

Vyjádření vedoucího doktorské disertační práce

Petr Podveský: Speech Recognition of Czech Using Finite-State Machines

Předložená práce se zabývá jazykovým modelováním pro češtinu, zejména těmi jeho aspekty, kterými se liší od angličtiny a které tedy způsobují problémy v systémech rozpoznávání řeči založených na technologii vyvinuté pro angličtinu. Autor je s problematikou velmi dobře seznámen, je třeba se zmínit i o jeho dlouhodobém pobytu v Center for Language and Speech Processing na Johns Hopkins University v Baltimore, MD, USA, které je v tomto oboru špičkovým světovým pracovištěm.

Práce je rozdělena do osmi kapitol a tří příloh. V první kapitole autor zdůvodňuje potřebu kvalitního automatického rozpoznávání řeči pro aplikace v průmyslu, rozebírá problémy s rozpoznáváním češtiny a krátce popisuje komponenty výsledného systému pro rozpoznávání řeči (zpracování signálu, akustický model, jazykový model, dekodér). V kapitole 2 se zabývá jazykovým modelováním; nejprve popisuje jeho funkci v systému rozpoznávání, definuje příslušné matematicko-statistické vzorce a zejména se zabývá problematikou tzv. vyhlazování, neboť to je vzhledem k permanentnímu nedostatku dat klíčové pro úspěšné nasazení rozpoznávačů v praxi (a při jejich laboratorním kvalitativním vyhodnocování). V téže kapitole popisuje konečné automaty, jejich pravděpodobnostní verzi a příslušné definice a dále jejich „nastavbu“ (konečné převodníky) používanou v modelech rozpoznávačů. Na závěr této kapitoly popisuje i metody pročištění jazykového modelu (pruning). Ve třetí kapitole popisuje další metody jazykového modelování používané v současné praxi. Krátká čtvrtá kapitola specifikuje přesně cíle předložené práce. Pátá kapitola se zabývá problémy češtiny jako přirozeného jazyka ve vztahu k automatickému rozpoznávání řeči. Konstatuje, že zejména bohatost forem slovní zásoby, volný slovosled a tzv. obecná čeština (a další její varianty) jsou nejslabšími místy při pokusech o vytvoření automatického rozpoznávače češtiny na základě současné technologie. V šesté kapitole popisuje tři zdroje dat (novinářské texty z Lidových novin, čtené zprávy z rozhlasu, a spontánní mluvenou řeč z korpusu projektu Malach), které používal k jazykovému modelování; se dvěma z nich (korpus čtených rozhlasových zpráv a korpus Malach, které jsou v originále nahrávky mluvených projevů) pak prováděl experimenty (a jejich vyhodnocení) popsané v dalších částech předložené práce. V sedmé kapitole popisuje, jak se vypořádal ve vytvořených jazykových modelech s problémem vysokého počtu slov chybějících v systémovém slovníku především z důvodů české flexe a používání obecné češtiny (což se zejména projevovalo v korpusu projektu Malach). Všechny vytvořené modely jsou experimentálně otestovány a vyhodnoceny na korpusu čtených zpráv a na korpusu projektu Malach. Osmá kapitola obsahuje autorovo shrnutí předložených výsledků. Přílohy pak obsahují vstupní data (ve tvaru tzv. lattice, která je v zásadě výstupem akustické části rozpoznávače), a příklady rozpoznávaných vět pro snadné porovnání části výsledků v obou testovaných korpusech (a také z nich lze snadno vyčíst meze, kterých za daného nastavení systému bylo vůbec možno dosáhnout – tzv. oracle accuracy).

Vlastní samostatný autorův přínos je popsán zejména v kapitolách 5 až 8 a v přílohách; vlastní jádro práce přitom spočívá v kapitole 6, která popisuje jazykové modely a experimenty a vyhodnocovací testy s nimi prováděné.

Hodnocení:

Celkově: výsledky autorovy práce jsou i z celosvětového hlediska originální. Autor navazoval na svou starší diplomovou práci, ale výsledky jsou podstatně rozšířené a doplněné o nové experimenty s podstatně odlišnými daty. Autor rovněž používá i pro související zpracování (akustický model apod.) nejnovější metody a poznatky. Výsledky uvedené v kapitole 5, a to na obou testovaných typech promluv, prokazují, že autorem vytvořené jazykové modely vykazují signifikantně lepší výsledky, než metody dosud používané. Úvod práce a zejména popis metod vyhlazování a pruningu uvedený v kapitole 2 je velmi pěkným shrnutím všech těchto poznatků z dostupné literatury a bude jistě sloužit i dalším doktorandům v tomto oboru jako základní informační a studijní materiál.

Za největší pozitivní přínos autora k problematice jazykového modelování v systémech automatického rozpoznávání řeči považuji

- návrh, specifikaci, implementaci a důkladné otestování (pomocí světově uznávaných standardních technik) několika variant jazykových modelů motivovaných problematikou češtiny jako flektivního a variantně bohatého jazyka;
- vytvoření datových zdrojů (zejména výslovnostních slovníků) pro další zájemce o tuto problematiku.

Jazyk práce: práce je psaná dobře anglicky a je velmi dobře čitelná. Drobné překlepy: „dependend“ má být „dependent“ (Abstrakt).

Po formální stránce (obsah, prohlášení a vlastní práci, literatura a její citování, seznamy obrázků a tabulek) je práce dle mého názoru v pořádku.

Závěr: celkově práci považuji za přínosnou k problematice automatického rozpoznávání mluvené češtiny a doporučuji, aby byla přijata a obhájena jako práce disertační; autor dostatečně prokázal schopnost k samostatné tvořivé práci i několika recenzovanými publikacemi na mezinárodním fóru (1 v časopise s IF, 2 další na mezinárodně recenzovaných celosvětových konferencích), z nichž jedna uspěla ve velmi přísné konkurenci na tzv. Student Research Worskhop na celosvětové konferenci Association for Computational Linguistics v Ann Arbor, Michigan, MI, USA v roce 2005.

