Music genres classification

Petr Lorenc

Oleksandra Liutova

# Popis projektu

Cílem projektu je vytvoření aplikace schopné klasifikovat vstupní nahrávku do jednoho z předem nadefinovaných žánrů. Definice žánrů probíhá tak, že je zvoleno několik klasických skladeb daného žánru, čímž se vytvoří referenční množina skladeb. Při dotazu jsou identifikovány nejpodobnější databázové skladby a na jejich základě je dotaz zařazen do daného žánru.

# Způsob řešení

Audio signál je popsatelný různými deskriptory. Známá sada deskriptorů, popisující audio z různých pohledů, jsou např. MPEG7 deskriptory. V současné době se ale začíná používat MFCC (Mel-frequency cepstral coeffients), který zvuk lépe popisuje z hlediska lidského sluchu. Jeho vznik zahrnuje Fourierovu transformaci, škálování pomocí Melovy stupnice a také diskrétní kosinovu transformaci. Způsob tvorby tohoto deskriptoru je mimo rozsah této práce.

Práce také zahrnuje grafický výstup. Jako výstup se proto používá webová stránka, kde je přehledně zobrazen výsledek našeho dotazu.

# Implementace

Celý projekt je implementován v jazyce Python.

Knihovny

* Librosa
* Scikit-learn
* Numpy
* Flask

K extrakci features z hudební skladby je používano knihovny librosa, která nám umožňuje extrahovat MFFC. Z trénovačích dat (1000 30-sekundových skladeb z 10 žánrů) vznikne matice 20x1293. Tuto množinu se snažíme redukovat na co nejmenší reprezentativní prvek. V literatuře1 můžeme najít třeba způsob který zprůměruje složky jednotlivých 1293-složkových vektorů. Tak nám vznikne prostor o 20 dimenzích, který už se hodí pro další zpracování, třeba za pomoci kNN metody.

V experimentální fázi se budeme zabývat i jinými příznaky, které nabízí knihovna librosa. Dále budeme zkoušet různé metriky pro měření vzdálenosti mezi skladbami.

Výstup bude zobrazen pomocí knihovny Flask, kde budou implementovány 2 stránky. Na první bude možnost volit skladbu a na druhé se zobrazí výsledky v podobě grafu (uvidíte v příkladu výstupu)

# Příklad výstupu

# Experimentální sekce

Čisté MFCC (zprůměrované po 1 sec – 600 složkový vektor)

Experimentálně bylo zjištěno, že 1 vteřina záznamu vygeneruje matici o velikosti 20x44. Toho bylo využito pro zprůměrování po sekundách a poté zprůměrování přes všech 20 řádků matice. Bohužel bylo dosaženo stejných výsledků jako když se jednotlivé řádky v matici (tj. 1293 členů) zprůměruje. Bylo proto využíváno tohoto druhého přístupu který je přehlednější a rychlejší. Pro výběr klasifikátoru a vhodnému K bylo využito Eukleidovy metriky.

Pro účely testování bylo využito techniky cross-validace. Celková množina 1000 skladeb byla rozdělena na trénovací část a testovacích. Po natrénování na trénovačích datech bylo zkoušena predikce na testovacích.

* Random Forest
  + Úspěšnost 44.5 procenta
  + Dobré určení Reggae, Hip Hop, Metal a Classical (přes 80%)
* KNN – K=5
  + Úspěšnost 42 procent
  + Dobré určení Blues, Jazz a Classical (přes 80%)
* KNN – K=7
  + Úspěšnost 47 procent
  + Dobré určení Blues, Classical (přes 80%)
* KNN – K=13
  + Úspěšnost 46 procent
  + Dobré určení Blues, Classical, Metal (přes 80%)

Pro další experimenty tedy byl vybrán 7NN klasifikátor. U něho bylo vyzkoušeno několik metrik:

* Kullback-Leibner distance
  + np.sum(np.where(p != 0, p \* np.log(p / q), 0))
  + Úspěšnost 15 %
* Minkowski 3
  + Úspěšnost 43%
* Minkowski 4
  + Úspěšnost 43,5%
* Chebyshev
  + Úspěšnost 43%

# Diskuze

# Závěr

# Literatura

1. http://www.cp.jku.at/research/papers/A%20fast%20audio%20similarity%20retrieval%20method.pdf