

Folk-rnn: využití rekurentní neuronové sítě v rámci kompozice folkové hudby

Daniel Kvak

Bankovníctví, internetové vyhledávače, dokonce termostaty v našich domovech: umělá inteligence (artificial intelligence - AI) postupně proniká do každodenního života. Algoritmy dokáží predikovat uživatelské chování a simulovat očekávané reakce v takové míře, že v různých odvětvích nahrazují lidský faktor. Využití umělé inteligence v kulturních oblastech ovšem budí rozpaky. Je vůbec nahrazení procesu kreativity bez znalosti kulturního a společenského pozadí možné?

Od náhody k algoritmu

Za počátek algoritmické kompozice můžeme označit Musikalisches Würfelspiel, hru z osmnáctého století, která využívá hod kostkou k náhodnému generování skladeb z předem sestavených možností. Právě náhodný proces zde řídí aspekt výběru melodického materiálu a opakované vyvolávání algoritmu přináší různé výsledky: v případě nejslavnějšího díla tohoto typu, Mozartova valčíku, se jedná o 11^{16} unikátních, leč často podobných kombinací.¹

Nejběžnější způsob kompozice skladby pomocí matematických modelů je stochastický proces, který je založen na náhodné události stejně jako hod kostkou. Tvůrčí proces může být skladatelem kontrolován pouze částečně, jak lze slyšet například v aleatorických skladbách Johna Cage či Iannis Xenakise.²

Významným příkladem stochastického algoritmu je Markovův řetězec. Ten popisuje diskrétní procesy, pro které platí, že výsledek v budoucím čase závisí pouze na současném stavu řetězce, nikoliv na předchozích stavech.³ Skryté Markovovy řetězce, které nezobrazují aktuální stav, ale pouze výstupní hodnoty,⁴ jsou jedním ze základních pilířů umělé neuronové sítě (ANN). Z důvodu skrytých vrstev a užití

¹ Noguchi, Hideo. Mozart: Musical game in C major, K.516f. *Mitteilungen der Internationalen Stiftung Mozarteum*, roč. 38, č. 1-4, 1990. Dostupné z <<http://www.asahi-net.or.jp/~rb5h-ngc/e/k516f.htm>>.

² Iverson, Jennifer. Statistical form amongst the Darmstadt school. *Music analysis* [online], roč. 33, č. 3, 2014. DOI: <https://doi.org/10.1111/musa.12037>

³ Ji, Ke - Sun, Runyuan - Shu, Wenhao - Li, Xiang. Next-song recommendation with temporal dynamics. *Knowledge-Based Systems* [online], roč. 88, 2015. DOI: <https://doi.org/10.1016/j.knsys.2015.07.039>

⁴ Každý stav má pravděpodobnostní vliv na výstup systému. Přívlástek skrytý se tedy vztahuje pouze na posloupnost vnitřních stavů.

různých topologií tvoří ANN spletitý systém vztahů, který je pro člověka nečitelný. Oproti tradičním algoritmům strojového učení, které jsou schopny zobrazit korelační analýzu, tak nedokáže popsat strukturu sítě ani závislosti mezi proměnnými.⁵

Každá neuronová síť je složena z jednotlivých neuronů (perceptronů),⁶ které jsou vzájemně propojeny: výstup jednoho neuronu je vstupem do dalších. Výsledná funkce sítě se nazývá topologie. Dle ní rozlišujeme, zdali síť obsahuje (recurrent) či neobsahuje (feedforward) zpětné vazby, které slouží k zachycení vnitřního stavu sítě.⁷

Praktické užití neuronových sítí

V květnu roku 2017 bylo publikum ve východním Londýně svědkem každoročního festivalu tradiční irské hudby. Tato událost byla ovšem jiná: melodie, které návštěvníci slyšeli, byly vytvořeny umělou inteligencí. Algoritmus vyvinutý vědci na Kingston University a Queen Mary University of London je jedním z mnoha projektů zkoumajících průnik AI a kreativního umění. Program zvaný folk-rnn byl přijat se smíšenými pocity; zatímco někteří přihlížející byli překvapeni autentičností projevu, jiní nezakrývali deziluzi z něčeho, co by mělo navždy zůstat doménou člověka.⁸

Folk-rnn, jak již z názvu vyplývá, pracuje na bázi rekurentní neuronové sítě (recurrent neural network, RNN). Namísto spoléhání se na předdefinovaná pravidla, folk-rnn zpracovává velké množství dat a vytváří číselnou reprezentaci vzorků. Korelace, které najde, pak ukládá a používá k plnění úkolů.⁹ Možnost ukládat stavy za pomoci paměťových bloků (zde se jedná o tři skryté vrstvy) je umožněno architekturou Long-Short Term Memory (LSTM). Folk-rnn byl postupně „nakrmen“ repertoárem 23 000 tradičních irských skladeb ve formátu ABC,¹⁰ než mohl začít generovat unikátní melodie. Od svého

⁵ Chernoded, Andrey - Dudko, Lev - Myagkov, Igor - Volkov, Petr. Deep Learning Neural Networks and Bayesian Neural Networks in Data Analysis. *EPJ Web of Conferences* [online], roč. 158, 2017. DOI: 10.1051/epjconf/201715806008

⁶ Perceptron je formální model biologického neuronu, který představuje všeobecný výpočtový prvek neuronové sítě.

⁷ Radová, Vlasta. *Typy umělých neuronových sítí a fáze jejich činnosti*. [online] 2017 [cit. 2020-10-27]. Dostupné z <<https://1.url.cz/EzS44>>.

⁸ Brown, Mike. An A.I. in London is Writing Its Own Music and It Sounds Heavenly. *Inverse* [online], 2017 [cit. 2020-10-23]. Dostupné z <<https://www.inverse.com/article/32276-folk-music-ai-folk-rnn-musician-sbest-friend>>.

⁹ Kim, Hyoung-Gook - Kim, Geen, Yeun - Kim, Jin, Young. Music Recommendation System Using Human Activity Recognition From Accelerometer Data. *IEEE Transactions on Consumer Electronics* [online], roč. 65, č. 3, 2019. DOI: 10.1109/TCE.2019.2924177. Dostupné z <<https://1.url.cz/jzS4Y>>.

¹⁰ ABC notace je zkrácená forma notového záznamu. V základní podobě používá k reprezentaci daných tónů písmena A až G. Tato forma se často užívá pro vývoj softwaru, protože je vyjádřitelná ASCII tabulkou.

založení v roce 2007 vygeneroval více než 100 000 písní, z nichž mnohé byly kompilovány do veřejně dostupných svazků.¹¹

Bob L. Sturm, jeden z autorů folk-rnn, se na svém blogu poměrně zajímavě vyjádřil: „*To be blunt, folk-rnn is a computational idiot savant and the Endless folk-rnn Traditional Music Session is a parlour trick. However, folk-rnn is not so idiotic that it is just reproducing material from the training data.*“¹² Tyto dvě věty jsou nesmírně důležité pro pochopení, jak tento program pracuje: zatímco kreativita je ovlivněna osobními zkušenostmi a sociokulturním prostředím, ve kterém se autor pohybuje, rekurentní algoritmus „jen“ analyzuje data a hledá korelace.¹³ Folk-rnn sráží společenskou hodnotu díla na pouhou součást algoritmu.

Umělou inteligenci chápeme z antropologického hlediska jako sociotechnický systém, jehož technologické aspekty jsou úzce spojeny se společenskými hodnotami. Problém ovšem nastává ve chvíli, kdy se snažíme porozumět sociokulturnímu dopadu výstupních dat bez porozumění společenskému světu, ve kterém je AI zabudováno. Můžeme předpokládat, že každá kultura má určité jádro, charakteristiky a neměnnou sadu přesvědčení.¹⁴ Tuto analogii pak můžeme aplikovat i na hudbu, kdy například folkové písně budou vykazovat podobné znaky.

Představme si absurdní příklad: folkový autor prožil své dětství v Japonsku, díky čemuž do své kompozice implementuje místní pentatonickou stupnici Hirajōshi. Dokázal by folk-rnn napodobit tuto kompozici? S vhodnými vstupními daty nejspíše ano. Dokázal by folk-rnn vytvořit takovou kompozici bez znalosti požadovaného výstupu? Ne, leč možno dodat prozatím. Otázkou zůstává, zdali od současného modelu můžeme vůbec očekávat něco jiného, když je naprosto oddělen od míst, kde hudba skutečně vzniká.

Folk-rnn si neklade za cíl nahrazení lidského faktoru v umění, snaží se spíše objevit nové způsoby, jak může člověk a stroj spolupracovat.¹⁵ Právě vysoká účinnost RNN, kdy i relativně jednoduchá neuronová síť dokáže stanovit přesnou korelaci,¹⁶ dala vzniknout startupům jako Jukedeck, Aiva nebo Amper, které

¹¹ Sturm, Bob, L. *What do these 5,599,881 parameters mean? An analysis of a specific LSTM music transcription model, starting with the 70,281 parameters of the softmax layer.* Charlotte, North Carolina, 2018. Dostupné z <<http://kth.diva-portal.org/smash/get/diva2:1260836/FULLTEXT01.pdf>>.

¹² Sturm, Bob, L. *On folk-rnn cheating and music generation.* [online] 2017 [cit. 2020-10-28]. Dostupné z <<https://highnoongmt.wordpress.com/2017/06/15/on-folk-rnn-cheating-and-music-generation/>>.

¹³ Tan, Ai-Girl - Perleth, Christoph. An introduction to the volume of creativity, culture and development. In *Creativity, culture, and development* [e-book]. Tan, Ai-Girl - Perleth, Christoph (eds.). Springer, 2015.

¹⁴ Göransson, Bo - Magnus, Florin (eds.). *Artificial intelligence, culture and language: On education and work.* Berlin: Springer, 1990.

¹⁵ Thompson, Clive. WE WILL BOT YOU. *Mother Jones* [online], roč. 44, č. 2, 2019. Dostupné z <<https://1url.cz/vzSDS>>.

¹⁶ Karpathy, Andrej. *The Unreasonable Effectiveness of Recurrent Neural Networks.* [online] 2015 [cit. 2020-10-28]. Dostupné z <<http://karpathy.github.io/2015/05/21/rnn-effectiveness/>>.

se snaží zaplnit mezeru mezi generováním kratších slok a generováním celé kompozice. V roce 2015 získala firma Jukedeck první sérii investic, které využila k najmutí týmu hudebníků. Ti mají za úkol vytvářet jednoduché, žánrově specifické melodie, m.j. pop, blues či folk. Tato data slouží stejně jako v případě folk-rnn ke strojovému učení a hledání korelací.

Úspěch firmy Jukedeck inspiroval i „velké hráče“ na poli technologického pokroku: Google stojí za vznikem open-source projektu Magenta, IBM se soustředí na svůj projekt zvaný Watson Beat, Spotify vlastní Spotify Creator Technology Research Lab. François Pachet, současný ředitel poslední zmíněné divize, je jedním z průkopníků rychle rostoucího oboru. V rozhovoru, který v loňském roce zveřejnila Evropská komise, trefně zmiňuje, že zatímco před dvaceti lety byly „základním kamenem“ výpočetních oborů algoritmy, dnes jsou vysoce ceněnou komoditou velká data, ve kterých můžeme korelace vyhledávat.¹⁷

Závěr

Na rozdíl od dřívějších modelů algoritmické kompozice, mezi které řadíme například Musikalisches Würfelspiel, nepracuje folk-rnn s pouhými náhodnými procesy, ale dokáže generovat unikátní kombinace, které se vždy liší od vstupních dat. Ve stochastických modelech je ovšem hudba výsledkem nedeterministických metod, které zanedbávají sociokulturní význam původních děl.

Folk-rnn tak sice analyzuje velké množství ABC souborů a následně vyhledává korelace mezi vstupními parametry, nedává si ovšem za cíl (a vlastně ani nemůže) porozumět kulturnímu impaktu folkové hudby. Historické, kulturní ani sociální pozadí díla nelze dle mého názoru vyjádřit algoritmem – alespoň tedy v současné době.

Seznam pramenů a literatury

Computer science in music: interview with François Pachet, director of the Spotify Creator Technology Research Lab. *European Commission* [online] 2018 [cit. 2020-10-23].

Dostupné z <<https://1url.cz/RzSyD>>.

Göranzon, Bo - Magnus, Florin (eds.). *Artificial intelligence, culture and language: On education and work*. Berlin: Springer, 1990.

Chernoded, Andrey - Dudko, Lev - Myagkov, Igor - Volkov, Petr. Deep Learning Neural Networks and Bayesian Neural Networks in Data Analysis. *EPJ Web of Conferences* [online], roč. 158, 2017. DOI: 10.1051/epjconf/201715806008.

¹⁷ Computer science in music: interview with François Pachet, director of the Spotify Creator Technology Research Lab. *European Commission* [online] 2018 [cit. 2020-10-23]. Dostupné z <<https://1url.cz/RzSyD>>.

Iverson, Jennifer. Statistical form amongst the Darmstadt school. *Music analysis* [online], roč. 33, č. 3, 2014. DOI: <https://doi.org/10.1111/musa.12037>

Ji, Ke - Sun, Runyuan - Shu, Wenhao - Li, Xiang. Next-song recommendation with temporal dynamics. *Knowledge-Based Systems* [online], roč. 88, 2015. DOI: <https://doi.org/10.1016/j.knosys.2015.07.039>

Karpathy, Andrej. *The Unreasonable Effectiveness of Recurrent Neural Networks*. [online] 2015 [cit. 2020-10-28]. Dostupné z <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>.

Kim, Hyoung-Gook - Kim, Geon, Yeun - Kim, Jin, Young. Music Recommendation System Using Human Activity Recognition From Accelerometer Data. *IEEE Transactions on Consumer Electronics* [online], roč. 65, č. 3, 2019. DOI: 10.1109/TCE.2019.2924177 Dostupné z <https://1url.cz/jzS4Y>.

Noguchi, Hideo. Mozart: Musical game in C major, K.516f. *Mitteilungen der Internationalen Stiftung Mozarteum*, roč. 38, č. 1-4, 1990. Dostupné z <http://www.asahi-net.or.jp/~rb5h-ngc/e/k516f.htm>.

Radová, Vlasta. *Typy umělých neuronových sítí a fáze jejich činnosti*. [online] 2017 [cit. 2020-10-27]. Dostupné z <https://1url.cz/EzS44>.

Sturm, Bob, L. *What do these 5,599,881 parameters mean? An analysis of a specific LSTM music transcription model, starting with the 70,281 parameters of the softmax layer*. Charlotte, North Carolina, 2018. Dostupné z <http://kth.diva-portal.org/smash/get/diva2:1260836/FULLTEXT01.pdf>.

Sturm, Bob, L. *On folk-rnn cheating and music generation*. [online] 2017 [cit. 2020-10-28]. Dostupné z <https://highnoongmt.wordpress.com/2017/06/15/on-folk-rnn-cheating-and-music-generation/>.

Tan, Ai-Girl - Perleth, Christoph. An introduction to the volume of creativity, culture and development. In *Creativity, culture, and development* [e-book]. Tan, Ai-Girl - Perleth, Christoph (eds.). Springer, 2015.

Thompson, Clive. WE WILL BOT YOU. *Mother Jones* [online], roč. 44, č. 2, 2019. Dostupné z <https://1url.cz/vzSDS>.