных результатов. Например, возможность выставления границ нот в абсолютных единицах вручную, добавление нот, не обнаруженных программой как корректировка результатов первых двух этапов. Это TS-WIDI PRO, ScoreCloud, AnthemScore, Melody Scanner и др.

Такие программы, как правило, обладают возможностью экспорта результатов в файл формата MIDI. С файлами такого формата умеют работать практически все нотные редакторы.

Задача АМТ может полностью решаться еще одной группой программ – цифровыми звуковыми рабочими станциями (англ. digital audio workstation, DAW). Это крупные системы, предусматривающие законченный цикл работ по производству музыки – от первичной записи до получения готового результата. Преобразование аудиозаписи в нотный текст, как правило, является дополнительной функцией таких систем. Она может присутствовать в системе изначально или в составе расширений, плагинов. Примеры программ – (Celemony) Melodyne, (Steiberg) Cubase и SONAR, Music Creator (Cakewalk).

4. СОЗДАНИЕ ПРИЛОЖЕНИЯ

Решение задачи АМТ может быть достаточно ресурсоемким процессом. В то же время большую часть целевой аудитории приложений для АМТ составляют начинающие музыканты. Необходимо упростить процесс использования приложения и снизить требования к устройствам пользователя. Архитектура веб-приложения успешно справляется с этой задачей. Веб-приложения не требуют установки, а ресурсоемкие процессы могут выполняться на стороне сервера. Подавляющее большинство перечисленных выше программ представляют из себя настольные приложения. Исключения составляют Tune Transcriber и Melody Scanner.

Второй этап решения задачи АМТ наиболее требователен к пользователю с точки зрения опыта. В решении задач с подобной спецификой хорошо показали себя технологии нейронных сетей. Такие технологии используются в AnthemScore.

Необходимо создать решение, сочетающее простоту использования, достижимую за счет применения веб-архитектуры, и опыт, полученный при обучении нейронной сети. Рассмотрим, как может быть построено такое приложение. Пусть оно позволяет автоматически выполнить первые два этапа АМТ.

4.1. ВХОДНЫЕ И ВЫХОДНЫЕ ДАННЫЕ

Входными данными приложения будет аудиозапись в одном из распространенных цифровых форматов. Выходными – описание момента начала и длительности, основной частоты, громкости каждой выделенной ноты. Такую информацию можно наглядно отобразить на спектрограмма – подобном графике.

Стоит отметить, что подавляющая часть современной музыки написана с использованием равномерного темперированного строя и настройкой высоты «ля» первой октавы на 440 Гц. Приняв это допущение для анализируемой музыки, мы можем легко перейти от частоты звука к его высоте в терминах музыкальной нотации. Тогда, зная высоту звуков, мы можем отобразить результат работы программы с помощью piano

roll – подобного графика, более привычного для музыкантов.

4.2. АРХИТЕКТУРА

На начальном этапе можно использовать двухзвенную клиент-серверную архитектуру. Таким образом, вычислительная нагрузка ложится на серверную часть.

4.3. ОПИСАНИЕ ПРОЦЕССА РАБОТЫ

Первый этап – создание спектрограммы. Наиболее общий метод – дискретное преобразование Фурье (ДПФ). Однако частоты, на которые раскладывается звук ДПФ, расположены линейно. В то время как частоты соседних по высоте нот отличаются фиксированным множителем ($\sqrt[12]{2}$). Использование ДПФ приведет к тому, что полученная сетка частот будет слишком плотной в высоком регистре и недостаточно плотной в низком. Увеличение разрешения по частоте происходит за счет уменьшения разрешения по времени. Желательно подобрать метод разложения, учитывающий особенности расположения частот нот. [2]

Второй этап – выделение нот. Если рассматривать полученную спектрограмму как изображение, эта задача сводится к распознаванию образов. Такие задачи решаются с помощью сверточных нейросетей. Однако размер спектрограммы по оси времени не фиксирован, так как не фиксирован размер аудиозаписи.

В вопросах определения входных и выходных данных для нейронной сети можно воспользоваться опытом разработчиков AnthemScore. Предлагается использовать нейросеть только для определения наличия нот на данном временном интервале. Таким образом, на входе нейросети фрагменты спектрограммы, обладающие данными по всему спектру частот во временном интервале фиксированной продолжительности. На выходе – вероятности для 88 высот нот [3].

Анализ с помощью нейросети всех возможных фрагментов спектрограммы может оказаться затратным по времени. Разумно было бы проводить анализ тех фрагментов, на которых происходят изменения – появляются или исчезают ноты. Для определения таких моментов можно использовать функцию

спектрального потока, описанную С. Диксоном в алгоритме определения темпа [4].

Для обучения нейросети необходимы фрагменты аудиозаписей с точной по времени информацией о начале и окончании нот в них. Таким условиям удовлетворяют нотный текст в формате MIDI и записи, созданные на его основе с помощью виртуальных музыкальных инструментов. Благодаря тому, что записи созданы на основе MIDI-файла, а не исполнены реальным человеком, моменты начала и окончания каждой ноты в них точно соответствуют моментам, описанным в MIDI-файле. Использование качественных виртуальных инструментов позволяет максимально приблизить звучание к реальному исполнению. Можно использовать, например, датасет MAPS. [5] Он содержит 120 аудиозаписей и соответствующих MIDI-файлов. Чтобы получить готовые данные для обучения, нужно получить спектрограмму каждой записи и извлечь фрагменты из записи и соответствующего MIDI-файла.

Таким образом, нейросеть позволяет для фрагмента спектрограммы оценить вероятности наличия конкретных нот в этом временном интервале. Имея такой инструмент, можно получить динамику этих вероятностей по времени и на основе нее построить предположения и моментах начала и окончания соответствующих нот.

ЗАКЛЮЧЕНИЕ

Рассмотрена задача АМТ, обоснована ее актуальность, перечислены области применения. В решении задачи выделено 3 этапа. Существующие решения разбиты на группы в зависимости от того, какая часть задачи АМТ решается автоматически. Предлагается создать веб-приложение, автоматически выполняющее первые 2 этапа задачи АМТ и использующее технологии нейронных сетей. Рассматриваются алгоритм работы приложения, место нейронной сети в нем. Предлагается использовать сверточную нейронную сеть для анализа ограниченных по времени фрагментов спектрограммы для получения вероятностей наличия нот. Предложен датасет для обучения нейронной сети.

СПИСОК ЛИТЕРАТУРЫ

1. **Kirchhoff H.** A user-assisted approach to multiple instrument music transcription. PhD thesis. QMUL, London, 2013. [Электронный ресурс] URL: <https://www.eecs.qmul.ac.uk/~simond/phd/HolgerKirchhoff-PhD-Thesis.pdf> (дата обращения 14.04.2019). [H. Kirchhoff (2013, Apr. 14). "A user-assisted approach to multiple instrument music transcription" [Online], Phd thesis, QMUL, London, 2013. Available: <https://www.eecs.qmul.ac.uk/~simond/phd/HolgerKirchhoff-PhD-Thesis.pdf>]
2. **Music** Transcription with Convolutional Neural Networks [Электронный ресурс]. URL: <https://www.lunaverus.com/cnn> (дата обращения 14.04.2019). [(2019, Apr. 14). Music Transcription with Convolutional Neural Networks [Online]. Available: <https://www.lunaverus.com/cnn>]
3. **Сапунов Г.** Введение в архитектуры нейронных сетей [Электронный ресурс]. URL: <https://habr.com/company/oleg-bunin/blog/340184/> (дата обращения 14.04.2019). [G. Sapunov (2019, Apr. 14). Introduction to neural networks architectures [Online], (in Russian). Available: <https://habr.com/company/oleg-bunin/blog/340184/>]
4. **Dixon S.** Evaluation of audio beat tracking system BeatRoot // Journal of New Music Research. 2007. 36(1). P. 39–51 [S. Dixon, "Evaluation of audio beat tracking system BeatRoot", in Journal of New Music Research, no. 36(1), pp. 39-51, 2007.]
5. **MAPS** Database – A piano database for multipitch estimation and automatic transcription of music [Электронный ресурс]. URL: <http://www.tsi.telecom-paristech.fr/aao/en/2010/07/08/maps-database-a-piano-database-for-multipitch-estimation-and-automatic-transcription-of-music/> (дата обращения 14.04.2019). [(2019, Apr. 14). MAPS Database – A piano database for multipitch estimation and automatic transcription of music [Online]. Available: <http://www.tsi.telecom-paristech.fr/aao/en/2010/07/08/maps-database-a-piano-database-for-multipitch-estimation-and-automatic-transcription-of-music/>]

ОБ АВТОРЕ

СВИЯЗОВА Екатерина Романовна, магистрант. каф. ВМиК.

METADATA

Title: Musical transcription using neural network technology

Authors: E. R. Sviyazova

Affiliation:

Ufa State Aviation Technical University (UGATU), Russia.

Email: sviyazovaer@gmail.com

Language: Russian.

Source: Molodezhnyj Vestnik UGATU (scientific journal of Ufa State Aviation Technical University), no. 1 (20), pp. 159–160, 2019. ISSN 2225-9309 (Print).

Abstract: Discusses the automatic music transcription problem. Music transcription task is split into main stages. An overview on existing solutions is provided. Proposed application development. Application architecture, algorithm and place of neural network in general work process are described.

Key words: music transcription; digital audio record; sheet music; spectrogram; convolutional neural network; discrete Fourier transform; piano roll; MIDI.

About authors:

SVIYAZOVA, Ekaterina Romanovna., master student 2 year, Ufa state aviation technical University