

Abstrakt

V poslední době nabídl výzkum strojového překladu nové metody pro zrychlení generování. Jedním z navrhovaných metod je takzvaný neautoregresivní neuronový strojový překlad. V klasických autoregresivních překladových systémech jsou výstupní pravděpodobnostní rozdělení modelována podmíněně na předchozích výstupech. Tato závislost umožňuje modelům sledovat stav překládání a obvykle vede ke generování velmi plynulých textů. Autoregresivní postup je však ze své podstaty sekvenční a nelze jej paralelizovat. Neautoregresivní systémy modelují pravděpodobnosti jednotlivých cílových slov jako navzájem podmíněně nezávislé, což znamená, že dekodování lze paralelizovat snadno. Nevýhodou je ovšem nízká kvalita překladu ve srovnání s modely autoregresivními. Cíl výzkumu neautoregresivních metod strojového překladu je zlepšit kvalitu překladu a zároveň uchovat vysokou rychlost dekodování. Naše práce předkládá rešerši publikovaných metod a poukazuje na některé nedostatky plynoucí z obecně přijímané evaluační metodologie. Popisujeme experimenty s neautoregresivními modely trénovaných pomocí takzvané „connectionist temporal classification“. Z našich výsledků plyne, že i když dosahujeme nejlepších výsledků mezi neautoregresivními modely na datech z WMT z roku 2014, při porovnání s nejnovějšími optimalizovanými autoregresivními systémy tyto modely pořád zaostávají.