

Metoda optymalizacji pracy eksperta przy anotowaniu niewiarygodnych treści medycznych

Aleksandra Nabożny

Bartłomiej Balcerzak

Mikołaj Morzy

Adam Wierzbicki

Cel pracy

Automatyczna lub półautomatyczna ocena wiarygodności tekstów medycznych.

Główne założenia

- Zdanie / stwierdzenie jako jednostka informacji.
- Wiarygodność, a nie prawdziwość.
- Ocenie poddawana jest wyłącznie treść tekstowa.

Istniejące rozwiązania

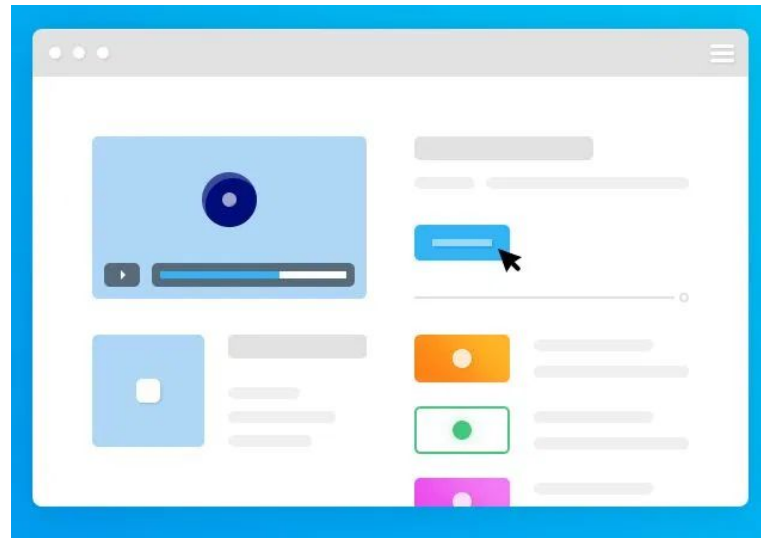
Afsana, F., Kabir, M. A., Hassan, N., & Paul, M. (2020). **Automatically Assessing Quality of Online Health Articles**. *IEEE Journal of Biomedical and Health Informatics*.

- Precyzja do 90%
- algorytm: SVM
- cechy: LIWC, wordCount, zgodność nagłówek z treścią, TF-IDF, NER, analiza połączeń w sieci (np. ranking Alexa)

Ograniczenia podejść opartych o analizę całych dokumentów WWW

Analiza źródeł: wiarygodne źródła zawierające nieprawdziwe informacje; niewiarygodne źródła, które czasem “mają rację”.

reklamy / analiza dyskursu / struktura strony: inne formy publikacji w Internecie (fora, media społecznościowe)



vs.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute ...

Nasza propozycja

Analiza krótkich fragmentów tekstu i ich oddzielna klasyfikacja.

Koncepcja

Stwierdzenia wiarygodne i niewiarygodne można grupować. Odległość nowej próbki (fragmentu tekstu) od klastra “wiarygodnego” lub “niewiarygodnego” może wskazywać na (nie)wiarygodność tej próbki.

Pierwsze wyzwanie: zebranie danych

Dlaczego to trudne?

1. \$\$
2. problem z kontekstem
3. niespójność ocen



Nasze rozwiązania

1. (\$\$) no cóż... :)

2. (*problem z kontekstem*)

słowa kluczowe + dostęp do zdań okalających

3. (*niespójność ocen*)

tagi, przykłady, dobry protokół...

SHOW INSTRUCTION

Keywords:

section

baby

delivery

emergency

vaginal

An emergency C-section might be done if: the placenta separates from the uterine wall too soon (called placental abruption).

- credible
- noncredible
- neutral/irrelevant

BACK

I NEED MORE CONTEXT

ABANDON

SAVE

SHOW INSTRUCTION

Keywords: [section](#) [baby](#) [delivery](#) [emergency](#) [vaginal](#)

Some C-sections are unexpected emergency deliveries done when complications arise with the mother and/or baby during pregnancy or labor.

An emergency C-section might be done if: the placenta separates from the uterine wall too soon (called placental abruption).

Some C-sections are unexpected emergency deliveries done when complications arise with the mother and/or baby during pregnancy or labor.

- credible
- noncredible
- neutral/irrelevant

BACK

I NEED MORE CONTEXT

ABANDON

SAVE

Protokół anotacyjny

Pytanie kontrolne

“If the patient asked you if he or she should trust this statement, would you say yes or no?”

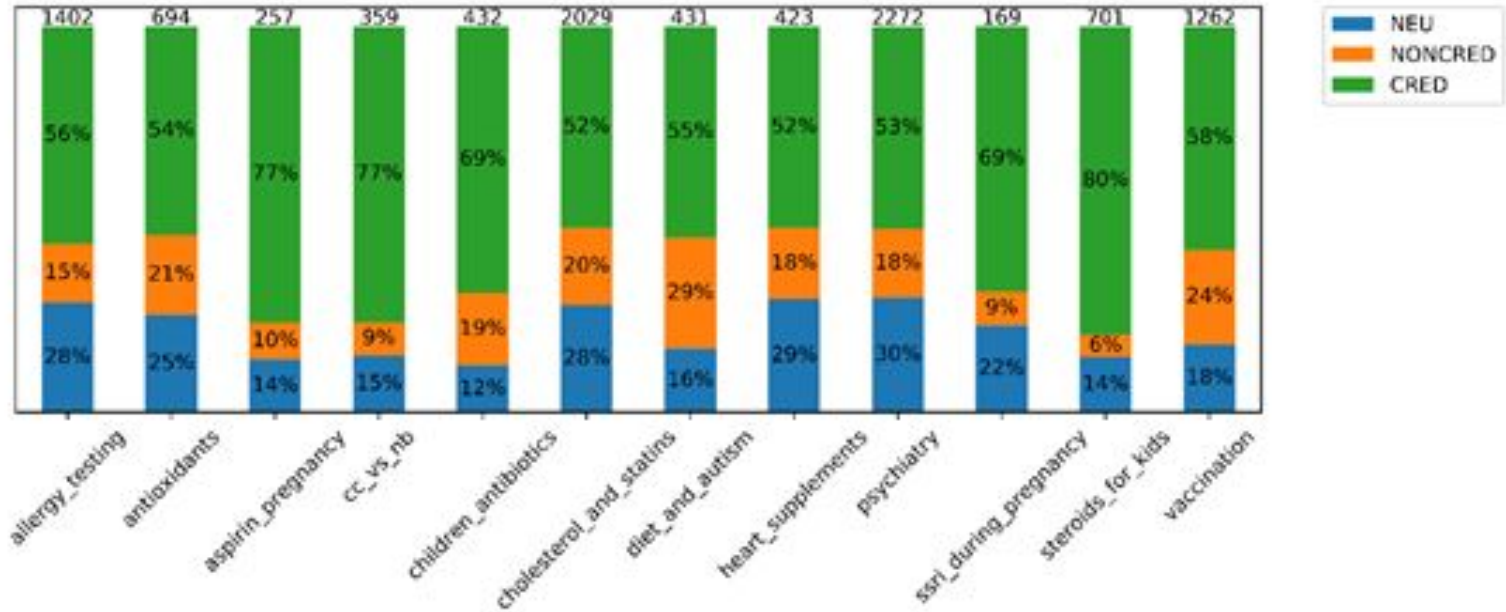
Etykiety

CRED (credible) - a sentence is reliable; does not raise major objections; contains verifiable information from the medical domain;

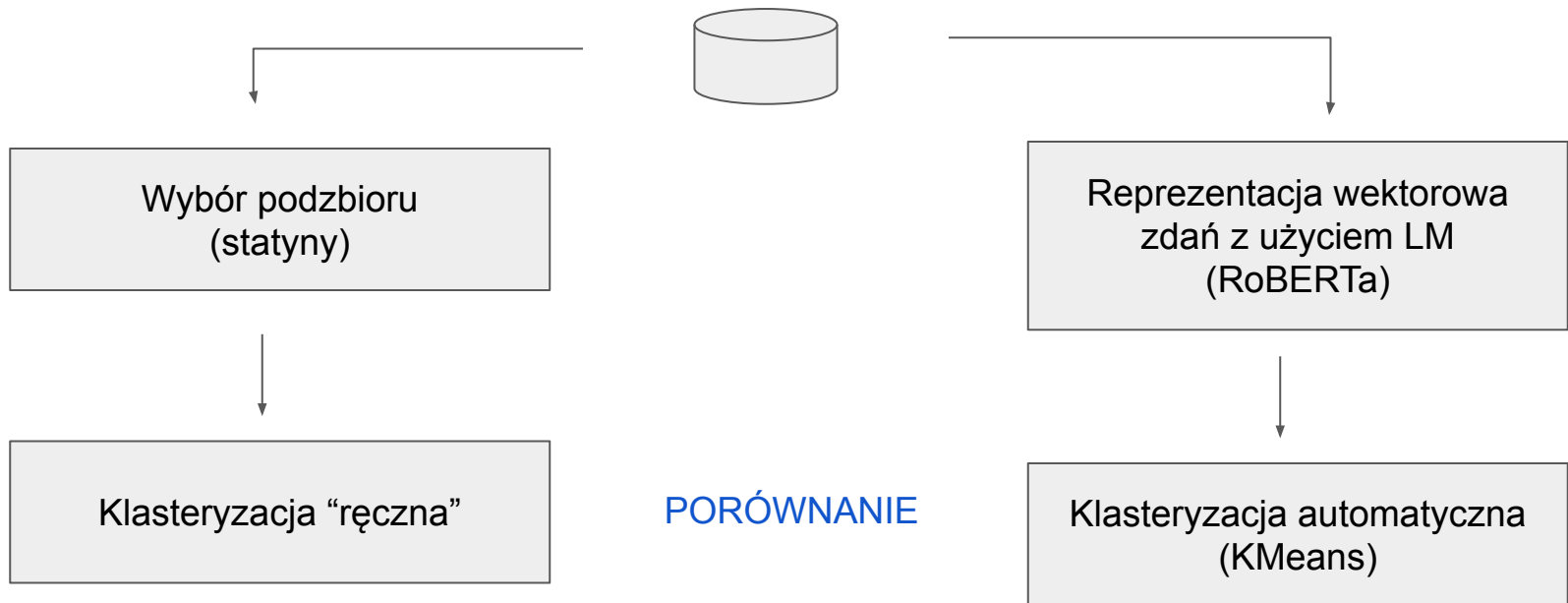
NONCRED (not credible) - a sentence contains false or unverifiable information; contains persuasion contrary to current medical recommendations; contains outdated information;

NEU (neutral) - a sentence does not contain factual information (e.g., is a question); is not related to medicine;

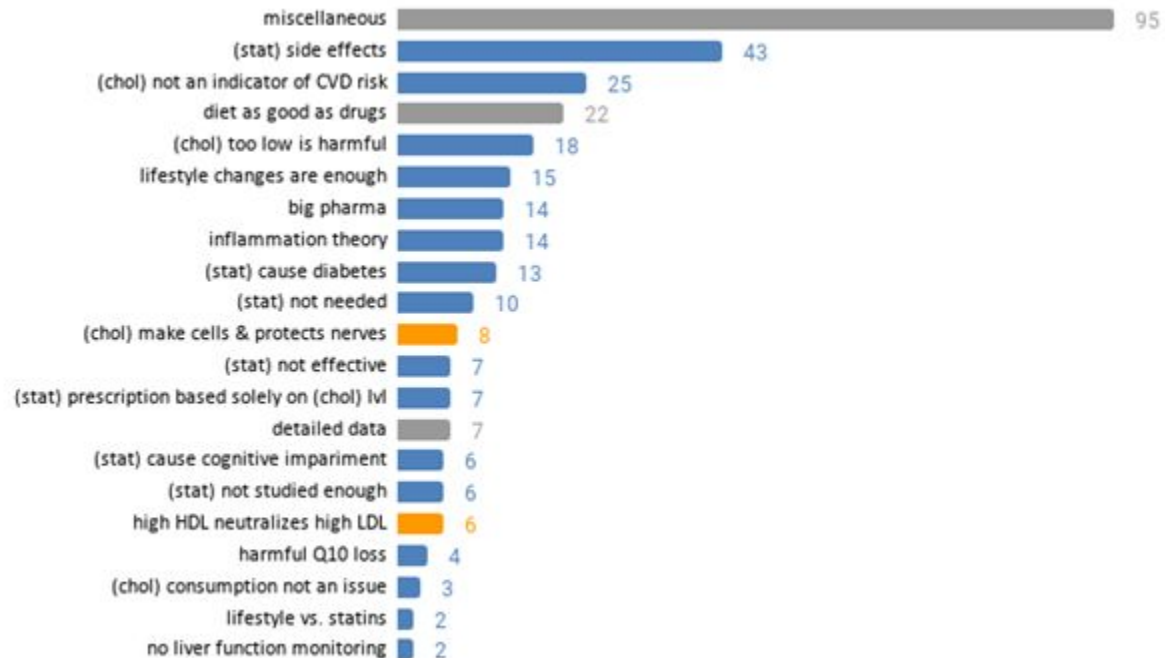
Podsumowanie zbioru



Analiza danych - jakościowa i ilościowa



Klasteryzacja ręczna



Klasteryzacja automatyczna i ręczna - porównanie

Modele językowe nie radzą sobie z reprezentacją negacji ani “pewnością” argumentacji.

Przykłady:

“Statyny są szkodliwe”

“Statyny nie są szkodliwe”

--

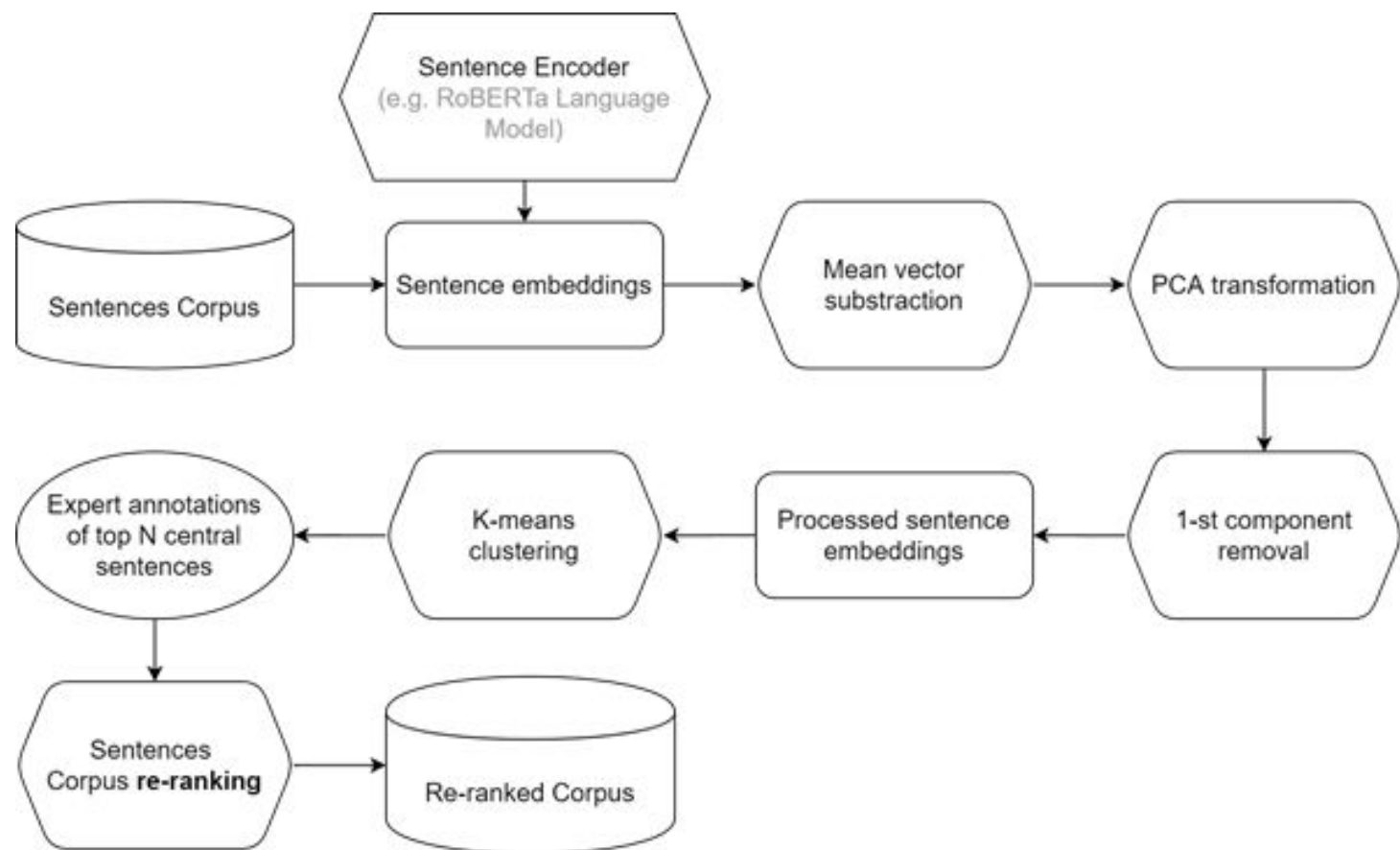
“Zażywanie statyn powoduje bóle mięśni.”

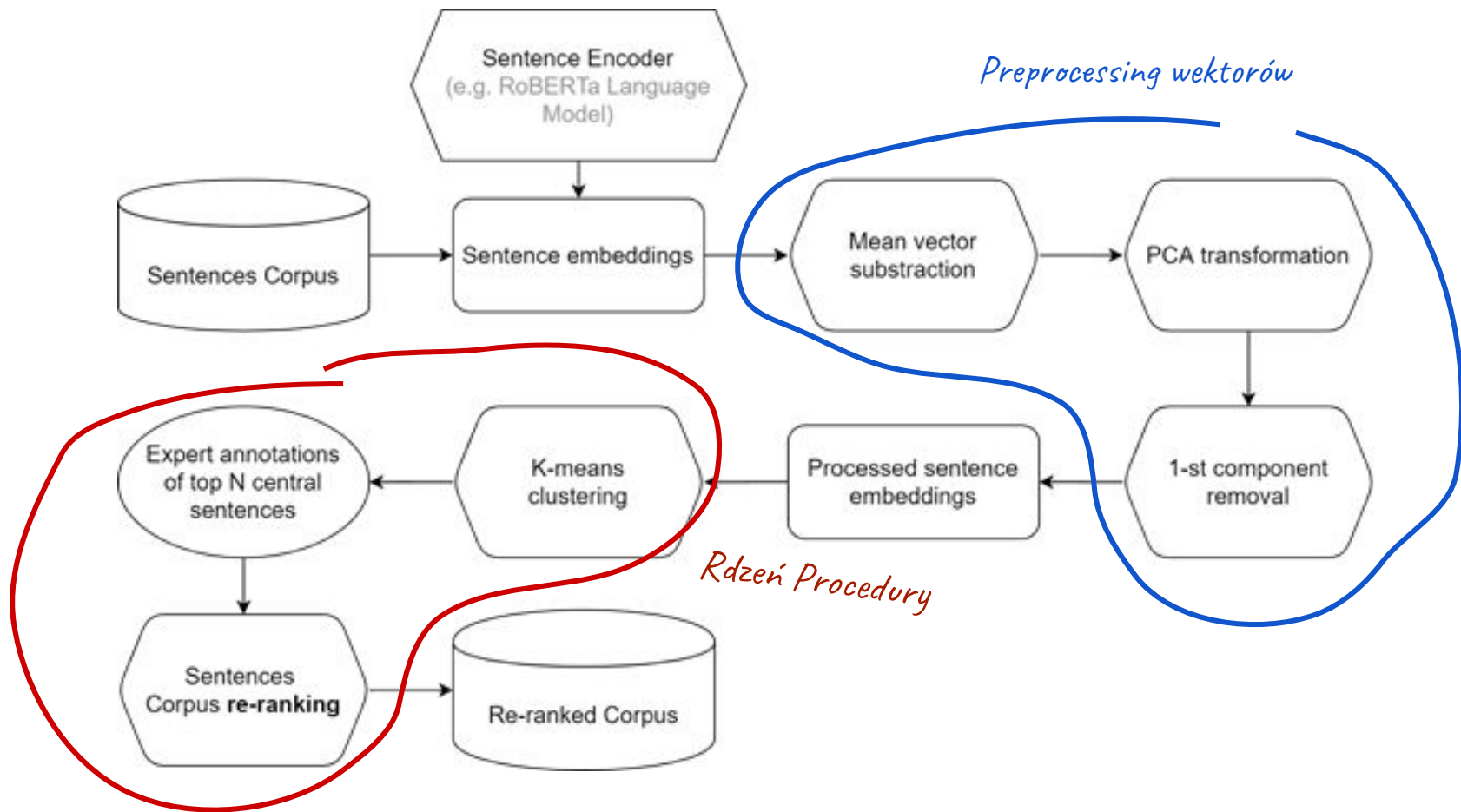
“Zażywanie statyn rzadko powoduje bóle mięśni”

Procedura szeregowania zdań poddawanych ocenie

Wektory dystrybucyjne “niedotrenowane” do rozpoznawania specyficznych zjawisk (negacja, słowo zmiękczone wydźwięk) nie mogą być użyte do automatycznej klasyfikacji wiarygodności. Ale czy mimo tego mogą być przydatne?

Tak!



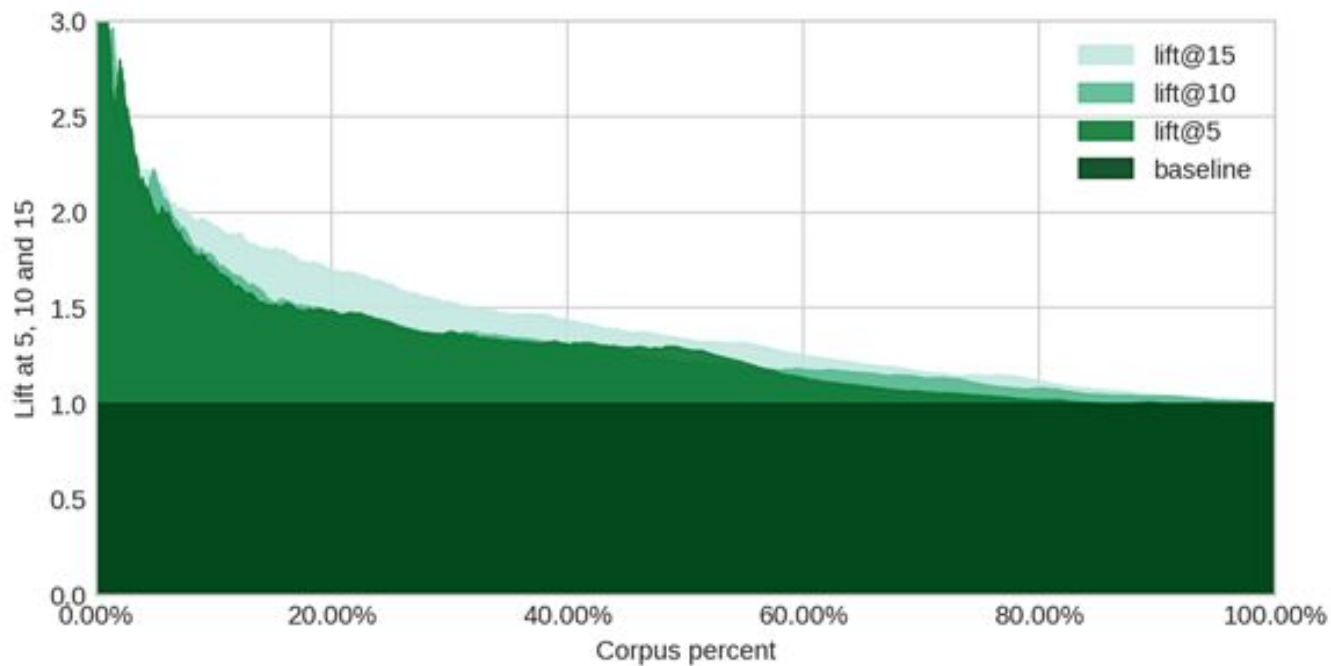


Wzór szeregujący i miara skuteczności

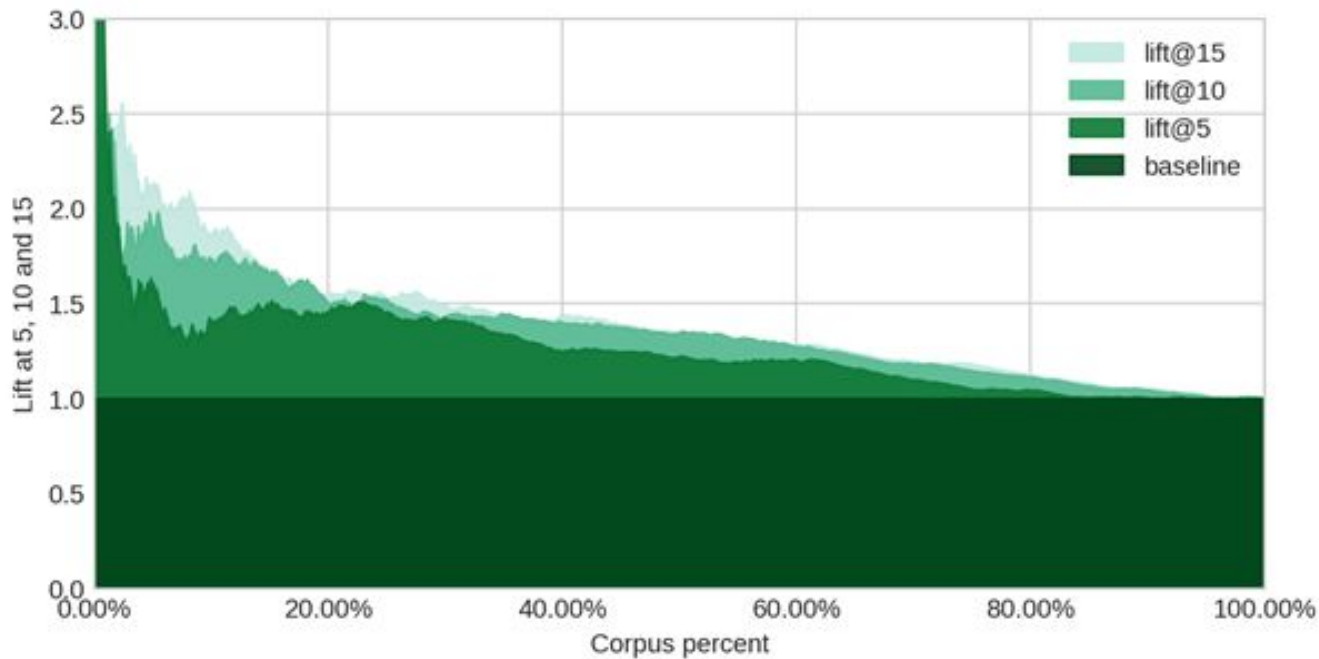
$$score@k = \frac{1}{e^{-p(n)+p(u)}} + \frac{1}{W^{p(u)+1}}$$

$$lift@p = \frac{N}{p} recall@p$$

Wyniki: pełen zbiór



Wyniki: zbiór “statynowy”



Analiza jakościowa - raz jeszcze

- Wgląd w ograniczenia związane ze stosowaniem wektorów dystrybucyjnych opartych o deep learningowe modele językowe.
- Wgląd w przyczyny niespójności ocen ekspertów.

Wykryte “narracje” (zniekształcenia):

- 1. Slippery slope**
- 2. Hedging**
- 3. Alleged negative consequences**
- 4. Twisting word**

Slippery slope

Zdanie jest generalnie prawdziwe, ale konsekwencje prezentowanego faktu są wyolbrzymione.

“Hence, while the drug might synergise with a statin to prevent a non-fatal (or minor) heart attack, it seems to increase the risk of some other equally life-threatening pathology, resulting in death”

“Cholesterol also helps in the formation of your memories and is vital for neurological function.”

Hedging

Zdanie jest dalekie od prawdy, ale jego część zmiękcza wydźwięk prezentowanego stwierdzenia.

“However, cholesterol content should be less of a concern than fat content” (CRED)

“Coenzyme Q10 supplements may help prevent statin side effects in some people, though more studies are needed to determine any benefits of taking it.” (CRED)

The FDA warns on statin labels that some people have developed memory loss or confusion while taking statins. (CRED)

Alleged negative consequences

Zdanie jest w większości prawdziwe, ale ekspert może mieć wątpliwość, czy prezentowanie go jako argumentu nie powstrzyma pacjenta od postępowania zgodnie z zaleceniami medycznymi.

“For starters, statin drugs deplete your body of coenzyme Q10 (CoQ10), which is beneficial to heart health and muscle function”

“Cholesterol is a waxy, fatty steroid that your body needs for things like: cell production.”

Twisting word

Pojawienie się jednego słowa/wyrażenia zmienia wydźwięk całego zdania.

*“Statins may **slightly** increase the risk for Type 2 diabetes, a condition that can lead to heart disease or stroke” (CRED)*

*“For example, it may be **enough** to eat a nutritious diet, exercise regularly, and avoid smoking tobacco products.” (NONCRED) vs.
“Eating a healthy diet and doing regular exercise **can help** lower the level of cholesterol in your blood.” (CRED)*

Wnioski

- Jeszcze wiele pracy powinno zostać włożonej w dostrojenie protokołu anotacyjnego.
- Metoda rankingująca może pomóc efektywniej wykorzystać czas i wysiłki ekspertów przy tagowaniu zbioru.
- Automatyczna ocena treści medycznych powinna wychodzić poza “fact checking”. Musi zawierać element wykrywania narracji.

Dalsze prace

→ RDF + MetaMap (UMLS) + BioBert?

→ Model zbiorczy + modele składowe?

Dziękuję za uwagę!

mgr inż. Aleksandra Nabożny
aleksandra.nabozny@pja.edu.pl
github.com/alenabozny/medical_credibility_corpus