УДК 004:89:004.93

ПЕРСПЕКТИВНІ ПІДХОДИ ДО РОЗВ'ЯЗАННЯ ЗАДАЧІ АВТОМАТИЗОВАНОГО ТРАНСКРИБУВАННЯ МУЗИЧНИХ КОМПОЗИЦІЙ

Арсенюк Ігор, Кучеровський Юрій

Вінницький національний технічний університет

Анотація

В даній статті запропоновано використання методів глибокого навчання нейронних мереж для розв'язання задачі автоматизованого визначення послідовностей акордів в музичних композиціях.

Abstract

In this paper methods of neural networks deep learning were suggested to solve the tasks of automated chord recognition in musical compositions.

Вступ

Аналіз музичних композицій є однією з областей, що активно вивчаються у даний час. Конкретизація завдання дослідження музики приводить до ряду напрямків, таких як класифікація музики за жанрами, визначення наявності музичних інструментів у композиції, розпізнавання голосу або звуків окремих музичних інструментів в аудіозаписі та вилучення (отримання) їх партії з аудіокомпозиції. Більш глибоке дослідження музики актуалізує завдання відновлення окремих нот музичного твору і автоматичної ідентифікації акордів у цифровому звуці. Завданням цієї роботи є автоматичне відновлення послідовності акордів по цифровому запису музичної композиції. Рішення цього завдання має на меті максимально деталізувати розбір музичного твору. Актуальність розв'язання цієї задачі досить очевидна.

Ціллю роботи є представлення перспективного підходу щодо розв'язання задачі автоматизованого транскрибування музичних композицій.

Стислий огляд перспективних підходів до вирішення задачі автоматизованого транскрибування музичних композицій

Задача автоматизованого транскрибування полягає в отриманні множини послідовностей акордів із зазначенням позиції кожного з них. Таке представлення може бути проміжним етапом у роботі інших алгоритмів, а також може представляти цінність само по собі: за його допомогою можна індексувати музичні твори для пошуку композицій за заданою послідовністю акордів, знаходити різноманітні аранжування однієї і тієї ж композиції. Дану інформацію можна використовувати також для визначення структури композиції, її розділення на більш крупні сегменти.

Найбільш вдалий метод вирішення вищеокресленої проблеми, який на сьогоднішній день знайшов широке застосування дозволяє забезпечити до 80% точності та використовує хроматичні профілі (хромограми) у якості ознак та Гауссові змішані моделі (Gaussian Mixture Models) [1]. Проте даний рівень точності, порівняно із сучасними системами розпізнавання мови, де мав місце якісний перехід до глибоких нейронних мереж (Deep Learning), є доволі низькою [2, 3].

В останні роки підходи, які використовують глибоке навчання, як спосіб створення ієрархічних представлень великої кількості даних, набули значного інтересу серед дослідників. Глибоке навчання було особливо вдало застосоване у системах розпізнавання мови та класифікації зображень. Техніки глибокого машинного навчання, з їх можливостями знаходити складні зв'язки у значному обсязі даних, без сумніву мають

значний потенціал для застосування їх з метою вирішення завдання отримання послідовностей акордів.

Особливості підготовки даних

Відомо, що для тренування нейронної мережі потрібно мати достатньо великий обсяг даних. Стандартного набору транскрибованих композицій "Бітлз", який складається зі ста вісімдесяти пісень не вистачає для повноцінного тренування. У всесвітній мережі Internet є сотні професійно виконаних табулатур, але усі вони не сегментовані, тобто для них не виконане часове вирівнювання назви акорду і відповідного сегменту композиції. Серед досліджень у галузі розпізнавання мови була представлена техніка CTC (Connectionist Temporal Classification) [4], яка наразі використовується у передових сервісах Baidu [5], Google [6] та ін. CTC може застосовуватися як softmax шар виводу мережі, який видає ймовірності позначень акордів відносно усіх можливих послідовностей позначень вхідних послідовностей, включаючи паузи та не потребує попередньої сегментації.

Використання часової структурованості даних

Так як і у природній мові, послідовності акордів досить сильно корелюють у часі. Для використання структурованості аудіоданих доречно використати архітектуру з керованими рекурентними нейронами GRU (Gated Recurrent Unit) запропоновану у роботі [7]. Порівняно з рекурентними нейронними мережами RNN (Recurrent Neural Network), GRU можуть зберігати та використовувати більш довготривалі взаємозалежності послідовностей. Крім того GRU стійкі до проблеми розмиття-розростання градієнтів. Порівняно з нейронами з тривалою короткочасною пам’яттю LSTM (Long Short Term Memory) GRU мають схожі властивості, але за рахунок внутрішньої структури використовують на 25% менше пам'яті та більш стабільні під час обробки довгих послідовностей [8, 9].

Використання перспективних методів композиції нейронної мережі.

Сучасні архітектури RNN/LSTM/GRU, які використовують часовий зв'язок у послідовностях даних зберігають інформацію попередніх часових кроків (t-1) і використовують її при обчисленні поточних (t) кроків. Групою дослідників у роботі [10] було запропоновано методи композиції шарів нейронної мережі при яких використовується не тільки інформація попередніх кроків (t-1) а й інформація з наступних (t+1) кроків.

Тестування на популярних наборах даних для задач розпізнавання мови показало високу ефективність двонаправлених LSTM (Bidirectional LSTM). Ці методи композиції можливо та доцільно було б перенести і на GRU архітектуру [11].

Отримання ознак

Глибокі нейронні мережі успішно отримують високорівневі ознаки з необроблених даних, при цьому їх якість часто перевищує якість ретельно підібраних "ручних" ознак. Для запропонованої архітектури доцільно відмовитися від ручного підбору ознак (хромограм), та дозволити глибокій нейронній мережі реалізувати її потенціал. Достатньо за допомогою віконного перетворення Фур'є з подальшою L2-нормалізацією створити спектрограму та застосувати метод аналізу головних компонент (Principal component analysis) для зменшення розмірності.

Висновок

Як показали дослідження, застосування запропонованих технік дало значний приріст якості розпізнавання мови, що дає усі підстави сподіватися, що їх використання у

задачі автоматизованого транскрибування музичних композицій також дозволить підвищити якість розпізнавання акордів.

Список використаних джерел:

1. Audio Chord Estimation Results 2016 [Електронний ресурс] : [Веб-сайт]. – Режим доступу: http://www.music-ir.org/mirex/wiki/2016:Audio_Chord_Estimation_Results (дата звернення 01.09.2016). – Назва з екрана.

2. Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning. An MIT Press book. [Electronic resource]. – Access mode: http://www.deeplearningbook.org/ (last access: 01.09.2016). – Title from the screen.

3. Alex Graves, Mohamed Abdel-rahman, Hinton, Geoffrey. Speech Recognition with Deep Recurrent Neural Networks: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on: 6645–6649. [Electronic resource]. – Access mode: http://www.cs.toronto.edu/~graves/icassp_2013.pdf (last access: 01.09.2016). – Title from the screen.

4. Alex Graves, Santiago Fernández, Faustino Gomez, Jürgen Schmidhuber. Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks. ICML 2006, Pittsburgh, USA, pp. 369-376. [Electronic resource]. – Access mode: http://www.cs.toronto.edu/~graves/icml_2006.pdf (last access: 01.09.2016). – Title from the screen.

5. Deep Speech 2: End-to-End Speech Recognition in English and Mandarin: Proceedings of The 33rd International Conference on Machine Learning, pp. 173–182, 2016 [Electronic resource]. – Access mode: http://jmlr.org/proceedings/papers/v48/amodei16.pdf (last access: 01.09.2016). – Title from the screen.

6. Haşim Sak, Andrew Senior, Kanishka Rao, Françoise Beaufays, Johan Schalkwyk. Google voice search: faster and more accurate. Google Research Blog 24.09.2015 [Electronic resource]. – Access mode: https://research.googleblog.com/2015/09/google-voice-search-faster-and-more.html (last access: 01.09.2016). – Title from the screen.

7. Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, Yoshua Bengio. Gated Feedback Recurrent Neural Networks. – 17.06.2015, pp. 1-9. [Electronic resource]. – Access mode: https://arxiv.org/pdf/1502.02367.pdf (last access: 01.09.2016). – Title from the screen.

8. Rafal Jozefowicz, Wojciech Zaremba, Ilya Sutskever. An Empirical Exploration of Recurrent Network Architectures: Proceedings of The 32nd International Conference on Machine Learning, pp. 2342–2350, 2015. [Electronic resource]. – Access mode: http://jmlr.org/proceedings/papers/v37/jozefowicz15.pdf (last access: 01.09.2016). – Title from the screen.

9. Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, Yoshua Bengio. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, 2014. [Electronic resource]. – Access mode: https://arxiv.org/pdf/1412.3555.pdf (last access: 01.09.2016). – Title from the screen.

10. Alex Graves, Navdeep Jaitly and Abdel-rahman Mohamed. Hybrid Speech Recognition With Deep Bidirectional LSTM. ASRU 2013, Olomouc, Czech Republic. [Electronic resource]. – Access mode: http://www.cs.toronto.edu/~graves/asru_2013.pdf (last access: 01.09.2016). – Title from the screen.

11. Florian Eyben, Sebastian Bock, Bjorn Schuller. Universal Onset Detection with Bidirectional Long-Short Term Memory Neural Networks. 11th Intern. Soc. for Music Information Retrieval Conference, ISMIR, Utrecht, Holland, pp. 589-594, August 2010. [Electronic resource]. – Access mode: http://ismir2010.ismir.net/proceedings/ismir2010-101.pdf (last access: 01.09.2016). – Title from the screen.

12. Бардаченко В. Ф. Таймерні нейронні елементи та структури. Монографія / В. Ф. Бардаченко, О. К. Колесницький, С. А. Василецький. – Вінниця : УНІВЕРСУМ-Вінниця, 2005. – 126 с.