

Efektívne modelovanie textu cez množinu atribútov

Samuel Revúcky
Mgr. Marek Šuppa

Transfer learning

- Je v dnešnej dobe štandardnou metódou na riešenie úloh (nielen) z oblasti spracovania prirodzeného jazyka
- Využíva pred-trénované jazykové modely
- Fine tuning na špecifickú úlohu
- Ukazuje sa, že takéto modely dosahujú lepšie výsledky, než trénovanie len na špecifických dátach ^[1]
- Pre každú úlohu potreba natrénovať nový model

[1] *Universal Language Model Fine-tuning for Text Classification*, Howard & Ruder, Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics

Riešenie viacerých úloh

- Prirodzený prístup je natrénovať samostatný model na každú úlohu
- Fine tuning je výpočtovo náročný
- Modely zaberaajú veľmi veľa pamäte
- Nevieme využiť možné užitočné informácie, ktoré medzi sebou dané úlohy zdieľajú

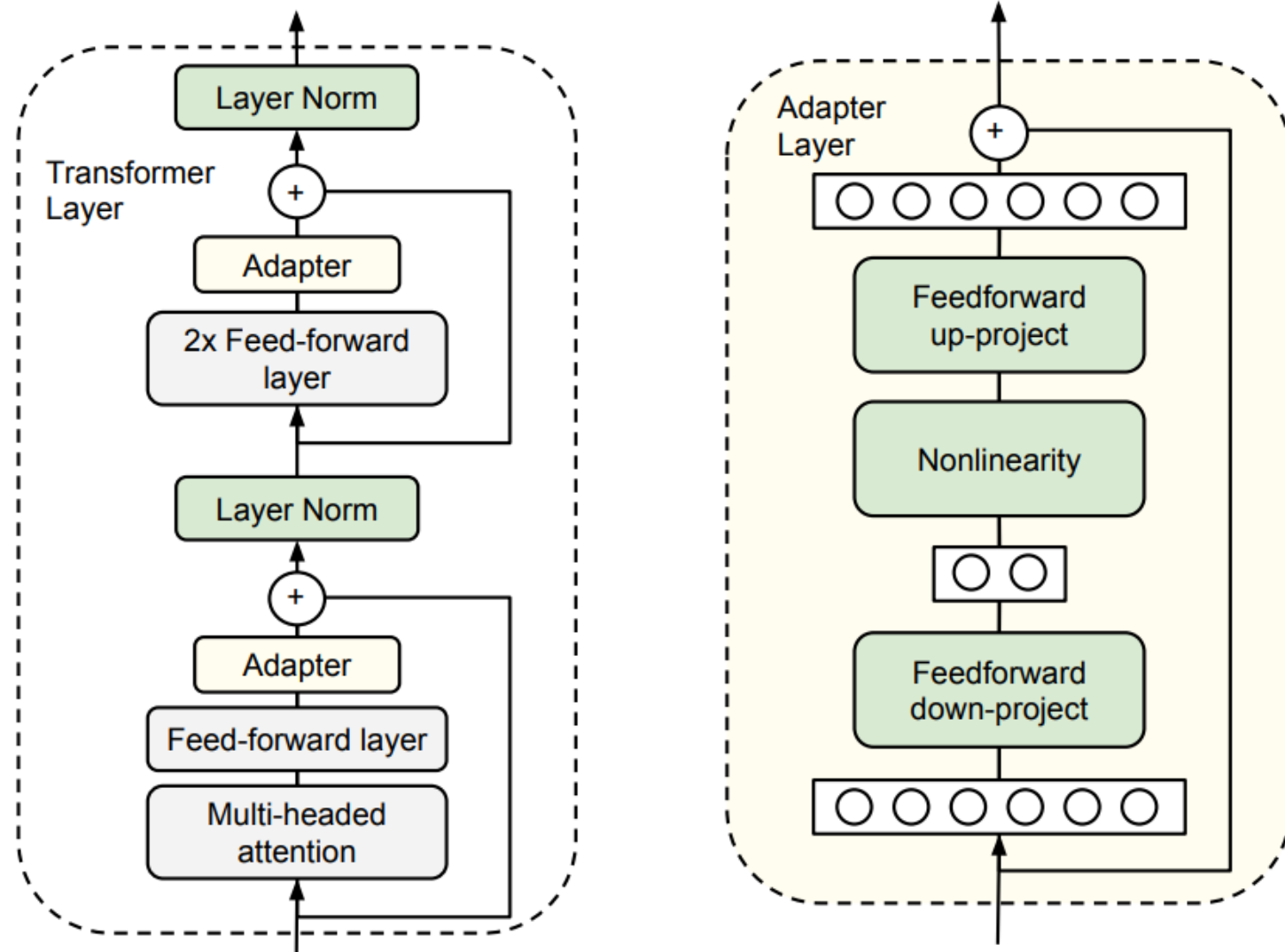
Sekvenčný fine tuning

- Začneme s pred-trénovaným modelom
- Postupne ho trénujeme na špecifických dátach pre každú z úloh
- Redukcia pamäte
- Časovo rovnaké, ako trénovanie samostatných modelov
- “Catastrophic forgetting”
- Nájst' správne poradie, v akom zoradiť úlohy na testovanie je netriviálne

Multi-task learning

- Na začiatku pred-trénovaný model
- Pri tréovaní sa na váhy aplikuje lineárna kombinácia chybových funkcií pre jednotlivé úlohy
- Pridanie novej úlohy vyžaduje trénovať úplne odznovu
- Zvykne overfitovať na nedostatočnom množstve dát a underfitovať na veľkom množstve dát

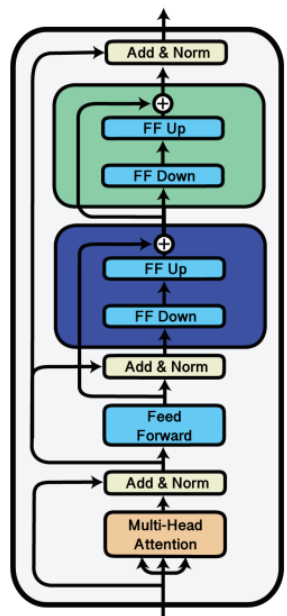
Adaptéry



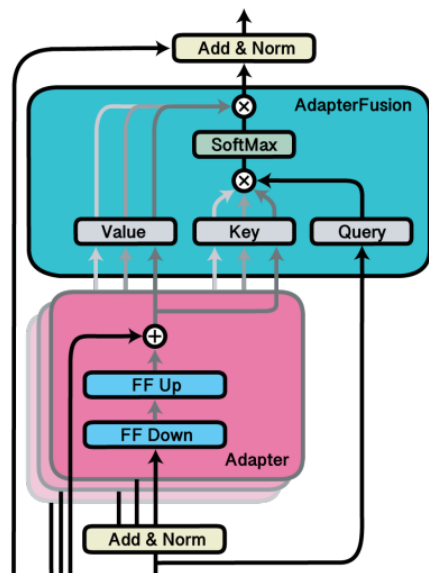
[2] *Parameter-Efficient Transfer Learning for NLP*, Houlsby et. al., Proceedings of the 36th International Conference on Machine Learning

Adaptéry

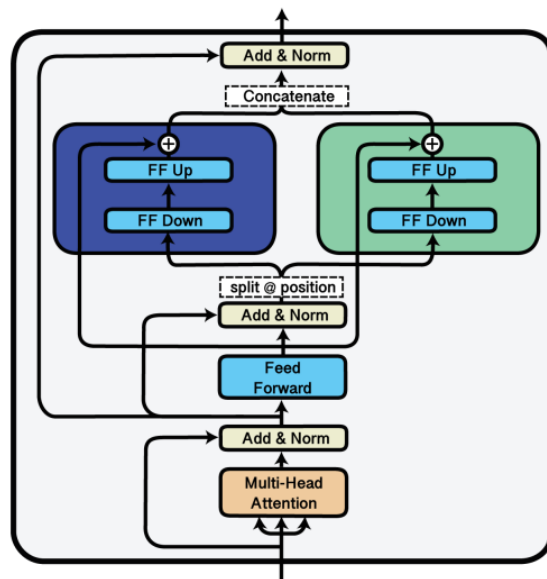
- Váhy pred-trénovaného modelu sú zamrznuté
- Počet trénovaných parametrov je rádovo menší ako pri LLM
- Umožňujú rôzne spôsoby kombinovania a skladania



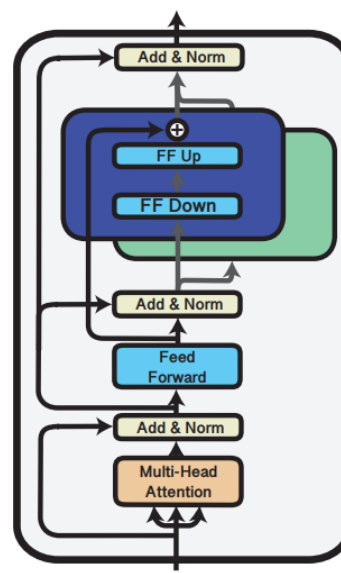
(a) Stack



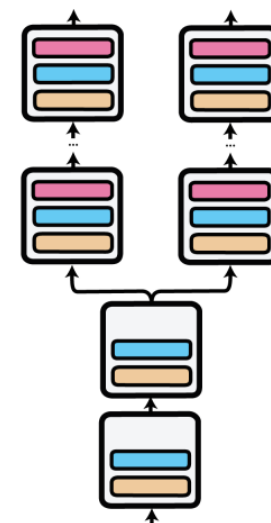
(b) Fuse



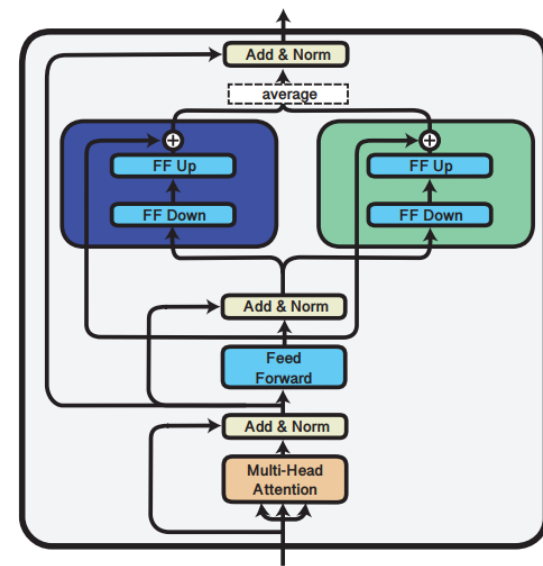
(c) Split



(d) BatchSplit



(e) Parallel



(f) Average

Dataset

Configuration	Train	Validation	Test
default	5828	728	729
filtered_theme_relevance	3316	413	403
filtered_geo_relevance	3727	455	452
filtered_direction	3097	395	378
filtered_target_ludia	2394	296	289
filtered_target_umoznovatelia	2317	275	276
filtered_target_politiky	2423	290	300

<https://cogsci.fmph.uniba.sk/MIMEDIS/index.html>

Ciele

- Navrhnuť a porovnať viacero modelov na klasifikáciu do atribútov z datasetu
- Využiť pri tom adaptéry a princípy z multi-task learningu a sekvenčného fine tuningu

Ďakujem za pozornosť