

## Posudek oponenta diplomové práce

**Student:** Nodžák Petr, Bc.  
**Téma:** Automatické rozpoznání akordů pomocí hlubokých neuronových sítí (id 22903)  
**Oponent:** Bidlo Michal, Ing., Ph.D., UPSY FIT VUT

**1. Náročnost zadání** **značně obtížné zadání**

Práce má za cíl automatickou identifikaci akordů ve skladbách prostředky umělé inteligence, konkrétně využitím různých variant neuronových sítí. Jedná se u výzkumné téma, kde bylo třeba skloubit několik disciplín (vybrané partie teorie hudby, neuronové sítě, strojové učení a další) a vyladit systém celkově tak, aby poskytoval akceptovatelné výsledky. Ačkoli bylo použito podpůrných knihoven, kde to bylo možné, celkově představuje návrh a realizace řešení komplexní problém. Náročný úkol představovala i samotná volba a sestavení datasetu pro trénování, validaci a evaluaci řešení. Z těchto důvodů hodnotím téma jako značně obtížné.

**2. Splnění požadavků zadání** **zadání splněno s drobnými výhradami**

Celkově bylo zadání splněno, za netriviální rozšíření považuji jazykový model schopný predikce akordů a jeho integraci do celkově vyladěného systému.

Drobnou výhradu mám ke splnění části bodu 5 týkající se vyhodnocení výpočetní náročnosti: v tabulkách s výsledky je uveden počet epoch potřebných k trénování modelů, ale chybí podrobnější rozbor a analýza toho, co tato data vypovídají o prezentovaných experimentech.

**3. Rozsah technické zprávy** **je v obvyklém rozmezí**

**4. Prezentační úroveň předložené práce** **85 b. (B)**

Celkově je práce psána srozumitelně, konkrétně mám ale k prezentační stránce několik výhrad:

Téma DP je zaměřeno na aplikaci hlubokých neuronových sítí, v textu ale zcela postrádám bližší informace, co je pro účely práce tímto pojmem myšleno. Dnes je to již dobře zavedený pojem, ale v praxi může mít řadu variant, takže bych zde aspoň stručně obecně pojednání rozhodně očekával. Další (již existující) popis měl z tohoto vycházet - popsané varianty (CNN, RNN atd.) jsou vlastně speciálními případy hlubokých neuronových sítí.

Nejasná potřeba jazykového modelu vedle modelu akustického a o co se vlastně jedná (popis jen jednou větou a obrázkem v kap. 6). Koncept zjevně převzat z literatury, ale ani v sekci 4.3.3, která se toho týká, tato základní informace není. Jazykový model přitom vidím jako významné rozšíření práce nad rámec zadání.

Přínejmenším některé z rovnic popisujících chromagram a jeho různé složky a parametry nejsou správně zapsány. Např. v rovnici (2.4) není jednoznačný rozsah indexu  $p$  (podobně ve vztahu (2.14)). Výraz (2.12) nevyjadřuje množinu. Podobně i v textu, zápis tvaru  $p$  náleží do  $[0 : 127]$  spíše připomíná práci se seznamy v jazyku Python, ale nejedná se o matematicky korektní popis množiny (je-li tím myšlena množina).

Text občas postrádá jednoznačné odkazy na obrázky a je tak nutno si domýšlet, k čemu se daný objekt vztahuje.

Některé objekty nekorespondují s označením v textu. Např. v obr. 3.1 je aktivační funkce perceptronu označena  $g$  (v textu  $y$ ) a bázová funkce  $f$  (v textu  $u$ ).

Za vážný (ale snadno napravitelný) nedostatek považuji až příliš malou výšku grafů s konvergenčními křivkami v kapitole 7, jež zhoršuje jejich čitelnost.

**5. Formální úprava technické zprávy** **85 b. (B)**

Text má vysokou formální úroveň, občas se vyskytují překlipy a gramatické chyby, ovšem ne v míře, která by byla pro hodnocení zásadní. Za zmínku stojí snad jen systematické komolení příjmení Josepha Fouriera, kdy je nesprávně uváděn zápis "Furier".

**6. Práce s literaturou** **90 b. (A)**

Výběr zdrojů je adekvátní řešenému tématu, citace by se na některých místech hodily podrobněji, ale i jinak k tomuto bodu nemám výhrad.

**7. Realizační výstup** **90 b. (A)**

V rámci DP vznikla sada nástrojů pro učení hlubokých neuronových sítí v úloze rozpoznávání a predikce akordů v hudebních korpusech a realizaci nezbytných činností týkajících se předzpracování dat a analýzy výsledků. Nezanedbatelným přínosem diplomanta je též sestavení datové sady pro korektní vyhodnocení experimentů a možnost srovnání s existujícím řešením.

Ačkoli nemám pochybnost o původu zdrojových kódů (při jejichž návrhu bylo též využito volně dostupných prostředků z oblasti strojového učení), přiložené soubory zcela postrádají identifikaci autora a jsou navíc téměř bez komentářů. Částečnou náhradou tohoto může být přiložené readme.pdf s přehledným návodem a popisem jednotlivých částí. Navzdory tomuto z funkční stránky není realizaci co vytknout.

### 8. **Využitelnost výsledků**

Práce přináší množství původních výsledků s publikačním potenciálem.

### 9. **Otázky k obhajobě**

1. Objasněte pojem "jazykový model" a jeho potřebu pro řešení - proč je třeba predikovat akordy, když ty tvoří samotnou skladbu a jejich rozpoznávání provádí neuronová síť (podobně jako navíc predikci v jazykovém modelu)?
2. Jaký závěr dokážete vyvodit z naměřených dat týkajících se výpočetní náročnosti?

### 10. **Souhrnné hodnocení**

**89 b. velmi dobře (B)**

S přihlédnutím k vysoké náročnosti práce, jejímu experimentálně-výzkumnému charakteru, množství původních výsledků, celkovému zpracování, ale i výše uvedeným nedostatkům, navrhuji hodnocení 89 - B.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 12. srpna 2020

Bidlo Michal, Ing., Ph.D.  
oponent