



Inżynieria wiedzy i symboliczne uczenie maszynowe - projekt

Kacper Jurek, Mateusz Łopaciński



Cel projektu

Wybraliśmy temat: *“Rozpoznawanie pytań na podstawie analizy głosu”*

Celem projektu jest to, aby na podstawie próbek nagrań audio można było stwierdzić czy wypowiedziane zdanie jest pytaniem czy też nie.

Dataset składa się wielu fragmentów audiobooków, które zostały zakwalifikowane do kategorii “questions” albo “others”.



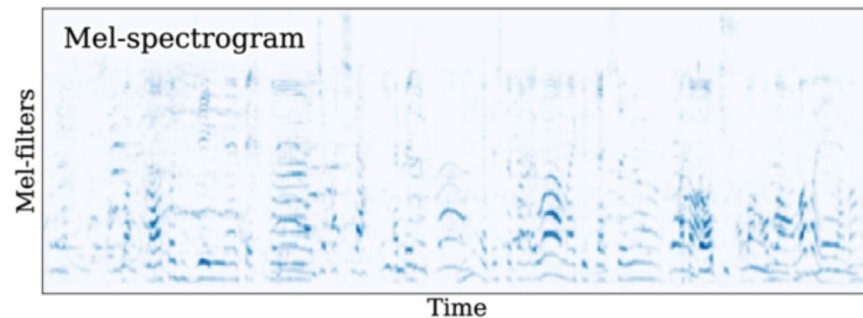
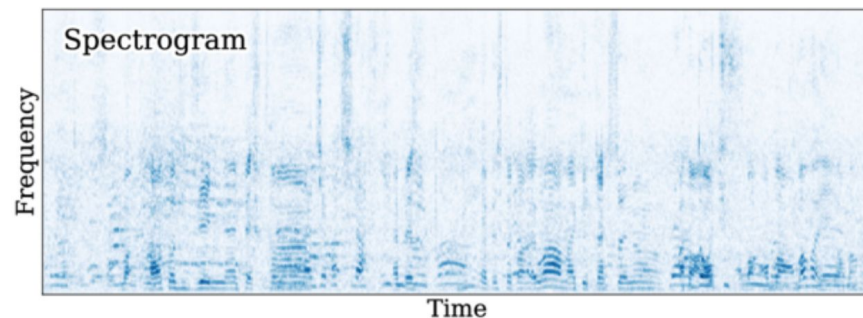
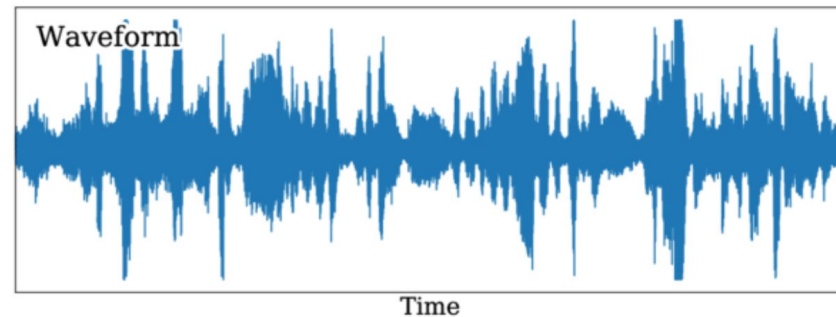
Pomysły aby rozwiązać problem

Wymyśliliśmy kilka możliwych rozwiązań naszego problemu:

Analiza intonacji na plikach audio	Analiza semantyczna wypowiedzi	Analiza spektrogramów
Intonacja czasem jednoznacznie wskazuje na pytanie, ale różne akcenty i sposób wypowiedzi może nie zawsze na to wskazywać	Istnieją słowa które wskazują na to, że dane zdanie jest pytaniem, ale język polski jest na tyle niejednoznaczny, że pytanie można zadać na wiele sposobów	Pozwala wyłapać subtelne cechy akustyczne, jednak wymaga dużo danych aby poprawnie mógł je wykrywać

Postanowiliśmy wybrać analizę spektrogramów jako rozwiązanie naszego problemu

Jest dużo typów spektrogramów które moglibyśmy wykorzystać. Po konsultacji z panem Jaroszem zdecydowaliśmy się na wybranie spektrogramów MEL - ta reprezentacja odzwierciedla w jaki sposób człowiek postrzega dźwięki. Nadaje się ona do rozpoznawania mowy, emocji, a może i pytań





Jaki jest obecny plan?

Mamy pomysł jaki ma być pipeline dla naszego problemu, na każdej fazie mamy jednak pytania:





Propozycje modeli

- CNN
 - Dobrze do wykrywania wzorców
 - Może nie zwracać uwagi na kontekst czasowy (nie uwzględnia, w którym momencie zdania się znajduje)
- CRNN
 - Popularne dla detekcji dźwięków i emocji w głosie
 - Rozpoznaje wzorce, a także zwraca uwagę na kontekst czasowy
- Transformer
 - Wymaga dużej mocy obliczeniowej i danych (można jednak spróbować czy coś uda się zrobić)
 - ...jednak może przynieść najlepsze efekty
- Lub jeszcze coś innego...

Dziękuję za uwagę!