

Posudek diplomové práce

Petry Tobolíkové

Detekce chyb v rozpoznávání mluvené řeči

Karlova Universita, Matematicko-fyzikální fakulta

Obsah diplomové práce

Práce Petry Tobolíkové se zaměřila na možnosti detekce chyb v rozpoznávání mluvené češtiny. Řešení této úlohy má praktické využití v reálných aplikacích rozpoznávání mluvené řeči, které je vždy zatíženo určitým procentem chyb. Schopnost detekovat nejslabší místa výstupu umožňuje nasazení dalších metod k upřesnění rozpoznání.

První kapitola práce rozebírá aktuální stav řečových technologií a hlavní úskalí rozpoznávání mluvené řeči.

V druhé kapitole autorka představuje současné metody rozpoznávání mluvené řeči a celý proces přípravy a využití akustických modelů využívajících technologie skrytých markovovských řetězců i zapojení jazykového modelu do procesu rozpoznávání.

Detekce chyb rozpoznávání je detailně rozebrána ve třetí kapitole. Ta obsahuje četné odkazy na nejnovější monografie a články v této oblasti. Jsou popsány detekce založené jak na pravděpodobnostních metodách a akustických datech, tak i nepravděpodobnostní metody pracující pouze s textem.

Autorka využila k detekci tří metod strojového učení, jejichž popisu věnovala čtvrtou kapitolu. Mezi metody strojového učení byly zařazeny logistická regrese, rozhodovací stromy a neuronové sítě.

V páté kapitole jsou popsány vlastní experimenty detekce chyb založené na třech uvedených metodách strojového učení a jejich výsledky ve formě ROC křivek. Jako příznaků bylo využito skóre hran lattice, které vygeneroval rozpoznávač na základě svého akustického a jazykového modelu, a především informací obsažených přímo v rozpoznávaném textu jako jsou pozice slova ve větě, délka slova, kategorie posledních tří, dvou a jednoho znaku, trigramová pravděpodobnost nad slovy, nad posledním jedním, dvěma a třemi znaky slov uskupených do kategorií. K experimentům bylo použito knihoven statistického softwaru R a výstupních lattice rozpoznávače ZČU vygenerovaných z vět korpusu CZBN. Do experimentů byla zahrnuta i autorčina implementace jazykového modelu s modifikovaným Kneser-Ney vyhlazováním natrénovaná nad vybranými daty ČNK.

Šestá kapitola slouží jako manuál k autorkou naprogramovanému prostředí usnadňujícímu experimenty – od přípravy dat, přes spuštění detekce, až po vyhodnocení a prezentaci výsledků.

Závěrečná kapitola obsahuje souhrnné vyhodnocení úspěšnosti zvolených metod, jejich kombinací a váhy příznaků v rámci jednotlivých experimentů. Nejúspěšnější metodou při detekci chyb se staly rozhodovací stromy, které daly největší váhu trigramovému pravděpodobnostnímu modelu slov, dále délce slova a trigramovému pravděpodobnostnímu modelu posledních tří znaků (v tomto pořadí). V závěru navržené budoucí experimenty odkazují na možnost testování další kombinací metod a atributů, případně při trénování pracovat s více hypotézami rozpoznávače řeči.

Příklad lattice a přehled směrodatných odchylek experimentů byly zařazeny do příloh. Příložené DVD obsahuje elektronickou dokumentaci a všechny instalační soubory i zdrojové kódy potřebné k zopakování experimentů.

## Hodnocení

Vybrané téma se zabývá již skutečným výzkumem, který si žádal jak nové nápady, tak i přesné a důsledné provádění experimentů. Autorka nejen že tyto předpoklady splnila a ukázala jednu z nadějných cest na vylepšení rozpoznávání mluvené řeči, ale navíc celou práci i svůj software napsala tak přehledným a čtivým stylem, že mohou ihned sloužit jako literatura uvádějící do problematiky a poskytnout experimentální prostředí pro další zájemce. Práci hodnotím jako výbornou.

Mgr. Nino Peterek, Ph.D.  
UK MFF, ÚFAL

V Praze 3.5. 2008

