MUSIC GENRE RECOGNITION

Jiří Vaněk
Master Degree Programme (2), FIT BUT
E-mail: xvanek26@stud.fit.vutbr.cz

Supervised by: Michal Hradiš
E-mail: ihradis@fit.vutbr.cz

ABSTRACT

This paper describes software to determine music genre of given song. The system does not use any textual information or sheet of music, the recognition is based on the analysis of the audio signal. For this purpose the statistical methods of classification are used. First step in the classification is feature extraction, the choosing of the right features is essential. The characteristics of the human hearing must be taken into account, a simple approximation is proposed. Second step is the choice of the classifier, which is based on the data distribution in the required classes. The experimental results of the proposed software are described in the end of the paper.

1. ÚVOD

Hudební žánry reprezentují kategorie, do kterých zařazujeme hudbu podle jejích obsahových charakteristik. Tato kategorizace je významná například v obchodech s hudbou. S rostoucím množstvím multimediálních dat v počítačích roste potřeba třídit tato data podle skutečného obsahu. Určení jakékoli obsahové charakteristiky je snadné pro člověka, zatímco pro počítač se jedná o složitý a prozatím velmi otevřený problém. Proto se dosud pro určení hudebního žánru dané písně používalo manuální označování člověkem. Pro větší množství dat by však bylo užitečné, kdyby toto označení bylo provedeno automaticky počítačem.

2. PRINCIP ŘEŠENÍ

Pro zjištění vysokoúrovňové informace, jako je přiřazení písně k určitému hudebnímu stylu, je nemožné využít čistě analytické metody. Můžeme však využít statistické metody klasifikace, které na základě dostupných trénovacích dat vytvoří modely tříd, do kterých chceme data zařazovat. Podle těchto modelů je pak možné určit pravděpodobnou příslušnost nových dat ke třídám. U statistické klasifikace je potřeba vyřešit dva oddělené problémy: Extrakce příznaků a klasifikace. Při extrakci příznaků z hudby je nutné vzít v úvahu, jakým způsobem vnímá hudbu člověk.

3. LIDSKÉ VNÍMÁNÍ HUDBY

Můj systém, stejně jako jiné systémy které zpracovávají signály, pracuje velmi často s frekvenčním spektrem signálu. Frekvenční analýza se používá proto, že odpovídá způsobu, jakým pracuje lidský sluchový systém. Ucho provádí také frekvenční analýzu. Mozek dostává informaci o tom, jak jsou ve slyšeném zvuku jednotlivé frekvence zastoupeny [1]. Matematicky provedená frekvenční analýza má však na výstupu spektrum s lineární frekvenční osou. To neodpovídá lidskému slyšení, které je logaritmické: v nižších frekvencích má lepší rozlišení a směrem k vyšším frekvencím rozlišení klesá.

Jednoduchou aproximaci tohoto jevu navrhuje systém pro detekci rytmu [2]. Metoda popsaná v uvedeném dokumentu spočívá v použití frekvenčních pásem s proměnnou šířkou. Takové spektrum je možné získat výpočtem frekvenčního spektra s vyšším počtem pásem, která následně budeme slučovat. Začneme od nejnižších frekvencí, kde sloučíme malý počet pásem, pokračujeme k vyšším frekvencím a slučujeme stále větší počet pásem. Tak dostaneme nová pásma s rostoucí šířkou, což odpovídá klesajícímu rozlišení lidského ucha. Metoda určení

šířek nových pásem je popsána v [2]. Pro veškerou práci s frekvenčním spektrem používám v mém systému toto upravené spektrum.

4. EXTRAKCE PŘÍZNAKŮ

Pro účely statistické klasifikace je důležitá volba vhodných příznaků. Můžeme použít fyzikální vlastnosti, jako je energie, průchody nulou atd. Stejně tak je ale vhodné využít i vlastnosti perceptuální, tj. ty, které jsou odvozeny z lidského vnímání, například rytmus nebo barva zvuku. Oddělit fyzikální vlastnosti od perceptuálních by bylo velmi obtížné, protože perceptuální vlastnosti jsou založeny na fyzikálních. Najít mezi nimi jasnou hranici není možné, a vzhledem k jejich úzké souvislosti by to nebylo ani užitečné.
Příznakový vektor pro náš systém tedy složíme z vybraných vlastností obou těchto druhů. Nabízí se přitom jiná možnost dělení příznaků do dvou skupin: První skupinu tvoří příznaky získané ze spektra a ze samotného signálu, v zahraniční literatuře někdy označované souhrnným termínem „musical surface features“ [3]. Druhou skupinu tvoří vlastnosti rytmické.
Pro určení hudebního žánru není nutné slyšet celou píseň, stačí krátký úsek. Jeho délka by měla být minimálně 3 vteřiny – experimenty ukázaly, že člověk je schopen na této délce hudební žánr rozpoznat [3]. Příznaky se tedy počítají pro tyto krátké úseky písní.

4.1. MUSICAL SURFACE FEATURES

V mém systému používám čtyřvteřinové úseky písní, na nichž počítám níže uvedené vlastnosti. Na celém úseku pak spočítám pro každou vlastnost střední hodnotu a standardní odchylku. Ty se pak použijí do příznakového vektoru, reprezentujícího daný úsek písně. Výjimku tvoří Low Energy - počítá se jednou pro celý úsek (pro jedno okno nelze určit).

Spectral Centroid - těžiště spektrálního tvaru.
Spectral Rollof - frekvence, pro kterou platí, že 85% energie ve spektru je pod touto frekvencí
Spectral Flux - míra změny spektrálního tvaru mezi sousedními okny.
Zero Crossing Rate - počet průchodů signálu nulou.
Low Energy - procentuální část oken, jejichž energie je pod průměrnou hodnotou.

Způsob výpočtu těchto příznaků je blíže popsán například v [3].

4.2. RYTMICKÉ VLASTNOSTI

Pro klasifikaci potřebujeme najít celkovou rytmickou charakteristiku daného úseku písně. K jejímu zjištění se hodí autokorelační funkce. Ta zjišťuje, jak je signál při posunutí podobný sám sobě. Hledá tedy v signálu určitou periodičnost – tu hledáme i v rytmu hudby.
Jako vstupní signál autokorelační funkce použijeme časový průběh energie jednoho frekvenčního pásma, vyhlazený klouzavým průměrem. Indexy největších koeficientů na výstupu autokorelační funkce udávají *lagy* odpovídající periodám mezi nejvýraznějšími beaty (s výjimkou indexu 0, zde je koeficient vždy největší). My vezmeme 3 největší koeficienty, jejich indexy uložíme do histogramu. Opakujeme celý výpočet pro všechna frekvenční pásma. Ve výsledném histogramu nalezneme dvě největší špičky na nenulové pozici. Ty odpovídají periodám dvou nejvýraznějších opakujících se beatů v daném úseku písně. Můžeme z nich určit průměrné tempo těchto beatů a další vlastnosti, které vyjadřují rytmickou charakteristiku úseku písně. Odtud tedy počítáme rytmické vlastnosti, které použijeme v našem příznakovém vektoru:

Relativní síla nejvýraznějšího beatu. Spočteme jako podíl hodnoty největší špičky v histogramu se sumou všech hodnot. Výsledek udává, jak výrazný je rytmus.
Relativní síla druhého nejvýraznějšího beatu – analogie předchozí vlastnosti.

Podíl indexů druhé a první největší špičky histogramu. Vyjadřuje poměr hlavního beatu k druhému nejvýraznějšímu. U písně se silným, pravidelným rytmem se bude hodnota blížit ke 2 nebo k 0,5.
Perioda hlavního beatu v BPM – spočteme z indexu nejvyšší špičky histogramu.
Perioda druhého nejvýraznějšího beatu v BPM – analogie předchozí vlastnosti.
Suma celého histogramu. Udává celkovou sílu rytmu.

5. KLASIFIKÁTOR

Pro volbu vhodného klasifikátoru je dobré znát přibližné rozložení dat, která chceme klasifikovat. Podle rozložení dat v jedné třídě můžeme vybrat klasifikátor, který využívá modely odpovídající tomuto rozložení.
Pro hudbu z hlediska žánrů je však těžké vyvodit jakékoli závěry o rozložení dat uvnitř tříd. Hudební žánry nemají pevně stanovené hranice, tvorba jednoho umělce často zasahuje do více různých žánrů a u některých písní se i lidé těžko shodnou na zařazení. Hudební žánry jdou navíc jen velmi těžko popsat, případný popis je vždy nepřesný. Lidé je tedy často definují pomocí typických představitelů. Tomuto způsobu definování tříd pomocí příkladů odpovídá funkčnost klasifikátoru KNN – k Nearest Neighbours. Jeho funkčnost bývá někdy označována jako „learning by example“, tedy učení podle příkladů. Tato funkčnost přesně odpovídá potřebám tohoto projektu. Klasifikátor KNN nestaví na žádných předpokladech o rozložení dat. Pouze uchovává polohu všech trénovacích dat ve vektorovém prostoru, a nová data klasifikuje na základě analýzy k nejbližších sousedů. Analýzou může být například vážený průměr, kde váha je vzdálenost od klasifikovaného objektu. Hodnota k – tj. kolik nejbližších sousedů bereme v úvahu, se dá určit experimentálně. Nejlepší výsledky tento systém podává s hodnotou k=3.

6. VÝSLEDKY

Systém byl natrénován pro rozpoznání tří žánrů: klasika, rock a taneční hudba.
Použitá data:
- Rock: 62 písní, 3,7 hodin. Výrazně zastoupeni: AC/DC, Led Zeppelin, Rolling Stones…
- Classic: 28 písní, 1,6 hodin. Výrazně zastoupeni: Mozart, Tschaikovsky, Vivaldi…
- Dance: 63 písní, 3,9 hodin. Výrazně zastoupeni: Shaun Baker, Mr. Vee…

Systém byl testován cross-validací na všech datech (5 dělení v poměru 1:4), průměrná úspěšnost byla 76,36%. Drobné zlepšení přinesla normalizace hodnot příznakového vektoru. Ta spočívala v dělení hodnoty každého příznaku sumou všech hodnot stejného příznaku na všech dostupných datech. Výsledná úspěšnost systému je nyní 77,7%.

7. ZÁVĚR

Tato práce navrhuje systém pro rozpoznání hudebních stylů. Navržený systém byl implementován a otestován pro rozpoznání tří žánrů, dosažená úspěšnost je 77,7%. Vzhledem k nemožnosti přesně definovat hranice jednotlivých žánrů je tento výsledek uspokojivý, systém je použitelný pro automatické třídění hudby. Další možný vývoj tohoto systému vidím v experimentování s extrakcí rytmických příznaků. Existuje totiž mnoho různých přístupů k detekci rytmu, poskytující různé typy výsledků. Jinou možností by bylo přidat další příznaky, například MFCC.

LITERATURA

[1] Harris, T.: How Hearing Works, 30. března 2001, HowStuffWorks.com.
Dokument dostupný na URL: http://health.howstuffworks.com/hearing.htm
[2] Patin, F.: Beat Detection Algorithms, 2003
Dokument dostupný na URL: http://www.gamedev.net/reference/articles/article1952.asp
[3] Kosina, K.: Music Genre Recognition [diplomová práce], Medientechnik und Design in Hagenberg, červen 2002