

Multimodalita ve strojovém překladu

Jindřich Libovický

Tradičně se většina úloh zpracování přirozeného jazyka řeší výhradně uvnitř jazyka, kdy modely spoléhají na distribuční vlastnosti slov. Hluboké učení se svojí schopností učit se vhodné reprezentace vstupních dat umožňuje využít více informací tím, že trénovací signál nepochází pouze z jazyka, ale o i z obrazové modality. Jednou z úloh, které se pokoušejí využít vizuální informace, je multimodální strojový překlad: překlad popisků obrázků, kdy je stále k dispozici původní obrázek, který lze využít jako vstup pro překladač.

Tato práce shrnuje metody společného zpracování jazykových dat a fotografií s využitím hlubokého učení. Uvádíme přehled metod, které se využívají k řešení multimodálního strojového překladu a popisujeme náš původní příspěvek k řešení této úlohy. Představujeme metody kombinování více vstupů z potenciálně různých modalit v modelech sekvenčního učení založených na rekurentních neuronových sítích a neuronových sítích s mechanismem sebezpozornosti. Uvádíme výsledky, kterých jsme dosáhli při řešení multimodálního strojového překladu a dalších úloh souvisejících se strojovým překladem. Na závěr analyzujeme, jak multimodalita ovlivňuje sémantické vlastnosti větných reprezentací, které v sítích vznikají, a jak sémantické vlastnosti reprezentací souvisí s kvalitou překladu.