Przetwarzanie i Rozpoznawanie Dzwięku

Klasyfikacja utworów muzycznych ze zbioru gtzan.

30 stycznia 2014

Autor:	**Paweł Rychły**	inf94362	ISWD	pawelrychly@gmail.com
	Dawid Wiśniewski	inf94387	ISWD	wisniewski.dawid@gmail.com

1 Wstęp teoretyczny

Problem klasyfikacji gatunku muzyki można sprowadzić do problemu uczenia maszynowego na odpowiednio przetworzonych danych będących próbkami dźwiękowymi określonej długości. Przetworzenie w tym przypadku polega na takiej transformacji utworów do pewnej przestrzeni współczynników, że możliwe jest trafne klasyfikowanie nowo przybyłych obiektów w oparciu o zgromadzoną wiedzę. W naszym projekcie jako źródła danych użyliśmy paczki ze strony Marsyas (GTZAN genre collection). Zbiór ten to ponad 1GB danych, podzielonych na 1000 plików, po 30 sekund każdy. Pliki te zaklasyfikowane są do 10 kategorii: Muzyka klasyczna, Blues, HipHop, Rock, Metal, Disco, Pop, Country, Jazz, Reggae - i dzielą zbiór danych na równe segmenty po 100 na każdą kategorię.

Zaletą korzystania z tego zbioru danych jest relatywnie wysoka długość próbek (30 sekund) oraz to, że istnieją publikacje analizujące te dane. Dzięki temu możemy odwołać się do wyników innych osób. Przykładem może być praca: "Musical Genre Classification of Audio Signals" napisana przez G. Tzanetakisa i P. Cooka. Jej autorom udało się osiągnąć precyzję oraz pokrycie na średnim poziomie ok 60

Jako biblioteki pomocniczej użyjemy w tym przypadku projektu Essentia przygotowanego przez Uniwersytet Pompeu-Fabra w Barcelonie do celów ekstrakcji danych oraz pakietu WEKA pomocnej przy klasyfikacji.

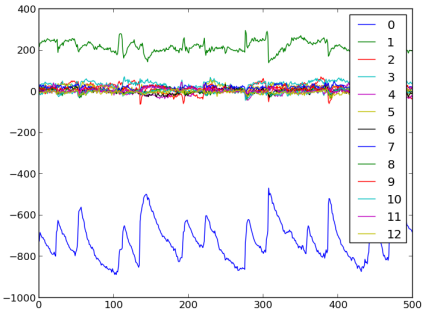
2 Przygotowanie danych

Wszystkie pliki muzyczne zakodowane są w formacie Au zaproponowanym przez firmę Sun. Biblioteka Essentia umożliwia nam wyciągnięcie informacji o poszczególnych próbkach z takiego pliku bez zagłębiania się w strukturę formatu co jest bardzo dużym ułatwieniem. Mając wczytany plik muzyczny należy wybrać pewien wektor cech, które możliwie dobrze będą reprezentować nasz problem i będą użyteczne z punktu widzenia klasyfikacji.

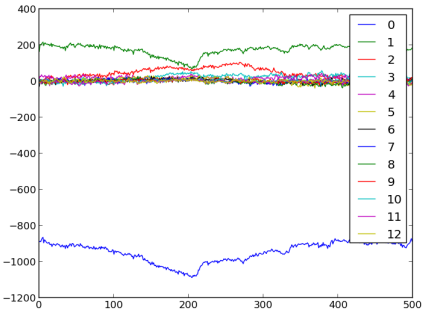
W projekcie zdecydowano się na zbadanie jakości klasyfikacji w oparciu o współczynniki melowe (MFCC). Cechy te opierają się na charakterystyce słyszenia ludzkiego ucha i są bardzo często wykorzystywane w problemach rozpoznawania mówców oraz klasyfikacji gatunków muzycznych. Współczynniki melowe obliczane były dla bardzo krótkich fragmentów czasu. Z tego powodu każdy analizowany plik dźwiękowy opisany został za pomocą kilku sekwencji liczb, odpowiadających wartością poszczególnych współczynników cepstralnych w kolejnych momentach czasu. W związku z tym, głównym problemem na jaki natrafiliśmy był sposób agregacji tych danych.

3 Analiza zmian wartości współczynników cepstralnych w czasie.

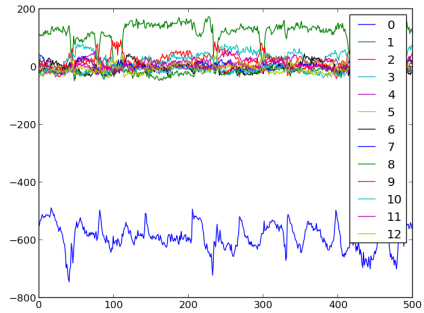
W celu dobrania najbardziej odpowiednich metod agregacji danych opisujących poszczególne próbki dźwięku, wygenerowano zbiór wykresów prezentujących zmiany wartości współczynników cepstralnych w kolejnych próbkach.



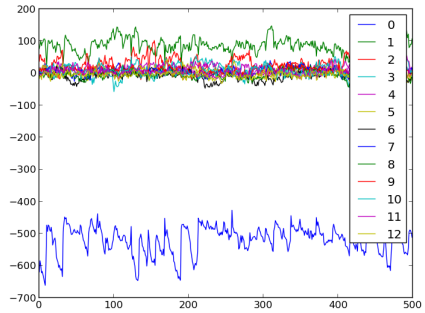
(a) blues



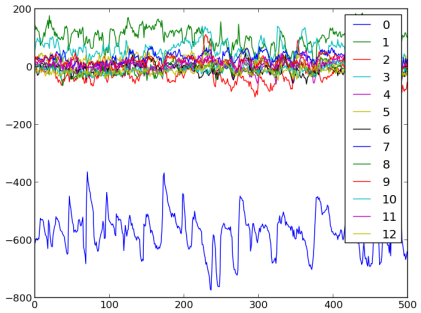
(b) classical



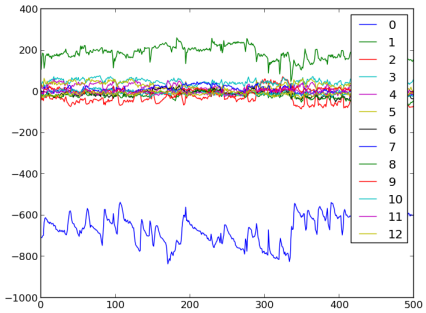
(c) country



(d) disco

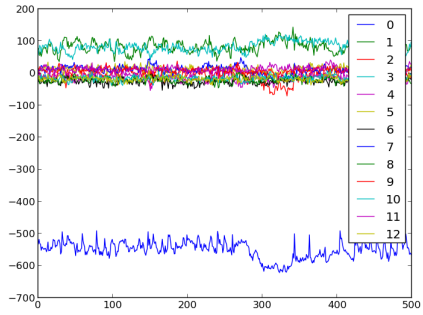


(e) hiphop

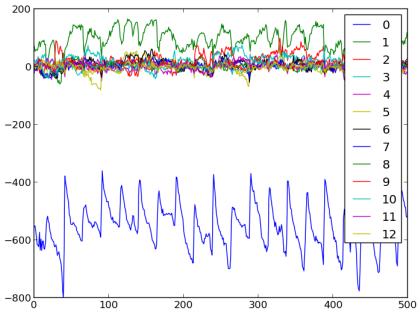


(f) jazz

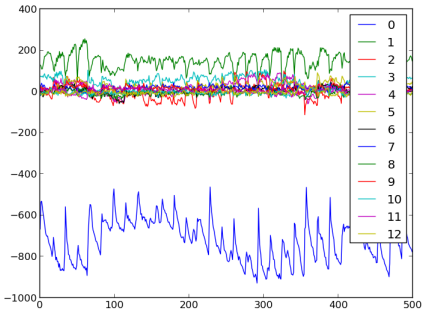
Rysunek 1: Porównanie zmian wartości współczynników cepstralnych dla przykładowych plików reprezentujących różne gatunki muzyczne.



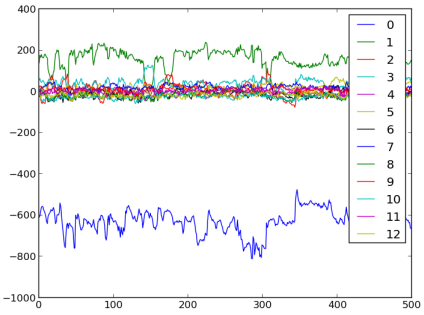
(a) metal



(b) pop



(c) reggae



(d) rock

Rysunek 2: Porównanie zmian wartości współczynników cepstralnych dla przykładowych plików reprezentujących różne gatunki muzyczne.

Jak można zauważyć, kształty wykresów ilustrujących zmiany współczynników cepstralnych, dla różnych gatunków muzycznych różnią się pomiędzy sobą. Wydaje się, że miara agregująca te dane, powinna odzwierciedlać w maksymalnym stopniu, nie tyle średnie wartości funkcji co ich kształt. Bardzo charakterystyczny jest wykres będący wynikiem analizy utworu muzyki klasycznej. Wartości współczynników cepstralnych zmianiały się tutaj w sposób zdecydowanie bardziej łagodny niż w przypadku innych wykresów.

4 Metody agregacji danych dotyczących współczynników cepstralnych.

W celu agregacji danych pochodzących z różnych okien zastosowano kilka miar. Były to takie podstawowe miary jak: wartość minimalna, wartość maksymalna, wariancja oraz średnia arytmetyczna. Ich znaczenie jest oczywiste, dlatego w dalszej części zawarto opis trzech mniej popularnych ocen sygnału. Dwie pierwsze metody wykorzystują Dyskretną transformatę Fouriera sygnału, oznaczaną literą F.

4.1 Spectral Flatness Measure

Miara ta opisuje jak bardzo analizowany sygnał zbliżony jest do szumu białego. Jeżeli wartość współczynnika jest duża, oznacza to, że wartość analizowanego sygnału (w tym przypadku wartość współczynnika cepstralnego) nie zmienia się lub, że zmienia się w sposób nieregularny. Mała wartość miary mówi o tym, że zmiany wartości współczynników cepstralnych mogłyby zostać opisane za pomocą kilku sinusoid.

$$sfm = \frac{e^{\overline{ln(F)}}}{\overline{F}} \tag{1}$$

4.2 The most significant frequency

Miara ta zawiera informacje o częstotliwości sinusoidy, która ma największy wpływ na kształt funkcji wykreślanej przez zmianę współczynnika cepstralnego w czasie.

$$f_{max} = argmax(F) \tag{2}$$

4.3 Maksymalna korelacja sygnału z własnym przesunięciem.

Ostatnia zastosowana miara, określa maksymalną korelację sygnału (Kolejnych wartości współczynnika cepstralnego), z własnym przesunięciem.

5 Wyniki

Uczenie maszynowe przeprowadzono z użyciem Baggingu oraz RandomForests (100 drzew)
Przeprowadzona analiza doprowadziła do otrzymania następujących wyników

5.1 Wyniki corssvalidacyjne

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,940	0,020	0,839	0,940	0,887	0,875	0,996	0,963	classical
0,670	0,017	0,817	0,670	0,736	0,714	0,953	0,790	hiphop
0,730	0,030	0,730	0,730	0,730	0,700	0,951	0,802	reggae
0,760	0,027	0,760	0,760	0,760	0,733	0,953	0,817	country
0,810	0,024	0,786	0,810	0,798	0,775	0,982	0,897	jazz
0,930	0,029	0,782	0,930	0,849	0,835	0,987	0,924	metal
0,730	0,036	0,695	0,730	0,712	0,680	0,946	0,776	disco
0,900	0,026	0,796	0,900	0,845	0,829	0,986	0,917	pop
0,470	0,031	0,627	0,470	0,537	0,500	0,900	0,561	rock
0,680	0,026	0,747	0,680	0,712	0,683	0,964	0,792	blues
0,762	0,026	0,758	0,762	0,757	0,732	0,962	0,824	

Rysunek 3: Podsumowanie klasyfikacji

a	b	c	d	e	f	g	h	i	j	<-- classified as
94	0	0	1	5	0	0	0	0	0	a = classical
0	67	13	2	0	2	4	7	4	1	b = hiphop
0	7	73	3	0	1	5	6	2	3	c = reggae
2	0	1	76	1	1	4	3	10	2	d = country
12	0	2	1	81	1	0	1	1	1	e = jazz
0	0	0	0	2	93	1	0	2	2	f = metal
1	5	3	5	0	4	73	5	3	1	g = disco
1	0	1	1	2	0	4	90	0	1	h = pop
2	1	3	6	6	10	12	1	47	12	i = rock
0	2	4	5	6	7	2	0	6	68	j = blues

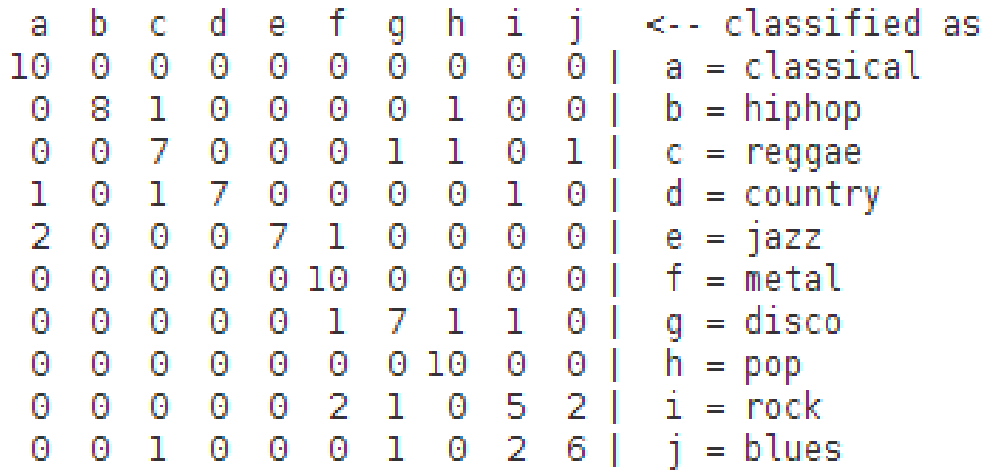
Rysunek 4: Macierz pomyłek

5.2 Wyniki z wydzielonym zbiorem testowym

Zbiór testowy równy 10% całej przestrzeni, zrównoważone klasy.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1,000	0,033	0,769	1,000	0,870	0,862	0,998	0,981	classical
0,800	0,000	1,000	0,800	0,889	0,885	0,994	0,959	hiphop
0,700	0,033	0,700	0,700	0,700	0,667	0,956	0,813	reggae
0,700	0,000	1,000	0,700	0,824	0,823	0,941	0,750	country
0,700	0,000	1,000	0,700	0,824	0,823	0,993	0,953	jazz
1,000	0,044	0,714	1,000	0,833	0,826	0,999	0,991	metal
0,700	0,033	0,700	0,700	0,700	0,667	0,996	0,962	disco
1,000	0,033	0,769	1,000	0,870	0,862	0,996	0,971	pop
0,500	0,044	0,556	0,500	0,526	0,478	0,918	0,645	rock
0,600	0,033	0,667	0,600	0,632	0,594	0,947	0,639	blues
0,770	0,026	0,787	0,770	0,767	0,749	0,974	0,866	

Rysunek 5: Podsumowanie klasyfikacji



Rysunek 6: Macierz pomyłek

6 Wnioski

Widzimy zatem, że użycie najprostszych metod jakimi są współczynniki cepstralne przynosi bardzo dobre rezultaty (przebijające nawet te, które uzyskali poprzednicy stosując GMMy). Patrząc na zróżnicowanie w jakiści wykrywania klas można dostrzec, że gatunki "gitarowe" takie jak Blues, Rock, Metal, Country są często mylone - pomysłem na jeszcze większe polepszenie wyników może być ekstrakcja informacji o średniej głośności utworu [metalowe piosenki powinny być głośniejsze niż bluesowe].