POSUDEK NA DISERTAČNÍ PRÁCI

Téma práce: Iterativní zdokonalování přepisu zvukových nahrávek s využitím zpětné vazby posluchačů

Doktorand: Mgr. Jan Oldřich Krůza

Posudek vypracoval: Doc. Ing. Petr POLLÁK, CSc.
ČVUT FEL K13131, Technická 2, 166 27 Praha 6

Předložená práce Mgr. Krůzy zpracovává interdisciplinární problematiku přepisu existujících zvukových nahrávek hovorů českého filozofa, ing. Karla Makoně. Jedná se o přepracovanou práci, ve které se autor snaží vypořádat s připomínkami v posudcích obou oponentů i s výhradami členů komise, které zazněly při neúspěšné obhajobě v září 2020.

Jak již bylo zmíněno z mé strany při prvním odevzdání, řešená úloha zpracování off-line audioarchivů představuje v současné době často řešenou problematiku pro různé světové jazyky včetně češtiny. Vždy se však jedná o výzkum vázaný na konkrétní nahrávky, kdy je nutné řešit mnoho různých problémů souvisejících zejména s konkrétním obsahem a různou kvalitou zpracovávaných záznamů, což přináší prostor pro disertabilní výzkum. To je případ i předložené práce, jejíž cíle stanovené v úvodní kapitole proto považuji i nadále za disertabilní.

Lze konstatovat, že své stanovené cíle autor v předložené práci naplnil a v přepracované verzi odstranil největší diskutované problémy. S některými připomínkami se nebylo možné zcela vypořádat, neboť není možné během jednoho roku práci kompletně přepracovat od počátku. Na druhou stranu je však trochu škoda, že autor nevyužil více možnosti, které se mu v čase k přepracování nabízely. To je zejména případ připomínek k metodám kompenzace horší akustické kvality zpracovávaných nahrávek, ke kterým se autor v průběhu minulé obhajoby v vyjádřil, v přepracovaném textu však závěry této diskuse nejsou zmíněny.

V následujících bodech je shrnutí nejvýznamnějších přínosů předložené práce a připomínek z mé strany.

- V první části práce autor shrnuje život Karla Makoně a obsah jeho díla, jímž je evidentně významně osloven. Zpřístupnění tohoto díla formou přepisů vnímám jako zajímavé z hlediska obecné filosofie, teologie resp. religionistiky, což lze považovat za jeden z autorem zmíněných interdisciplinárních přínosů dané práce. Celkový význam Makoňova díla však neposuzuji, to je na odbornících ze zmíněných filosofických oborů.

  Dále autor realizuje v kapitole 2.2 rozbor resp. analýzu obsahu Makoňova díla, včetně automatické identifikace témat textovým vyhledáváním. Realizovaný popis a přehled jednotlivých témat je logickým prvním krokem, významným pro následné využití při anotaci nahrávek v korpusu resp. možnosti přizpůsobení automatického přepisu danému tématu. Na základě předchozích připomínek byl seznam mírně zestručněn. Využití těchto kategorií lze nalézt v rozšířené kapitole 7 popisující vyhledávání v realizovaných přepisech.
- Důležitým úkolem práce byla příprava používaných dat, v první řadě nahrávek Makoňova korpusu, kde se autor musel vypořádat s poměrně složitým problémem, a to nižší kvalitou zpracovávaných audio nahrávek různého stáří. Autor analyzuje akustické vlastnosti záznamů, v první řadě na časových průbězích a spektrogramech, které však přinášejí jen velmi hrubě ilustrativní informaci.

Vhodnějším způsobem analýzy akustické kvality signálů je použití objektivní metriky na bázi Kullback-Leiblerovy divergence, kterou autor převzal a kterou definuje pojmem "akustická vzdálenost". Její funkčnost potvrzují realizovaná srovnání napočítané vzdálenosti mezi více a méně zkreslenými nahrávkami.

Následně autor realizuje hierarchické shlukování signálů, výsledek provedeného shlukování byl na základě předchozích připomínek doplněn, je však zmíněn trochu nevhodně. Autor uvádí obrázek 3.10 "Velikosti clusterů během hierarchickkého shlukování", který je nedostatečně popsaný a bez jakéhokoliv dalšího komentáře v textu samotném. Očekávaným výsledkem by měla být informace, zda shlukování vedlo k rozlišení popisovaných zkreslení či jaká je akustická vzdálenost mezi záznamy v jednotlivých shlucích.

- Velmi zásadním úkolem byla kompletace trénovacích dat, neboť nahrávky Makoňova korpusu mohou být použity jen pro případné rozšíření standardní trénovací množiny. Autor shromáždil data z několika veřejných zdrojů, kde jádrem jsou zejména přepisy jednání poslanecké sněmovny. Na tvorbě toho korpusu se autor podílel, což lze považovat za významný vedlejší přínos jeho vědeckých aktivit.

  V kap. 5.6 je popsaná tvorba jazykového modelu (LM), který je další důležitou součástí systému významně ovlivňující přesnost přepisu. Autor kompletuje zdrojový korpus pro trénování LM selektivním výběrem ze dvou zdrojů, a to z Makoňových spisů a z korpusu WMT, obsahujícího obecné české texty. Ten je ale popsán velmi nedostatečně jen stručným odkazem. Pro vlastní tvorbu LM používá autor volně dostupný nástroj KenLM. Dosažené výsledky s různými nastaveními parametrů výběru textů z obou korpusů jsou popsány v doplněné tab. 5.3 resp. na obr. 5.10.

- Jádrem práce je realizace automatického přepisu daného korpusu. Autor vycházel z mnoha prací popisujících přepis audio resp. multimediálních archivů v širším mezinárodním měřítku, nejvýznamnější práce jsou citovány. V přepracované verzi autor doplnil popis používaných systémů na bázi GMM-HMM a DNN-HMM včetně stručného srovnání aktuálně dosahovaných výsledků různými autory. Některé části v této kapitole jsou nyní možná až příliš podrobné (např. detailní popis vzorkování či krátkodobé segmentace), to však považuji za menší problém, než v minulé práci chybějící alespoň stručný popis používaných systémů rozpoznávání řeči.

  Autor používá pro implementaci rozpoznávače na bázi GMM-HMM nástroje HTK v poměrně standardní konfiguraci pro trénování modelů a dekodér Julius pro realizaci přepisů. Rozpoznávání na bázi DNN je realizováno s volně dostupným systémem DeepSpeech a autorův přínos je v použití manuálně přepsaných dat Makoňova korpusu, která jsou doplněna o další shromážděná trénovací data z veřejně dostupných zdrojů.

  Již při první obhajobě byla diskutována nešťastná volba nástrojů pro rozpoznávání, což v přepracované verzi práce nebylo možné zcela změnit. Použití GMM-HMM architektury má své logické důvody zejména v prvních krocích, autor zmiňuje snažší implementaci, možnost natrénování s menším množstvím dat, použití pro fonetickou segmentaci. V přepracované verzi práce autor rozšířil využití systému DeepSpeech. Přepisy pomocí DNN systému dosáhly dle očekávání nejlepších výsledků na zvolené testovací množině.

- Kompenzaci akustických nedostatků dat v dostupných záznamech se autor pokusil řešit různými metodami. V první řadě na úrovni samotného signálu pomocí spektrálního odečítání s obecně používaným nástrojem "sox", kdy z vybraných úseků získá vzorek pozadí, které se následně ve spektrální oblasti odečítá. Dosažené výsledky jsou dle oče-

kávání nepříslíš uspokojivé. Druhou zvolenou metodou transformace signálů je použití neuronové sítě typu CycleGAN, což autor vyzkoušel na dvou kategoriích dat, bohužel opět bez výsledků umožňujících praktické použití.

Na úrovni příznaků se autor pokoušel implementovat kepstrální normalizaci a aktivního učení s výběrem dat na bázi míry spolehlivosti. Obě metody však ke zvýšení přesnosti přepisu nevedly, což je vzhledem k charakteru zkreslení resp. variabilitě akustických podmínek očekávatelný výsledek a pro realizaci přepisů nebyly nakonec použité.

Zmíněné problémy byly diskutovány v průběhu předchozí obhajoby, autor toto v přepracované verzi práce však nezohlednil a tyto části zůstaly beze změny. V tabulce 3.1 je stejný výsledek jako v předchozí verzi práce s problematickou vypovídací schopností. Podobné je to u ostatních popisovaných metod. Je škoda, že v přepracované práci chybí alespoň zpětná vazba na tuto proběhlou diskusi.

- Hlavní přínos práce vidím ve skutečnosti, že autor se k automatickému přepisu nahrávek Makoňova korpusu v akceptovatelné kvalitě dopracoval. Celkově byl automatický přepis vytvořen u více než 1000 hodin záznamů, u téměř 110 hodin záznamů byly přepisy ručně zkorigovány. To považuji za dobrý výsledek.

  Klíčovým krokem k relativně úspěšnému přepisu je iterativní vylepšování použitých systémů na základě korigovaných přepisů. Samotná tato metodika zcela neznámá není, avšak její konkrétní aplikace na přepis zvoleného korpusu zahrnující mnoho formální a organizační práce související s manuální korekcí automatických přepisů Makoňova korpusu představuje významné kroky, bez kterých by nebylo možné se k dosaženému výsledku dopracovat a které mohou být aplikovatelné v podobných projektech.

  Dosažená přesnost přepisů je dobrá, autor nyní uvádí nejlepší WER 10,3 % pro DNN systém s pentagramovým LM na standardní testovací množině. V přepracované práci byl doplněn přehled zpřesňování přepisů na základě rostoucí množiny trénovacích dat z Makoňova korpusu resp. pro systémy s různým nastavením. Nejlepší výsledek s GMM systémem byl 46,3 % WER na standardní testovací množině, což potvrzuje použitelnost tohoto systému jen pro počáteční fázi přepisů případně pro fonetickou segmentaci.

  V tab. 5.5 jsou uvedeny výsledky na dalších testovacích sadách a různých variantách trénování, dosažená chybovost na sadě obsahující i zkreslené záznamy byla 22,2 %, což je spíše dobrý výsledek. Výsledky pro GMM systém a zkreslené záznamy chybí, zde lze však očekávat vysokou chybovost. Chybí mi výsledky pro variantu DNN systému trénovaného jen na obecných datech, neboť zahrnutí dat z Makoňova korpusu by pak v principu ukazovalo možnost přizpůsobení DNN systému těmto záznamům. Výsledky v tab. 4.1 jsou pravděpodobně dosažené s jinými testovacími množinami.

- Jednoznačným přínosem předložené práce je také vytvořené Webové rozhraní s dostupným přepisem Makoňova korpusu, možnostmi fulltextového vyhledávání a editace přepisů. Autor se inspiruje jinými zpřístupněnými korpusy a programy pro anotaci signálů. S dostupnými WEBovými technologiemi pak vytváří finální aplikaci se zaměřením na efektivní ruční korekci aktuálně dostupných přepisů, vyhodnocení přesnosti realizovaných ručních přepisů či efektivní dostupnost odpovídajících audiosouborů. Detaily implementace neposuzuji, z uživatelského pohledu je na základě osobního vyzkoušení aplikace funkční a použití velmi srozumitelné. Při ilustrativním poslechu několika úseků ve zpřistupněném korpusu jsem v poslechnutých částech zaregistroval spíše jen menší změnu smyslu sdělení, což odpovídá dosahované relativně nízké chybovosti na testovací množině.

Publikační výstupy autora jsou slabší, publikoval 6 prací související s tématem disertace, kde významnější jsou 2 publikace na konferenci Text, Speech, and Dialogue (zaindexované ve Web of Science). Oproti stavu v minulém roce přibyly 2 publikace popisující novou verzi korpusu ParCzech, což hodnotím pozitivně. Vzhledem ke skutečnosti, že všechny publikace byly prezentovány na mezinárodních fórech (byť některé byly spíše lokálního významu), lze tyto výstupy akceptovat jako potvrzení originálního přínosu práce autora.

Po formální stránce je zpracování práce průměrné. Autor v přepracované verzi nepatrně změnil pořadí kapitol k lepšímu a logičtějším sledu popisu realizovaného výzkumu. Na druhou stranu řadu formálních prohřešků zmíněných v dřívějším posudku autor neodstranil, např. časté formální chyby v citacích či opravu některých dříve zmíněných velmi neobvyklých formulací. Ještě bych přidal výhradu k velké nejednotnosti obrázků, a to jak ve velikosti a typu fontů, tak ve velikosti samotné, viz především přidané obr. 5.1, 5.2, 5.3, 5.5, apod. Uvedené formální prohřešky snižují celkový dojem z práce a zejména v případě přepracované práce si mi to zdá být opravdu škoda.

Přes zmíněné výhrady považuji předloženou práci za disertabilní, a to zejména díky její šíři a interdisciplinaritě. Jednoznačně pozitivním výsledkem je, že autor dokázal realizovat přepisy s dobrou přesností a dosáhnout konkrétního a hmatatelného výsledku v podobě zpřístupnění Makoňova korpusu ve WEBové aplikaci. Od minulé obhajoby množství ověřených zpřístupněných přepisů vzrostlo, lze tedy předpokládat, že projekt přepisu Makoňova korpusu je živý a že dosažené výstupy budou nadále využívány.

Na základě všech výše uvedených skutečností práci **doporučuji** k obhajobě za účelem získání vědecké hodnosti doktora na Matematicko-fyzikální fakultě Karlovy Univerzity.

V diskusi bych se ještě zeptal:

- Budou práce na zpracování Makoňova korpusu pokračovat i nadále?
- Pokud ano, předpokládáte pro realizaci přepisů možnost použití i jiné implemetace použitého rozpoznávače?

V Praze dne 8. září 2021