

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy v Praze

Autor práce Marek Židek
Název práce Generování hudby pomocí neuronových sítí
Rok odevzdání 2017
Studijní program Informatika
Studijní obor IOI

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

K celé práci

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Obtížnost zadání	X			
Splnění zadání	X			
Rozsah práce <i>... textová i implementační část, zohlednění náročnosti</i>	X			

Řešitel vytvořil experimentální model pro generování hudby „z ničeho“, a to s poměrně zdařilými výsledky. Použil přitom aktuální techniky využívající rekurentní neuronové sítě, přičemž vymyslel, implementoval a vyhodnotil inovaci, která doposud v literatuře o generování hudby nebyla vyzkoušena a která poměrně jednoznačně zlepšila výsledky: „řidké zkratky“ v rekurentních sítích (sparse skip-connections).

Úvodní nečíslovaná kapitola slouží jako rozšířený abstrakt. V kapitole 1 autor podrobně formuluje zadání problému: zavádí používané hudební termíny, definuje výstupní reprezentaci a motivuje ji; na základě této analýzy problému generování hudby pak v závěru kapitoly precizně formuluje cíle práce. V kapitole 2 popisuje autor standardní neuronové sítě pro modelování sekvenčních dat: rekurentní sítě, konvoluční sítě, a na závěr metodu generativních adversariálních sítí (GAN), s ohledem na jejich využití v práci, a zavádí tak technickou část terminologie. Kapitola 3 popisuje současný stav poznání v generování hudby. V kapitole 4 autor přikročí k definování modelů, které chce ke generování hudby použít: začíná jednoduchou sítí typu Long Short-Term Memory (LSTM), a postupně na základě průběžných výsledků motivuje „přídavky“: vzorkování výstupu, zkratky přes časové kroky, a zlepšovací GAN. Kapitola 5 pečlivě popisuje zdroje dat a přípravu trénovací sady. Nezbytnou odbočkou z technické části práce je pak kapitola 6, ve které autor poměrně podrobně rozebírá možnosti, jak pokusy vyhodnocovat, a popisuje a motivuje schéma evaluací, které použil. V kapitole 7 jsou dodány detaily experimentálního nastavení, především přesné parametry optimalizačního procesu. Kapitola 8 popisuje výsledky; autor nespokojil s „jedním číslem“, a snaží se o širší studii toho, jak vygenerovaná hudba působí, což jde dle mého názoru vysoko nad rámec evaluací v citované literatuře. Součástí je i meta-evaluace: reflexe samotného evaluačního schématu. Nečíslovaná závěrečná kapitola rekapituluje nejdůležitější výsledky a naznačuje směr, kterým se chce autor dále ubírat.

Zdrojový kód a dokumentace pro replikování pokusů jsou poskytnuty v příloze, stejně jako sada vygenerovaných vzorků z různých vyzkoušených modelů.

lepší OK horší nevyhovuje

Nejsilnější stránkou práce je množství provedených experimentů a z nich plynoucí inovace generujícího modelu, která vedla ke zlepšení výsledné hudby. Řešitel samostatně nastudoval příslušnou literaturu, a navíc prováděl zcela samostatně i většinu pokusů, ještě než mne v dubnu požádal o vedení práce (kvůli odchodu předchozího vedoucího z MFF).

Druhou velice silnou stránkou je důsledná snaha o podrobnou evaluaci. Evaluace generování hudby je klíčová a přitom obtížná a sporná záležitost, kde interpretace výsledků velice záleží na přesném postupu vyhodnocování: zde je nutné ocenit ochotu řešitele se této problematice věnovat velmi důsledně, a jednoznačně nad rámec zadání, které bylo v tomto směru opatrné. Výsledná metodologie není ještě zcela přesvědčivá (jak je ostatně diskutováno v kapitolách 6 a 8), nicméně řešitel v průběhu práce dokázal do problematiky evaluací vygenerované hudby proniknout, vypracovat podrobnou metodologii, a vyhodnotit výsledky – do té míry, do které to omezený počet respondentů v dotaznících umožňoval. Evaluační schéma je navržené robustně a klade správné otázky. Pouze míra nejistoty výsledků by měla být v části 8.4 reflektována explicitněji.

Nejslabší stránkou práce je samotný text. Ačkoliv text obsahuje všechny potřebné informace a má přehlednou strukturu, anglický jazyk řešitel nakonec nedokázal ovládnout natolik, aby odstranil všechna obtížně srozumitelná místa, a v práci je také poměrně hodně gramatických chyb. Práce jako taková však srozumitelná je.

Vygenerovaná hudba je vůči deklarovanému cíli „libozvučnosti“ ještě poněkud vzdálena. Harmonicky jasnější části se střídají s plochami, které postrádají směr. Hudba má v rámci jednoho vygenerovaného vzorku poměrně konzistentní texturu, často však příliš dlouho; v tomto směru přinesl model s řídkými zkratkami největší zlepšení. Souzvukový svět vygenerované hudby připomíná nejvíce avantgardní klavírní hudbu první poloviny 20. století. Model tedy rozhodně nekopíruje klasicistní či romantický styl trénovacích dat. To je sice částečný neúspěch použitých metod, ale model zajímavý aplikačně: jeho výstup může sloužit jako „surový materiál“ pro další kompoziční zpracování. Naopak úspěchem je, že poměrně často modely produkují alespoň na začátku generovaných vzorků harmonicky uzavřená čtyřtaktí, tj. středněúrovňovou hudební strukturu, ačkoliv bývá tato struktura skrytá pod disonantními „notami navíc“. (Jedná se o bakalářskou práci a tedy nelze očekávat, že budou experimentální výsledky srovnatelné s aktuálním stavem poznání; tím pádem skutečnost, že vygenerovaná hudba nedosahuje většinou kvalit hudby skutečné, nepovažuji za zásadní pro závěrečné hodnocení práce.)

Práce má potenciál vzhledem k současnému stavu poznání v hudební informatice směřovat ke třem různým vědeckým publikacím. Zaprvé, je možné rozpracovat nápad se sparse skip-connections: tato modifikace architektury má zřetelný vliv na strukturu vygenerované hudby, a elegance tohoto nápadu snadno umožňuje navrhnout a provést pokusy, které podrobněji zachytí chování modelu v závislosti na parametrech „zkratek“. Zadruhé, řešitel zajímavě využívá metodu trénování generativních modelů Generative Adversarial Networks (GANs) pro vylepšení vygenerované hudby. Aplikace metody GAN je nejvíce ambiciózní směr navazujícího výzkumu, avšak zároveň má největší potenciál vést k inovativním výsledkům. Zatřetí, samotná evaluační metodologie a analýza problémů spojených s návrhem příslušných dotazníků je pro podobor generování hudby zajímavý a přínos, ačkoliv také vyžaduje další rozpracování, a především vyšší počet respondentů.

Celkově se jedná o zajímavý experimentální příspěvek k nadstandardně obtížnému tématu, s originálním přínosem, který by při rozpracování např. formou diplomové práce mohl vést k publikování na mezinárodních konferencích. Za velké plus považuji také evaluační metodologii, která je navržena tak, aby autorovi poskytla užitečné informace pro další vývoj vč. aplikací. Řešitel prokázal schopnost vykonávat výzkumnou práci v souladu s oborovými standardy, především navrhopvat a implementovat smysluplné inovace modelů hlubokého učení. Z těchto důvodů navrhuji práci hodnotit stupněm **Výborně**.

Textová část práce

lepší OK horší nevyhovuje

Formální úprava <i>... jazyková úroveň, typografická úroveň, citace</i>			X	
Struktura textu <i>... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</i>	X			
Analýza		X		
Vývojová dokumentace		X		
Uživatelská dokumentace		X		
<p>Práce má logickou strukturu a převážně splňuje náležitosti odborného stylu, obsahuje však množství gramatických chyb, místy chyby stylistické, a některé věty jsou nesrozumitelné. V kapitole 8 by asi bylo vhodnější popsat interpretaci houslových grafů přímo v textu, před Fig. 8.1, než v jejím popisku. Vyskytují se také drobné formální nedostatky (např. číslování v kapitole 5: 5.0.1 místo 5.1, poznámky pod čarou začínající malým písmenem).</p> <p>Rozsah textu je větší (73 stran), což je adekvátní objemu práce.</p> <p>Uživatelská dokumentace je psaná stručně a jasně, obsahuje všechny pasáže nutné pro replikaci pokusů. Pro podobné pokusné skripty není vývojová dokumentace kritická, resp. její funkci plní popis pokusu v samotné práci, avšak by ještě bylo vhodné opatřit alespoň jednotlivé třídy odvozené od <code> theano.Op</code> popisem toho, jak se využívají.</p>				

Implementační část práce

lepší OK horší nevyhovuje

Kvalita návrhu <i>... architektura, struktury a algoritmy, použité technologie</i>	X			
Kvalita zpracování <i>... jmenné konvence, formátování, komentáře, testování</i>		X		
Stabilita implementace		X		
<p>Vzhledem k tomu, že se jedná o experimentální práci a cílem nebylo vytvořit žádné softwarové dílo, je smysluplné hodnotit implementaci pokusů z hlediska jejich replikovatelnosti. Replikovat pokusy na základě poskytnutého softwaru lze. Řešitel poskytuje experimentální skripty jako balíček virtuálního prostředí <code>conda</code>, používá jazyk Python a knihovnu <code> theano</code>, masově využívanou pro hluboké učení; tyto volby jsou v souladu s doporučenými postupy v oboru.</p> <p>Samotný dataset bohužel nebylo možné poskytnout vzhledem ke copyrightovým omezením části trénovacích dat. Řešitel ale poskytl alespoň soupis všech použitých souborů, a na požádání poskytuje i nástroje na replikování datasetu; toto je v podobných situacích běžný postup.</p> <p>Instalace proběhla hladce. Oproti běžným implementacím experimentálních balíčků je slabší uživatelské rozhraní hlavního skriptu <code>main.py</code> : chybí např. možnost přesměrovat výstup do uživatelem definovaného adresáře místo defaultního <code>output/</code>, a chybí kontroly, které zkušenější výzkumníci často používají (např. proti přepsání dříve uložených modelů). Nejkritičtější část pokusného „potrubí“, extrakce dat, je implementována adekvátně; stejně tak modely samotné. Dobře jsou používány inline komentáře, které v kódu popisují details, jejichž smysl není na první pohled zřejmý.</p>				

Celkové hodnocení Výborně

Práci navrhuji na zvláštní ocenění Ne

Datum: 25. 8. 2017

Podpis: