FAKULTA APLIKOVANÝCH VĚD ZÁPADOČESKÉ UNIVERZITY V PLZNI

Rozpoznávání hudebních žánrů

Kateřina Štollová
Katedra informatiky a výpočetní techniky

Vedoucí práce:
Ing. Kamil Ekštein, Ph.D.

Motivace

Se zvyšujícím se počtem digitálně uchovávaných zvukových nahrávek roste potřeba automaticky rozpoznávat jejich obsah. Cílem této práce bylo prozkoumat možnosti automatického určování hudebních žánrů a vytvořit klasifikátor, který dané nahrávky bude zařazovat do skupin dle hudebního žánru. Takto získaná informace může být použita pro automatické nastavení ekvalizéru a zlepšení posluchačského zážitku.
Určování hudebních žánrů je složitá úloha, zejména z toho důvodu, že hudebních žánrů existuje mnoho a rozřazování skladeb je silně subjektivní. Pro potřeby práce bude proto definováno menší množství hudebních žánrů a výsledek klasifikátoru bude porovnán s dostatečně velkou referenční skupinou dobrovolníků.

Myšlenka

Řešení úlohy je založeno na myšlence, že jednotlivé hudební žánry se liší jak rytmem, tak použitými nástroji, které se projevují v různých frekvenčních pásmech.
Pro klasifikaci jsou jako příznaky využívané zejména spektrální charakteristiky signálu zkoumající použité nástroje. Rytmické příznaky jsou použity doplňkově.

Rytmické charakteristiky

Pro vyjádření rytmické charakteristiky je využíván **spektrální tok**, který zjišťuje změny v jednotlivých frekvencí mezi dvěma časovými okny. Protože se rytmus-určující nástroje (bicí, baskytara, apod.) projevují v nižších frekvencích, je spektrální tok počítán pouze pro frekvence 0 – 1 kHz. Výpočet spektrálního toku zobrazuje rovnice

$$F_t = \sum_{n=1}^{N} (N_t[n] - N_{t-1}[n])^2, \qquad (1)$$

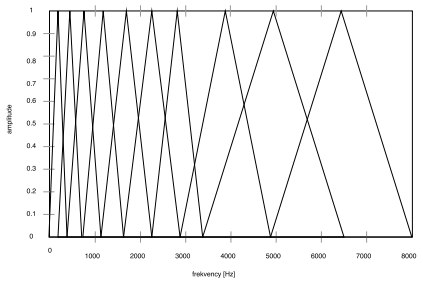
kde $N_t[n]$ je normalizovaná frekvence n-tého koeficientu FFT v čase t.

Spektrální charakteristiky

Spektrální charakteristiky popisují tvar obálky výkonnové spektrální hustoty signálu získané Fourierovou transformací (FFT).

Mel-frekvenční kepstrální koeficienty

Mel-frekvenční kepstrální koeficienty (MFCC) reprezentují frekvenční spektrum signálu v logaritmické stupnici, která klade důraz na frekvence tak, jak jsou vnímané lidským uchem. Pro převod do logaritmické stupnice se na signál aplikují trojúhelníkové filtry (viz obrázek 1), pod kterými se signál integruje. Pro každý filtr je získán jeden koeficient.



Obrázek 1: Příklad trojúhelníkových filtrů.

Spektrální centroid

Spektrální centroid je definován jako střední hodnota frekvence signálu vážená amplitudou spektra. Vypočte se dle rovnice

$$C_t = \frac{\sum_{n=1}^{N} M_t[n] \cdot f_n}{\sum_{n=1}^{N} M_t[n]}, \qquad (2)$$

kde $M_t[n]$ je amplituda n-tého koeficientu FFT v čase t a f_n je jeho frekvence.

Spektrální rolloff

Spektrální rolloff je definován jako hodnota R_t, pro kterou platí rovnice

$$\sum_{n=1}^{R_t} M_t[n] = T_h \cdot \sum_{n=1}^{N} M_t[n], \qquad (3)$$

kde hodnota T_h typicky odpovídá 0.8 – 0.95.

Parametry realizace

Zpracování nahrávek

1. Rozdělení nahrávky na 2 s vzorky.
2. Zpracování vzorku po 250 ms oknech s 50 % překryvem.
3. Spočtení spektrálních (rytmických) charakteristik pro každé okno.
4. Vytvoření příznakového vektoru pro vzorek (střední hodnoty, rozptyly).
5. Určení žánru pro vzorek.
6. Zvolení žánru pro nahrávku – na základě zařazení většiny vzorků.

Klasifikátor

- 6 hudebních žánrů,
  - klasická hudba
  - folk,
  - rap,
  - jazz,
  - rock,
  - metal,
- 4 příznakové vektory,
  - pouze MFCC,
  - spektrální,
  - rytmický,
  - kombinovaný,
- 9 hod trénovacích dat,
  - 1 hod 30 min/žánr,
- 2 klasifikační algoritmy,
  - neuronová síť (NN),
  - k-nejbližších sousedů (K-NN).

Výsledky

Výsledky klasifikátoru byly porovnány s výsledky testu referenční skupiny. Dobrovolníci měli za úkol zařadit každou ze 48 nahrávek právě do jednoho ze 6-ti žánrů. Jako referenční pro hodnocení klasifikátoru byl vybrán žánr, který získal většinu hlasů.
Obecně lepších výsledků dosáhl klasifikátor s použitím algoritmu NN, na druhou strau K-NN byl rychlejší při učení.

žánr \ klasifikace	klasika	folk	rap	jazz	rock	metal	%
klasika	**8**	0	0	1	0	0	$88,\overline{8}$
folk	2	**4**	0	0	0	0	$66,\overline{6}$
rap	0	1	**7**	0	0	0	87,5
jazz	1	1	0	**5**	0	0	71,43
rock	3	1	0	0	**6**	0	60
metal	0	0	0	0	0	**8**	100

Tabulka 1: Výsledky klasifikátoru pro spektrální příznaky

Tabulka 1 zobrazuje výsledky nejlepšího příznakového vektoru (spektrální). Každý řádek představuje nahrávky daného žánru, sloupce pak žánr, do kterého byly zařazeny klasifikátorem. Celková úspěšnost pro algoritmus neuronové sítě je 79 %. Algoritmus k-nejbližších sousedů dosáhl o něco nižší úspěšnosti 75 %.