

Problémy hodnocení kvality strojového překladu

Vít Baisa
RASLAN 2009

- Úvod
- Přehled metod
- LAMENT
- Závěr

Úvod

- Co je to kvalita překladu?
 - přesnost (accuracy, adequacy)
 - srozumitelnost (intelligibility, fluency)
 - ruční vs. automatické hodnocení kvality
 - paradox související s hodnocením
-

Metody hodnocení kvality SP

- referenční překlady
- označení
 - s - zdrojová věta
 - c - kandidátní překlad zdrojové věty
 - $r_1 \dots r_n$ - referenční překlady

BLEU

- nejstarší metoda (2001)
- n-gramové porovnávání c a $r_1 \dots r_n$
- uni-, bi-, tri- a kvadrigramy
- trestání za krátkost
- modifikovaná n-gramová přesnost

NEVA

- drobné vylepšení metriky BLEU
 - vhodné i pro krátké věty
 - ve výsledku velmi podobná
 - krátké věty se zase tak často nevyskytují
-

WAFT

- editační vzdálenost mezi c a $r_1 \dots r_n$
- smazání, prohození, vložení *slova*
- skóre počítáme pomocí r_i s nejnižší EV

TER & HTER

- TER velmi podobný metodě WAFT
- dovoluje prohodit sekvence slov
- HTER varianta
 - ruční hodnocení vs.
 - vytvoření nového referenčního překladu, následné hodnocení pomocí metody TER

METEOR

- mapování slov z c na slova z $r_1 \dots r_n$
- synonyma z WordNetu
- bez synonym ~ unigramová přesnost BLEU
- menší počet referenčních překladů
- vysoká korelace s ručním hodnocením

Problémy a nedostatky n-gramů

- unigramy odpovídají přesnosti
- n-gramy odpovídají srozumitelnosti
- předpoklad
 - dobrý překlad c je podobný překladu r_i
- námitka
 - silně závisí na kvalitě a množství $r_1 \dots r_n$
 - nikdy nepokryjeme všechny rozumné možnosti referenčními překlady

Problémy a nedostatky editační vzdálenosti

- předpoklad
 - čím méně dodatečných úprav překladu, tím vyšší jeho kvalita
- námitka
 - není chyba jako chyba
 - různé výsledky pro různé jazyky

Kam se poděla zdrojová věta?

- žádná metoda nebere v úvahu zdrojovou větu
- přepoklad
 - ručně vytvořené referenční překlady jsou významově stejné nebo podobné zdrojové větě a jsou dobře srozumitelné
- námitka
 - neošetřujeme chybné vstupy
 - případ vícenásobného překladu

LAMENT - úvod

- nástin myšlenek metody LAMENT
 - Language model and meaning based evaluation of translation
 - jazykový model ~ korpus
 - mapování slov se stejným významem
-

LAMENT - srozumitelnost

- čím větší část věty najdeme v korpusu, tím větší můžeme očekávat srozumitelnost
 - čím větší srozumitelnost c , tím lepší můžeme očekávat překlad
 - čím nižší srozumitelnost s , tím mírnější by hodnocení kvality překladu mělo být
 - skóre srozumitelnosti pro s a c
-

LAMENT - přesnost

- těžší část
- mapování podobné jako METEOR
- místo synonym hledáme překladové dvojice
- ideálně nesmíme dělat stejné chyby jako strojový překlad
- mapování lemmat
 - chyby lemmatizace, chyby mapování
- $M(_, x)$, $M(x, _)$, $M(x, yz)$

LAMENT - výsledné skóre

- Jak to skloubit dohromady?
 - hledání vhodného vzorce
 - Nutně platí, že
 - čím vyšší pokrytí **c** v korpusu, tím lepší
 - čím vyšší pokrytí **c** mapováním, tím lepší
 - čím nižší pokrytí **s** v korpusu, tím bychom měli být mírnější při hodnocení
-

Závěr

- Jak měřit kvalitu měření kvality?
 - Co s paradoxem?
 - Jaký je smysl toho všeho?
 - Shrnutí
-