# **Názov práce:** Parametrizácia zvukového signálu pomocou rôznych typov príznakov

**Zadávateľ:** Ing. Maroš Jakubec PhD.

**Popis:** Študent navrhne systém, ktorý extrahuje vybrané typy príznakov (časové a frekvenčné) z rôznych audio nahrávok a tieto príznaky využijú na jednoduchú klasifikáciu zvukov (napr. reč, hudba, šum).

**Cieľ Práce:** Cieľom tejto práce je navrhnúť a implementovať jednoduchý systém na klasifikáciu hudobných žánrov na základe extrakcie príznakov zo zvukových nahrávok. Študent sa zameria na výber a výpočet vybraných časových a frekvenčných príznakov, ktoré zachytávajú kľúčové charakteristiky zvukového signálu. Na základe týchto príznakov bude realizovaná klasifikácia do viacerých hudobných žánrov, ako sú rock, reggae, pop, metal, jazz, hip-hop, disco, country, classical a blues. Hlavným zámerom je overiť, do akej miery je možné pomocou jednoduchej sady príznakov a základných metód strojového učenia rozlíšiť jednotlivé hudobné žánre.

**Kód:** <https://github.com/ManuelMaslonka/extractionOfFeatures> , sú tam aj nahrávky, grafy.

**Teória:**

Zvukový signál je v surovej forme neštruktúrovaný. Ide o časovo závislý priebeh amplitúdy, ktorý samostatne poskytuje veľmi málo informácií použiteľných pre strojové spracovanie alebo analýzu. Parametrizácia zvukového signálu je proces, pri ktorom sa tento signál prevádza do kompaktnejšej a informatívnejšej podoby prostredníctvom extrakcie príznakov.

Príznaky predstavujú numerické charakteristiky, ktoré opisujú rôzne aspekty zvuku. V podstate ide o výber a výpočet takých parametrov, ktoré umožňujú zachytiť podstatné vlastnosti signálu a zároveň znížiť jeho zložitosť pre ďalšie spracovanie.

Tieto príznaky slúžia ako vstupy pre rôzne analytické a klasifikačné systémy. Pomocou nich je možné:

• **rozpoznať reč** (speech recognition),

• **klasifikovať typ zvuku** (napr. hudba, reč, šum),

• **extrahovať emócie zo zvuku** (v emocionálnej analýze reči),

• **detegovať udalosti v prostredí** (napr. výstrely, alarmy, kroky),

• alebo **analyzovať hudbu** (napr. na účely odporúčania skladieb).

Príznaky použité v tejto práci pokrývajú spektrum charakteristík zvukového signálu. Od jednoduchých časových a energetických parametrov až po komplexnejšie frekvenčné a tonálne popisy. Každý z týchto príznakov prispieva k lepšiemu porozumeniu štruktúry a obsahu hudby, čím vytvára vhodný základ pre následnú klasifikáciu.

Základné časové a energetické príznaky, ako napríklad RMS alebo zero-crossing rate, pomáhajú opísať dynamiku a rytmický charakter nahrávky. Frekvenčné príznaky, ako sú spectral centroid, spectral contrast či spectral bandwidth, zachytávajú rozloženie energie v spektre a umožňujú rozlíšiť žánre s rozdielnou zvukovou textúrou.

Veľmi dôležitú úlohu zohrávajú aj tonálne príznaky, ako sú chroma, tonnetz alebo MFCC, ktoré poskytujú informácie o harmónii, melodike a zvukovej farbe. Tieto parametre sú obzvlášť užitočné pri rozlišovaní hudobných žánrov s rôznou hudobnou štruktúrou – napríklad klasickej hudby, jazzu či bluesu.

Okrem toho boli použité aj rytmické príznaky, ako napríklad tempo alebo tempogram, ktoré sú kľúčové pri identifikácii žánrov s charakteristickým tempom a rytmom, ako je disco, hip-hop alebo reggae.

Výber a kombinácia týchto príznakov umožňuje vytvoriť viacrozmernú reprezentáciu každého hudobného diela. V tejto forme sa dá zvuk efektívne porovnať, analyzovať a nakoniec klasifikovať do jednotlivých žánrov pomocou algoritmov strojového učenia.

**k-NN (k-Nearest Neighbors)** je neparametrický algoritmus strojového učenia, ktorý sa používa predovšetkým na klasifikáciu, ale aj na regresiu. Pracuje na veľmi intuitívnom princípe – pri rozhodovaní o tom, do ktorej triedy patrí nový vstupný bod, sa pozrie na k najbližších susedov v trénovacích dátach (napr. podľa Euklidovskej vzdialenosti) a vyberie triedu, ktorá sa medzi nimi vyskytuje najčastejšie. Výhodou k-NN je jednoduchosť, flexibilita a schopnosť dobre pracovať s nelineárnymi dátami bez potreby trénovania modelu. Nevýhodou môže byť pomalosť pri veľkých dátových množinách a citlivosť na výber vhodného k, ako aj normalizáciu vstupných dát. V oblasti audiospracovania sa môže použiť napríklad na klasifikáciu hudobných žánrov na základe extrahovaných príznakov.

**SVM (Support Vector Machine)** je silný a presný algoritmus určený na klasifikáciu aj regresiu, ktorý hľadá optimálnu hranicu (tzv. hyperrovinu), ktorá čo najlepšie oddelí body rôznych tried. SVM sa snaží maximalizovať vzdialenosť medzi touto hranicou a najbližšími bodmi (support vectors), čo zvyšuje robustnosť a generalizáciu modelu. Pomocou tzv. kernelových funkcií dokáže SVM pracovať aj s nelineárne oddeliteľnými dátami, pričom dáta prevádza do vyšších dimenzií, kde je možné nájsť lineárnu hranicu. Vďaka tejto schopnosti je SVM obľúbený v úlohách s vysokou komplexnosťou a rozmerovosťou, ako napríklad rozpoznávanie reči, textová klasifikácia alebo detekcia zvukových udalostí.

**Postup práce:**

V úvode mojej práce som sa zameral na výber vhodného datasetu, ktorý by bol dostatočne rozsiahly, reprezentatívny a diverzifikovaný z hľadiska hudobných žánrov. Po zvážení viacerých dostupných možností som sa rozhodol pre GTZAN Genre Collection, ktorý sa dlhodobo využíva ako štandardný benchmark v oblasti hudobnej klasifikácie.

Dataset GTZAN obsahuje 1000 hudobných nahrávok, pričom každá stopa má dĺžku 30 sekúnd. Tieto nahrávky sú rozdelené rovnomerne do 10 hudobných žánrov, teda každý žáner je zastúpený 100 skladbami. Všetky súbory sú vo formáte .wav, v kvalite mono, 16-bit, so vzorkovacou frekvenciou 22 050 Hz.

Zastúpené žánre sú nasledovné:

• rock

• reggae

• pop

• metal

• jazz

• hip-hop

• disco

• country

• classical

• blues

**Schéma postupu:**A screenshot of a computer screen

AI-generated content may be incorrect.

Na účely klasifikácie hudobných žánrov bol zvukový signál najprv spracovaný tak, aby bolo možné získať z každého audio súboru číselné reprezentácie príznaky. Tento proces zahŕňal niekoľko základných krokov:

1. **Rozdelenie signálu na rámce (frames)** – Keďže hudobný signál sa v čase mení, bol rozdelený na kratšie segmenty. Dĺžka rámca 46,4 ms. Prekryv medzi rámcami 11.6 ms.
2. **Výpočet príznakov pre jednotlivé rámce** – Z každého rámca boli vypočítané príznaky
3. **Uloženie do CSV** – Výsledné príznaky boli uložené do štruktúrovaného CSV súboru, kde každý riadok reprezentuje jeden hudobný súbor a každý stĺpec predstavuje konkrétny príznak.

Pri návrhu architektúry na extrakciu príznakov som sa rozhodol využiť dizajnový vzor **Strategy Pattern**, ktorý umožňuje flexibilné a modulárne pridávanie alebo úpravu jednotlivých metód extrakcie. Každý príznak je reprezentovaný samostatnou triedou, ktorá implementuje spoločné rozhranie definované v abstraktnej triede base.py. Toto rozhranie zaručuje jednotnú štruktúru a spôsob volania všetkých extraktorov.

Všetky extraktory boli následne inicializované v jednom poli, ktoré bolo odovzdané triede AudioProcessor. Ten zabezpečuje spracovanie jednotlivých audio súborov volaním metódy process(), ktorá aplikuje všetky extraktory postupne.

Po úspešnej extrakcii príznakov z jednotlivých nahrávok bolo nevyhnutné zabezpečiť ich správnu normalizáciu, aby boli porovnateľné v rámci celého datasetu. Na tento účel som využil StandardScaler z knižnice scikit-learn, ktorý príznaky štandardizuje tak, že každá hodnota má nulový priemer a jednotkovú smerodajnú odchýlku. Bez tejto štandardizácie niektoré príznaky s veľkým rozsahom dominovali modelu a spôsobovali zlé výsledky počas trénovania.

Ďalším dôležitým krokom bolo priradenie labelu ku každému audio súboru. Keďže dátová štruktúra neobsahovala tieto informácie priamo, musel som ich extrahovať z názvov alebo adresárovej štruktúry, kde každá zložka reprezentovala jeden hudobný žáner. Tento label bol následne pripojený ku každému záznamu vo výstupe CSV.

**Hodnotenie extrahovaných príznakov pomocou modelov KNN a SVM**

V rámci extrakcie som najprv vytvoril jednu rozsiahlu množinu príznakov obsahujúcu všetky extrahované charakteristiky, ako napríklad MFCCs, spektrálne vlastnosti, tonálne, rytmické aj časové charakteristiky. Táto množina bola uložená vo forme súboru features\_s\_labelmi.csv.

Na účely testovania a porovnania som následne vytvoril niekoľko tematických podskupín príznakov, aby som zistil, ktoré typy vlastností prispievajú najviac k presnosti klasifikácie. Tieto podskupiny sú nasledovné:

|  |  |  |
| --- | --- | --- |
| **Trieda príznakov** | **Popis** | **Príklady príznakov** |
| **Časové príznaky** | Zachytávajú zmeny signálu v čase, bez transformácie do frekvenčného spektra. | rms, zero\_crossing\_rate, onset\_strength\_mean, onset\_strength\_std, harmonic\_percussive\_ratio |
| **Frekvenčné príznaky** | Popisujú rozloženie frekvencií v spektre zvuku. | spectral\_centroid, spectral\_bandwidth, spectral\_flatness, spectral\_rolloff, spectral\_contrast\_1 až spectral\_contrast\_7 |
| **Tonálne príznaky** | Súvisia s harmóniou, tonálnosťou a tóninou hudby. | chroma\_1–12, chroma\_cqt\_1–12, chroma\_cens\_1–12, tonnetz\_1–6 |
| **Mel-frekvenčné koeficienty** | Používajú sa najmä v rečových a hudobných rozpoznávacích systémoch. | mfcc\_1–13, delta\_mfcc\_1–13 |
| **Rytmické príznaky** | Charakterizujú tempo a rytmus hudby. | tempo, tempogram\_1–384, onset\_strength\_mean, onset\_strength\_std |
| **Zložené príznaky A** | Vlastný výber príznakov kombinujúci spektrálne, tonálne a MFCC charakteristiky. | rms, mfcc\_1, tonnetz\_1, spectral\_centroid, chroma\_cqt\_1 |
| **Zložené príznaky B** | Kombinácia príznakov s dôrazom na energiu, spektrálny kontrast a dynamiku zvuku. | rms, mfcc\_1, mfcc\_2, spectral\_contrast\_1, delta\_mfcc\_1, onset\_strength\_mean |

Pri metóde KNN bol použitý optimalizovaný model s metrikou Manhattan, váhovaním podľa vzdialenosti a počtom susedov k = 7. V prípade metódy SVM bol nasadený klasický lineárny model bez ďalšieho ladenia hyperparametrov.

Každý model bol trénovaný na jednotlivých príznakových sadách v štyroch variantoch:

1. Plná sada príznakov bez PCA,
2. Plná sada príznakov s PCA (so zachovaním 99 % variability),
3. Výber 50 najvýznamnejších príznakov (pomocou SelectKBest) bez a s PCA,
4. Výber všetkých príznakov (v rámci konkrétnej sady) bez a s PCA.

Výsledky ukázali, že SVM vo väčšine prípadov prekonáva KNN z hľadiska presnosti, najmä pri plných dátach bez PCA alebo pri výbere všetkých príznakov. Naopak, KNN model niekedy profitoval z aplikácie PCA, čo možno pripísať redukcii dimenzionality a zníženiu rizika pretrénovania.

Z tabuliek vyplýva, že:

* Najvyššiu presnosť dosiahlo SVM pri sade features\_s\_labelmi.csv s výberom 50 príznakov bez PCA – až 72,50 %, zatiaľ čo KNN v tejto konfigurácii dosiahol 62,00 %.
* V prípade PCA pri KNN sa presnosť zvýšila až na 65,00 %, čo naznačuje, že KNN vie efektívne ťažiť z redukcie dimenzií, obzvlášť pri väčšom počte príznakov.
* Spektrálne príznaky sa ukázali ako relatívne stabilné pre oba modely – dosahovali nadpriemerné hodnoty (napr. SVM so spektrálnymi príznakmi a PCA: 59,5 %).
* Rytmické a časové príznaky boli najslabšie – oba modely s nimi dosahovali nízku presnosť (napr. SVM s rytmickými: len 41 % – 35,5 %).
* Kombinované sady ako kombinacia\_b alebo kombination\_a dosahovali mierne lepšie výsledky ako izolované MFCC alebo tonal, no stále výrazne zaostávali za plnými spektrálnymi alebo rozšírenými sadami.

SVM sa javil ako robustnejší model pri práci s plnými príznakovými sadami. KNN mohol byť zlepšený aplikáciou PCA, najmä pri veľkom počte vstupných príznakov. PCA malo pri niektorých príznakových sadách pozitívny vplyv, najmä pri features\_s\_labelmi.csv a rytmických príznakoch. Najefektívnejšie príznaky boli tie, ktoré kombinovali viacero vlastností (napr. features\_s\_labelmi.csv a spectral). Rytmické a časové znaky boli zrejme slabo diskriminačné pre účely klasifikácie v oboch modeloch.

KNN výsledky

A graph of different colored bars

AI-generated content may be incorrect.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Feature Set** | **Accuracy (k=50, no PCA)** | **Accuracy (k=50, PCA)** | **Accuracy (all, no PCA)** | **Accuracy (all, PCA)** | **Original Features** | **Selected Features (k=50)** | **Selected Features (all)** | **PCA Features (k=50)** | **PCA Features (all)** |
| features\_s\_labelmi.csv | 0.6200 | 0.6500 | 0.4200 | 0.4250 | 469 | 50 | 469 | 37 | 178 |
| spectral | 0.5450 | 0.5550 | 0.5450 | 0.5550 | 12 | 12 | 12 | 9 | 9 |
| kombinacia\_b | 0.5200 | 0.5300 | 0.5200 | 0.5300 | 6 | 6 | 6 | 6 | 6 |
| time | 0.4800 | 0.4800 | 0.4800 | 0.4800 | 5 | 5 | 5 | 5 | 5 |
| tonal | 0.4450 | 0.4550 | 0.4450 | 0.4550 | 43 | 43 | 43 | 25 | 25 |
| MFCC | 0.4650 | 0.4500 | 0.4650 | 0.4500 | 27 | 27 | 27 | 25 | 25 |
| kombinazione\_a | 0.3700 | 0.4050 | 0.3700 | 0.4050 | 5 | 5 | 5 | 5 | 5 |
| Rhythmic | 0.3500 | 0.3600 | 0.3150 | 0.3300 | 386 | 50 | 386 | 39 | 132 |

SVM výsledky

A graph of different colored bars

AI-generated content may be incorrect.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Feature Set** | **Accuracy (k=50, no PCA)** | **Accuracy (k=50, PCA)** | **Accuracy (all, no PCA)** | **Accuracy (all, PCA)** | **Original Features** | **Selected Features (k=50)** | **Selected Features (all)** | **PCA Features (k=50)** | **PCA Features (all)** |
| features\_s\_labelmi.csv | 0.7250 | 0.6950 | 0.5900 | 0.5800 | 469 | 50 | 469 | 37 | 178 |
| spectral | 0.5750 | 0.5950 | 0.5750 | 0.5950 | 12 | 12 | 12 | 9 | 9 |
| MFCC | 0.5350 | 0.5250 | 0.5350 | 0.5250 | 27 | 27 | 27 | 25 | 25 |
| kombinacia\_b | 0.5100 | 0.5100 | 0.5100 | 0.5100 | 6 | 6 | 6 | 6 | 6 |
| time | 0.5050 | 0.5050 | 0.5050 | 0.5050 | 5 | 5 | 5 | 5 | 5 |
| tonal | 0.4900 | 0.5000 | 0.4900 | 0.5000 | 43 | 43 | 43 | 25 | 25 |
| kombinazione\_a | 0.4600 | 0.4600 | 0.4600 | 0.4600 | 5 | 5 | 5 | 5 | 5 |
| Rhythmic | 0.4100 | 0.4100 | 0.3550 | 0.3400 | 386 | 50 | 386 | 39 | 132 |