Posudek oponenta k disertační práci

Autor práce: Ing. Kateřina Žmolíková

Název práce: Neurální extrakce řeči cílového řečníka

Předložená disertační práce se zabývá zpracováním záznamů řeči za účelem potlačení interferencí (konkurenčních hlasů a jiných směrových zdrojů zvuku) a ostatního ruchu. Metodikou je trénování hlubokých neuronový sítí (DNN) s využitím velkého množství příkladů. Hlavním tématem je takzvaná extrakce řečníka, kdy cílem je získání čistého signálu řeči konkrétní osoby. To obnáší řešení problémů způsobených neurčitostí úlohy především pak určením, která ze složek pozorované směsi signálů je ta cílová. Metodika se primárně zaměřuje na zpracování signálu z jednoho mikrofonu a rozšiřuje se na použití více mikrofonů. Problém neurčitosti je řešen pomocí identifikačních znaků cílového řečníka, které jsou získány předem z krátké a nezarušené promluvy tohoto řečníka. Téma je velmi aktuální. Rozvoj metodiky DNN nabízí v oblasti nový potenciál a slibné praktické výsledky. Jedná se experimentální přístup k řešení, ve kterém je prostor pro rozličné nápady na zlepšení a řešení dílčích problémů.

Práce obsahuje tři úvodní kapitoly, kde jsou úlohy separace a extrakce definovány, jsou popsány základní principy řešení pomocí neuronových sítí, další příbuzné úlohy jako je například dereverberace a celkově je shrnut současný stav poznání v této oblasti. Následují tři kapitoly věnující se zpracování signálu z jediného mikrofonu, z více mikrofonů a aplikacím metodiky za účelem zlepšení automatického rozpoznávání řeči a diarizace. Sedmá kapitola obsahuje závěr.

Z hlediska formálního má práce velmi vysokou úroveň. Je napsána plynulou a bezchybnou angličtinou. Text je konzistentní, skvěle organizovaný do kapitol a podkapitol a lehce srozumitelný. Přehled literatury je velmi bohatý, relevantní ke všem možným detailům a prokazuje tak perfektní seznámení autorky se stavem poznání. Typografické úpravě a kvalitě obrázků a tabulek není co vytknout.

Přínos práce ke stavu problematiky je velmi významný, což dokládá související publikační aktivita autorky a ohlas komunity v množství citací. Publikace byly prezentovány na nejvýznamnějších konferencích v oboru: ICASSP, INTERSPEECH a ASRU. Nejvýznamnější publikace vyšla

v prestižním časopise IEEE Journal on Selected Topics in Signal Processing, což je výsledek, ke kterému autorce srdečně blahopřeji. Zmíněné výsledky a publikace jsou prací několika autorů. Autorka je prvním autorem v pěti případech ze sedmi. V práci sice postrádám explicitní popis vlastního podílu autorky na výsledcích, avšak vzhledem ke zvyklostem v oboru (pořadí autorů na publikacích) nepochybuji, že **podíl autorky je dostatečně významný a naplňuje obecné požadavky na přínos a originalitu disertační práce**. Přesto doporučuji u obhajoby podíl autorky upřesnit.

K technickým nápadům, řešením a způsobu vědecké práce **nemám připomínky, které by smysl a kvalitu práce zpochybňovaly**. Oceňuji navržené způsoby řešení jako velmi zajímavé a relevantní nápady, což ostatně dokládají výsledky. Oceňuji vysokou kvalitu zpracování experimentů, které jsou u použité metodiky zásadní, porovnání s relevantními a konkurenčními metodami či alternativami vlastních řešení, vyhodnocení relevantními kritérii a kritické zhodnocení výsledků. V práci trochu postrádám srovnání s metodami, které jsou odvozeny bez učení na základě matematických modelů (např. NMF, ICA, SCA) a které metodám založených na trénování neuronových sítí v mnoha ohledech konkurují. Uznávám ale fakt, že nasazení těchto metod pro objektivní srovnání s neurální extrakcí není triviální a samotné srovnání představuje problém, který lze v současné době považovat za nevyřešený (úlohou je srovnat trénované a netrénované metody). Ostatní připomínky a dotazy k některým detailům uvádím v příloze posudku.

Závěrem konstatuji, že **práce splňuje nároky na udělení akademického titulu Ph.D. a doporučuji ji k obhajobě**.

V Liberci dne 3. června 2022

prof. Ing. Zbyněk Koldovský, Ph.D.

Ostatní připomínky k práci a dotazy k obhajobě:

- V kapitole 4.3 je popis způsobů jak ovlivnit neuronovou síť tak, aby extrahovala vybraného řečníka. Je důležitá, protože se týká významné části celkového přínosu práce. Ačkoliv se autorka snaží o zavedení značení a detailní popis postupu, není tento popis úplný, spoléhá na informace a detaily, které v práci nejsou konkretizovány. Brání to v pochopení podstaty těchto řešení a neumožňuje práci reprodukovat.
- Obrázek 4.11 na straně 54 demonstruje bez mála perfektní diskriminaci 18 mluvčích zobrazením příznaků ve 2D. Vkrádá se otázka, jak je to možné, že to takto vyšlo? Vždyť přidáním dalšího mluvčího už by se nový příznakový vektor musel nutně krýt s jiným mluvčím. Byla množina mluvčích vybrána záměrně tak, aby mezi nimi byla dostatečná diverzita?
- Na straně 64 popisujete metodu, která je trénována optimalizací kritéria SNR. SNR počítané pro všechny frekvence nebo pro celou časovou oblast by ale těžko bylo relevantním kritériem (například lze dosáhnout velkého SNR, pokud je signál extrahovaný jen ve velmi úzkém pásmu, kde není šum a v ostatních pásmech je výstup nulový). Jak je zde kritérium SNR konkrétně definováno, aby se dosáhlo kvalitního SDR?
- Tabulka 5.2 ukazuje, že MVDR je ve smyslu kritéria SDR lepší než MWF. MWF ale v principu optimalizuje kvadratické kritériu obsažené v definici SDR. Tedy dává signál, který se co nejvíce podobá (ve smyslu kvadratické vzdálenosti) cílovému signálu a za určitých okolností musí MWF dávat SDR, které již nelze překonat. Jak tedy tento rozpor vysvětlujete?

Prof. Ing. Zbyněk Koldovský, Ph.D. | proděkan pro VaV a DS| +420 732 700 585 | zbynek.koldovsky@tul.cz
Technická univerzita v Liberci | Fakulta mechatroniky, informatiky a mezioborových studií | ITE
Studentská 1402/2, 461 17 Liberec 1 | www.fm.tul.cz