Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Tomáš Pavlín		
Název práce	Dance Recognition from Audio Recordings		
Rok odevzdání	2020		
Studijní program	Informatika	**Studijní obor**	Umělá inteligence

Autor posudku	Mgr. Josef Moudřík	**Role**	Oponent
Pracoviště	KTIML MFF UK		

Text posudku:

Cílem práce bylo prozkoumat možnosti vytvoření automatického systému pro rozpoznávání hudebních žánrů (například z krátkých zvukových nahrávek) pomocí konvolučních neuronových sítí. Problém rozpoznávání hudebních žánrů byl v práci omezen na rozpoznávání 10 tanečních stylů (valčík, rumba, tango, samba, ..) - velice praktický problém každého žáka taneční školy.

Co se modelu / použitého přístupu týče, jádro práce je v podstatě aplikace běžného řešení - transformace zvuku na obrázek (spektrogram) a použití CNN klasifikátoru. Hlavní přínos práce vidím v tom, že si řešitel musel projít celým cyklem aplikace AI modelu: od studia literatury a současného state-of-the-art (jak v konkrétní doméně rozpoznávání hudby, tak v širší doméně CNN), přes (částečně i ruční) sběr dat, až po train-eval-improve iterace a experimenty a ve finále i praktické nasazení ve webové aplikaci. Na práci mi přišlo zajímavé hlavně použití aktuálních architektur CNN (DenseNet, ResNet, ResNeXt) v experimentech, spolu s použitím Transfer Learning.

Práce je psána dobrou angličtinou, v textu jsem našel jen jednu textovou chybu - zmatené objasnění v 4.1.1., 2 krát zvýrazněná "aggregation is used".

Jisté výhrady mám vůči testování & evaluaci. Testovací i validační množiny jsou malé (každá z 10 tříd má 6 vzorků v testovací a 6 vzorků ve validační množině) a jsou obě z jedné distribuce, která je jiná než training data. Epocha, na které má validační množina nejlepší Acc je označena za nejlepší a použita pro testování. Malá velikost testovací & validační množiny spolu s vysokou korelací scores testovací a validační množiny a relativně velkým počtem bodů kdy se nejlepší epocha vybírá (100) ale podle mě znamená, že tato procedura výběru nejlepšího modelu je v podstatě hrubé fitování právě na validační (a tím i testovací) množinu. O takovém testování se pak podle mého názoru nehodí přemýšlet jako o testování modelu na nezávislých datech, ale spíše jako na "dotrénování & evaluace modelu pro YT studiový dataset s malým počtem vzorků". To bych ale neoznačoval jako testování/validaci. Oporu pro to poskytuje nepřímo i Tabulka 4.8, kde mají data samplovaná ze stejné distribuce jako training data (s disjunktními samply) horší skóre než data z testovací množiny (tzn, model zde může klidně být víc overfitovaný na "nezávislou" testovací množinu, než na neviděná data z training distribuce).

Práce ale poměrně důkladně testuje takto vybraný model i na jiných datech (Extended Ballroom dset, low-quality data, Star Dance, ..), takže dostatečně objektivní představu o

výkonu modelu si z práce udělat lze.

Pro nezávislé srovnání kvality výsledků a podložení tvrzení "our results probably achieve state-of-the-art" by však bylo třeba více srovnání s ostatními pracemi. Řešitel přímo porovnává svou metodu jedině s metodou MASSS [MARCHAND, PEETERS] na Extended Ballroom datasetu (Table 4.5.), ale protože výsledek MASSS nepoužívá Accuracy (ačkoliv to řešitel chybně tvrdí), ale mean-recall (viz Table 1 v MASSS paperu), je Table 4.5. v práci nadsazená v neprospěch MASSS paperu (který ale je i tak podle těchto metrik lepší). Užitečné by bylo práci srovnat například i na "Cretan Dances" datasetu (viz MASSS paper), nebo nějakém jiném musical-genre classification datasetu.

Práci doporučuji k uznání, ale na obhajobě doporučuji výše uvedené prodiskutovat. Zajímalo by mě i:

Q: Jak by vypadaly výsledky na Test/Validační množině, pokud by se nejlepší epocha vybírala podle accuracy na datech z 4.6.1. - "Separation from training dataset"?

Q: Proč se omezovat s validací jen na tance, když music genre classification je typově zcela shodná úloha a navíc je k dispozici množství datasetů a prior art? Jak by si metoda vedla pro nějaký takový dataset (třeba včetně trénování a test/validate procedury)?

Q: Pro robustnost jakékoliv aplikace je zásadní confidence modelu; jak by šel přístup rozšířit tak aby rozpoznal, že e.g. vstup je jiný žánr, nebo, že vstup vůbec není hudba, atp?

Práci doporučuji **k obhajobě.**

Práci nenavrhuji **na zvláštní ocenění.**

Pokud práci navrhujete na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).

Datum 24.1.2020 **Podpis**