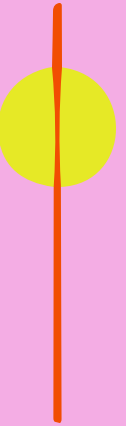


DETEKTOR SARKAZMU



Karol Kusper | Szymon Jagła | Jakub Kubica




AGENDA



- Krótko o sarkaźmie
- Jak rozpoznajemy sarkazm?
- Trudności
- Zasada działania detektora sarkazmu
- Jak stworzyć detektor sarkazmu
- Przykłady zastosowań
- Podsumowanie



Krótko o sarkazmie



Sarkazm to wyrażenie językowe, które często służy do komunikowania przeciwnych intencji w stosunku do tego, co się mówi, zazwyczaj mając na celu wyrażenie czegoś bardzo nieprzyjemnego z zamiarem obrażenia lub wyśmiania. Sarkazm nie zawsze objawia się w ten sam sposób, co sprawia, że nawet ludzie nie zawsze go rozumieją.

JAK ROZPOZNAJEMY SARKAZM?



Nie ma jednego sposobu na rozpoznawanie sarkazmu, ponieważ każdy używa go na swój sposób



TRUDNOŚCI PRZY WYKRYWANIU SARKAZMU

- Kontekst wypowiedzi
- Sposób akcentowania pewnych słów
- Nawiązania do popkultury
- Mowa ciała
- Lub...brak jasnego przekazu o użytym sarkazmie



ZASADA DZIAŁANIA

- Z góry ustalone wzorce lub reguły, takie jak określone zasady gramatyczne, etykiety czy reguły metaforyczne.
- Użycie Semantycznych korelacji, czyli mierzenie jak znaczenia słów są ze sobą powiązane, z wykorzystaniem WordNet bazy danych językowych używanej do analizy relacji między słowami.
- Zastosowanie tradycyjnych metody uczenia maszynowego: SVM (maszyny wektorów nośnych), drzewa decyzyjne, czy naiwny klasyfikator Bayes'a, który może analizować częstotliwość słów lub fraz w sarkastycznych i niesarkastycznych tekstach, aby obliczyć prawdopodobieństwo, że dane zdanie jest sarkastyczne.
- Wprowadzenie pretrenowanych modeli BERT i GPT: Zwiększyło to możliwości analizy emocji i kontekstu w zadaniach przetwarzania języka naturalnego.
- Zastosowanie grafowych sieci konwolucyjnych (GCN): Wykorzystywane do klasyfikacji tekstów poprzez reprezentację relacji emocjonalnych i składniowych jako grafów.

1. Załadowanie zestawu danych np. z Kaggle.
2. Tokenizowanie zdań za pomocą NLTK.
3. Lematyzacja słowa Np. "Trick or treating, in general, is just weird [...]", został przekształcony na „tricktreatgeneralweird”.
4. Przetwarzanie danych - regresja logistyczna (LR) oraz maszyna wektorów nośnych (SVM).
5. Aby uzyskać różne reprezentacje cech, dane zostały poddane przekształceniom TF-IDF oraz Word2Vec.
6. Po przetworzeniu danych, uczenie maszynowe zostało zastosowane do analizy tekstu i klasyfikacji danych jako sarkastyczne lub niesarkastyczne..



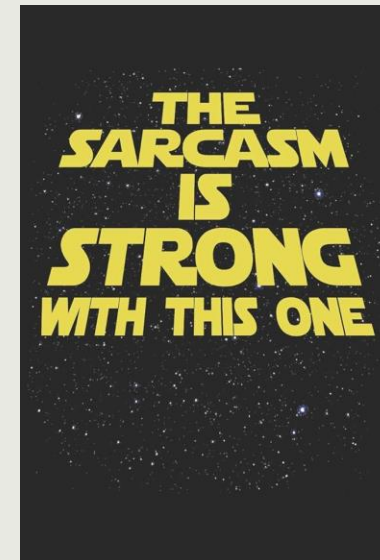
JAK STWORZYĆ DETEKTOR SARKAZMU? – PRZYKŁADOWE PODEJŚCIE

PRZYKŁADY ZASTOSOWAŃ

- Rozpoznawanie rzetelnych, niesatyrycznych źródeł wiedzy
- Klasyfikowanie wypowiedzi użytkowników stron internetowych
- Automatyczna analiza wypowiedzi różnych osób (np na podstawie transkrypcji)
- Wytrenowanie modelu do generowania żartów i wypowiedzi sarkastycznych
- Analizowanie wypowiedzi osób przesłuchiwanym lub w czasie rozpraw sądowych
- Poprawa automatycznych moderatorów stron, w celu rozróżniania mowy nienawiści od sarkazmu

PODSUMOWANIE

Sarkazm, charakteryzujący się celowym używaniem słów w celu wyrażenia przeciwieństwa ich dosłownego znaczenia, stanowi złożoną formę komunikacji. Niemniej jednak, to rozeznanie tematu przedstawia solidną ramę do utworzenia narzędzia, które pozwoli na wykrycie. Poprzez dokładne identyfikowanie przypadków sarkazmu, można przyczynić się do lepszego zrozumienia złożonych wzorców komunikacji w interakcjach online i wspomóc bardziej efektywne podejmowanie decyzji.



DZIĘKUJEMY ZA UWAGĘ

