

Tvorba hudby pomocí počítače

Martin Švanda

Úvod	1
Metoda strojového vytváření hudby	1
Příklady vygenerované hudby	2
Vlastní práce	2
Shrnuté výsledků a diskuze	3
Shrnutí	4
Reference	4

Úvod

Hudbu často považujeme za čistě lidské umění. Jak by mohl stroj stvořit hudbu? Někdo vám, řekne že na to potřebujete sluch, kreativitu, inovativitu a někdo vám možná řekne, že dokonce potřebujete duši. Díky nejmodernějšímu výzkumu v oblasti umělé inteligence, ale dokážou stroje i skládat hudbu. Jsou to stroje, které se umí učit a učí se podobně jako my: studiem předchozích děl. Program se naučí tisíce nahrávek od reálných umělců a následně je schopen napodobit jejich styl i hlas. Nevytváří ale hudbu tak, že by poskládal kousky té již existující, ale vytváří zcela novou hudbu, kterou ještě nikdy nikdo neslyšel. Pokud vás toto zaujalo, tak se můžete dále podívat na příklady takto vytvořené hudby a i se dočíst, jak si takovou hudbu vytvořit sám.

Metoda strojového vytváření hudby

Způsobů, jak může počítač vytvořit hudbu je více. Jeden může například být, že mu poskytneme nahrávku, kterou si pro sebe „rozseká“ na malé kousky. Ty se potom naučí přeházet za sebou tak, aby vznikla hudba srovnatelná s originálem. Metoda, kterou jsem ale využil funguje jinak. Zde se programu opět poskytnou nějaké nahrávky, ale místo toho aby je „sekal“, tak se z nich naučí, jak hudba vypadá a potom vygeneruje hudbu zcela novou.

Konkrétně jsem použil jeden z nejmodernějších programů na generování hudby a to Jukebox^{1,2} od společnosti OpenAi, která se dlouhodobě zabývá vývojem umělé inteligence. Tento program nabízí navíc něco velmi převratného a to dokonce přidat to vygenerované hudby i zpívaná slova. Tento program je volně přístupný. V rámci této práce byla využita uživatelsky nejpřívětivější způsob interakce s programem a to předpřipravený program v prostředí google collaboratory³.

Při generování hudby je potřeba nastavit několik parametrů. Konkrétně „teplotu“ (0-1), která určuje jak moc se má vygenerovaná hudba podobat originálu na kterém se program učil. Autoři doporučují 0,98 až 0,99, autora a žánr (je možné s i vybrat z velké nabídky¹) a délku nahrávky, která velmi výrazně ovlivňuje časovou náročnost a tak byla v rámci této práce pouhých 45-60 sekund.

Příklady vygenerované hudby

Pokud vás zajímají možnosti, jaké tato technologie nabízí, je možné si prohlédnout seznam příkladů vygenerovaný samotnými autory⁴.

Můžete si vybrat z několika špičkových výtvorů, kde je již velmi těžké rozpoznat, že nejde o originál:

Classic pop, Frank Sinatra: https://soundcloud.com/openai_audio/jukebox-265820820

Country, Alan Jackson: https://soundcloud.com/openai_audio/jukebox-86115728

Pop, Katy Perry: https://soundcloud.com/openai_audio/jukebox-novel_lyrics-78968609

Ale toto jsou již opravdu vrcholové výkony. Průměrné výtvořky vypadají spíše takto:

Classic pop, Frank Sinatra: <https://jukebox.openai.com/?song=787818868>

Country, Alan Jackson: <https://jukebox.openai.com/?song=789018742>

Pop, Katy Perry: <https://jukebox.openai.com/?song=788110930>

I toto je velmi dobrý výkon, nicméně slyšíme, že výslovnost a srozumitelnost již často pokulhává. Ale vzhledem k náročnosti napodobení lidského hlasu je to velmi dobře funkční. Hudba samotná má málokdy chyby, občasné šumění, které se někdy objevuje je důsledkem nartikulovaných slov, nebo tzv. nekompletně unsamplované nahrávky.

Vlastní práce

Pro vlastní práci a vlastní generování hudby jsem si zvolil za cíl přenést naši školní hymnu do několika různých žánrů. Vzhledem k tomu, že ke generování jsem neměl k dispozici velkou výpočetní kapacitu, tak jsou všechny nahrávky krátké a jejich kvalita tím také utrpěla. Odkazy na všechnu vygenerovanou hudbu lze najít [zde](#). Dále budu uvádět nejlepší příklady.

První pokusy proběhly s latinským textem:

Gaudeamus igitur,

Juvenes dum sumus;

Post icundum iuventutem,

Post molestam senectutem

Nos habebit humus.

Vivat academia,

Vivant professores,

Vivat membrum quodlibet,

Vivat membra quaelibet;

Semper sint in flore!

Hip Hop, Eminem, teplota 0,98: <https://drive.google.com/file/d/11tDXQF5h97rAX-00-nJB4VmJ1zWBUBbl/view?usp=sharing>

Hip Hop, Eminem, teplota 0,99, vzorek nebyl plně unsamplován:

<https://drive.google.com/file/d/1104sjrStEAJMbpQ31rwtC7nQYiBDeTOB/view?usp=sharing>

Rock, Dawid Bowie, teplota 0.98, vzorek nebyl plně unsamplován:

https://drive.google.com/file/d/10vTIFliw_cEftPeHn0A5TCmxISRJYV6R/view?usp=sharing

Další práce probíhala již s anglickým textem:

Let us therefore rejoice,
While we are young;
After our youth,
After a troublesome old age
The ground will hold us.

Long live the university,
Long live the teachers,
Long live each male student,
Long live each female student;
May they always flourish!

Soundtrack, Hans Zimmer, teplota 0,98, vzorek nebyl plně unsamplován:

<https://drive.google.com/file/d/10tvKvOt5MRDmF4bFr-EipIL7cSkY04oY/view?usp=sharing>

Country, Alan Jackson, teplota 0,98, vzorek nebyl plně unsamplován:

https://drive.google.com/file/d/11TOiwzq6XKvY40Q1aCyJqd_bdr5YWCKV/view?usp=sharing

https://drive.google.com/file/d/10hBV_kacITq-tzRwiKZLbJffVs8mq4ea/view?usp=sharing

https://drive.google.com/file/d/10rfsMJXzyR7_eA1dDc9IZABM7zcAninh/view?usp=sharing

Shrnutí výsledků a diskuze

Závěry z vygenerovaných nahrávek a jejich vysvětlení:

- Latinský text je nesrozumitelný - Vzhledem k tomu, že program byl trénován na angličtině, tak není příliš překvapivé, že nezvládá jiné jazyky.
- Nahrávky jsou velmi zašuměné a text je těžko srozumitelný - Toto je důsledek nekompletního unsamplování nahrávek (viz. další bod).
- Až na jeden případ jsou všechny nahrávky neunsamplované - Zde je vidět hlavní problém nižší výpočetní kapacity a využití googel collaboratory (která má omezený čas běhu jednoho programu a to pouze 12 hodin). Naprostá většina programů se tak spadla a vůbec se nedostala ke generování užitečné hudby, anebo ji již nestihla unsamplovat, než byl program přerušen.
- Když ignorujeme šum a špatnou výslovnost, tak je hudba relativně dobrá - Hudba samotná nebyla nedostatečným unsamplováním tolik ovlivněna. Tento výsledek není natolik překvapivý, generovat zpívající lidský hlas je výrazně těžší než hudbu.

Vygenerované nahrávky mají bohužel pouze velmi špatnou kvalitu, především zpěv. Nicméně i u vrcholových nahrávek (viz. výše) jsme slyšeli, že výslovnost není zcela dokonalá a že věrohodně napodobit lidský hlas je problém i pro profesionální vědce zabývající se tímto tématem. Když tedy uvážíme technické limitace laického přístupu, tak se výsledky dají považovat za velmi dobré.

Shrnutí

Program Jukebox od OpenAi je revoluční nástroj, který umožňuje generovat hudbu, včetně zpívaného textu, ve stylu určeného žánru a umělce. Vrcholové výtvořky tohoto programu mají stále menší problémy s výslovností, ale jinak je jejich kvalita až překvapivá. Pro laické

použití je tento program ale značně omezen hlavně z důvodu velké výpočetní kapacity. Tento problém by mohl být vyřešen spuštěním programu na místním zařízení (tedy ne google collaboratory), to je ale ještě více uživatelsky nepřívětivé. V případě, že nemáte k dispozici lepší výpočetní kapacitu, tak bych doporučoval omezit se na netextovou hudbu, která funguje velmi dobře. Pokud vás tento program zaujal, můžete využít odkazů níže a vygenerovat si například zcela novou hudbu od Beethovena, kterou ještě nikdo neslyšel.

Reference

- [1] <https://openai.com/blog/jukebox/>
- [2] Prafulla Dhariwal and Heewoo Jun and Christine Payne and Jong Wook Kim and Alec Radford and Ilya Sutskever, Jukebox: A Generative Model for Music, [arXiv:2005.04141](https://arxiv.org/abs/2005.04141)
- [3] https://colab.research.google.com/github/openai/jukebox/blob/master/jukebox/Interacting_with_Jukebox.ipynb
- [4] <https://jukebox.openai.com/>