

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Jiří Balhar

Název práce Extrakce melodie pomocí hlubokého učení

Rok odevzdání 2019

Studijní program Informatika **Studijní obor** Obecná informatika

Autor posudku Jan Hajič jr. **Role** vedoucí

Pracoviště Ústav formální a aplikované lingvistiky

Text posudku:

Cílem předložené práce je vylepšit metody automatické extrakce melodie z nahrávek hudby. Melodie je jedna z nejdůležitějších charakteristik hudební skladby, možná právě tím hlavním rysem, který si z poslechu odnáší člověk, a tím pádem je zároveň dominantním deskriptorem užitečným prakticky pro všechny další úlohy zpracovávající hudební signál: od ztotožňování cover verzí po vyhledávání pomocí broukání (query-by-humming). Extrakce melodie sestává z dvou kroků: výpočet salienční funkce, která v každém časovém okamžiku udává pravděpodobnosti (nebo obecná skóre) pro každé frekvenční pásmo, zda se melodie nachází právě na dotyčné frekvenci, a v druhém kroku výběr melodie včetně označování úseků bez melodie z výsledků salienční funkce. Práce se zaměřuje na hlavní část úlohy extrakce melodie: výpočet salienční funkce. Autor také analyzuje obtíže s úlohou spojené: jak nejednoznačnost samotného pojmu melodie (který tón je melodie, když hrají dva nástroje melodii o oktávu od sebe?), tak obtíže spojené se sčítáním vyšších harmonických frekvencí jednotlivých znějících nástrojů ve vícehlasých nahrávkách.

Je třeba ocenit, že práce přináší kompletní přehled souvisejících metod (kap. 2) a datasetů (kap. 3), které jsou pro úlohu k dispozici, a podrobně popisuje a diskutuje vhodnost jednotlivých evaluačních metrik (kap. 4). Obzvlášť přehled (Tab. 3.1) a kritika datasetů je cenným přínosem.

V kapitole 5 leží těžiště celé práce: experimenty s výpočtem salienční funkce pro extrakci melodie. Student pracuje se třemi různými architekturami neuronových sítí. Vychází ze state-of-the-art metody CREPE a z architektury WaveNet a obě zlepšuje pečlivým výběrem hyperparametrů (diskretizace výšek not, „změkčování“ cílového vektoru, rozlišení první konvoluční vrstvy, atd), které příslušně diskutuje (viz např. obr. 5.5) a kromě cílení na co nejlepší výsledek se také snaží (úspěšně) najít co nejmenší a tím pádem nejrychlejší modely se srovnatelnými výsledky (5.1.2, 5.2.1, atd.). Pečlivá práce s hyperparametry například zlepšila oproti baseline výsledky architektury WaveNet z Raw Pitch Accuracy (RPA) 0.583 na 0.673.

Největší přínos práce je v sekci 5.3: autor na základě existující práci se vstupní reprezentací

spektrogramů vymyslel způsob, jak využít strukturu vyšších harmonických tónů v logaritmickeém spektrogramu skrz celou konvoluční architekturu. Výsledná architektura HCNN zlepšuje výsledky a navíc potřebuje pouze cca desetinu parametrů oproti modelu, ze kterého čerpá. Autor zkoumá optimální nastavení svého modelu HCNN a dosáhne dalších zlepšení.

V kapitole 6 jsou prezentovány výsledky. Student si dal záležet na tom, aby byly přímo porovnatelné s aktuálním stavem poznání. **Ze 6 měřených datasetů dosahuje model HCNN přesvědčivě nejlepších výsledků na čtyřech z nich.** Kvalitativní výsledky jsou podrobně diskutovány. Porovnávané metody student zpřístupňuje jako online demo.

Nejslabší stránka práce je místy nejasný popis toho, jaké nastavení bylo použito pro které pokusy. Vzhledem k množství pokusů s každou architekturou bych býval ocenil stručnou rekapitulaci toho, jak která sada pokusů s hyperparametry ovlivnila výsledek, na konci každé ze sekcí 5.1, 5.2 a 5.3, a úplný přehled použitých hyperparametrů v tabulkách v sekci 6.1.

Nejsilnější stránkou bakalářské práce jsou původní výsledky přesvědčivě zlepšující stav poznání ve výpočtu salienční funkce pro extrakci melodie, a následně v extrakci melodie jako takové. Práci považuji za srovnatelnou s nejlepšími diplomovými pracemi.

Po formální stránce je práce v pořádku. Je přehledně strukturovaná, psaná adekvátní češtinou, pouze občas by bylo vhodné zvolit jiné prostředky dělení vět (srov. závěr, odst. 2, věta 3: „...výstup této funkce však nezpracováváme, obvyklým krokem...” – zde je lepší nahradit čárku středníkem či rozdělit věty tečkou), citace a bibliografické údaje jsou použity v souladu s oborovými zvyklostmi; práce cituje více než 60 zdrojů, včetně prakticky všech prací o extrakci melodie.

Práci doporučuji k obhajobě.

Práci navrhuji na zvláštní ocenění.

Práce (bakalářská!) přináší výsledky významně porážející state-of-the-art v oboru. Student na klíčové inovace zlepšující výsledky přišel samostatně na základě kritického čtení aktuálních publikací, a inovací přinesl systematickou experimentální práci několik. Práce je jak výsledky, tak metodologií a nápaditostí na úrovni oceňovaných publikací v oborech music information retrieval a strojového učení (a k publikování směřuje, skrze nadcházející účast na soutěži v extrakci melodie v rámci zavedené série hudebně-informatických soutěží MIREX).

V Praze dne 18. 6. 2019

Podpis: