

I. IDENTIFIKAČNÍ ÚDAJE

| | |
|-----------------------------------|---|
| Název práce: | Rozpoznávání spojitě řeči s pokročilými strukturami hlubokých neuronových sítí |
| Jméno autora: | Martin Šubert |
| Typ práce: | diplomová |
| Fakulta/ústav: | Fakulta elektrotechnická (FEL) |
| Katedra/ústav: | Katedra radioelektroniky |
| Oponent práce: | Ing. Petr Mizera, Ph.D. |
| Pracoviště oponenta práce: | Omilia Natural Language Solutions Ltd |

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

| | |
|---|-------------------|
| Zadání | náročnější |
| <i>Hodnocení náročnosti zadání závěrečné práce.</i> | |
| Zadání diplomové práce považuji za náročnější, jelikož se student musí seznámit s open source nástrojem KALDI, s principy trénování akustických modelů na bázi hlubokých neuronových sítí a realizací experimentů pomocí výpočetního klastru. | |

| | |
|--|----------------|
| Splnění zadání | splněno |
| <i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i> | |
| Diplomant splnil podmínky zadání v plném rozsahu. | |

| | |
|--|-------------------|
| Zvolený postup řešení | vynikající |
| <i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i> | |
| Student realizoval experimenty dle standardu KALDI nástrojů. Nejprve ověřil funkčnost KALDI nástrojů pomocí dostupných trénovacích skriptů pro anglickou databázi WSJ na interním výpočetním klastru a šikovně adaptoval tyto skripty pro realizaci experimentů. | |

| | |
|--|--------------------|
| Odborná úroveň | A - výborně |
| <i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i> | |
| Práce je logicky členěna. Teoretická část práce popisuje významné architektury akustických modelů na úrovni, která splňuje požadavky na diplomovou práci. Experimentální část práce je vhodně zvolena, analyzuje přesnost různých akustických modelů, rychlost trénování a dekodování. | |

| | |
|---|------------------------|
| Formální a jazyková úroveň, rozsah práce | B - velmi dobře |
| <i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i> | |
| Práce je po formální a typografické stránce v pořádku a rozsahem splňuje požadavky diplomové práce. | |

| | |
|---|--------------------|
| Výběr zdrojů, korektnost citací | A - výborně |
| <i>Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.</i> | |
| Student cituje důležité publikace a velmi oceňuji rozsah provedené rešerše. | |

| |
|--|
| Další komentáře a hodnocení |
| <i>Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.</i> |

Vložte komentář (nepovinné hodnocení).

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Předloženou práci považuji za velmi povedenou. Diplomant se dokázal zorientovat v problematice automatického rozpoznávání řeči, provedenými experimenty prokázal programátorské dovednosti a získal zkušenosti s trénováním akustických modelů na bázi neuronových sítí na výpočetním klastru.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně**.

Otázky ke zodpovězení autorovi při obhajobě práce:

1. V kapitole 5.3.1 "Model processing time" porovnáváte nároky na trénování a dekodování vybraných architektur akustických modelů. "Chain akustické modely" na bázi LF-MMI dosahují výrazně rychlejších trénovacích a dekodovacích časů v porovnání s AM trénovanými pomocí cross entropy loss. V tabulce 5.10 uvádíte, že TDNN chain model je výrazně pomalejší. Jaké nastavení hyper parametrů neuronové sítě jste při trénování použil a jaký hardware byl pro trénování použit?
2. Real-time Factor (RTF) se běžně používá pro uvádění rychlosti dekodování. Jakých RTF hodnot dosahují rozpoznávače s analyzovanými akustickými modely v diplomové práci?
3. Kapitola 4.1.5 "Chain Models" shrnuje důležité vlastnosti tohoto trénovacího postupu. V textu jsem nenašel informaci o použitých fonetických jednotkách pro chain model. Jaké fonetické jednotky byly použity a kolik sloučených stavů bylo modelováno?

Datum: 31.5.2021

Podpis: