Automatyczna ocena poprawności wymowy w języku angielskim

Autor: Jakub Gucik

Akademia Górniczo-Hutnicza w Krakowie
Wydział Inżynierii Mechanicznej i Robotyki
Inżynieria Akustyczna
Specjalizacja: Inżynieria Dźwięku w Mediach i Kulturze
Opiekun: dr inż. Marcin Witkowski

Kraków, 24.01.2024 r.

Wprowadzenie – Goodness of Pronunciation

01 Cele pracy

1. Analiza dostępnych metod oraz baz danych,
2. Implementacja i uruchomienie prototypu wybranej metody GoP,
3. Ocena skuteczności zaimplementowanej metody.

02 Motywacja

1.Zagadnienie mogące pomóc w diagnozie/leczeniu schorzeń i defektów akustycznych,

2.Tematyka nowa, rozwijająca się, nie rozpowszechniona szeroko na rynku,

3.Temat w zakresie zainteresowań.

Klasyfikacja i algorytm GoP

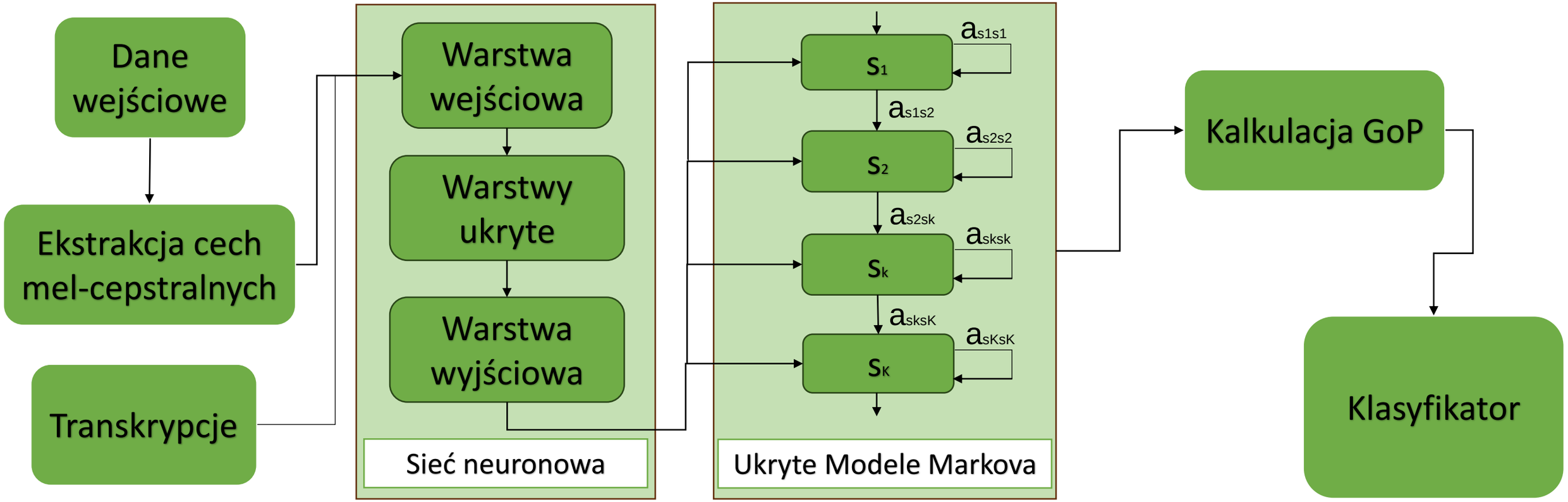
$$GoP(p) = \log \frac{LPP(p)}{max_{q \in Q} LPP(q)} \quad (1)$$

$$LPP(p) \approx \frac{1}{t_e - t_s + 1} \sum_{t=t_s}^{t_e} \log(P(p|o_t)) \quad (2)$$

$$P(p|o_t) = \sum_{s \in p} P(s|o_t) \quad (3)$$

p – obecny analizowany fonem
LPP – Log Phone Posterior
Q – całkowity zestaw fonemów
q – fonem ze zbioru Q
t_s – początek trwania fonemu
t_e – koniec trwania fonemu
s – senon
o_t – obserwacja wejściowa

Schemat systemu



Rysunek 1: Schemat systemu automatycznej oceny wymowy

Wykorzystane bazy danych i narzędzia

01 Speechocean762

1. Korpus mowy stworzony do zadań oceny wymowy
2. Nagrania w nienatywnym języku angielskim

02 Librispeech

1. Zestaw nagrań bazujący na audiobookach
2. Korpus zawierający pełne transkrypcje

03 Narzędzia i zasoby

1.Zestaw narzędzi Kaldi
2.Infrastruktura PLGrid

Architektura sieci neuronowej

Architektura 1B	Architektura 1C
Dane wejściowe	Dane wejściowe
Warstwa wejściowa (MFCC – wymiar 40)	Warstwa wejściowa (MFCC – wymiar 40)
Warstwy ukryte 5 warstw ReLU z normalizacją wsadową	Warstwy ukryte warstwa ReLU z normalizacją wsadową 15 warstw TDNN-F
Warstwa wyjściowa (Softmax)	Warstwa wyjściowa (Softmax)

Wprowadzone zmiany w treningu ASR

Architektura sieci neuronowej

Długość treningu (Epoki)

Podział na podzbiory treningowe (Mini Batch)

Tempo poszukiwania minimum funkcji kosztu (Learning Rate)

Wprowadzenie normalizacji

Ewolucja architektury w trakcie treningu

Losowość próbek i wielkość zbiorów obliczeniowych

Wyniki – Accuracy i MSE

W ocenie wyników GoP wykorzystano miarę Mean Square Error oraz parametr Accuracy.

Tabela 1: Wyniki najlepszych modeli dla algorytmu GoP na danych testowych

Model	Accuracy	MSE
1B (Basic)	0,43	0,65
1C (S)	0,50	0,57
1C (F)	0,53	0,54
1C (FNorm)	**0,55**	**0,53**

Wyniki – GoP (miary kasyfikacyjne)

Przeprowadzono również szczegółową analizę z wykorzystaniem miar klasyfikacyjnych oraz analizę wyników sieci neuronowej.

Tabela 2: Wyniki najlepszych modeli dla algorytmu GoP na danych testowych - parametry klasyfikacyjne

Parametr	Precision			Recall			F1-Score		
Model \ Klasa	0	1	2	0	1	2	0	1	2
1B (Basic)	0,29	0,05	0,99	0,42	0,74	0,41	0,34	0,09	0,58
1C (S)	0,27	0,06	0,99	0,39	0,72	0,50	0,32	0,11	0,66
1C (F)	0,26	0,06	0,98	0,37	0,69	0,53	0,31	0,11	0,69
1C (FNorm)	**0,28**	**0,06**	**0,98**	**0,41**	**0,67**	**0,54**	**0,33**	**0,11**	**0,70**

Wyniki – wnioski

Skuteczność algorytmu GoP

1. Rozbudowa i zmiana architektury sieci neuronowej
2. Wprowadzenie większej ilości parametrów (modyfikacja)
3. Wprowadzenie normalizacji CMVN

Potencjalne problemy związane z prototypem GoP

1. Niezbalansowany pod kątem klas korpus Speechocean762
2. Wpływ progów klasyfikacyjnych na poprawność wyników
3. Dostosowanie struktury systemu pod cel

Podsumowanie

Osiągnięto wszystkie z trzech przedstawionych celów.

Poprawiono skuteczność algorytmu GoP w kolejnych próbach.

W wyniku analizy danych wyszczególniono elementy odpowiadające za podwyższenie skuteczności miary GoP.

Wykryto również problemy, których poprawa pomogłaby w pracy nad GoP.

Dziękuję za uwagę!

Wyniki – modele ASR (WER)

Dane	Dane testowe – proste				Dane testowe - zaawansowane			
Model / Model językowy	LM4	LM3	PLM3	SLM3	LM4	LM3	PLM3	SLM3
1B (Basic)	5,04	5,16	6,27	7,02	12,51	13,14	15,44	16,94
1C (S)	4,84	4,95	5,89	6,51	12,55	13,02	15,09	16,24
1C (F)	4,81	4,98	5,89	6,40	12,57	12,85	14,95	15,99
1C (FNorm)	**4,89**	**4,97**	**5,95**	**6,62**	**12,40**	**12,84**	**14,97**	**16,07**

Dane	Dane testowe – proste				Dane testowe - zaawansowane			
Model / Model językowy	LM4	LM3	PLM3	SLM3	LM4	LM3	PLM3	SLM3
1B (Referencyjny)	5,00	5,22	6,40	7,14	12,56	13,04	15,58	16,88
1C (Referencyjny)	4,91	4,99	5,93	6,49	12,94	13,38	15,11	16,28