

Tomasz Dwojak

Streszczenie rozprawy doktorskiej

Wydajne algorytmy dla hybrydowego modelu neuronowego tłumaczenia maszynowego

15 grudnia 2020

W ramach pracy doktorskiej przedkładam zbiór artykułów opublikowanych w recenzowanych czasopismach naukowych bądź recenzowanych materiałach z konferencji o zasięgu międzynarodowym. Wszystkie artykuły omawiają zagadnienie współczesnych metod tłumaczenia automatycznego, a w szczególności – relacji pomiędzy podejściem statystycznym i neuronowym. Są to zatem prace ściśle związane z tematem pracy doktorskiej.

Ważnym elementem mojej działalności naukowej były analizy porównawcze obu podejść. Pierwszą analizę przeprowadziłem w 2016 roku, gdy tłumaczenie neuronowe nie było jeszcze metodą dominującą. W wyniku analizy postawiłem hipotezę badawczą o zasadności rozwiązania łączącego oba podejścia. W 2016 roku wraz z promotorem pomocniczym zaproponowałem model hybrydowy, który wykorzystywał neuronowy model tłumaczenia maszynowego z atencją w podejściu statystycznym. Model został zaprezentowany w konkursie (ang. shared task) na konferencji Workshop for Machine Translation (WMT) 2016 i zajął pierwsze miejsce w kategorii rosyjsko-angielskiej. W 2017 roku opublikowałem pracę dotyczącą efektywnych algorytmów tłumaczenia automatycznego. W tym samym roku opublikowałem pracę o wykorzystaniu informacji syntaktycznych w procesie tłumaczenia. Jestem współautorem dwóch narzędzi do neuronowego tłumaczenia automatycznego: AmuNMT i Marian NMT, których opisy zostały opublikowane na międzynarodowych konferencjach. Swoje badania spiąłem kłamrą w 2019 roku, kiedy wykonałem kolejną analizę porównawczą obu podejść — tym razem z punktu widzenia zastosowań komercyjnych. Przeprowadzona analiza potwierdziła aktualność postawionej przeze mnie hipotezy badawczej.

Tomasz
Dwojak