

Oponentský posudek na diplomovou práci:

E. JANÁSKOVÁ: STATISTICAL MACHINE LEARNING WITH APPLICATION IN MUSIC

Předložená práce se zabývá použitím metod strojového učení na generování melodií ve stylu Beatles. První kapitola obsahuje obecný úvod a stručný přehled vybraných statistických metod rozdělených na kategorie ‘supervised’ (s učitelem) a ‘unsupervised’ (bez učitele). Druhá kapitole nepatrně podrobněji (na osmi stránkách) popisuje různé modely neuronových sítí. Ve třetí kapitole je opět velice stručný přehled literatury týkající se generování hudby pomocí neuronových sítí.

Vlastní příspěvek spočívá v přípravě dat (tj. úpravě vstupních MIDI souborů na ‘monofonní melodie’), volbě potřebných parametrů, vygenerování 124 nových melodií pomocí tří různých algoritmů a porovnání výsledných skupin melodií pomocí popisných statistik (podle článku Yang, Lerch (2018)). Práce končí konstatováním, že ‘uměle vygenerované melodie nepůsobí tak skutečně a příjemně jako ty původní’.

K praktické části mám následující připomínky:

- Styl písní skupiny Beatles se v průběhu jejich kariéry poměrně výrazně měnil. Nebylo by vhodné použít melodie rozdělit do více skupin podle jejich ‘stylu’? Lze očekávat, že učení podle ‘směsi melodií v různých stylech’ nemůže vést k dobrému výsledku.
- Úloha generování hudby se v práci popisuje jako ‘predikce’. Není mi však jasné, jestli se zde postupuje podobně jako například při predikování budoucích hodnot náhodně vygenerovaných z normálního rozdělení, kdy je dobrým prediktorem odhad střední hodnoty: vlastnosti takového prediktoru se pak obvykle výrazně liší od vlastností predikovaného náhodného výběru (na který se můžeme dívat jako na obdobu melodie). Je někde v práci vysvětleno, jestli se použitý algoritmus snaží jako pokračování skladby dosadit nejpravděpodobnější hodnotu (tj. obdobu odhadu střední hodnoty) nebo jestli je cílem pokračování skladby generovat náhodně, např. z odhadnutého pravděpodobnostního rozdělení možných dalších not?
- Porovnání vygenerovaných a původních melodií je založeno pouze na popisných statistikách. Bylo by možné doplnit uvedené statistiky i vhodnými testy? Lze například rozeznat původní melodie ‘zamíchanou’ mezi ty uměle vygenerované?
- Ve čtvrté kapitole jsou popisné statistiky vypočítané vždy pouze pro jedno nastavení vstupních parametrů pro každou ze tří použitých metod. Bylo by možné prozkoumat, jak změna těchto vstupních parametrů ovlivňuje kvalitu vygenerovaných melodií?
- Proč jsou čísla 1 až 16 v sekci 4.3.2 zakódována pomocí šesti bitů? Nestáčí čtyři?

Připomínky k úvodním přehledovým kapitolám:

Str. 3 Mezi kreativními úkoly lze možná zmínit i počítač Watson navržený pro hru Jeopardy.

Figure 1.1 Jaká funkce Φ je použita na obrázku?

Sekce 1.2.1 Opakovaně prohozené symboly Λ a Γ , nejasná definice vektoru průměrů.

Sekce 1.2.2 Označuje zde slovo ‘sample’ pozorování?

Str. 13 Proč se místo značení používaného v první kapitole (např. X_i, Y_i, \hat{Y}_i) přechází k malým písmenům (tj. x_i, y_i, \hat{y}_i)? Liší se performance measure P zavedená v předchozí kapitole (str. 5) od loss function L (str. 13 dole)?

Str. 15 Musí být matice W symetrická?

V prvních dvou přehledových kapitolách autorka vlastními slovy popisuje vybrané mnohorozměrné statistické metody a modely neuronových sítí. Popis každé metody je velmi stručný (až povrchní), v závěru však lze nalézt odkazy na další studijní literaturu (nejčastěji na knihu Efron, Hastie (2016)). Zdá se, že všechny obrázky v prvních dvou kapitolách jsou převzaté z literatury. Ve druhé kapitole postrádám komentář možných souvislostí popisovaných metod s generováním hudby. Popis tří variant nastavení programu Magenta ve čtvrté kapitole (str. 23) se zdá být zkrácenou verzí příspěvku ‘Generating Long-Term Structure in Songs and Stories’ na webové stránce:

<https://magenta.tensorflow.org/2016/07/15/lookback-rnn-attention-rnn>.

Hlavním cílem bylo generování melodií ve stylu Beatles; podobné pokusy lze ale najít například na stránce www.flow-machines.com, kde píseň ve stylu Beatles vyrobili již v roce 2016, viz například qz.com/790523/daddys-car-the-first-song-ever-written-by-artificial-intelligence-is-actually-pretty-good/. Aplikace zpracovaná v programu Magenta následuje bezprostředně po třech úvodních spíše přehledových kapitolách a text tedy neobsahuje žádnou teoretickou (matematickou) část. Závěrečné porovnání popisných statistik založené na článku Yang, Lerch (2018) pak také nevede k ničemu objektivnímu. Otázka tedy je, jestli předložená práce obsahuje vůbec něco přínosného.

SHRnutí: Uchazečka se očividně snažila dodržet zadání diplomové práce, ale přehledová část mohla být zpracovaná lépe. Nikde nejsou uvedeny žádné teoretické vlastnosti popisovaných algoritmů a závěrečné porovnání výsledných melodií je založeno pouze na popisných statistikách. Domnívám se, že předložená práce nesplňuje požadavky kladené na diplomové práce na MFF UK a nedoporučuji ji k obhajobě.

Doc. RNDr. Zdeněk Hlávka, Ph.D.