

Review of Master's Thesis

Student: Kocour Martin, Bc.
Title: Automatic Speech Recognition System Continually Improving Based on Subtitled Speech Data (id 22041)
Reviewer: Veselý Karel, Ing., Ph.D., UPGM FIT VUT

- 1. Assignment complexity** **considerably demanding assignment**
Jedná se o výzkumné zadání, které zahrnuje sběr, přípravu a automatické čištění řečových dat s titulky. Vyčištěná data jsou pak použita pro trénování/adaptaci rozpoznávače řeči, který má přesnější výstup díky zvětšené sadě trénovacích dat. Přitom dopředu není jasné, jaký postup bude nejlepší. (Jak čistit data? Jak vybrat data?)
- 2. Completeness of assignment requirements** **assignment fulfilled with enhancements**
Nad rámec zadání jsou provedeny experimenty s výběrem před-čištěných dat podle následujících heuristik: a) podle podobnosti s cílovou doménou rozpoznávání (Frobeniova norma rozdílu kovariančních matic nad signály z před-výstupní vrstvy neuronové sítě), b) přidáním balíku promluv s 'podobným' konfidenčním skóre, kde se autor snaží identifikovat balík, který pomáhá nejvíce. Tato rozšíření přináší zajímavý vhled do řešeného problému.
- 3. Length of technical report** **within minimum requirements**
I když text práce není rozsáhlý, obsahuje vše podstatné pro pochopení experimentů, které byly provedeny.
- 4. Presentation level of technical report** **80 p. (B)**
Logická struktura je velmi přehledná. Na druhou stranu, výsledný text by bylo možné lépe provázat a jednotlivé kapitoly opatřit krátkým shrnutím spojeným se zamyšlením a uvedením výsledků do širšího kontextu. Některé části z teoretického úvodu by bylo dobré lépe vysvětlit 2.2, 2.6. Tyto ovšem nejsou zásadní pro pochopení experimentální části.
- 5. Formal aspects of technical report** **85 p. (B)**
Místo se vyskytují drobné chyby v angličtině. Typograficky je práce zdařilá díky sazbě v LaTeXu.
- 6. Literature usage** **100 p. (A)**
Autor hojně cituje relevantní prameny.
- 7. Implementation results** **100 p. (A)**
Realizačním výstup je přehledný dobře strukturovaný kód, který lze použít jako rozšíření toolkitu Kaldi.
- 8. Utilizability of results**
Práce přináší nové poznatky, které bude možné využít pro další výzkumnou činnost.
- 9. Questions for defence**
Q1: Můžete okomentovat, proč je na obr. 6.3 lepší 'iterativní' trénování, když později v obr. 7.4 je lepší trénování 'od začátku = from scratch'?
Q2: Jakým způsobem by se dalo odhadnout, kolik řečových dat je nejméně potřeba pro 'dostatečně dobrý' odhad kovarianční matice z před-výstupní vrstvy neuronové sítě?
- 10. Total assessment** **90 p. excellent (A)**
I přes drobné nedostatky hodnotím práci velmi kladně. Rozpoznávání řeči je nelehký obor a je vidět, že bylo provedeno velké množství experimentů, které není snadné zpracovat a jednoznačně interpretovat.

In Brno 6. June 2019

Veselý Karel, Ing., Ph.D.
reviewer