

Review of Bachelor's Thesis

Student: Pavlus Ján
Title: Learning Speech Separation Using Spatial Cues (id 23153)
Reviewer: Mošner Ladislav, Ing., DCGM FIT BUT

- 1. Assignment complexity** **more demanding assignment**

Zadání považuji za obtížnější, neboť pro jeho řešení je potřeba se seznámit s rekurentními neuronovými sítěmi. Dále je třeba porozumět dvěma přístupům k trénování neuronových sítí pro separaci hlasů řečníků ("deep clustering, DC" a "permutation invariant training, PIT"). Rovněž je potřebné obeznámení s možnostmi, jakými lze získat prostorou informaci z vícekanálových nahrávek pro následné trénování neuronových sítí.
- 2. Completeness of assignment requirements** **assignment fulfilled**

Poznámku mám pouze k 5. bodu zadání. Ten požaduje diskuzi možností zlepšení výsledků. Student pouze v jednom odstavci uvedl přístup založený na míře konfidence, který byl diskutován ve článku, ze kterého BP vychází. Vzhledem k obtížnosti práce k tomuto nedostatku nepřihlížím.
- 3. Length of technical report** **in usual extent**

Výtku mám k poslední kapitole týkající se experimentů. Působí dojmem, že chtěl student dosáhnout vyššího počtu stránek. Kapitola zahrnuje pět obrázků masek a porovnání aplikace masek na spektrum sečtených signálů s originálními spektry. Ty jsou vždy uvedeny podobným popisem v textu bez zaměření se na specifika jednotlivých obrázků. Dále kapitola obsahuje několik neúčelných prázdných míst.
- 4. Presentation level of technical report** **75 p. (C)**

Text považuji za logicky strukturovaný a rozsahy kapitol jsou standardní. Text je obohacen hned o několik pěkných grafů a obrázků, které přidávají na názornosti práce. Z hlediska pochopitelnosti je práce na horší úrovni v důsledku nedůsledného popisu metod. V uváděných vzorcích a definicích často chybí popis symbolů (např. rovnice 2.7, 2.11). Pro tuto práci důležitá metoda DC je popsána problematicky kvůli notaci a zejména popisu objektivní funkce. Stejně tak se vyskytují nepřesnosti v popisu LSTM. Objektivní funkce střední kvadratické odchylky a binární vzájemné entropie jsou definované jako vektory hodnot spočítaných pro jednotlivé trénovací příklady.
- 5. Formal aspects of technical report** **68 p. (D)**

Práce je psána v anglickém jazyce. Často se však objevují gramatické chyby a formulace občas brání porozumění významu. Z typografického hlediska by šlo zlepšit následující nedostatky. Popisky tabulek jsou v práci uváděny až za samotnou tabulkou. Několik obrázků není odkazováno z textu. U řady referencí chybí mezera, která by ji měla oddělovat od okolního textu. V Obrázcích 6.5, 6.6 a 6.13 jsou zaměněné spektrogramy. Popisky os u spektrogramů jsou matoucí, protože zobrazované hodnoty jistě nejsou frekvence v Hz a čas v sekundách.
- 6. Literature usage** **90 p. (A)**

Až na výjimky (jako například citování Neural networks in material science, knihovny PyTorch u definice objektivních funkcí) používá student relevantní a současnou literaturu. Myšlenky čerpané z literatury jsou odlišeny citacemi. Citace [18] je duplikátem citace [17].
- 7. Implementation results** **75 p. (C)**

Student vytvořil kód rozumné kvality pro provádění experimentů pro separaci hlasů řečníků. Využil i převzatého kódu pro výpočet objektivní funkce pro DC a pro práci s modelem cACGMM (využitým pro modelování prostorové informace). Závislost na balíčku pb_bss je uvedena v dokumentaci, ne však v kódu. Ta by měla být doplněna. Dosažené výsledky poukazují na existenci problémů, ale po jejich vyřešení by mohl být kód zajímavý pro širší komunitu.
- 8. Utilizability of results**

Práce staví na článku "Bootstrapping single-channel source separation via unsupervised spatial clustering on stereo mixtures". Ten se zaměřuje (i když ve větších detailech) pouze na DC. Oproti tomu se bakalářská práce zabývá nejen DC, ale i PIT. I když dosažené neintuitivní výsledky nekorelují s výsledky ze článku, programová část i právě díky implementaci PIT může být využita po odladění k dalšímu experimentování v kontextu separace řečníků. Tato oblast výzkumu je aktivní a téma je aktuální.
- 9. Questions for defence**
 - Odůvodněte, proč byly nejhorší výsledky dosaženy pro přístup využívající neuronové sítě trénované s využitím správných masek a nelepší výsledky pro přístup trénování bez učitele, který byl dále použit pro

odhad masek (pro následné trénování nerunové sítě).

- Porovnejte vstupy neuronové sítě při trénování se správnými a odhadovanými maskami. Zdůvodněte rozdíly.

10. Total assessment

75 p. good (C)

Ačkoli výsledky experimentů nedopadly dle očekávání, student vytvořil kód, na kterém lze stavět další experimenty po vyřešení aktuálních problémů. Ukázal, že je schopen získat potřebné informace a natrénovat některé modely. Dle mého názoru není dokumentace na dobré úrovni. Zejména se jedná o jazykovou stránku a dále o nepřesné či neúplné popisy napříč prací.

In Brno 23 June 2020

Mošner Ladislav, Ing.
reviewer