

Projekt "Genre Sherlock" – system służący rozpoznawaniu gatunków muzyki

Ł. DUDEK, A. BODURKA, M. CISZEWSKI, M. FAŁOWSKI, A. GRABOWSKI,
M. JASKÓŁA, M. KANTOR, W. KITKA, K. MUCHA, K. OLSZOWY,
A. SKIRZYŃSKI, K. STACHOŃ, M. STELMASZCZUK, A. SZELIGA

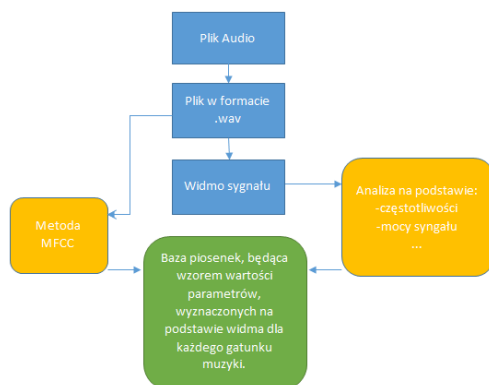
Akademia Górniczo-Hutnicza

Streszczenie

Rozpoznawanie gatunków muzyki może nie być najpotrzebniejszym programem codziennego użytku lecz samo jego stworzenie daje wiele możliwości poznania sposobów rozpoznawania i określania dźwięku. Metody zastosowane w opisywanym projekcie mają zastosowanie m.in. w rozpoznawaniu mowy, a więc na pewno można je zaliczyć do istotnych.

I. WSTĘP

Genre Sherlock został stworzony przez zespół 14 osób na Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie. Zadaniem aplikacji jest rozpoznawanie gatunku muzycznego załadowanego utworu MP3 spośród 6 "gatunków bazowych": rock, jazz, metal, hip-hop, muzyka klasyczna oraz muzyka elektroniczna.



Rys. 1. Schemat tworzenia bazy danych

Aplikacja back-end została napisana w języku R, w środowisku RStudio, natomiast front-end ...

Pierwszym krokiem było stworzenie bazy danych (Rys. 1). Baza danych jest zbiorem różnych parametrów uzyskanych za pomocą zaimplementowanych metod.

Nasz zespół starannie wyselekcjonował 20 typowych utworów muzycznych dla każdego z wyżej podanych gatunków. Każda ścieżka dźwiękowa została skonwertowana na plik w formacie WAV (ang. WAVE form audio format) w celu dekompresji. Następnie dane zostały podzielone na 20ms fragmenty oraz poddane dyskretniej transformacji Fouriera (DFT) i przekazane do metod służących wyłuskaniu parametrów. Na ich podstawie stworzona została baza danych każdego z sześciu gatunków muzycznych.

W celu rozpoznania załadowanego utworu dokonywana jest tożsama operacja, lecz otrzymane parametry są porównywane z utworzoną wcześniej bazą danych.

II. METODY

Genre Sherlock, w procesie identyfikacji wykorzystuje dwie metody służące otrzymaniu parametrów danego utworu. Są to: autorska funkcja GetMeanAndStd (ang. Get mean and standard

deviation) oraz powszechnie stosowana MFCC (ang. Mel-frequency cepstral coefficients).

I. Uzyskanie Widma

Jak opisano we wstępie, obydwie zastosowane metody badają widmo sygnału. Plik WAV jest zamieniany na utwór MONO a do badań wycinane są 2 minuty ze środka utworu. Powyższe operacje są możliwe dzięki bibliotece tuneR. Następnie sygnał dzielony jest na ramki o określonej długości, które częściowo mogą się na siebie nakładać [2]. Aby uniknąć zniekształceń wybierane są różne typy okien czasowych, w tym przypadku najczęściej wykorzystuje się okno Hamminga.¹

Widmo amplitudowe dla każdej ramki wyznaczane jest przy pomocy dyskretnej transformaty Fouriera. DFT przekształca ciąg próbek sygnału na w ciąg harmonicznym zgodnie ze wzorem (1).

$$A_k = \sum_{n=0}^{N-1} a_n w_n^{-kn} \quad (1)$$

gdzie:

$$w_n = e^{i\frac{2\pi}{N}}$$

$$k \in [0, N-1]$$

k – numer harmonicznnej

A_k – wartość k -tej harmonicznnej

n – numer próbki sygnału

a_n – wartość n -tej próbki sygnału

N – liczba próbek

Ponieważ aplikacja została napisana w języku R, w celu wyznaczenia spektrogramu (widmo dla wielu ramek czasowych) skorzystano z gotowej funkcji specgram z biblioteki signal.

II. GetMeanAndStd

Pierwsza z zastosowanych metod jest bardzo prosta. Dla każdej ramki oblicza średni poziom sygnału [dB] dla wszystkich częstotliwości oraz odchylenie standardowe tych wartości. Następnie wyniki są uśredniane dla wszystkich okien czasowych. Jest to wykonane za

pomocą funkcji rowMeans oraz rowSds z biblioteki matrixStats.

III. MFCC

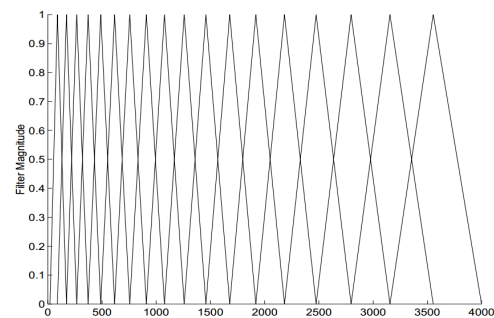
Metoda MFCC wykorzystywana jest w szeroko pojętych systemach rozpoznawania mowy oraz klasyfikacji gatunków muzycznych. Jako parametry charakteryzujące dany gatunek stosuje się często średnią oraz odchylenie standardowe MFCC^[1].

Wyznaczanie MFCC^[2;3]:

- Podział sygnału dźwiękowego na ramki
- Dyskretna transformata Fourier'a (DFT)
- Wyznaczenie logarytmu z widma amplitudy
- Przeliczenie na skalę melową (mel scale)
- Dyskretna transformacja cosinusowa (DTC)

Ponieważ pierwsze dwa kroki mające na celu uzyskanie widma zostały omówione powyżej skupmy się na kolejnych.

Następnym krokiem jest wyznaczenie logarytmu. Ze względu na to, że ludzkie ucho jest lepiej przystosowane do rozpoznawania zmian w zakresie niskich niż wysokich częstotliwości [4], wprowadzona została skala melowa, która wydaje się bardziej naturalna. Oparta jest ona na związku pomiędzy częstotliwością czystego tonu harmonicznego, a częstotliwością postrzeganą przez człowieka.



Rys. 2. Skala melowa^[4]

Zależność pomiędzy częstotliwością sygnału, a wartością tej częstotliwości wyrażonej w skali melowej opisana jest wzorem (2)^[2].

¹kkkkkjhgj

$$f_m = 2595 \cdot \log_{10}(1 + \frac{f}{100}) \quad (2)$$

gdzie:

f – częstotliwość sygnału

f_m – częstotliwość sygnału wyrażona w skali melowej

Aby wyznaczyć współczynniki cepstralne należy wykonać odwrotną szybką transformację Fouriera (IFFT). Ze względu na to, że widmo jest w skali logarytmicznej, symetryczne oraz rzeczywiste zadanie to sprowadza się do zastosowania dyskretnej transformacji cosinusowej. Współczynniki cepstralne wyznaczone przy pomocy DTC opisane są zależnością (3).^[6]

$$C_n = \sqrt{\frac{2}{N}} \sum_{i=1}^N \log(S_i) \cdot \cos[\frac{\pi n}{N}(i - \frac{1}{2})] \quad (3)$$

gdzie:

C_n – n -ty współczynnik cepstralny

S_i – i -ty współczynnik po przetworzeniu sygnału przez filtry

N – ilość filtrów

Analiza współczynników cepstralnych pozwala na ocenę zależności pomiędzy częstotliwościami składowych spektralnych w sygnale^[6].

Najczęściej stosuje się 26 filtrów melowych. Większa ilość nie prowadzi do znaczącej poprawy jakości rozpoznawania^[7].

III. RESULTS

Tabela 1: *Example table*

Name		
First name	Last Name	Grade
John	Doe	7.5
Richard	Miles	2

Tu nic nie ma i nie będzie dopóki nie zrobimy #testy.

IV. DISCUSSION

I. Subsection One

tu też

II. Subsection Two

i tu

LITERATURA

LITERATURA

- [1] [Automatic music genre classification using modulation spectral contrast feature, 2007] Chang-Hsing Lee, Jau-Ling Shih, Kun-Ming Yu, Jung-Mau Su.
- [2] [Music Classification based on MFCC Variants and Amplitude Variation Pattern: A Hierarchical Approach, 2012] Arijit Ghosal, Rudrasis Chakraborty, Bibhas Chandra Dhara, Sanjoy Kumar Saha.
- [3] [Genre classification and the invariance of MFCC features to Key and Tempo, 2011] Tom LH. Li, Antoni B.
- [4] [On The Mel-scaled Cepstrum] H.P. Combrinck, E.C. Botha
- [5] <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>
- [6] <http://winntbg.bg.agh.edu.pl/rozprawy/9873/full9873.pdf>
- [7] <http://winntbg.bg.agh.edu.pl/rozprawy2/10009/full10009.pdf>