



POLITECHNIKA WROCŁAWSKA
WYDZIAŁ PODSTAWOWYCH PROBLEMÓW TECHNIKI

Praca inżynierska

**Rozwój mobilnej aplikacji do rozpoznawania
języka migowego z wykorzystaniem
transformerów**

Development of mobile application for sign language recognition
using transformers

Autor: Maciej Grzesik

Opiekun: Prof. uczelni dr hab. Sebastian Kraszewski

Wrocław 2024

Spis treści

1	Wprowadzenie	2
1.1	Słowa kluczowe	2
1.2	Cel pracy	2
1.3	Związek z inżynierią biomedyczną	2
2	Teoria	3
2.1	Zasada działania	3
	Bibliografia	4

Rozdział 1

Wprowadzenie

1.1 Słowa kluczowe

Rozpoznawanie języka migowego, aplikacja mobilna, flutter, uczenie głębokie, transformer, wizja komputerowa

1.2 Cel pracy

Celem pracy jest stworzenie krossplatformowej aplikacji mobilnej w środowisku Flutter opartym o język programowania Dart, która wykorzystuje architekturę transformerów do rozpoznawania i tłumaczenia języka migowego na tekst w czasie rzeczywistym.

1.3 Związek z inżynierią biomedyczną

Osoby z upośledzeniem słuchu zmagają się z wykluczeniem społecznym, ze względu na brak możliwości komunikacji za pomocą konwencjonalnej mowy. Na podstawie badań Pauliny Malczewskiej wynika, że aż 90% ankietowanych osób głuchych doświadcza alienacji oraz dyskryminacji ze strony osób słyszących [1].

Bariery komunikacyjne między osobami głuchoniemymi a tymi, którzy nie znają języka migowego, stanowią znaczący problem społeczny i medyczny. Przyczynia się to do izolacji społecznej, trudności w dostępie do usług zdrowotnych i ogólnie niższej jakości życia.

Stworzenie aplikacji wspierającej osoby głuche nie tylko ułatwi komunikację, ale także przyczyni się do zwiększenia inkluzywności społecznej, zapewniając równe szanse i redukując poziom dyskryminacji.

Rozdział 2

Teoria

Transformery to zaawansowane modele uczenia maszynowego, szeroko stosowane w przetwarzaniu języka naturalnego, wizji komputerowej, ale także w audio i przetwarzaniu multimedialnym.

W odróżnieniu od poprzednio stosowanych modeli uczenia maszynowego opierających się na podejściu rekurencyjnym i mechanizmie uwagi, naukowcy pracujący dla firmy Google zaproponowali rozwiązanie wykorzystujące jedynie mechanizm uwagi. Skutkiem tego podejścia jest znaczne przyspieszenie czasu uczenia modelu przy zachowaniu wysokiej precyzji [3].

2.1 Zasada działania

Działanie transformera opiera się na warstwie kodującej oraz na warstwie dekodującej. Zadaniem warstwy kodującej jest zakodowanie argumentów wejściowych (np. tekstu, obrazów) do postaci numerycznej. Warstwa dekodująca, wykorzystuje przekazane z warstwykodującej zakodowane informacje do przetworzenia ich na wartości wyjściowe (np. do tekstu czy wideo).

Istotnym elementem, który różnicuje transformery od rekurencyjnych modeli uczenia maszynowego jest mechanizm uwagi. W tradycyjnych modelach rekurencyjnych sieci neuronowych (RNN) dane przetwarzane są sekwencyjnie tzn., że są analizowane krok po kroku (patrz Rysunek 1) [2] co skutkuje ograniczeniami m.in. w równoległości przetwarzaniu danych. Uczenie transformerów przebiega w określonej stałej ilości równolegle wykonywanych operacji, modele RNN wymagają natomiast sekwencyjnie wykonywanych operacji, co sprawia, że są mniej wydajne przy dłuższych sekwencjach.

Bibliografia

- [1] P. Malczewska. Izolacja społeczna osób z uszkodzonym słuchem jako wspólny obszar badań pedagogiki i antropologii. *Pedagogika a etnologia i antropologia kulturowa. Wspólne obszary badań*, 10(2):128, 2011. [2](#)
- [2] M. Mamczur. Czym jest i jak działa transformer sieć neuronowa?, March 2020. Dostęp: 23 maj 2024. [3](#)
- [3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023. [3](#)