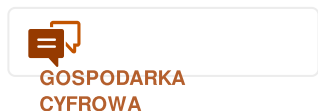


Komputery uczą się języka maltańskiego

Naukowcy wytrenowali model językowy na bazach tekstów w języku maltańskim, nauczyli go rozpoznawać wydźwięk emocjonalny wypowiedzi i oznaczać w tekstach nazwy własne.



© graphicwithart/Shutterstock.com

Technologie językowe oparte na sztucznej inteligencji wspierają komunikację cyfrową we wszystkich językach europejskich. Brakuje jednak narzędzi i zasobów w zakresie technologii językowych dla języka maltańskiego. Od momentu rozpoczęcia prac w 2021 roku zespół finansowanego ze środków Unii Europejskiej projektu LT-BRIDGE zajmuje się rozwiązaniem tego problemu.

W przetwarzaniu języka naturalnego wykorzystuje się sieci neuronowe do trenowania modeli językowych tak, aby umiały powiązać słowa w określonym kontekście. Naukowcy z Uniwersytetu Maltańskiego, który pełni rolę koordynatora projektu LT-BRIDGE, wytrenowali model BERTu na maltańskich danych tekstowych.

Uzupełnianie luk

Czym dokładnie są modele językowe? Kurt Micallef, doktorant z Uniwersytetu Maltańskiego, przybliżył ten temat w [artykule](#) opublikowanym niedawno w serwisie „Times of Malta”: „Modele językowe opierają się na abstrakcyjnym rozumieniu języka. Można je uznać za «intuicję» w kontekście językowym. Na przykład, jeśli zadanie polegałoby na uzupełnieniu luki w zdaniu «Jien _____ il-gazzetta» (Ja _____ gazetę), można by wstawić tam słowo «qrajt» (czytał*m) lub «xtrajt» (kupił*m), natomiast słowa «kilt» (zjadł*m) lub «karozza» (samochód) pasowałyby tam w znacznie mniejszym stopniu”.

Jednym ze sposobów szkolenia takich modeli językowych jest stosowanie metody maskowanego modelowania języka – losowe słowa w tekście są ukrywane, a model

ma za zadanie je odgadnąć. „W powyższym przykładzie model powinien wybrać słowo «qrajt»”, wyjaśnia Micallef. Proces powtarza się dla wielu zdań, aby model językowy mógł nauczyć się języka maltańskiego. Sieć neuronowa jest aktualizowana z każdym zdaniem za pomocą algorytmów uczenia maszynowego, a słowom, które mogą pasować do zdania, przypisywane jest prawdopodobieństwo.

Inne zadania

Dwa inne zadania, do których wykonywania szkolono model BERTu, to analiza sentymentu i rozpoznawanie nazw własnych. „Analiza sentymentu to proces rozpoznawania wydźwięku emocjonalnego danego tekstu”, tłumaczy badacz w [kolejnym artykule opublikowanym w „Times of Malta”](#) . „Jej najprostszą formą jest określanie, czy w danym fragmencie tekstu pojawiają się pozytywne bądź negatywne odczucia w odniesieniu do konkretnego tematu lub koncepcji. Można na przykład określić, czy komentarz dotyczący planów dysponowania budżetem Malty jest przychylny czy krytyczny. Tego typu zadanie nazywamy problemem klasyfikacyjnym, ponieważ do tekstu, który otrzymujemy na wejściu, przypisuje się etykietę klasyfikacyjną (w tym przykładzie – pozytywną lub negatywną)”.

Uczony kontynuuje opisywanie drugiego zadania: „Rozpoznawanie nazw własnych to zadanie z zakresu tagowania, w którym do każdego słowa w tekście wejściowym przypisywana jest etykieta. Zadaniem jest określenie, które etykiety odnoszą się do nazw własnych i jakich podmiotów dotyczą. W porównaniu z analizą sentymentu, zadanie to jest dość niskopoziomowe i zazwyczaj jest wykorzystywane jako uzupełnienie innych systemów językowych. Na przykład można by zidentyfikować imiona i nazwiska osób w sklasyfikowanych danych i je zanonimizować, aby chronić poufne dane”.

Zespół badaczy dostosował wstępnie wytrenowany model BERTu do wykonywania tych zadań poprzez dodanie dodatkowej warstwy do każdego zadania, a następnie uruchomienie algorytmów uczenia maszynowego na zbiorze danych w celu poznania parametrów dodatkowej warstwy. Zaobserwowano, że BERTu osiąga lepsze wyniki niż inne modele językowe, czasami o ponad 20 %. Aktualnie model ten umożliwia wykonywanie bardziej złożonych zadań związanych z rozumieniem języka maltańskiego. Projekt LT-BRIDGE („Bridging the technology gap: Integrating Malta into European Research and Innovation efforts for AI-based language technologies”) dobiegnie końca w grudniu 2023 roku.

Więcej informacji:

[strona projektu LT-BRIDGE](#) 

Słowa kluczowe

LT-BRIDGE

sztuczna inteligencja

język

maltański

model językowy

BERTu

tekst

dane tekstowe

słowo

Powiązane projekty



PROJEKT

**“Bridging the technology gap:
Integrating Malta into European
Research and Innovation efforts for AI-
based language technologies”**

LT-BRIDGE

7 Kwietnia 2025

Powiązane artykuły



WIADOMOŚCI

POSTĘPY NAUKOWE

Tworzenie podstaw godnej zaufania SI



2 Listopada 2022



WYNIKI W SKRÓCIE

**Spersonalizowana aplikacja oparta na SI
uczy języków obcych poprzez rozmowy**



20 Marca 2020



W jaki sposób maszyny interpretują ludzki język?



18 Marca 2022



Pionierska platforma pozwala zwiększyć produktywność



18 Marca 2019

Ostatnia aktualizacja: 2 Grudnia 2022

Permalink: <https://cordis.europa.eu/article/id/442665-computers-learn-maltese/pl>

European Union, 2025