Posudek oponenta diplomové práce

Student: Gajdár Matúš, Bc.
Téma: Nízko-dimenzionální faktorizace pro "End-To-End" řečové systémy (id 23195)
Oponent: Grézl František, Ing., Ph.D., UPGM FIT VUT

1. Náročnost zadání — obtížnější zadání

Práci hodnotím jako obtížnější vzhledem k nutnosti seznámist se s několika nástroji pro rozpoznávání řeči a práci s neuronovými sítěmi. Práce také svým zaměřením cílý na vytvoření state-of-the-art rozpoznávače řeči.

2. Splnění požadavků zadání — zadání splněno

Práce splnila všechny body zadání. Požadavky zadání jsou překročeny v testovací části, kdy student nejen že implementuje maticovou faktorizaci do LSTM neuronových sítí, ale i nahrazuje tyto architektury TDNN sítěmi.

3. Rozsah technické zprávy — je v obvyklém rozmezí

4. Prezentační úroveň předložené práce — 75 b. (C)

Text logicky navazuje, struktura práce je vhodně zvolená.
Text je hutný bez hluchých míst. To se sice v úvodní kapitole projevuje výrazným zjednodušením a některými nepřesnostmi, ale z pohledu celkové práce je to jen drobný nedostatek.

Na duhou stranu kvalitu práce snižuje uvedení obrázků, na které není přímo odkazováno v textu a uvádění vícero označení pro stejnou věc a stejného písmene pro různé věci (i když v různých kapitolách).

5. Formální úprava technické zprávy — 75 b. (C)

Práce je psaná slovensky takže nemohu plně posoudit jazykovou stránku práce, nicméně neshledávám žádné problémy.
Typografická úprava je v pořádku, jen v některých případech je pár řádků textu nad tabulkou případně mezi tabulkami. Toto nepůsobí dobře a někdy je třeba pokračování textu hledat. Tabulky pokud možno dávejte k sobě a na začátek nebo konec stránky.

6. Práce s literaturou — 85 b. (B)

Práce je řádně ocitována a jsou uváděny zdroje převzatých obrázků. citace obsahují řadu současných konferenčních článků, což je důvodem uděleného lepšího hodnocení.

7. Realizační výstup — 85 b. (B)

Práce implementuje maticovou faktorizaci do prostředí pytorch.
Nad rámec zadání je implementována TDNN architektura neuronové sítě a doimplementována podpora faktorizace v LSTM sítích.
Veškeré implementace jsou otestovány a v případě problémů jsou implementovány změny zaručující funkčnost modelů.

8. Využitelnost výsledků

Výsledky jsou využitelné v práci skupiny Speech@FIT případně pro celou komunitu zabývající se rozpoznáváním řeči a zpracováním přirozeného jazyka

9. Otázky k obhajobě

- V pytorch implementaci jste použil jinou topologii TDNN(-F) sítě než je topologie v Kaldi. Můžete porovnat výsledky Kadi a pytorch topologie?
- Pytorch implementace TDNN se dle vašich výsledků trénuje 10x pomaleji než je tomu v Kadi. V čem jsou hlavní důvody tohoto rozdílu?
- Můžete uvést čas potřebný pro natrénování sítě TDNN-F+BLSTMP a BLSTMP (tabulka 6.19)? Krátce uveďte rozdíly mezi oběma topologiemi.

10. Souhrnné hodnocení — 85 b. velmi dobře (B)

Práci hodnotím nadprůměrně z důvodu náročnějšího zadání a kvalitně provedené experimentální části. Testy jsou navíc provedeny na dvou datových sadách, kdy menší je využita pro více testů a hledání vhodné konfigurace a větší slouží k dalšímu otestování nejlepší konfigurace.

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

V Brně dne: 30. června 2020

Grézl František, Ing., Ph.D.
oponent