

## Vyjádření vedoucího k disertační práci

### David Klusáček: New Methods in Statistical Speech Recognition

David Klusáček se ve své práci zaměřil na (jak se všeobecně soudí) slabé místo dnešních automatických rozpoznávačů (přepisovačů) mluvené řeči, a to na akustický front-end. Navrhuje modifikovanou strukturu takového front-end (kterou nazývá NUFIBA). Nejpodstatnějšími změnami proti současným state-of-the-art systémům jsou kompenzace ozvěny, segmentace zvuku na řečníka a pozadí, a průběžné sledování SNR, které v součinnosti s akustickým modelem zabraňuje propagaci chyb.

Práce má 226 stran a je členěna na obvyklý úvod, který charakterizuje nedostatky současných rozpoznávačů a návrhy na jejich odstranění, dále čtyři kapitoly (2-5) popisují současný stav automatického rozpoznávání řeči (ASR) – od teoretických základů ASR, přes základy teorie pravděpodobnosti a teorie informace a teorii zpracování signálu až po popis dnešních state-of-the-art front-endů rozpoznávačů. Tato část je i na disertační práci velmi rozsáhlá, je matematicky velmi fundovaně a úplně popsána a mohla by sloužit prakticky jako důkladná učebnice pro teorii i praxi akustické části rozpoznávačů mluvené řeči. V šesté kapitole pak D. Klusáček popisuje vlastní práci – front-end NUFIBA, opět velmi podrobně od designu, přes architekturu, jádro tohoto nového front-endu, až po shrnutí vlastního přínosu (tj., navržených podstatných modifikací současných front-endů). V další kapitole krátce popisuje akustický model a jeho vazbu na navržený front-end. V práci je popsána i metoda clusteringu, jejíž nový velmi rychlý algoritmus a jeho implementace je rovněž součástí práce (bylo použito v akustickém modelu, ale dosah toho modulu je širší – lze jej používat pro jakýkoli clustering, např. pro jazykové modely). I v této (osmé) kapitole postupuje při výkladu D. Klusáček systematicky, včetně kompletního popisu optimalizací a „triků“, které použil pro (asymptotické) zrychlení clusteringového algoritmu. V deváté kapitole popisuje experimenty, které s novými modely v oblasti front-endu provedl na úrovni fonémů (s korpusem SwitchBoard 1). Práce končí závěrem, poděkováními včetně grantů, a seznamem použitých matematických symbolů a seznamem literatury.

Hodnocení:

Práce je psána velmi dobrou angličtinou. Úvodní přehledové kapitoly jsou na úrovni publikovaných knih (učebnic) o speech recognition (jako např. knihy F. Jelineka Statistical Speech Recognition z r. 1999), a to dotažené (kap. 5) až do skutečně nejnovějších postupů a algoritmů. Vlastní teoretická část práce D. Klusáčka týkající se akustického front-endu, popsaná v kap. 5, je rovněž na vysoké úrovni, a obsahuje ne jednu, ale několik zcela nových myšlenek (v absolutním smyslu), rigorózně matematicky a algoritmicky popsaných. Kapitola o MMI, kde D. Klusáček zavedl nové optimalizace vedoucí k podstatnému zrychlení a tím rozšíření možností obecného clusterování na úlohy, kde to dříve nebylo možné, je vynikající (a plně implementovaná). Seznam literatury je rozsáhlý a zcela postačující. Za podstatný nedostatek práce však považuji to, že vzhledem k termínu odevzdání v limitní době studia nebyly dokončeny experimenty, které by komplexně zhodnotily přínos navržených metod pro standardní metriky v oblasti ASR, tj. WER na slovech ve standardní evaluační úloze. Je třeba poznamenat, že D. Klusáček je si tohoto nedostatku vědom, a v příslušných pasážích tento nedostatek sám zdůrazňuje.

Závěr:

Teoretickou část práce, navržené nové postupy ve front-endu, i vlastní text práce hodnotím velmi vysoko. Pokud by D. Klusáček prokázal alespoň u obhajoby, že navržené nové postupy efektivně zlepšují kvalitu rozpoznávání (ať už plnými výsledky experimentů, nebo alespoň jejich částí, která by porovnávala vlastní front-end – jádro práce - se současnými state-of-the-art výsledky), považuji ji za práci splňující kritéria disertace v oboru Matematická lingvistika na MFF UK v Praze.



Praha, 26. 8. 2012, Jan Hajič, ÚFAL MFF UK