

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Marek Židek

Název práce Controlled Music Generation with Deep Learning

Rok odevzdání 2020

Studijní program Informatika **Studijní obor** Umělá inteligence

Autor posudku Filip Matzner **Role** oponent

Pracoviště Katedra softwaru a výuky informatiky

Text posudku:

Práce se zabývá kontrolovatelným generováním hudby pomocí hlubokých neuronových sítí. Problematika generování hudby skýtá krom jiného dvě fundamentální překážky. První překážkou je neexistence objektivního měřítka pro kvalitu hudby a druhou překážkou jsou dlouhé a závislé časové horizonty (např. sloka, refrén, sloka, refrén...), které bývají pro generativní modely stále problém. Autor k těmto překážkám přistupuje originálním způsobem a práci rozděluje do dvou rozsáhlých částí. V jedné z nich se zabývá tvorbou datasetu s hudebními vzorky a v druhé z nich generováním nových vzorků, které se skládají dohromady způsobem, na který jsme zvyklí v naší hudební kultuře.

Samotná tvorba datasetu vyžadovala rozpoznat podobnost a odlišnost v hudbě, což je těžký a subjektivní úkol i pro člověka. Autor vytvořil smysluplné formální definice této problematiky a chytře využil algoritmy hledající nejmenší editační vzdálenost. Při generování nových vzorků využil state-of-the-art v seq2seq modelech neuronových sítí a ukázal, že umí porozumět moderní literatuře v tomto oboru.

Díky neexistenci objektivních měřítek pro kvalitu hudby byla tato práce vysoce experimentální s nejasným výsledkem, nicméně autor ukázal systematický postup výzkumu a dobře podkládal své myšlenky. Nagegenerované ukázky dávaly hudebně smysl, nicméně je třeba mnoho dalšího výzkumu, aby si člověk mohl s chutí pustit večerní hudební generátor místo svého oblíbeného skladatele.

Autorovi práce se povedlo ukázat, že tento originální směr dává smysl a splnil zadání diplomové práce. Práci proto doporučuji k obhajobě.

Otázky pro autora:

1. Při hledání editační vzdálenosti se akordy považují za neuspořádanou množinu not a nahrazení libovolné noty má stejnou váhu. Myslí si autor, že nahrazení či smazání např. tercie a kvinty z kvintakordu má stejný vliv na hudební formu jako vyměnění základního tónu?

Např. nahrazení kvintakordu G dur jediným tónem G versus výměna D za E ve stejném kvintakordu (což z něho udělá kvintakord e moll)?

2. Transformer často potřebuje velké množství dat k rozumnému natrénování. Zkoušel autor natrénovat tento model na různě velkých podmnožinách datasetu a zjistit tak, jestli by mu pomohlo dataset zvětšit?
3. Plánuje autor práci publikovat ve formě článku?

Poznámky k textu:

1. Jedná se o rozsáhlou práci, kde samotný text má přes 100 stran (vč. referencí a seznamů).
2. Text je psán velmi dobrou angličtinou a až na výjimky je srozumitelný s malým množstvím chyb a překlepů.
3. Obrázky jsou většinou nereferecovány. Myslím, že s referencováním všech obrázků a tabulek by pro čtenáře bylo snazší přesně vědět, kdy se na obrázek podívat.
4. Číselné reference na kapitoly se občas jen tak objeví v textu a není vůbec jasné, co takové číslo znamená.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 25. 1. 2020

Podpis: