Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy v Praze

Autor práce	Petr Vincena
Název práce	Trojrozměrná vizualizace hudby
Rok odevzdání	2020
Studijní program	Informatika
Studijní obor	IOI

Autor posudku	Jan Hajič jr.	Oponent
Pracoviště	Ústav formální a aplikované lingvistiky	

K celé práci	lepší	OK	horší	nevyhovuje
Obtížnost zadání	X			
Splnění zadání	X			
Rozsah práce *... textová i implementační část, zohlednění náročnosti*	X			

Řešitel vytvořil systém pro vizualizaci hudebního signálu. Vstupem pro vizualizaci jsou extrahované rysy vstupního signálu, od poměrně jednoduchých (energie) přes složitější (tonalita, rytmus) po vysoce netriviální (emoce). „Odbornou" úlohou a hlavní ambicí práce je extrakce především těchto netriviálních rysů z hudební nahrávky; software navíc výsledek vizualizuje ve 3D pomocí platformy Unity. Práci považuji po obou těchto stránkách za kvalitní. Vizualizace funguje a postup extrakce hudebních rysů z nahrávek je logický, dobře zdůvodněný a dostatečně analyzovaný a vyhodnocovaný; z práce je zřejmé, že řešitel provedl více než dostatečné množství pokusů s extrakcí emocí. Za moudré rozhodnutí považuji také ve vizualizaci nakonec rozpoznané emoční kategorie nepoužít – na základě pečlivého vyhodnocení příslušných pokusů, které prokazuje, že řešitel zvládl základy vědecké práce v oblasti strojového učení (především práci s daty) a které považuji za nejhodnotnější část práce.

Práci doporučuji k obhajobě a navrhuji hodnocení stupněm výborně. K obhajobě mám následující dotazy:

1. Daly by se definovat hranice emocí jinak, aby byl klasifikační dataset lépe vyvážený?
2. Jaké alternativní způsoby trénování a vyhodnocení nad takto neurčitým a zároveň nevyváženým datasetem by se daly použít?
3. Byla použita regularizace vah modelu? Pokud je oblast pozitivní valence reprezentována v polárních souřadnicích úhly s malou absolutní hodnotou, může být drift směrem k pozitivní valenci při použití polárních souřadnic způsoben případně touto regularizací?

Textová část práce	lepší	OK	horší	nevyhovuje
Formální úprava *... jazyková úroveň, typografická úroveň, citace*	X			
Struktura textu *... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu*	X			
Analýza	X			
Vývojová dokumentace		X		
Uživatelská dokumentace			X	

Práce je psaná jasně a přehledně, kvalitní angličtinou s minimem chyb. Má logickou strukturu a povětšinou splňuje náležitosti odborného stylu (stylistické prohřešky jsou drobné, např. v kap. 6: "Now we need to think about how we visualize individual features"). Její délka dobře odpovídá jejímu reálnému rozsahu a obtížnosti. Po formální stránce je práce v pořádku.

Stručný úvod práci motivuje a vytyčuje její cíle. Kapitola 1 uvádí potřebné termíny z hudební domény, zpracování zvuku a zběžně i strojového učení. Kapitola 2 pak popisuje a přesně definuje konkrétní extrahované rysy. Kapitoly 3, 4 a 5 popisují, jak autor řešil úlohu rozpoznávání emocí v hudbě. Kapitola 3 uvádí, jaký model emocí byl v práci používán, a jaký dataset. Oceňuji příklady rozptylu lidských anotací! Kapitola 4 popisuje provedené pokusy s klasifikací emocí. Řešitel postupoval logicky od jednodušších pokusů ke složitějším na základě průběžných výsledků a analýzy problémů. Zajímavý nápad je použití polární reprezentace. Obrázky s křivkami přesnosti by měly mít popsanou vertikální osu jako Error, nikoliv Accuracy, a osa by také měla mít souměřitelný rozsah. Pro replikovatelnost pokusů by také měla být přesně popsána architektura použitých modelů. Kapitola 5 pokusy s klasifikací emocí shrnuje. Oceňuji, že se pouští do kritického zhodnocení literatury a dat, nejen učících algoritmů samotných.

Bibliografie obsahuje 27 položek a je relativně aktuální (8 položek vydaných v posledních třech letech). Je pokryta základní literatura o rozpoznávání emocí v hudbě jako Krumhansl (1990) či Thayer (1990). Po formální stránce je bibliografie v pořádku. Co však chybí, aby práce byla po vědecké stránce zcela vyhovující, je o něco kompletnější přehled o souvisejících pracích. S ohledem na ne vždy přesvědčivý výsledek beat trackingu reflektovaný v závěru práce bych řešitele rád odkázal na práce vyhravší soutěž v beat trackingu ze série MIREX (především práce Sebastiana Böcka, CP JKU Linz). Mohly by také být zmíněny alternativní modely emocí v hudbě, různé datasety a novější literatura o MER než článek z roku 2009 – toto by se hodilo v kapitole 3. Stejně tak práce není první, která vizualizaci hudby řeší – jaký zajímavý software pro tyto účely existuje? Čím je mu práce podobná, čím se liší?

Zatímco kód je dobře komentován, uživatelská dokumentace vizualizačního softwaru je minimální a skrývá se v sekci 6.4 s názvem Implementation. Uživatelská dokumentace pro extrakci rysů chybí.

Implementační část práce

	lepší	OK	horší	nevyhovuje
Kvalita návrhu *... architektura, struktury a algoritmy, použité technologie*	X			
Kvalita zpracování *... jmenné konvence, formátování, komentáře, testování*	X			
Stabilita implementace	X			

Software je logicky navržený, technologie pro jednotlivé části práce byly zvoleny vhodně. Kód je řemeslně v pořádku, nadstandardně dobře komentovaný, pouze (jak je pro bakalářské práce obvyklé) chybí testy. Součástí distribuce je i DLL pro extrakci hudebních featur. Je pěkné, že je rovnou open-source (odkazy na github jsou součástí bibliografie práce).

Co mi v příloze poněkud chybí je příklad vyrenderovaný jako video – doufám, že demo bude součástí obhajoby.

Celkové hodnocení Výborně

Práci navrhuji na zvláštní ocenění Ne

Datum Podpis