

Posudek oponenta diplomové práce

Autor práce: **Bc. Lukáš VLADAŘ**

Název práce: **Analýza velikosti dat pro neurální syntézu řeči**

Jazyková a grafická úprava

Nadprůměrné

Formální a obsahová stránka práce

Průměrné

Vhodnost použitých metod

Průměrné

Způsob zpracování a vyhodnocení

Nadprůměrné

Správnost získaných výsledků

Průměrné

Vlastní přínos

Nadprůměrné

Doplnění hodnocení, připomínky:

Diplomová práce se zabývá analýzou vlivu množství trénovacích dat a inicializace trénování modelu na kvalitu syntetické řeči, kdy inicializací je myšleno buďto trénování modelu "od nuly" (náhodná inicializace) nebo "fine tuning" na nějak předtrénovaném modelu (inicializace modelem jednoho řečníka nebo multi-speaker modelem). Pro vlastní hodnocení kvality je pak použit buďto MUSHRA poslechový test nebo objektivní MCD míra.

Před diplomantem stál nelehký úkol vyhodnotit velké množství experimentů s mnohdy protichůdným či obtížně interpretovatelným výsledkem, byť určité očekávané tendence se potvrdily. Z výsledků je evidentní, že některé z experimentů skončily velmi jasným výsledkem, který je ale v rozporu s očekáváním - např. zvýšení kvality u nižšího množství trénovacích dat (1.5 hodiny u obr. 6.2 nebo 45 minut u obr. 6.3). I přes snahu diplomanta shrnout výsledky a dát určitá doporučení (být' na základě velmi omezené množiny hlasů), mi u těchto případů chybí nějaká další analýza výsledků z experimentů s neočekávaným nebo protichůdným výsledkem.

Dotazy

Dělal jste si nějakou podrobnější analýzu u experimentů s neočekávaným výsledkem, kdy např. model trénovaný na větším množství dat získal nižší MUSHRA hodnocení? Např. použití 6 vs. 1.5 hodiny dat v obr. 6.2 nebo 3 vs. 1.5 u obr. 6.3, případně kolísání kvality v obrázcích 6.6/6.7? Něco, co se již do práce nedostalo nebo nebylo možné jednoduše vysvětlit? Například poslechem promptů s cílem identifikace příčiny nízkého hodnocení (výskyt nějakého nepřírozeného jevu?). Jsou výsledku MUSHRA testů v souladu s Vaším subjektivním hodnocením v těchto případech?

Mohla by být nižší kvalita modelu neprofesionálního řečníka "fine-tuněná" na modelu profesionálního řečníka způsobena odlišností charakteru použitých hlasů? Fine-tuning na multi-speaker modelu již dosahuje obecně lepších výsledků. To je nicméně v rozporu s profesionálním řečníkem (i multi-speaker model byl z profesionálů), kde jsou ale výsledky poměrně protichůdné

Splnění bodů zadání

úplně

Doporučení k obhajobě

ANO

Hodnocení: 1 - Výborně

V _____ dne _____

Ing. Daniel Tihelka, Ph.D.