

Posudek oponenta diplomové práce

Autor práce: **Zdeněk Janeček**

Název práce: **Generátor podkresové hudby**

Diplomant zkoumá možnosti rekurentních neuronových sítí pro generování polyfonní sekvence. Detailně popisuje teoretické principy fungování moderních “hlubokých” neuronových sítí (aka. deep learning) včetně několika architektur (v množině RBM a RNN) a představuje metody jejich trénování a regularizace. Implementace je stavěna na knihovně Torch.

Ačkoli patří vybrané téma k aktuálním problémům řešeným v deep learning komunitě, čili záběrem odpovídá vědecké práci, při bližším prozkoumání je vidět, že navrhované řešení není nové, jak sám autor přiznává na straně 44. Strukturou práce není bohužel možné rozlišit, kde končí existující známá řešení (state of the art) a co je samotná nová práce diplomanta. Tím nemyslím programovou realizaci v jiném jazyce, ačkoliv autor tvrdí, že jeho implementace je “více modulární” oproti existující implementaci ve frameworku Theano. Tomuto tvrzení bohužel chybí argumentace.

Druhý, velmi zásadní problém je nedostatečné vyhodnocení. Autor ukazuje křivky ztrátové funkce při učení sítě, ale v tabulce 6.2 (str. 53) chybí další uvedené metriky (precision, recall, F-measure). Dále není jasné, co přesně se vyhodnocuje. Jelikož se jedná o generativní model, měl by být vyhodnocen jeho výstup, autor ale “porovnával predikovaný vektor a trénovací”. To si mohu vysvětlit tak, že síť byla natrénována N trénovacími maticemi (písněmi) (piano-roll je sekvence vektorů, alespoň to vyplývá z tabulky 5.1 a sekce 5.2, která bohužel popisuje vstup velmi vágně), a pak každou matici (píseň) poslat na vstup sítě a porovnat s výstupem (píseň). To by ale měřilo jen schopnost sítě jako jakéhosi autoencoderu a nedávalo moc smysl. Bohužel detailní popis toho, jak se vyhodnocují modely generování hudby, chybí, což je pro práci zásadní.

Dále pak není nikde uvedeno, jak funguje samotné generování, jen že stačí zadat počáteční stav (str. 46). Z přiloženého kódu lze vyčíst těžko (snad řádek 5: náhodný vzorek z nějakého poolu?).

Chybí taktéž jakýkoliv hlubší popis úlohy, např. rozdělení na training/devel/test data, velikost těchto dat, atd. Množství dat je zásadním předpokladem pro úspěšné nasazení deep learning přístupu. Zároveň bych ocenil i kvalitativní analýzu chyb, čili zjistit, co je vůbec na výstupu a dává-li to hudebně nějaký smysl.

Po formální stránce je práce vysázena na dobré úrovni. Jak již bylo zmíněno výše, jako hlavní nedostatek vidím míchání teoretického základu s implementačními detaily (např. na straně 16 se zčistajasna objeví funkce v Torch, podobně na str. 20, 33, 36).

Diplomant splnil zadání až na chybějící smysluplnou evaluaci.

Dotazy k práci

Str. 20: convolutional neural networks (CNN). Autor zmiňuje, že generování z nich je “náročné a nespolehlivé” a proto z nich upustil. Co tím bylo myšleno a existuje pro toto tvrzení nějaká relevantní práce? Pokud jsem správně pochopil, chtěl autor použít CNN pro sequence-to-sequence learning, což je ale v principu nevhodné, protože CNN, pokud je míněno jako součást feed-forward sítí, potřebuje pevnou velikost vstupu a výstupu (něco se dá částečně vyřešit tzv. paddingem na vstupu).

Navrhuji hodnocení známkou **velmi dobře** a práci doporučuji k obhajobě.

Darmstadt, 7.6.2016

Ing. Ivan Habernal, Ph.D.

**SOUHLASÍ
S ORIGINÁLEM**

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
katedra informatiky a výpočetní techniky