

Odhad kvality strojového překladu předpovídá skóre kvality pro překlady vytvořené systémy strojového překladu na vstupních a výstupních segmentů. Systémy odhadu kvality jsou trénované pomocí učení s učitelem pomocí trénovacích dat, která obsahují překlad vytvořený jedním nebo více systémy strojového překladu. Výběr trénovacích dat pro strojový překlad má tedy vliv na to, jak dobře funguje systém odhadu kvality.

Tato práce studuje vztah mezi systémy strojového překladu a systémy pro odhad kvality na úrovni vět. S využitím našich definic výkonu systému strojového překladu a výkonu systému odhadu kvality jsme provedli experimenty, které zahrnují trénování systémů strojového překladu a odhadu kvality různých výkonů. Prezentovali jsme výsledky hodnocení systémů odhadu kvality na testovacích sadách různých domén překládaných strojovými překladači různé síly. Zjistili jsme, že (i) systémy hodnocení kvality vyškolené na překladech nižší kvality mají vyšší výkon, než systémy hodnocení kvality trénované na překladech vyšší kvality; (ii) hodnocení vysoce kvalitních systémů strojového překladu je výzvou pro systémy hodnocení kvality všech výkonů; (iii) vysoce výkonné systémy pro odhad kvality fungují lépe pro domény odlišné od trénovací, než systémy pro odhad kvality s nízkým výkonem