

Recenzja rozprawy doktorskiej

Kamila Kanclerza,

zatytułowanej:

Metody spersonalizowanego rozpoznawania tekstów obraźliwych

1. Problem badawczy i jego znaczenie

Praca doktorska zatytułowana "Metody spersonalizowanego rozpoznawania tekstów obraźliwych" porusza niezwykle istotne zagadnienie związane z dynamicznie rozwijającym się obszarem badawczym, jakim jest przetwarzanie języka naturalnego (NLP) oraz szczególny przypadek analizy sentymentu jakim jest wykrywanie tekstów obraźliwych. W dobie powszechnego dostępu do platform społecznościowych oraz narzędzi komunikacyjnych, problem obraźliwych i agresywnych treści staje się coraz bardziej palący. Skuteczne metody identyfikacji takich treści mają kluczowe znaczenie nie tylko dla zapewnienia bezpiecznego środowiska online, ale również dla ochrony użytkowników przed potencjalnymi zagrożeniami psychicznymi i emocjonalnymi.

Autor pracy, dążąc do zaproponowania nowatorskich rozwiązań w zakresie wykrywania tekstów obraźliwych, stawia sobie za cel opracowanie metod, które będą uwzględniały indywidualne preferencje i wrażliwość użytkowników. Tego rodzaju spersonalizowane podejście może znacznie zwiększyć skuteczność systemów filtrujących, dostosowując je do specyficznych potrzeb i wymagań różnych grup odbiorców. Jednocześnie takie podejście znacząco komplikuje samo zadanie od strony technicznej. Dlatego podejście do tego tematu od tej strony było bardzo dobrze wybranym zagadnieniem naukowym na pracę doktorską.

W recenzowanej pracy przedstawione są szczegółowe analizy i eksperymenty, które pozwoliły na opracowanie i przetestowanie proponowanych metod na wybranych korpusach danych i w określonych reżimach anotowania poziomu obraźliwości. Autor wykazuje się nie tylko głęboką wiedzą teoretyczną, ale także umiejętnością praktycznego zastosowania zdobytych informacji w celu rozwiązania rzeczywistych problemów. Niniejszy wstęp stanowi jedynie wprowadzenie do szerokiego spektrum zagadnień poruszanych w pracy, które zostaną omówione w dalszych częściach recenzji.

Jaki jest najważniejszy problem rozważany w rozprawie?

Najważniejszym problemem rozważanym w rozprawie doktorskiej zatytułowanej "Metody spersonalizowanego rozpoznawania tekstów obraźliwych" jest poprawa jakości predykcji metod rozpoznawania tekstów obraźliwych poprzez uwzględnienie indywidualnych cech użytkowników. Autor bada, jak różne podejścia, takie jak spersonalizowane modele sieci neuronowych oraz kontekstowe metody reprezentacji tekstu, mogą wpłynąć na skuteczność rozpoznawania obraźliwych treści. Kluczowe hipotezy sugerują, że uwzględnienie reprezentacji użytkownika może znacząco poprawić jakość predykcji, a spersonalizowane modele sieci neuronowych cechują się wyższą skutecznością w porównaniu z metodami uogólnionymi. Kontekstowe metody reprezentacji tekstu, takie jak architektury typu transformer, są bardziej skuteczne niż metody bezkontekstowe, a język oraz

rodzaj treści (komentarze, tweety, dłuższe wypowiedzi) wpływają na jakość predykcji. Ponadto, metody spersonalizowanego rozpoznawania tekstów obraźliwych charakteryzują się lepszą jakością predykcji niezależnie od rozmiaru zbioru uczącego, a spersonalizowane metody aktywnego uczenia pozwalają na osiągnięcie zbliżonej jakości predykcji przy mniejszym zbiorze danych. W zadaniu rozpoznawania obraźliwych treści generatywne modele ogólnego przeznaczenia, takie jak ChatGPT-3.5, mogą mieć mniejszą skuteczność niż dedykowane modele nauczone na specyficznych zbiorach danych.

Autor dąży do wykazania, że podejście spersonalizowane, które uwzględnia indywidualne cechy i preferencje użytkowników, może znacząco poprawić skuteczność wykrywania obraźliwych treści w porównaniu do tradycyjnych, uogólnionych metod. To zagadnienie ma kluczowe znaczenie w kontekście zapewnienia bezpieczniejszego środowiska online oraz ochrony użytkowników przed nieodpowiednimi treściami.

Czy ma on charakter naukowy?

Zarówno sam temat rozprawy, biorąc pod uwagę jego mocne zawężenie, jak i podjęte problemy badawcze mają charakter naukowy. Naukowy charakter mają również przeprowadzone pracochłonne badania eksperymentalne połączone z rzetelną analizą ich wyników.

Czy ma on znaczenie praktyczne?

Zarówno temat, faktyczny cel pracy jak i wyniki przeprowadzonych badań eksperymentalnych mają znaczący wymiar praktyczny, szczególnie dla środowiska inżynierii języka naturalnego a także dla osób zajmujących się badaniem mowy nienawiści w internecie od strony nauk społecznych.

2. Wkład autora

Autor niniejszej pracy doktorskiej w znaczący sposób przyczynił się do rozwoju metod spersonalizowanego rozpoznawania tekstów obraźliwych, wykorzystując zaawansowane techniki uczenia maszynowego oraz głębokich sieci neuronowych. Celem jego badań było zbadanie, jak indywidualne preferencje użytkowników mogą wpłynąć na dokładność wykrywania obraźliwych treści w tekstach. Badania te obejmowały zastosowanie zarówno klasycznych technik, jak i nowoczesnych modeli generatywnych, takich jak transformery.

Autor wykazał, że uwzględnienie reprezentacji użytkownika, jego preferencji oraz kontekstu, w jakim odbiera on daną wypowiedź, znacząco poprawia skuteczność predykcji. Przedstawione metody pozwoliły na bardziej precyzyjne rozpoznawanie tekstów obraźliwych w porównaniu do tradycyjnych podejść uogólnionych, które nie brały pod uwagę indywidualnych różnic pomiędzy użytkownikami. Zastosowanie metod aktywnego uczenia oraz modeli generatywnych również znacząco wpłynęło na poprawę efektywności klasyfikacji, szczególnie w kontekście mniejszych i bardziej zróżnicowanych zbiorów danych.

Autor przeprowadził liczne eksperymenty, w których testował różne metody generowania reprezentacji wektorowych tekstu, liczbę foldów uczących oraz wykorzystanie aktywnego uczenia. Wnioski z tych badań wskazują, że zaawansowane metody, takie jak BERT i DeBERTa, przewyższają klasyczne techniki, takie jak CBOW i Skipgram, zarówno pod względem dokładności predykcji, jak i zdolności do uchwycenia kontekstu tekstu. Ponadto, badania dotyczące liczby foldów uczących wykazały, że wraz ze wzrostem liczby foldów jakość modeli predykcyjnych znacząco się poprawia. Metoda UserEmb osiągnęła najlepsze wyniki przy ośmiu foldach, uzyskując wartość miary F1 macro na poziomie 81.68 na zbiorze WikiDetox: Toxicity. Aktywne uczenie było kluczowym elementem badań, a różne strategie aktywnego uczenia, takie jak VarRatio i Ratio Distance, pozwalały na dynamiczne dostosowanie modelu do nowych danych.

Wnioski z eksperymentów potwierdzają, że spersonalizowane metody aktywnego uczenia znacząco poprawiają jakość predykcji, szczególnie w kontekście ograniczonych zbiorów danych. Dodatkowo,

autor wykazał, że generatywne modele językowe, takie jak ChatGPT-3.5, mogą być użyteczne w kontekście spersonalizowanego rozpoznawania treści obraźliwych, choć ich efektywność jest niższa w porównaniu z wyspecjalizowanymi modelami trenowanymi na specyficznych zbiorach danych.

Podsumowując, praca ta stanowi istotny wkład w rozwój metod spersonalizowanego przetwarzania języka naturalnego, a zwłaszcza w rozpoznawanie obraźliwych treści tekstowych. Wykorzystanie reprezentacji użytkownika oraz personalizacji pozwala na lepsze dostosowanie systemów moderacji treści do potrzeb użytkowników, co przyczynia się do bardziej efektywnego eliminowania niepożądanych treści w przestrzeni internetowej.

3. Poprawność

Czy stwierdzenia zawarte w rozprawie są godne zaufania?

Zasadnicza struktura samej rozprawy jak i procesu badawczego została zaprojektowana prawidłowo. Badania eksperymentalne zostały przeprowadzone zgodnie ze sztuką, a przy przyjętych założeniach osiągnięte wyniki są wiarygodne. Autor wykazał dużą staranność w prowadzeniu badań oraz w analizie wyników, co dodatkowo wzmacnia zaufanie do przedstawionych konkluzji. Wyniki uzyskane w trakcie eksperymentów były analizowane przy użyciu odpowiednich narzędzi statystycznych, co potwierdza ich rzetelność i poprawność. Wszystkie te aspekty składają się na solidne podstawy, które przemawiają za wiarygodnością stwierdzeń zawartych w pracy.

Czy uzasadnienia są poprawne?

Uzasadnienia zawarte w pracy doktorskiej są poprawne i solidnie poparte przeprowadzonymi badaniami. Autor przedstawił wyniki eksperymentów w sposób szczegółowy, używając licznych danych liczbowych, metod oraz różnych korpusów tekstowych. Choć takie podejście zapewnia precyzyjność i rzetelność przedstawionych wyników, to może również stanowić wyzwanie dla czytelników. Ze względu na obszerność wyników i złożoność zastosowanych metod, przyswojenie i pełne zrozumienie wszystkich przedstawionych danych wymaga znacznego wysiłku. Ogrom wyników liczbowych, różnorodność korpusów oraz różnorodność metod sprawiają, że lektura pracy jest wymagająca. Niemniej jednak, takie podejście jest konieczne dla zapewnienia naukowej rzetelności i wiarygodności przedstawionych konkluzji w tematyce wybranej do ocenianej pracy doktorskiej. Autor wykazał się umiejętnością dokładnego i skrupulatnego przedstawiania wyników, co potwierdza poprawność uzasadnień i wniosków zawartych w pracy.

Uwagi szczegółowe:

Str. 23 W opisie wzoru przydatnym byłoby zamieszczenie charakterystyki poszczególnych wag warstw gęstych i omówienie ich specyfiki.

Str. 25. Literówka: wartości, a nie “wartości”.

Str. 28 Wyjaśnienia “ReLU oznacza funkcję aktywacji ReLU” czy “Softplus oznacza funkcję Softplus” nic nie wnoszą. W takiej sytuacji lepiej nic nie pisać albo zrobić krótkie wyjaśnienie celu użycia danej metody, np Softplus pozwala operować na dodatnich wartościach.

Str. 53 Błędne formatowanie, cytowanie - “(Liu” wchodzi na margines.

Str. 98 Eksperymenty w swoim charakterze niczego nie udowadniają. To robi odpowiedni dowód matematyczny lub analiza wszystkich przypadków, która w tym przypadku nie wydaje się możliwa. Eksperymenty wskazują, oznaczają itd.

4. Wiedza kandydata

W kontekście współczesnych badań nad NLP, istotnym aspektem jest zespołowy charakter prac, co sprawia, że ocena wiedzy kandydata jedynie na podstawie lektury pracy doktorskiej jest trudna. Prace nad NLP często angażują liczne zespoły badawcze, które wspólnie rozwijają i udoskonalają nowe technologie oraz metody, co patrząc po publikacjach Autora miało w tym przypadku miejsce. W takim środowisku indywidualny wkład kandydata może być trudny do jednoznacznego wyodrębnienia bez szerszego kontekstu dotyczącego zespołowej współpracy. Nie jest to jednak negatywna ocena kandydata, a raczej stwierdzenie faktu wynikającego z realiów współczesnych badań nad NLP, gdzie kolaboracja i wspólne działania stanowią kluczowy element sukcesu. Warto zatem uwzględnić te okoliczności podczas oceny kompetencji i wkładu kandydata, aby uzyskać pełniejszy obraz jego umiejętności i osiągnięć. Jednocześnie, należy zwrócić uwagę że samo doświadczenie w pracy w zespole stanowi wiedzę Autora którą mógł w ten sposób nabyć.

Autor przedstawił liczne i różnorodne metody NLP, którymi sprawnie się posługiwał w trakcie prowadzonych eksperymentów. Koncepcja pewnego rodzaju kompletności i kompleksowości przetestowania możliwości w zakresie spersonalizowanego rozpoznania tekstów obraźliwych także wskazuje na wysoką wiedzę kandydata na temat projektowania procesu szukania rozwiązań NLP określonego problemu. Miałem także możliwość przeprowadzić krótką rozmowę z Autorem, której przebieg potwierdzał zaawansowaną znajomość omawianych metod i korpusów danych.

Jaka jest opinia recenzenta o bibliografii? Na ile bibliografia jest kompletna?

Bibliografia jest bogata i dobrze dobrana. W mojej ocenie, biorąc pod uwagę współczesne realia badań nad NLP i różnego rodzaju istniejących systemów i bibliotek trudno mówić o kompletności przeanalizowania bibliografii. Ponownie, nie jest to jednak negatywna ocena Autora i jego pracy a jedynie komentarz do istniejących realiów.

5. Podsumowanie

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez Prawo o szkolnictwie wyższym i nauce (Dz.U.2023 poz. 742) moja ocena rozprawy pod względem podstawowych kryteriów jest następująca:

- rozprawa zawiera oryginalne rozwiązanie problemu naukowego, które może mieć zastosowanie w sferze gospodarczej i społecznej,
- kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie,
- kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej.

Podsumowując, praca mieści się w dyscyplinie Informatyka Techniczna i Telekomunikacja i spełnia w dostatecznym stopniu formalne i zwyczajowe wymagania stawiane rozprawom doktorskim. Wnoszę o dopuszczenie Doktoranta do dalszych faz przewodu doktorskiego oraz rozważenie oceny z wyróżnieniem.

Podpis