Politechnika Wrocławska

Wrocław, 27.06.2024

Projekt - Sieci neuronowe

Klasyfikacja gatunków muzycznych

Autorzy:

Jan Gromadzki [263657]

Karol Popiel [249188]

Prowadząca:

mgr inż. Monika Wasilewska

W sprawozdaniu przedstawiliśmy implementację programu służącego do klasyfikacji utworów do gatunków muzycznych. Każdy utwór muzyczny posiada wiele cech na podstawie których można określić do jakiego gatunku muzycznego należy. Klasyfikacja wykorzystuje sieć neuronową oraz metody uczenia maszynowego na podstawie ekstrahowanych cech utworów muzycznych takich jak BPMs (uderzenia na minutę), cechy chromatyczne i inne cechy które zostaną omówione w artykule. Omówimy również bazę danych zawierającą badane utwory, sposób przygotowania danych, zaprezentujemy je w formie graficznej, omówimy sposób działania naszej sieci neuronowej oraz zaprezentujemy wyniki klasyfikacji.

1. Wstęp

Utwory muzyczne składają się z wielu cech, które można ekstrahować oraz opisać za pomocą wartości liczbowych. Klasyfikacja utworów do gatunków muzycznych może być subiektywna aczkolwiek pewne konkretne cechy utworów muzycznych są obiektywne a ich przynależność do danych gatunków muzycznych jest oparta o wspomniane, obiektywne cechy. Dzięki temu możemy w obiektywny sposób znaleźć wspólne cechy danych utworów muzycznych i przypisać je do konkretnych gatunków, czyli zbiorów wspólnych cech.

2. Dane

Badania klasyfikacji prowadziliśmy używając danych z ogólnodostępnej bazy danych GTZAN. Zawiera ona 1000 ścieżek dźwiękowych, każda o długości 30 sekund. Zawiera 10 gatunków muzycznych, każdy reprezentowany przez 100 utworów. Wszystkie ścieżki to 16-bitowe pliki audio 22050 Hz Mono w formacie .wav. Dodatkowo posiada dwa zbiory cech utworów, 10 gatunków muzycznych, dla próbek 3 oraz 30 sekundowych. Każdy ze zbiorów jest zapisany w pliku w formacie .csv. Zbiór próbek 3 sekundowych zawiera 9990 próbek utworów a zbiór próbek 30 sekundowych zawiera 1000 próbek utworów. Każda z próbek posiada 60 cech.

Dataset shape: (9990, 60)

Count of each label:

	label	count
0	blues	1000
1	jazz	1000
2	metal	1000
3	pop	1000
4	reggae	1000
5	disco	999
6	classical	998
7	hiphop	998
8	rock	998
9	country	997

Dataset shape: (1000, 60)

Count of each label:

	label	count
0	blues	100
1	classical	100
2	country	100
3	disco	100
4	hiphop	100
5	jazz	100
6	metal	100
7	pop	100
8	reggae	100
9	rock	100

Rys 1. Prezentacja liczności danych ze zbiorów cech utworów. Po lewej stronie próbki 3 sekundowe a po prawej stronie 30 sekundowe

Na załączonym rys 1. zaprezentowaliśmy liczność próbek 3 oraz 30 sekundowych, utworów muzycznych z 10 gatunków muzycznych do których należą blues, jazz, metal, pop, reggae, disco, muzyka klasyczna, hiphop, rock oraz country. W celu badania klasyfikacji użyliśmy zbioru cech próbek 3 sekundowych utworów ze względu na większą liczbę próbek.

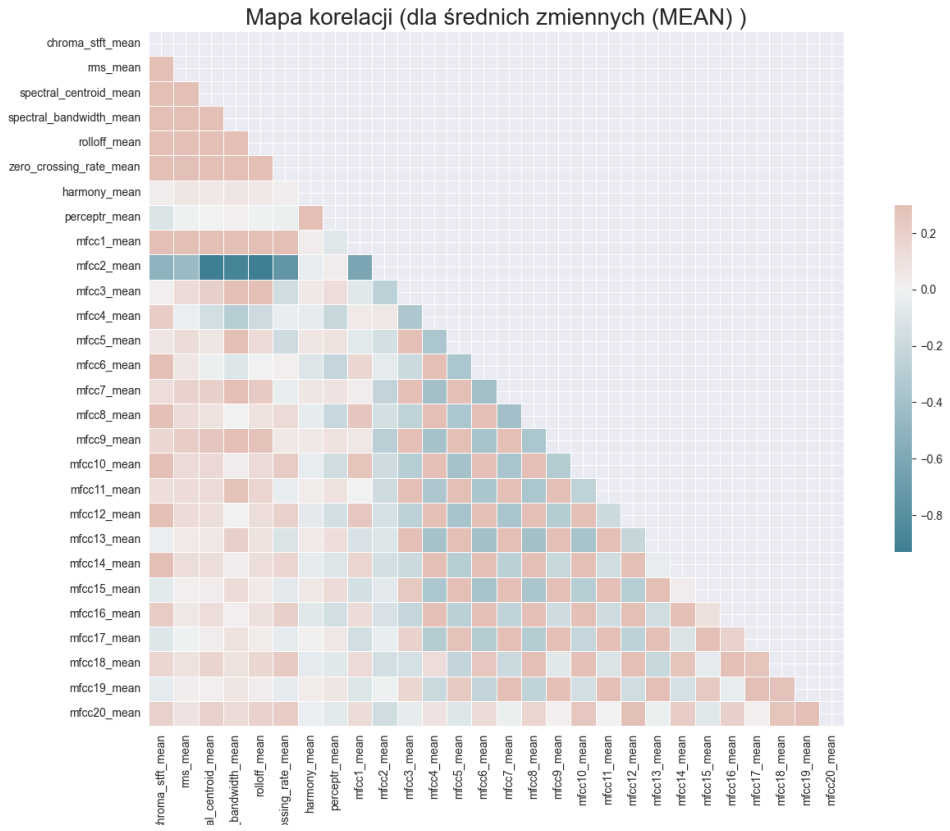
	filename	length	chroma_stft_mean	chroma_stft_var	rms_mean	rms_var	spectral_centroid_mean	spectral_centroid_var	spectral_bandwidth_mean
0	blues.00000.0.wav	66149	0.335406	0.091048	0.130405	0.003521	1773.065032	167541.630869	1972.744388
1	blues.00000.1.wav	66149	0.343065	0.086147	0.112699	0.001450	1816.693777	90525.690866	2010.051501
2	blues.00000.2.wav	66149	0.346815	0.092243	0.132003	0.004620	1788.539719	111407.437613	2084.565132
3	blues.00000.3.wav	66149	0.363639	0.086856	0.132565	0.002448	1655.289045	111952.284517	1960.039988
4	blues.00000.4.wav	66149	0.335579	0.088129	0.143289	0.001701	1630.656199	79667.267654	1948.503884

Rys 2. Przykładowe cechy 4 próbek 3 sekundowych utworów

Na rys 2. przedstawiliśmy 7 przykładowych cech dla 5 próbek utworu z gatunku muzycznego blues. Zaprezentowane cechy to cechy chromatyczne, średnia oraz wariancja mocy RMS, wariancja centroidu widmowego (spectral centroid) oraz średnia szerokość pasma widmowego (spectral bandwidth).

3. Ekstrahowanie cech

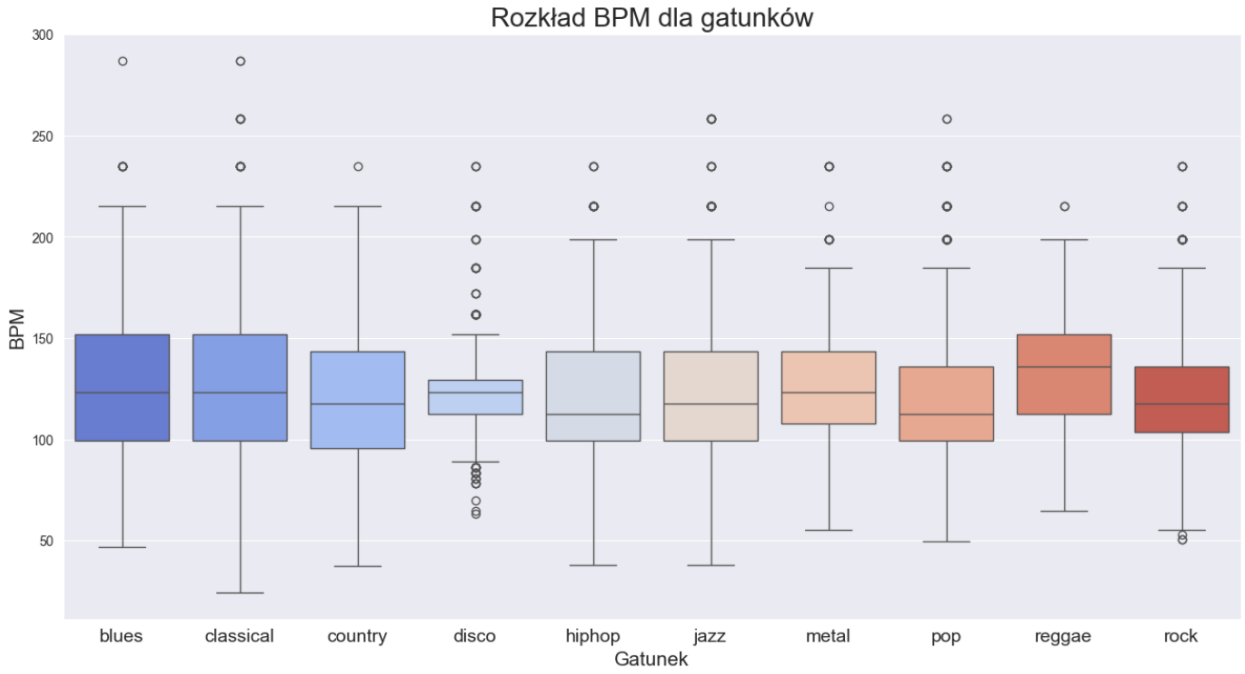
Na wejście naszej sieci neuronowej podawaliśmy wartości ekstrahowanych cech utworów z bazy danych, omówionej w poprzednim punkcie. Wygenerowaliśmy mapę korelacji dla przykładowych 56 cech średnich, aby zwizualizować korelację, czyli stopień w jakim konkretne cechy są do siebie podobne.



Rys 3. Mapa korelacji dla cech średnich

Mapa korelacji widoczna na rys 3. pokazuje zależności we wzajemnym podobieństwie między cechami. Występuje między nimi podobieństwo, ponieważ większość używanych cech to cechy parametrów Mel-cepstralnych, czyli reprezentacji krótkoterminowej mocy spektrum dźwięku, opartej na liniowej cepstralnej reprezentacji logarytmicznego spektrum mocy na nieliniowej skali częstotliwości Mel. Skala Mel jest skalą psychoakustyczną, która odzwierciedla, jak

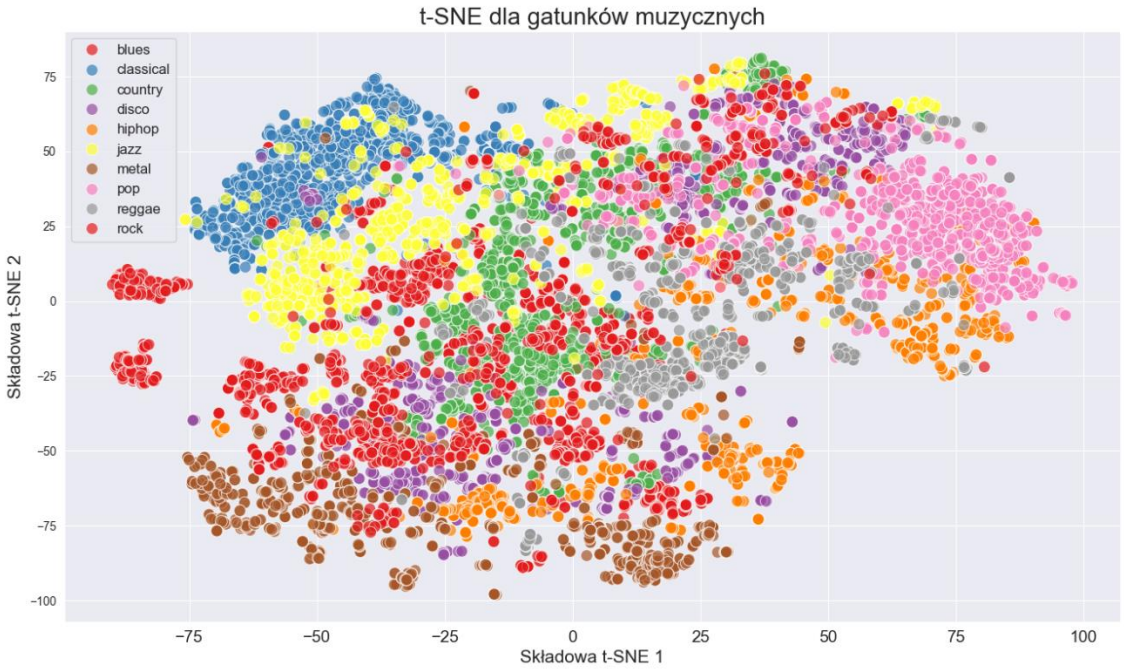
ludzkie ucho postrzega częstotliwości. Dodatkowo występuje również korelacja między wcześniej omówionymi cechami, zaprezentowanymi na rys 2. Niektóre cechy można w skuteczny sposób zwizualizować w postaci graficznej. Poniżej przedstawiamy wykres prezentujący zależność BPMów dla badanych gatunków muzycznych.



Rys 4. Rozkład BPM dla badanych gatunków muzycznych

Na rys 4. zaprezentowaliśmy rozkład BPM dla badanych gatunków muzycznych. Na podstawie rozkładu, można zauważyć, że spektrum BPM dla większości gatunków muzycznych jest bardzo podobne. Najwęższe, przez co najbardziej charakterystyczne jest dla muzyki z gatunku disco. Jest to jedna z wielu cech, która umożliwia klasyfikację gatunków muzycznych.

Aby zobrazować rozkład wszystkich cech dla gatunków muzycznych, wygenerowaliśmy wykres rozkładu t-SNE, który redukuje wymiarowość danych, obliczając podobieństwo wszystkich ekstrahowanych cech a następnie prezentuje je w postaci 2D.



Rys 5. Rozkład t-SNE dla wszystkich ekstrahowanych cech utworów z każdego gatunku muzycznego

Na rys 5. zaprezentowaliśmy rozkład t-SNE dla wszystkich ekstrahowanych cech utworów z każdego gatunku muzycznego. Dzięki temu możemy zaobserwować zależności międzygatunkowe. Na podstawie powyższego wykresu zaobserwowaliśmy, że niektóre gatunki muzyczne jak np. muzyka klasyczna albo pop, wyjątkowo odróżniają się od reszty, ale między każdym gatunkiem muzycznym zachodzi mniejsza lub większa korelacja. Dlatego ekstrahowanie tak wielu cech jest niezbędne w celu poprawnej klasyfikacji danego utworu do gatunku muzycznego.

4. Sieć neuronowa

W naszym projekcie wykorzystaliśmy sieć neuronową typu Multi-Layer Perceptron (MLP) do klasyfikacji utworów muzycznych na podstawie ich cech. Sieć MLP została wybrana ze względu na jej zdolność do przetwarzania danych o wysokiej wymiarowości, takich jak cechy muzyczne ekstrahowane z utworów. MLP jest wszechstronną architekturą sieci neuronowych, która jest dobrze dostosowana do problemów klasyfikacyjnych, w szczególności tych, które nie wymagają analizy lokalnych wzorców, jak to ma miejsce w przypadku sieci konwolucyjnych (CNN). Poniżej (Rys. 6) przedstawiono tabelę tworzenia modelu MLP:

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 512)	29,696
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131,328
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32,896
dropout_2 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 64)	8,256
dense_4 (Dense)	(None, 10)	650

Rys 6. Budowa modelu MLP

Architektura sieci

Nasza sieć MLP składa się z kilku warstw w pełni połączonych neuronów:

- **Warstwa wejściowa**: Przyjmuje wektor cech o wymiarze 60, które zostały wyekstrahowane z utworów muzycznych.
- **Warstwy ukryte**: Składają się z trzech warstw o różnej liczbie neuronów, każda z nich używa funkcji aktywacji ReLU (Rectified Linear Unit), która pomaga w efektywnym uczeniu się nieliniowych zależności.

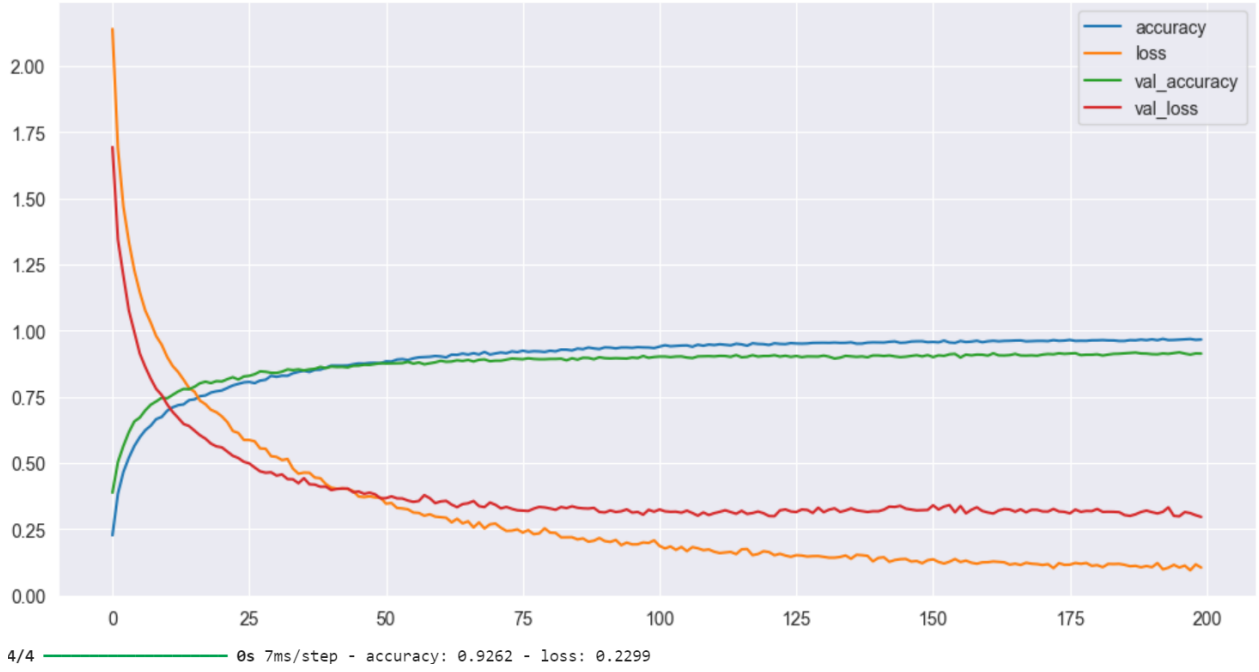
- **Warstwa wyjściowa**: Składa się z 10 neuronów, odpowiadających 10 gatunkom muzycznym, z funkcją aktywacji softmax, która przekształca wyniki w prawdopodobieństwa przynależności do poszczególnych gatunków.

Proces trenowania

Do trenowania sieci wykorzystaliśmy algorytm backpropagation z optymalizatorem Adam, który zapewnia szybkie i efektywne uczenie. Strata była mierzona za pomocą funkcji krzyżowej entropii, która jest standardowym wyborem w problemach klasyfikacyjnych.

5. Wyniki

Aby ocenić efektywność naszej sieci neuronowej, przeprowadziliśmy kilka eksperymentów z różnymi konfiguracjami hiperparametrów, takimi jak liczba neuronów w warstwach ukrytych, współczynnik uczenia się oraz wielkość batcha. Poniżej przedstawiamy przykładowe wyniki dokładności (accuracy):



Rys 7. Wykres wynikowy sieci MLP

Na powyższym wykresie widzimy, że sieć osiąga stabilną dokładność po około 100 epokach trenowania, z wartościami oscylującymi w granicach 90-93%. Wysoka wartość dokładności wskazuje, że model MLP jest w stanie prawidłowo sklasyfikować większość utworów muzycznych w zbiorze testowym, co potwierdza jego skuteczność. Niska wartość strat sugeruje, że model dobrze nauczył się rozpoznawać wzorce w danych treningowych, co przekłada się na skuteczność klasyfikacji.

Poniżej (Rys. 8) przedstawiono spis zawierający kilka kluczowych informacji dotyczących działania oraz skuteczności naszej sieci neuronowej w kontekście klasyfikacji utworów muzycznych do odpowiednich gatunków

```
Przewidywany gatunek dla indeksu 50: blues
Wartości procentowe dla wybranego przykładu:
blues: 99.99%
classical: 0.00%
country: 0.00%
disco: 0.00%
hiphop: 0.00%
jazz: 0.00%
metal: 0.01%
pop: 0.00%
reggae: 0.00%
rock: 0.00%

Przypisane gatunki dla wartości 0-9:
0: blues
1: metal
2: hiphop
3: jazz
4: rock
5: disco
6: country
7: reggae
8: classical
9: pop
```

Rys 8. Analiza skuteczności

Widzimy, że model został przetestowany na konkretnym przykładzie, w którym przewidywał gatunek muzyczny dla utworu o indeksie 50. Model przewidział, że

utwór ten należy do gatunku "blues". Z powyższych wartości wynika, że model jest niemal całkowicie pewny swojej predykcji, przypisując 99.99% prawdopodobieństwa gatunkowi "blues". Inne gatunki mają wartości bliskie zeru, co świadczy o wysokiej pewności modelu w tej konkretnej klasyfikacji. Jest to pozytywny wynik, wskazujący na zdolność modelu do zdecydowanego i trafnego rozpoznawania gatunków muzycznych.

6. Wnioski

Podsumowując, wyniki uzyskane przez naszą sieć neuronową typu MLP są bardzo obiecujące. Niska wartość strat oraz wysoka dokładność wskazują, że model skutecznie uczy się rozpoznawać wzorce w danych muzycznych i poprawnie klasyfikuje utwory do odpowiednich gatunków. Przykład predykcji dla konkretnego utworu pokazuje, że model potrafi z dużą pewnością przypisać gatunek muzyczny, co dodatkowo potwierdza jego efektywność.

Wysoka dokładność klasyfikacji i pewność predykcji sugerują, że model MLP może być z powodzeniem wykorzystany w praktycznych aplikacjach do automatycznej klasyfikacji utworów muzycznych.