

Posudek vedoucího diplomové práce

Jméno a příjmení autora posudku: RNDr. Ondřej Bojar, Ph.D.

Jméno a příjmení autora práce: Bc. Matouš Macháček

Název práce: Measures of Machine Translation Quality

Vlastní text:

Diplomová práce Matouše Macháčka zkoumá metody hodnocení strojového překladu. Postupuje přitom od metod ručních k automatickým a zabývá se využitím těchto metod pro prosté porovnání systémů i pro ladění vah statistických frázových modelů. Kromě návrhu vlastní ruční metody hodnocení (SegRanks, anotátoři v něm srovnávají krátké úseky vět místo vět celých) práce zahrnuje přehled aktuálních automatických metrik a výsledků ze soutěže WMT v automatickém hodnocení strojového překladu. Práce je členěna do šesti kapitol. Po úvodu (kap. 1) je popsána Matoušova metoda SegRanks (kap. 2) včetně vyhodnocení mezianotátorské shody. Ve třetí kapitole Matouš využívá získané anotace pro tři vytyčené cíle: měření korelace SegRanks se standardním hodnocením, měření kvality neviděných systémů pomocí existujících anotací SegRanks a konečně optimalizace překladového systému pomocí těchto anotací. Čtvrtá kapitola popisuje mnoho cizích technik i soutěž v metrikách. Pátá kapitola shrnuje volně související práce; přístupy, které s Matoušovým úkolem souvisejí těsněji už byly průběžně uváděny v jednotlivých kapitolách. Za závěrem v šesté kapitole nalezneme seznam literatury a přílohy především s uživatelskou a programátorskou dokumentací anotačního prostředí SegRanks. Text je psán dobrou angličtinou s velmi malým počtem překlepů nebo chyb v gramatice.

Vlastní pokusy ve druhé a třetí kapitole mají smíšené výsledky. Je velmi dobře, že SegRanks dosahuje lepší mezianotátorské shody než standardně užívaná metoda, kde lidé srovnávají celé věty, na druhou stranu jsme čekali zlepšení výraznější. Získaná hodnocení dobře korelují se standardním postupem, a Matoušova metoda anotace je tak přínosem zejména z hlediska rychlosti a atraktivity procesu pro uživatele-anotátory. Bohužel se nepodařilo využít získané anotace pro neviděné systémy, ale Matouš velmi pečlivě vysvětlil důvody tohoto neúspěchu: systémy mají tendenci se shodnout ve formulacích správných výstupů, počet schválených úseků u výstupu nového systému je tak nadhodnocen. Navíc je hodnocení věty jako celku velice ovlivněno predikátem, který však krátké úseky nemusí vždy zachytit. Experimenty s laděním systémů pomocí získaných anotací dávají opět smíšené, ale relativně nadějně výsledky: automatické hodnocení takto vyladěných systémů není příznivé, v ručním hodnocení jedna ze tří zkoumaných konfigurací vychází lépe.

Velmi přínosný je i přehled existujících automatických metrik a podrobná analýza některých z nich. Zde Matouš těží z možnosti, která se mu naskytla: již dvakrát organizoval soutěž v automatických metrikách strojového překladu v rámci workshopu WMT. Přitom nejen zajistil hladký průběh soutěže, sběr i analýzu výsledků, ale též v technických podrobnostech posunul tradiční postupy dále: zavedl intervaly spolehlivosti a intenzivně se zabýval způsobem interpretace soutěžních anotací (typ korelačního koeficientu a též použitý vzorec ovlivňující zejména vliv shodně hodnocených systémů na celkové skóre). Jediným problémem této kapitoly je skutečnost, že neodlišuje zcela přesně, které části textu byly přímo přejaty ze souhrnného soutěžního článku (Macháček & Bojar, Proc. of WMT, ACL, 2014.) Článek je v úvodu kapitoly 4 řádně citován a jeho autorem je z naprosté většiny skutečně Matouš, takže Matouš přejímá vlastní text. I přesto pro vyloučení veškerých pochybností bylo třeba tuto informaci uvést explicitně, např. formulací "most of this chapter is copied verbatim / taken over from the paper...".

Celkově jsem s Matoušovou prací velmi spokojen. Tématu se věnoval dlouhodobě (viz dva ročníky WMT), zadání splnil beze zbytku a jeho vlastní navržená metoda je minimálně z hlediska mezianotátorské shody a využití pro ladění systémů nadějná. Vytvořené anotační rozhraní je velmi intuitivní a snadno ovladatelné a stojí za zvážení, zda ho nezveřejnit i samostatně. Nejvíce si cením podrobných diskusí a analýz v celé práci, např. při ladění Matouš zkoumá, jak se optimalizované skóre navržených variant algoritmu vyvíjí v průběhu iterací.

Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

Vynikající práce vhodná pro soutěž studentských prací	ANO <input type="checkbox"/>
---	------------------------------

Seznam soutěží studentských prací, viz <http://www.mff.cuni.cz/studium/bcmgr/prace/>

Pokud jste výše zaškrtnli ANO, zdůvodněte prosím svůj návrh, případně uveďte konkrétní soutěž, pro kterou je práce vhodná (rámeček lze nechat prázdný, pokud za dostatečné zdůvodnění považujete text posudku):

V Praze dne: 1. 9. 2014

Podpis:**

** nehodící se škrtněte (vymažte)*

*** do SISu vkládejte formulář nepodepsaný (ve formátu PDF), podpis je potřeba doplnit až na vytištěný posudek.*