

Předcvičeným jazykovým modelům byla v posledních letech věnována značná pozornost. Je však stále náročné začlenit předcvičený model, jako je BERT, do úloh generování přirozeného jazyka. Tato práce zkoumá nedávnou metodu zvanou adaptéry jako alternativu k doladování. Adaptéry jsou slibným přístupem, který umožňuje doladování jen velmi malého zlomku předcvičené sítě. Ukazujeme, že při správné inicializaci mohou adaptéry pomoci dosáhnout lepšího výkonu než tréninkové modely od nuly při tréninku podstatně menšího množství závaží než původní model. Dále ukazujeme, že i s náhodně nastavenými závažími používanými jako základní modely pro doladování můžeme dosáhnout podobného výkonu jako jeden ze základních modelů a obejít tak potřebu trénovat stovky milionů závaží v předcvičební fázi. Dále studujeme účinnost adaptérů v modelu Transformer pro úlohy strojového převodu. Adaptéry umístíme buď do enkodéru, nebo pouze do dekodéru, a také se snažíme zmenšit velikost předtrénovaného modelu, abychom snížili nároky na paměť GPU. Zjistili jsme, že začlenění adaptérů do samotného enkodéru odpovídá výkonu nastavení, když zahrneme adaptéry do enkodéru i dekodéru. Nakonec naše studie zmenšení zjistila, že použití pouze poloviny původních předtrénovaných závaží může pozitivně ovlivnit výkon při doladění pomocí adaptérů. Naše experimenty ukazují, že po doladění vrstvy křížové pozornosti můžeme dosáhnout téměř stejného výkonu jako původní model BERT.