

Strojový překlad s využitím syntaktické analýzy

Martin Popel

Tato práce popisuje zlepšení anglicko-českého a česko-anglického strojového překladu pomocí metod, které lze použít i na další jazyky. V první části je popsáno několik zlepšení hloubkově-syntaktického překladače TectoMT, například rozšíření pro další jazyky a domény nebo implementace nového typu překladových modelů využívajících kontext a různé metody strojového učení. V druhé části je popsán neuronový překladač Transformer a jeho vylepšení. Po detailní analýze vlivu různých hyperparametrů, bylo optimalizováno trénování systému tak, že dosáhl o 1.0 BLEU lepšího překladu než nejlepší systém v soutěži WMT2017. Využitím jednojazyčných dat cílového jazyka pomocí nového typu zpětného překladu bylo dosaženo dalšího zlepšení kvality překladu o 2.8 BLEU. Využitím doménové adaptace zohledňující „překladštinu“ (translationese) – tedy zohledněním toho, zda paralelní data jsou původně psána česky, nebo anglicky – byl výsledný systém vylepšen o dalších 0.2 BLEU. Tento výsledný neuronový překladač byl signifikantně lepší ($p < 0.05$) než všechny ostatní anglicko-české a česko-anglické překladače v soutěži WMT2018. Podle výsledků ručního hodnocení byla kvalita tohoto strojového překladu dokonce vyšší než kvalita lidského referenčního překladu.