

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Vladislav Stankov
Název práce Speech-Informed Inverse Text Normalization
Rok odevzdání 2024
Studijní program Informatika **Studijní obor** Umělá inteligence

Autor posudku Ondřej Plátek **Role** Oponent
Pracoviště Ústav formální a aplikované lingvistiky

Text posudku:

Obsah práce

Práce se zabývá rozšířením úlohy “Inverse Text Normalization” (ITN), tedy normalizací textového výstupu rozpoznávače řeči. Autor navrhuje nový model pro kombinování signálu z řeči a textu na vstupu neuronového modelu pro lepší predikci diakritiky a kapitalizaci počátečních písmen slov.

Kromě úvodu a závěru autor rozdělil práci na kapitoly “Background” (Přehled), “Data”, “Training Speech Recognition Model” a “Experiments”.

V úvodu motivuje důležitou, ale ne tak známou úlohu ITN a vymezuje rozsah práce. Kapitola Background popisuje jednotlivé techniky a architektury trénování neuronových sítí. Kapitola Data detailně popisuje náročnou práci přípravy vhodných dat pro ITN z podkladových dat korpusu ParCzech 3.0, který byl připraven pomocí nového rozpoznávače řeči a metod pro časové zarovnání přepisu s mluvenou řečí. Kapitola Experiments je strukturovaná do Sekci, kde každá sekce se věnuje detailům jednotlivých rozšíření neuronových modelů pro ITN.

Práce je rozsáhlá (92 stran) a psaná anglicky. Trénovací skripty jsou zveřejněny na GitHubu. Nepodařilo se mi dohledat data nově připraveného korpusu na veřejně přístupném místě.

Autor reportuje signifikantní vylepšení dle mého názoru velmi silného modelu ITN pro doplnění nejčastějších interpunkce a kapitalizace.

Hodnocení práce

Autor prokázal dobré porozumění problematice moderních systému ITN. V práci popisuje holistický přístup k problému, kdy se věnoval jak přípravě dat, zvolení vhodných metrik, implementaci netriviálních rozšíření Transformer modelu a provádí řadu zajímavých experimentů, které rigorózně vyhodnocuje. Kvalitativní analýza jednotlivých příkladů jednak doplňuje kvantitativní výsledky a zároveň přibližuje čtenáři data a úlohu, která byla v práci řešena.

Na druhou stranu text práce nemá vždy dobrou strukturu. Text je zbytečně zdouhavý a zejména v kapitole “Background” chybí motivace a lepší provázání jednotlivých sekcí a subsekcí. Autor také mění míru detailu. Nicméně těžiště práce je v kapitole Experiments,

kteřá je psaná věcně a je doplněná o dobře zvolené příklady a diagramy.

Otázky:

1. Víte o vhodném datasetu pro evaluaci text ITN a speech+text ITN?
2. Proč jste Váš systém neporovnal s existujícími systémy na na již existujícím datasetu nebo nezkusili aplikovat modely dobře fungující na jiných datasetech než na Vašem rozšířeném ParCzech 3.0 datasetu?
3. Zveřejníte Vámi připravený rozšířený dataset ParCzech 3.0?

Práce po všech stranách splňuje požadavky kladené na diplomovou práci, a proto ji **doporučuji k obhajobě.**

Práci nenavrhuji na zvláštní ocenění.

Datum 15.5.2024 v Praze

Podpis