

Streszczenie

Ocena wiarygodności treści medycznych pojawiających się w Internecie staje się w XXI wieku coraz bardziej palącą potrzebą. Z jednej strony, niezliczone ilości danych publikowanych codziennie online nie pozwalają na ręczną ocenę treści przez ekspertów dziedzinowych. Z drugiej strony, decyzje podejmowane w oparciu o fałszywe zalecenia medyczne mogą być na tyle poważne w skutkach, że ostateczną klasyfikację wiarygodności powinien podejmować człowiek.

Niniejsza rozprawa doktorska zawiera opis prac mających na celu usprawnienie procesu oceny medycznych treści internetowych przez ekspertów. Prace te stanowią przygotowanie do stworzenia wspieranego przez ekspertów, półautomatycznego systemu wychwytywania i tagowania niewiarygodnych stwierdzeń medycznych pojawiających się w Sieci. Przeprowadzono trzy eksperymenty, których celem było zebranie danych (treści medyczne wraz z ocenami eksperckimi). Następnie przeprowadzono cztery analizy, każda opisana w oddzielnym artykule. Artykuły stanowią załączniki do niniejszej rozprawy. Pierwszy eksperyment polegał na ocenie pojedynczych zdań w dwóch trybach - z i bez znajomości kontekstu całego artykułu. Wyniki analizy wykazały dużą trudność ekspertów w ocenie zdań bez kontekstu, w związku z czym przeprowadzono drugi eksperyment. W drugim eksperymencie przetestowano cztery różne metody wzbogacania kontekstu pojedynczego zdania. W jego wyniku zdefiniowana została jednostka tekstowa dla ocenianych treści - trójka zdań wraz ze słowami kluczowymi. Pierwszy artykuł stanowi opis dwóch wyżej wymienionych eksperymentów oraz analizy danych. W drugim artykule opisano trzeci eksperyment - stworzenie zbioru danych, w którym wybrane jednostki tekstowe, wyodrębnione z popularnonaukowych artykułów medycznych, ocenione zostały przez ekspertów dziedzinowych (psychiatria, ginekologia, kar-

diologia i pediatria). Dane udostępniono. Drugi artykuł opisuje również analizę polegającą na wykryciu schematów retorycznych, które wprowadzają ekspertów w błąd, wypaczając ich ocenę wiarygodności. Trzeci artykuł opisuje stworzone klasyfikatory filtrujące, mające na celu maksymalizację efektywności eksperta pracującego przy anotacji. Czwarty artykuł opisuje badanie możliwości interpretacyjnych wyników zwracanych przez algorytmy filtrujące.

Wyniki analizy jakościowej wykazują istnienie powtarzalnych schematów retorycznych, które pojawiają się w niewiarygodnych treściach medycznych. Istnieją schematy zarówno podobne do tych rozpoznawanych w dezinformacji pozadzielnowej, jak i specyficzne dla popularnonaukowych treści medycznych. Wykorzystanie klasyfikatorów pozwala na wstępną filtrację, która przyspiesza wykrywanie treści niewiarygodnych przez anotującego eksperta średnio dwukrotnie, w zależności od dziedziny. Możliwości objaśniające klasyfikatorów zależą od stopnia kompresji danych wejściowych. Lepsza generalizacja wyników uniemożliwia wgląd w decyzje związane z atrybutami semantycznymi, podczas gdy mniejsza generalizacja to umożliwia, ale wymaga budowy oddzielnych klasyfikatorów dla wąskich tematycznie dziedzin.

Teoretyczny system, w którym odpowiednio szerokie grono ekspertów oceniałoby wszystkie dane publikowane w Internecie w czasie rzeczywistym w praktyce nie jest możliwy do wykonania. Dlatego moje wysiłki skupione były wokół maksymalizacji przepustowości wspieranego przez eksperta systemu ocen. Udało się uzyskać polepszenie przepustowości dwutorowo. Po pierwsze, wyniki eksperymentów pozwoliły wyodrębnić fragmenty tekstów medycznych - trójki zdań - które są na tyle małe, aby stanowiły sensowną jednostkę do zbierania danych za pomocą crowd-sourcingu, jednocześnie będąc na tyle złożone, aby ekspert oceniający nie tracił kontekstu potrzebnego do oceny. Po drugie, stworzone klasyfikatory fil-

trujące sprawiają, iż ekspert etykietujący tę samą liczbę treści co bez korzystania z filtracji, jest w stanie wyłapać średnio dwukrotnie więcej niewiarygodnych przykładów.

Ponadto, analiza algorytmów filtrujących pozwala na dobranie takich parametrów, aby uzyskać pożądaną informację zwrotną dla użytkownika końcowego.

Analiza jakościowa uzyskanych etykiet wiarygodności wskazuje na to, iż ekspercka ocena jest w pewnym stopniu wypaczana poprzez skrzywienia poznawcze. Wnioski te definiują nowe kierunki badań z zakresu psychologii dezinformacji, wymagane do stworzenia rzeczzonego systemu.