

Streszczenie rozprawy doktorskiej

Niniejsza rozprawa doktorska prezentuje techniki uczenia maszynowego wspomagające analizę stroniczości tekstów w internetowych mediach informacyjnych. Z uwagi na ilość informacji, które są na bieżąco udostępniane przez portale informacyjne, przeciętnemu czytelnikowi może być trudno dokładnie śledzić bieżące wydarzenia. Jest on wystawiony na wiele celowych lub niezamierzonych form manipulacji. W związku z tym, niezbędne jest utworzenie takich narzędzi które wspomagały by rozróżnienie treści obiektywnych oraz stroniczych.

Praca składa się z 3 części i 6 rozdziałów. W pierwszej części umieszczono 3 rozdziały. Pierwszym rozdziałem jest wstęp, gdzie opisano cel, motywację, główne osiągnięcia, dokładną zawartość pracy oraz użyte narzędzia. Drugi rozdział zawiera przegląd literatury. W rozdziale 3 zamieszczono opis najważniejszych algorytmów uczenia maszynowego, które mogą być wykorzystywane do analizy tekstu w kontekście stroniczości mediów.

Druga część zawiera eksperymenty. Znajdują się w niej 3 kolejne rozdziały. W rozdziale 4 opisano pomocnicze eksperymenty dotyczące stroniczości ze względu na widoczność i agendę. Pierwsza sekcja rozdziału 4 dotyczy wykrywania wystąpień encji w czasie i ich powiązania z wydarzeniami. Druga sekcja dotyczy wykrywania artykułów, które opisują ten sam temat. Dodatkowo przedstawiono wykrywanie źródła artykułu.

W rozdziale 5 przedstawiono główne eksperymenty tezy dotyczące wykrywania wydźwięku na temat danej jednostki w artykułach informacyjnych. Przedstawiono utworzony dwujęzyczny zbiór danych SEN dotyczący wykrywania wydźwięku jednostek w nagłówkach artykułów informacyjnych. Opisano szereg eksperymentów i modeli sieci neuronowych opartych o duże modele językowe. W eksperymentach dla porównania wykorzystano również inne zbiory danych. Porównano utworzone modele sieci neuronowych z innymi sieciami neuronowymi, które dotychczas osiągały najlepsze wyniki. Wykazano, że dla większości przypadków, czyli wariantów zbiorów danych, utworzone modele osiągają lepsze wyniki, a w pozostałych przypadkach porównywalne. Ponadto wykryto, że modele ucząc się wykrywać wydźwięk same stają się stronicze wobec niektórych jednostek. Wobec tego zaproponowano kilka rozwiązań problemu, z których niektóre poprawiły wyniki sieci.

Rozdział 6 zawiera pomocnicze eksperymenty dotyczące wykrywania technik manipulacji w paragrafach artykułów informacyjnych. Utworzono wielozadaniową hierarchiczną sieć neuronową opartą na dużych modelach językowych. Sieć przetestowano na kilku językach, dla których osiągała lepsze wyniki niż podstawowy model. Rozwiązanie zostało wykorzystane w prestiżowym konkursie Semeval23.

Trzecia część pracy stanowi podsumowanie rozprawy.