Projekt "Genre Sherlock" – system służący rozpoznawaniu gatunków muzyki

Ł. Dudek, A. Bodurka, M. Ciszewski, M. Fałowski, A. Grabowski,
M. Jaskuła, M. Kantor, W. Kitka, K. Mucha, K. Olszowy,
A. Skirzyński, K. Stachoń, M. Stelmaszczuk, A. Szeliga

Akademia Górniczo-Hutnicza

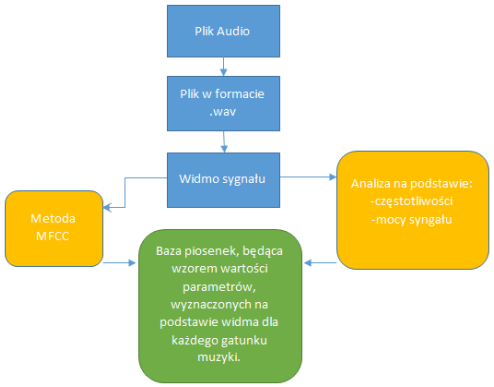
Abstrakt

Rozpoznawanie gatunków muzyki jest bardzo złożonym zagadnieniem, ponieważ dotyka wielu różnych aspektów analizy dźwięku: od badania widma krótkich fragmentów utworu po obserwacje powtarzalności na przestrzeni całości jego trwania. Od obliczania współczynników korelacji bazowanych na paśmie przenoszenia mocy po znajdowanie analogii dla wielu różnych przedziałów częstotliwości (liczonych w różnych skalach) dla wielu różnych utworów. Prezentowana w niniejszym artykule aplikacja, Genre Sherlock, na pewno nie wyczerpuje tematu, ale może być solidną podstawą pod dalsze badania na polu identyfikacji i klasyfikacji gatunków muzyki, jak również kilku pokrewnych.

Keywords: *rozpoznawanie gatunków muzyki; MFCC, badanie widma; RStudio; Shiny;*

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 5 "gatunków bazowych": rock, jazz, metal, hip-hop oraz muzyka klasyczna.



Rys. 1. Schemat tworzenia bazy danych

Celem prac było powstanie aplikacji rozpoznającej podzielonej na dwie części: front-end oraz back-end. Aplikacja back-end została napisana w języku R, w środowisku RStudio, natomiast front-end w R i HTML 5 z wykorzystaniem JavaScript i CSS, w środowisku Shiny przeznaczonym do tworzenia paneli użytkownika do aplikacji w R. Część widoczna (front-end) jest przeznaczona do umieszczenia na serwerze z uwagi na technologie wykorzystane do jej stworzenia.

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 dyskretnej transformacie Fouriera (DFT) i prze-

kazane do metod służących wyłuskaniu parametrów. Na ich podstawie stworzona została baza danych każdego z pięciu gatunków muzycznych.

W celu rozpoznania załadowanego utworu dokonywana jest tożsama operacja, lecz otrzymane parametry są porównywane z utworzoną wcześniej bazą danych. Sposób reprezentacji wyników znajduje się na Rys. 5.

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-friequency cepstral coeficcients).

II.1. 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.

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

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

gdzie:
$w_n = e^{i\frac{2\pi}{N}}$
$k \epsilon [0, N-1]$
k – numer harmonicznej
A_k – wartość k-tej harmonicznej
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.2. 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`.

II.3. MFCC

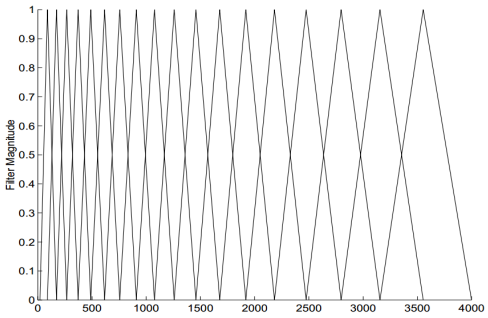
Metoda MFCC wykorzystywana jest w szeroko pojętych systemach rozpoznawania mowy oraz klasyfikacji gatunków muzycznych. Jako parametry charakteryzujące dany gatunek stosuje się często średnią oraz odchylenie standardowe MFCC [1].

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. Okna trójkątne rozłożone zgodnie ze skalą melową[4]

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

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

gdzie:
f – częstotliwość sygnału
f_m – częstotliwość sygnału wyrażona w skali melowej

Aby wyznaczyć współczynniki cepstralne[1] należy wykonać odwrotną szybką transformatę 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]

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

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].

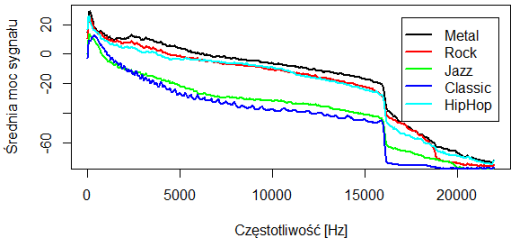
Powyższe operacje wykonywane są za pomocą funkcji `melfcc` z biblioteki `tuneR`.

Posiadając 26 współczynników obliczamy ich średnią dla wszystkich ramek czasowych danego utworu tak, aby otrzymać jeden wektor 26-ciu parametrów (wektor średnich). W analogiczny sposób otrzymujemy wektory odchyleń standardowych, współczynników przyrostowych i odchyleń współczynników przyrostowych.

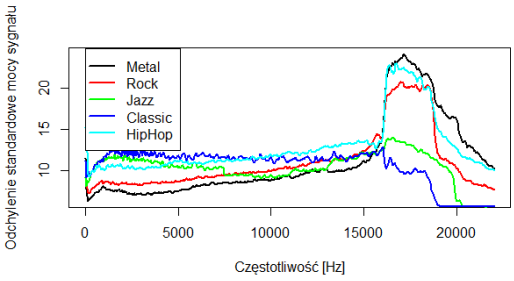
II.4. Uzyskanie bazy danych

Baza danych jest generowana poprzez uśrednienie poszczególnych parametrów uzyskanych dla każdego utworu wchodzącego w skład uprzednio wyselekcjonowanych utworów typowych dla danego gatunku.

Na poniższych rysunkach znajduje się porównanie bazy danych wygenerowanej przez metodę GetMeanAndStd dla różnych gatunków.



Rys. 3. Średnia moc sygnału



Rys. 4. Odchylenie standardowe sygnału

Wykresy dla metody MFCC nie zostały przedstawione na rysunkach ze względu na ich małą czytelność.

[1] Czy wiesz, że: słowo *cepstrum* jest anagramem słowa *spectrum*?

III. Wyniki

Pierwsze testy miały za zadanie sprawdzić korelację wsteczną. Po stworzeniu bazy danych analizie poddano utwory wchodzące w skład bazy. W Tab. 1. zawarto średnie wyniki różnicy między badaną piosenką a bazą danych dla pięciu gatunków (im mniej tym lepiej).

Utwór \ Baza	Metal	Rock	Jazz	Klasyczna	Hip-Hop
Metal	4.83	5.52	10.70	12.82	5.96
Rock	5.99	5.37	9.12	10.90	5.91
Jazz	10.83	8.9	5.22	5.76	9.01
Klasyczna	12.53	10.19	4.78	3.78	10.19
Hip-Hop	5.89	5.38	9.26	10.94	4.94

Tab. 1. Średnie różnice bazowych utworów od baz danych

Jak wynika z powyższego testu, badając utwory służące do generacji bazy danych otrzymujemy niezłe wyniki. Wprawdzie różnice między podobieństwem do różnych baz danych (np. Metal porównywany z bazą Metal, Rock oraz Hip-Hop) są niewielkie to jednak wybór gatunku według minimalnej różnicy jest poprawny.

Następnie zbadane zostało działanie układu dla losowych utworów z różnych gatunków nie wchodzących w skład piosenek bazowych. Wyniki zawarte w Tab. 2. zawierają średnie dopasowanie (na podstawie 3 piosenek) do każdego z garunku (im więcej tym lepiej).

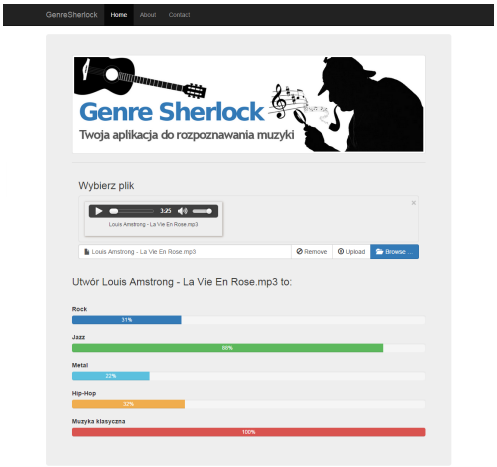
Utwór \ Gatunek	Metal	Rock	Jazz	Klasyczna	Hip-Hop
Metal	53.36	51.39	28.27	23.20	54.57
Rock	53.72	51.92	30.14	24.65	50.69
Hip-Hop	35.50	43.18	37.24	34.49	44.10

Tab. 2. Średnie przyporządkowanie bazowych utworów do baz danych

Użyte zostały piosenki:

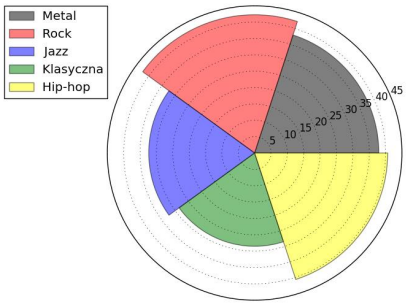
- **Metal:** Godsmack - Straight out of line, Metallica - Frantic, Metallica - St. Anger
- **Rock:** Muse- Starlight, Billy Talent - White Sparrows, Bon Jovi - Story of my life
- **Hip-Hop:** 50 Cent - My gun go off, Black Eyed Peas - Hey mama, 50 Cent - Amusement park

Wyniki są reprezentowane na panelu front-end. Użytkownik może wprowadzić jeden utwór wykorzystując pole wyboru pliku. Po wprowadzeniu pliku i odczekaniu chwili zostaną wygenerowane wykresy przedstawiające przynależność do poszczególnych gatunków muzyki. Dodatkowo użytkownik może skorzystać z podglądu wprowadzonego pliku i odsłuchać go w całości, bądź w kawałku.



Rys. 5. Okno aplikacji

Opcjonalnie zostanie użyty inny system wyświetlania informacji przedstawiony na poniższym rysunku.



Rys. 5. Okno aplikacji

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