

Posudek oponenta diplomové práce

Jméno a příjmení autora posudku: David Mareček

Jméno a příjmení autora práce: Dušan Variš

Název práce Automatic Error Correction of Machine Translation Output

Text posudku

Předložená diplomová práce se zabývá automatickou opravou chyb ve výstupech strojového překladu. Narozdíl od jejího pravidlového předchůdce Depfixu autor využívá pouze statistických metod, což umožňuje jednoduchou přenositelnost metody na další jazyky. Experimenty provádí na překladech z angličtiny do češtiny a z angličtiny do němčiny.

Obsah práce

Práce je rozdělena do osmi částí. V úvodní motivační části autor zdůvodňuje význam posteditačních nástrojů, jejich vývoj a související práce. Ve druhé části popisuje nástroje, které bude používat pro syntaktickou analýzu vět, pro generování nových slov a pro strojové učení chybových a opravovacích modelů. Třetí kapitola se zabývá zdroji dat, ze kterých je možné statistické modely učit. Autor zde uvádí čtyři různé zdroje a konstatuje, že data nejsou ideální, protože ruční posteditace často větu úplně přeformulují, než aby opravili pouze pár gramatických chyb. Ve čtvrté kapitole autor podrobně specifikuje, co bude jeho nástroj opravovat a jakým způsobem bude modely učit z dostupných dat. Navrhuje Oracle klasifikátor, který kromě vstupní a přeložené věty vidí i referenční (nebo manuálně posteditovaný) překlad. Pomocí něho pak definuje, jak bude z trénovacích dat extrahovat trénovací instance a určovat, které z nich jsou chybové. Pátá kapitola se věnuje trénování modelů. Autor zkouší několik různých učících algoritmů a testuje je na několika různých datasetech přes různé domény, různé systémy strojového překladu a různá zlatá data: referenční překlady, lidské posteditace a automatické posteditace generované systémem Depfix. Zkouší i různé metody selekce rysů. Trénuje čtyři modely pro predikci různých skupin morfologických rysů a vyhodnocuje jejich úspěšnost. Šestá kapitola se zabývá evaluací. Používá automatickou metriku BLEU i manuální evaluaci, kdy lidé hodnotí zlepšení nebo zhoršení automaticky posteditovaných vět. Na základě těchto evaluací dokazuje, že jeho systém umí překlady vylepšit. Z 252 změněných vět lidé pozorovali zlepšení ve 166 případech, zhoršení ve 43, u zbylých to nebylo možné určit. V sedmé kapitole se tento statistický systém snaží aplikovat na anglicko-německý překlad. Autor zde popisuje nástroje na analýzu a generování němčiny. Z vyhodnocení výsledků ovšem plyne, že systém německé věty spíše zhoršuje. Osmá kapitola práci uzavírá a shrnuje její výsledky. K práci je přiloženo CD s daty, skripty a s návodem na instalaci nástroje MLFix.

Hodnocení

Práce je psaná dobře srozumitelnou angličtinou, je přehledně strukturovaná, čistý text bez referencí a příloh čítá 50 stránek. Všechny použité nástroje a související práce jsou správně citovány.

V některých částech (například druhá kapitola) se ovšem zdá, že práce byla dopisována ve spěchu, je tam množství překlepů některé hůře formulované věty. Někde bych uvítal více příkladů, například jak vypadá tag a rysy v intersetu. Popisy tabulek jsou často příliš obecné a čtenář musí hledat v širším textu významy některých čísel.

V kapitole 4.2 “Feature Extraction” není zřejmé, jaké všechny rysy se používají. Extrahují se například ze všech okolních uzlů i formy, lemmata a všechny morfologické rysy z intersetu? Zde bych očekával daleko více informací nebo příkladů.

Přestože experimenty po přenosu z češtiny do němčiny vykazují spíše zhoršení, na češtině dosahují velmi slibných výsledků. Autor provedl velké množství experimentů a musel se popasovat s problémy jako jsou nedokonalá trénovací data, extrakce instancí z trénovacích dat a jejich nevyváženost. Oceňuji zejména zavedení orákula pro vývoj metod k extrakci a klasifikaci instancí.

Dotazy a připomínky

- Využil jste ruční post-editace z WMT datasetů? V kapitole 3.4 píšete, že existují, později ale například u dat WMT 2010 uvádíte, že používáte pouze reference. Jsou tyto posteditace vůbec významné a kolik jich je?
- Píšete, že z dat extrahujete asi 1500 rysů. Z textu není jasné, jakou roli v rysech hrají formy a lemmata, ale při velikosti vašich trénovacích dat mi napadá, jestli nedochází k přetrénování modelů. Zkoušel jste formy a lemmata úplně vypustit?
- Je zřejmé, že statistický systém trénovaný na výstupech z pravidlového systému Depfix postrádá hlavní výhodu přenositelnosti na jiné jazyky. I takový systém může být ale užitečný. Očekával bych v práci ale příslušnou evaluaci, tedy porovnání, zda může být systém natrénovaný na Depfixu lepší než Depfix. V tabulce 6.3 bohužel systémy trénované na Depfixu chybí. V předchozích tabulkách jsou zase systémy trénované na Depfixu evaluované také na datech vygenerovaných Depfixem, takže také nejdou porovnat.
- Bylo by zajímavé ručně porovnat i MLFix a Depfix a například potvrdit nebo vyvrátit, že MLFix se hodí lépe na neuronový překlad, jak se zdá z výsledků automatické evaluace.

Závěr

Diplomová práce splnila vytyčený úkol vytvořit trénovatelný nástroj automatické posteditace přenositelný mezi jazyky. Autor prokázal schopnost analyzovat a řešit komplexní úlohu, provedl velké množství experimentů a dosáhl vylepšení strojového překladu z angličtiny do češtiny. Slabší stránkou práce je její kratší délka, některé části by si zasloužily více detailů nebo příkladů.

Doporučení k obhajobě

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

Soutěž studentských prací

Vynikající práce vhodná soutěže studentských prací: **NE**.

V Praze dne 31. 8. 2016

Podpis: