

Název práce: Schopnosti učení neuronových sítí Transformer

Autor: Dušan Variš

Katedra: Ústav formální a aplikované lingvistiky

Vedoucí práce: doc. RNDr. Ondřej Bojar, Ph.D.,
Ústav formální a aplikované lingvistiky

Abstrakt:

Přestože současné neuronové sítě, inspirované biologickými neurony, byly v posledních letech schopny dosáhnout lidské úrovně na mnoha úlohách, proces jejich optimalizace (učení) je stále velmi odlišný od procesů pozorovaných u lidí. Tato práce zkoumá různé aspekty učení současných neuronových sítí Transformer, převládající architektury pro zpracování přirozeného jazyka. V první části zkoumáme úroveň generalizace v Transformerech pomocí analytických experimentů založených na myšlence adversariální evaluace. V části druhé pak zkoumáme jejich potenciál pro kontinuální učení s použitím regularizace založené na elastické konsolidaci vah. V závěru práce navrhneme modulární rozšíření stávající sítě Transformer umožňující výběr podsítí podmíněný zpracovaným vstupem spolu s demonstrací vlastností této síťové modularizace. Naše hypotézy testujeme především v kontextu neuronového strojového překladu a vícejazyčného překladu, přičemž naměřené výsledky odhalují limity původního Transformeru i metody regularizace pomocí elastické konsolidace vah. Navíc prezentujeme slibné výsledky navržené modulární architektury Transformeru.

Klíčová slova: neuronový strojový překlad, katastrofické zapomínání, modulární neuronové sítě, navazující učení, generalizace