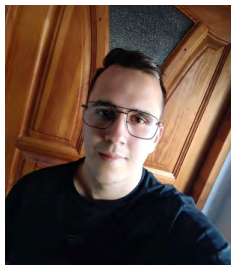


Czy rozumiesz, co do Ciebie migam?

8



Wiktor Barańczyk

Student II roku studiów inżynierskich na kierunku Informatyka na Wydziale Elektrotechniki, Elektroniki, Informatyki i Automatyki Politechniki Łódzkiej. Finalista pierwszej edycji programu *Excellence in Engineering: Talents in Research with Opportunities Programme* (E2TOP).

Komunikacja dla wszystkich?

Komunikacja między ludźmi nigdy nie była taka prosta jak teraz. W ciągu sekund możemy przesłać wiadomość na drugi koniec świata i porozumieć się z osobą, która nawet nie zna tego samego języka co my. Obecnie mamy do dyspozycji mnóstwo aplikacji tłumaczących teksty z jednego języka na inny, z najpopularniejszym tłumaczem Google'a na czele. Obsługuje on aż 133 różne języki i aktualnie trwają prace nad dodaniem kolejnych 103¹. Należy również podkreślić fakt, że z roku na rok jakość tłumaczonych tekstów jest coraz lepsza.

Prawdopodobnie, kiedy to czytasz, uważasz te informacje za oczywiste. Jeszcze niedawno każdy z nas siedział w domu i bardzo dobrze odczuwał, jak działa komunikacja przez Internet. Mimo wszystko ta ewolucja technologiczna nie objęła pewnej specyficznej, ale bardzo dużej grupy języków, jakimi są języki migowe. Czym różnią się one na przykład od języka polskiego, skoro tłumacz Google dalej ich nie obsługuje?

Podstawowym problemem jest sposób przekazu. Języki foniczne, oprócz mowy, która dalej jest czasami dość problematyczna dla algorytmów sztucznej inteligencji (nazywanych również modelami), posiadają ustrukturyzowany sposób zapisu. Sprawia on, że zebranie danych pozwalających na opracowanie tłumacza jest stosunkowo łatwe. W przypadku języków migowych nie mamy takiej dogodności (choć w historii były podejmowane próby opracowania transkrypcji dla nich, np. HamNoSys²). Dlatego, żeby ich tłumacz miał jakikolwiek sens, najpierw musi istnieć możliwość wykrywania miganych słów. Właśnie ten problem próbuję rozwiązać w projekcie „Czy rozumiesz co do

Ciebie migam?” pod kierunkiem dr inż. Piotra Ducha, realizowanym w ramach programu Politechniki Łódzkiej E2TOP. Docelowo w jego wyniku ma powstać aplikacja na telefon, która, wykorzystując obraz z kamery, ma wyświetlać tłumaczenie miganych zdań w formie tekstu.

Jak komputery postrzegają świat?

Dla większości z nas naturalne jest widzieć i interpretować to, co mamy przed oczami. Na przykład



Zaznaczone punkty charakterystyczne oraz otoczona czerwonym prostokątem dłoń, która pokazuje znak oznaczający zero lub 0.

1 „Google Translate”. https://en.wikipedia.org/wiki/Google_Translate. Dostęp: 9 stycznia 2023.

2 „HamNoSys”. <http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/hamnosys-97.html>. Dostęp: 9 stycznia 2023.



Kolejno znaki przedstawiające litery S oraz O (lub zero). Są do siebie bardzo podobne, dlatego model nie zawsze widzi między nimi różnice.

zdjęcie jest dla nas po prostu zbiorem kolorów. Czym jest jednak ono dla komputera? Po prostu liczbami. Od kilkudziesięciu lat (co dla informatyki jest niesamowicie długim czasem) naukowcy wymyślali lepsze oraz gorsze sposoby na to, jak sprawić, żeby programy mogły interpretować ten zestaw liczb. Jednak przez dłuższy czas żadne rozwiązanie nie dawało zadowalających efektów. Rewolucją w tym temacie było opracowanie sieci neuronowych. W przypadku zdjęć mówimy dokładnie o ich specyficznym rodzaju, jakim są sieci konwolucyjne. Temat sieci neuronowych jest jednak bardzo rozległy, nadający się na książkę albo nawet kilka książek. Dlatego w tym artykule nie ma miejsca na to, żebym je omówił. W Internecie znajdziecie wiele materiałów na ich temat, również po polsku. Przez wzgląd na to, że sztuczna inteligencja pojawia się praktycznie wszędzie, polecam zapoznać się chociaż z jej podstawowymi zagadnieniami.

Wszystko zaczyna się od danych

Każdy model sztucznej inteligencji potrzebuje danych, na których może się uczyć. W przypadku złożonych problemów, których przykładem jest rozpoznawanie znaków, potrzeba tych danych bardzo dużo. Na przykład GPT-3, czyli algorytm generujący tekst na bazie fragmentu innego tekstu opracowany przez OpenAI, do treningu potrzebował aż 300 miliardów słów, co przełożyło się na 570 gigabajtów danych³.

Naturalnie pierwszym pomysłem, jaki przychodzi do głowy, jest poproszenie o pomoc Głuchych. Niestety to nie wystarczy. Osób, które posługują się językiem migowym (w tym przypadku skupiłem się na Polskim Języku Migowym (PJM)), jest mało. Do tego, przez wzgląd na stosunkowo późne pojawienie się technologii, która swobodnie pozwala na komunikowanie się przy pomocy PJM, te same słowa w różnych zakątkach Polski migane są w bardzo zróżnicowany sposób. Jest to problemem

dlatego, że dla algorytmu inaczej pokazywany znak, mimo że może mieć to samo znaczenie, jest kolejną informacją, której musi się nauczyć. Tym samym zmniejsza się liczba osób, która może nam pomóc w zbieraniu danych wykorzystywanych do rozpoznawania konkretnego słowa. Dodatkowo, jeśli okroić tę grupę tylko do chętnych do współpracy, zostają nam pojedyncze jednostki.

W takim razie może wystarczy zrobić dużą liczbę zdjęć oraz nagrać wiele filmów przy pomocy tej małej grupy? Niestety nie. Algorytmy łatwo rozpoznają znaki pokazywane przez dłonie, które występowały w treningu, ale dalej mogą źle rozpoznawać słowa migane przez „obcą” rękę, czyli taką, która nie występowała w danych użytych do treningu.

Co w takim razie można zrobić? Zaangażować osoby, które nie znają PJM. Oczywiście istnieje wtedy ryzyko, że słowa nie będą migane odpowiednio. Jednak testy w pierwszym etapie projektu, tj. rozpoznawanie znaków ze zdjęć, pokazują, że model uczony na fotografiach, które przedstawiają słowa pokazywane przez osoby nieznające języka migowego, bardzo dobrze rozpoznaje znaki pokazywane przez ludzi, którzy używają ich na co dzień.

Jak wykorzystać dane?

Kiedy już mamy rozwiązany problem danych, czas na samo opracowanie algorytmu. W przypadku słów w PJM ważne jest ułożenie dłoni oraz ich położenie względem osoby mówiącej, dlatego system musi brać je pod uwagę. Dodatkowo nie może być on wrażliwy na to, w którym miejscu względem kamery będzie stać migający. Przez wzgląd na to zdecydowałem się, że pierwszym krokiem algorytmu będzie wykrywanie punktów charakterystycznych ciała (co zostało przedstawione na rysunku 2). Następnie obliczane będzie, jak oddalone są od siebie punkty. Czemu akurat takie operacje? Stosunki odległości między punktami są podobne niezależnie od tego, gdzie ktoś stoi, a dalej na ich podstawie możliwe jest wywnioskowanie ułożenia dłoni oraz ich położenia w przestrzeni.

Jeśli uda się opracować model, to czy będzie mógł być on stosowany tylko dla Polskiego Języka Migowego? Oczywiście, że nie! Z dużym prawdopodobieństwem, jeśli zostanie zebrana odpowiednia ilość poprawnych danych, algorytm będzie w stanie rozpoznawać słowa z innych języków migowych. Dodatkowo możliwe będzie wykorzystanie go do sterowania gestami i to nawet ich skomplikowanymi sekwencjami.



Logo Polskiego Związku Głuchych. Budynek oddziału Łódzkiego jest ważnym miejscem dla mojego projektu, ponieważ to w nim mogłem zebrać potrzebne dane oraz zaczerpnąć wiedzy o PJM.

³ „HamNoSys”. <http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/hamnosys-97.html>. Dostęp: 9 stycznia 2023.