

V mezinárodním a vysoce mnohojazyčném prostředí se často stává, že řeč, dokument nebo jakýkoli jiný vstup musí být přeložen do velkého počtu jazyků.

Není to vždy možné, mít zvláštní systém pro každou možnou dvojici jazyků vzhledem k tomu, že takový druh překladatelských systémů je výpočetně náročný.

Kombinace více cílových jazyků do jednoho překladového modelu obvykle způsobuje snížení kvality výstupu pro každý jeho směr překladu.

V této práci provádíme experimenty s kombinací cílových jazyků, abychom zjistili, zda jejich nějaké konkrétní seskupení může vést k lepším výsledkům, v porovnání s náhodným výběrem cílových jazyků.

Využíváme výsledky nejnovějších výzkumích prací o trénování vícejazyčného Transformer modelu beze změny jeho architektury: přidáváme značku cílového jazyku do zdrojové věty.

Natrénovali jsme řadu dvojjazyčných a vícejazyčných Transformer modelů a vyhodnotili jsme je na několika testovacích sadách z různých domén. Zjistili jsme, že ve většině případů seskupení souvisejících cílových jazyků do jednoho modelu způsobuje lepší výkon ve srovnání s modely s náhodně vybranými jazyky.

Zjistili jsme však také, že doména testovací množiny, stejně jako domény dat vybraných do trénovací množiny, má obvykle významnější vliv na zlepšení nebo zhoršení kvality překladu vícejazyčného modelu ve srovnání s dvojjazyčnou.