

Abstrakt

V posledných rokoch viedol rýchly rozvoj systémov umelej inteligencie a čoraz väčších jazykových modelov k výraznému pokroku v širokom spektre úloh spracovania prirodzeného jazyka. Napriek ich vysokej výkonnosti sú však tieto prístupy často spojené so značnými výpočtovými nárokmi, dlhým časom tréningovania a silnou závislosťou od veľkého objemu anotovaných dát, čo obmedzuje ich praktickú využiteľnosť v prostrediach s nízkymi a strednými zdrojmi. Prevládajúci trend smerujúci k neustálemu zväčšovaniu modelových architektúr tieto obmedzenia ďalej prehĺbuje a vyvoláva otázky týkajúce sa efektívnosti, dostupnosti a udržateľnosti.

Táto práca sa zameriava na preskúmanie súčasných prístupov k zvyšovaniu efektívnosti a robustnosti doladovania modelov v podmienkach obmedzených dátových a výpočtových zdrojov. Dôraz kladie na stratégie založené na informovanom výbere dát, adaptívnych tréningových postupoch a na kombinovaní komplementárnych metód namiesto výhradného spoliehania sa na škálovanie modelových architektúr. Navrhuje jednotný rámec, ktorý integruje výber vzoriek založený na neurčitosti, štruktúrovanú organizáciu dát a dynamické plánovacie mechanizmy, s voliteľnými rozšíreniami pre pokročilejšie modelovanie neurčitosti.

Prostredníctvom série empirických experimentov naprieč viacerými jazykmi a nastaveniami práca ukazuje, že vhodne navrhnuté kombinácie dátovo efektívnych učebných techník dokážu výrazne znížiť potrebu anotácií pri súčasnom zachovaní alebo zlepšení predikčnej výkonnosti a stability tréningového procesu. Získané poznatky naznačujú, že synergická integrácia moderných učebných stratégií predstavuje sľubný smer pre vývoj efektívnejších, dostupnejších a spoľahlivejších systémov umelej inteligencie v prostrediach s obmedzenými zdrojmi.

Kľúčové slová: Aktívne učenie, nízko-zdrojové NLP, výber vzoriek podľa neurčitosti, Epistemické neurónové siete, Aglomeratívne zhlukovanie, dynamické plánovanie, Analýza sentimentu, Rozpoznávanie pomenovaných entít, BERT