

Oponentský posudek doktorské disertační práce Ing. Michala Fapša

VYHLEDÁVÁNÍ VÝRAZŮ V ŘEČI POMOCÍ MLUVENÝCH PŘÍKLADŮ

Posuzovaná disertační práce Ing. Michala Fapša se zabývá vyhledáváním výrazů v řeči pomocí mluvených příkladů (angl. query-by-example spoken term detection, QbE STD). Práce celkem čítá 92 stran včetně dvou příloh a je psána v anglickém jazyce. Autor v úvodu stručně uvádí motivaci práce a jasně formuluje cíle práce. V dalších kapitolách pak podává přehled a rozbor metod a evaluačních metrik používaných v oblasti systémů QbE STD a dále popisuje nastavení použité pro provedené experimenty, včetně akustických dat a jejich zpracování. Jádrem práce je kapitola 7, ve které autor popisuje vývoj vlastního systému QbE STD založeného na srovnávání grafů hypotéz pomocí konečných automatů (WFST). V závěrečném shrnutí disertant popisuje dosažené výsledky, kterých dosáhl pro čtyři velmi odlišné jazyky (český, anglický, maďarský a arabský), vždy ve dvou scénářích – jazykově závislém a jazykově nezávislém, podle toho, zda byl fonémový rozpoznávač natrénován na stejném jazyce, pro který bylo provedeno vyhodnocení. Výsledky navrženého systému porovnává s jinými existujícími přístupy: sekvenčním statistickým modelováním (GMM/HMM) a srovnáváním vzorů v příznacích (DTW) a rovněž s běžným systémem STD vyhledávajícím textové výrazy, akustickým detektorem klíčových slov (AKWS) a navrhovaným přístupem WFST s referenční „slovníkovou“ výslovností (WSFTdict). Výsledky ukazují, že v jazykově závislých experimentech je navrhovaný QbE STD systém lepší než systém využívající GMM/HMM, srovnatelný se zavedeným přístupem DTW a jen o něco málo horší než referenční systém AKWS. V jazykově nezávislých experimentech byla úspěšnost jednotlivých systémů odlišná v závislosti na počtu použitých příkladů. V závěru autor naznačuje, že úspěšnost finálního systému QbE STD by mohla být dále zvýšena vhodnou kombinací jednotlivých dílčích systémů detekce.

Z popisu navržené metodiky a postupu řešení je zřejmé, že se jedná o velmi rozsáhlou a komplexní problematiku, kdy bylo zapotřebí provést velké množství práce, analýz a výpočtů. V neposlední řadě to znamenalo rovněž implementovat velké množství metod a provést řadu experimentů. To vše jako celek přesahuje možnosti jednoho člověka a rámec jediné disertační práce.

Ve smyslu požadavků na kladených na disertační práci mohu konstatovat:

- Námět práce odpovídá oboru disertace a je aktuální z hlediska současného stavu oboru.
- Práce vykazuje původní přínosné části. Originální přínos je zejména v implementaci metody pro QbE STD založené na konečných automatech (kapitola 7). Základní inspirace pro tento postup sice pochází z práce „*Query-by-example spoken term detection for OOV terms*“ (Parada et al., ASRU 2009), ale implementace popsaná v předložené disertační práci obsahuje několik netriviálních modifikací již známého základního algoritmu. V první řadě byla metoda rozšířena pro práci s tzv. „confusion networks“ (CN) – původní metoda pracovala pouze s mřížkami (tato varianta je v disertantově implementaci také zachována). Dále disertant na základě experimentálního a teoretického rozboru úlohy navrhl několik inovativních postupů pro kombinování jednotlivých příkladů („vzorů“ pro vyhledávání) a zacházení s překrývajícími se výskyty hledaných slov, což spolu s výše uvedenou podporou CN vedlo k podstatnému zlepšení detekční schopnosti systému. Disertant dále uvádí jako jeden ze svých hlavních přínosů zavedení nové evaluační metriky unppFOM (kapitola 2.3.4.3). Tuto míru kvality systému bych však považoval spíše za zajímavý teoretický koncept než skutečně prakticky použitelnou metriku, protože předpokládá ideální kalibraci detekčního prahu pro každou jednotlivou promluvu, aniž by autor zároveň nabízel metodu, jak něčeho takového dosáhnout (kromě ručního nastavení). Navíc tato metoda (ze své podstaty) uměle snižuje

množství falešných poplachů (což disertant v práci otevřeně přiznává). Nicméně, jak bylo uvedeno výše, originálních výsledků je v práci i tak dostatek.

- Z rozboru vědecké činnosti disertanta vyplývá, že jádro disertační práce bylo publikováno na dostatečné úrovni – v impaktovaném časopise *ACM Transactions on Information Systems* (disertant uveden jako druhý autor), problematika související s řešenou disertační prací byla publikována ve sbornících několika prestižních mezinárodních konferencí a workshopů (např. ACM Multimedia International Conference nebo IEEE Spoken Language Technology Workshop).
- Seznam vědecké činnosti uchazeče je velmi obsáhlý – publikační činnost zahrnuje 31 odborných článků (z toho jednu publikaci v již zmíněném impaktovaném časopise a více než deset publikací ve sbornících prestižních mezinárodních konferencí a workshopů), což jednoznačně dokládá vědeckou erudici Ing. Fapša i jeho schopnost informovat o dosažených výsledcích vědeckou komunitu. K výčtu publikací mám však následující připomínky:
 - Až na výjimky je Ing. Fapšo uváděn jako spoluautor; postrádám významnější publikaci, u níž by byl disertant hlavním autorem.
 - Poslední udávaná publikace je z roku 2012. Proč nebyl v posledních dvou letech, kdy bylo podle všeho tvořeno jádro disertační práce, publikován žádný výsledek?
- Z formálního pohledu je práce na solidní úrovni, a to jak po grafické, tak po lexikální stránce. Vyniká velkým množstvím přehledných obrázků a tabulek a je psána velmi dobrou angličtinou. Přesto lze v práci najít několik formálních chyb, které naznačují, že finalizaci práce mohla být věnována větší pečlivost:
 - Seznam použitých zkratk nepokrývá všechny zkratky (např. ROC, ATWV, MTWW nebo UBTWW).
 - Některé matematické operátory jsou chybně psány kurzívou (např. max a argmax na straně 19 elektronické verze disertační práce*).
 - Na straně 55 je nesoulad mezi obr. 7.1, v němž je příklad slova CAT, a textem, který na obrázek odkazuje, v němž je příklad slova HELLO.
 - V obrázcích 7.1 a 7.3 je ve slově CAT místo fonému /k/ uveden foném /c/.
 - Text na str. 63 odkazuje na obr. 6.5. Podle mého názoru by však měl odkazovat na obr. 7.8.
 - Popisek obr. 7.13 obsahuje nadbytečné slovo WFSTdict.
 - Popisek obr. 8.3 neodpovídá obsahu obrázku.
- Práce je dobře strukturovaná a logicky členěná, což napomáhá dobré čitelnosti a srozumitelnosti jinak hutného odborného textu. Snad jen kapitoly 4 až 6 o celkovém rozsahu 14 stran, které tvoří vlastní jádro disertační práce, mohly být součástí jediné kapitoly, popř. přehledu metod QbE STD.
- Citační odkazy jsou používány správně, relativně rozsáhlý výběr použitých zdrojů literatury svědčí o autorově dobré orientaci v řešené problematice.

Připomínky a dotazy do diskuse v průběhu obhajoby:

- 1) V kapitolách 3 až 6 je uvedeno, že velká část těchto kapitol byla napsána někým jiným než disertantem, např. v kap. 5 na str. 39 je v originálním znění: „*Major part of this chapter was written by Igor and Javier ...*“. Podobně je zde uvedeno, že některé prezentované experimenty byly provedeny někým jiným než disertantem. Protože disertační práce má být odborná monografie vyjadřující originální tvůrčí poznatky disertanta, považuji uvedená vyjádření za nešťastná. Věřím, že vhodnějšími slovy, které chtěl disertant použít, jsou vyjádření ve smyslu „*Some portions of this chapter were adopted from ...*“ a že vlastní texty práce byly napsány samotným disertantem. Přesto požaduji, aby disertant během obhajoby vysvětlil, které texty disertační práce jsou jeho vlastními a které experimenty jsou výsledkem jeho vlastní tvůrčí činnosti.

* Stránkování tištěné a elektronické verze disertační práce se poněkud liší. Čísla stránek, na které v posudku odkazují, odpovídají elektronické verzi.

- 2) V práci bohužel není podrobněji rozebrán algoritmus pro generování mřížek z fonémových posteriorgramů. Na str. 52 je pouze uvedeno, že byl použit dekodér HVite z HTK toolkitu. Vzhledem k tomu, že je následně v kapitole 7.6.1 poměrně obsírně (ale pouze slovně) rozebírána ztráta informace vzniklá při převodu z posteriorgramu na mřížku, byla by detailnější znalost mechanismu tohoto převodu velmi přínosná.
- 3) Výsledky ukazují, že ve většině případů vede použití pěti příkladů v porovnání s jedním příkladem ke zvýšení úspěšnosti prezentovaných systémů. Co bylo důvodem pro volbu právě pěti příkladů – kromě toho, že se stejný počet vyskytuje i v článku [Parada et al. 09]? Předpokládám, že použití více příkladů je (minimálně z uživatelského hlediska) méně komfortní. Byly provedeny experimenty i s jiným počtem příkladů? Nevykazuje např. použití 2-3 příkladů podobnou úspěšnost jako použití pěti příkladů?
- 4) Omezení na jednoslovné dotazy s minimální délkou 4 fonémy je opět (zřejmě) inspirováno článkem [Parada et al. 09]. To není v zásadě problém, jen jsem se chtěl zeptat, zda disertant nepovažuje tuto restrikcí za příliš omezující pro praktické nasazení?

Závěr:

Závěrem mohu konstatovat, že cíle disertační práce byly beze zbytku splněny. Disertant prokázal schopnost se dobře orientovat ve stavu řešené problematiky a pracovat tvůrčím způsobem. Za největší a původní přínos považuji představení a implementaci zásadním způsobem vylepšené metody pro QbE vyhledávání v mřížkách a zejména v „confusion networks“ a také rozsáhlou a dobře zdokumentovanou experimentální část srovnávající tuto novou metodu se špičkovými existujícími technikami. Práci proto **doporučuji k obhajobě**, v průběhu obhajoby však žádám o vysvětlení zásadní připomínky 1).

V Plzni, 24. listopadu 2014


doc. Ing. Jindřich Matoušek, Ph.D.