

Politechnika Warszawska  
Wydział Elektryczny  
Kierunek Informatyka

# SPRAWOZDANIE KOŃCOWE SYSTEMU Z SIECIĄ NEURONOWĄ DO ROZPOZNAWANIA JĘZYKA DOKUMENTU

Wykonał: Aliaksandr Karolik

Warszawa, 15.05.2019

## Spis treści

1	Wprowadzenie	2
2	Zaimplementowany algorytm	2
3	Wpływ liczby iteracji na poprawność rozpoznawania języka tekstu	2
4	Podsumowanie projektu	4

## 1 Wprowadzenie

Dokument jest specyfikacją końcową projektu dotyczącego systemu służącego do rozpoznawania języka tekstu, w oparciu o sztuczną sieć neuronową. Dokument będzie zawierał obserwacje dotyczące zachowania programu w różnych przypadkach, a także wnioski wypływające z zebranych danych.

## 2 Zaimplementowany algorytm

Zgodnie ze specyfikacją implementacyjną dla przeprowadzenia uczenia sztucznej sieci neuronowej został zaimplementowany algorytm wstecznej propagacji błędu (ang. Backpropagation).

Algorytm został poprawnie zaimplementowany. Poszczególne funkcje tworzące algorytm zostały przetestowane za pomocą testów jednostkowych.

## 3 Wpływ liczby iteracji na poprawność rozpoznawania języka tekstu

Aby zaobserwować jak wpływa liczba iteracji na wyniki systemu, zostały przygotowane zbiory danych uczących. Dane zostały dobrane w taki sposób, żeby każdy plik tekstowy miał prawie taką samą liczbę znaków. Średnio każdy plik uczący miał dwa tysiące znaków.

Program był uruchomiany w następującej konfiguracji:

```
python3 main.py -t -f engt.txt engt2.txt engt3.txt germt.txt  
germt2.txt plt.txt plt2.txt -a 1 1 1 0 0 0 0 -n 1000
```

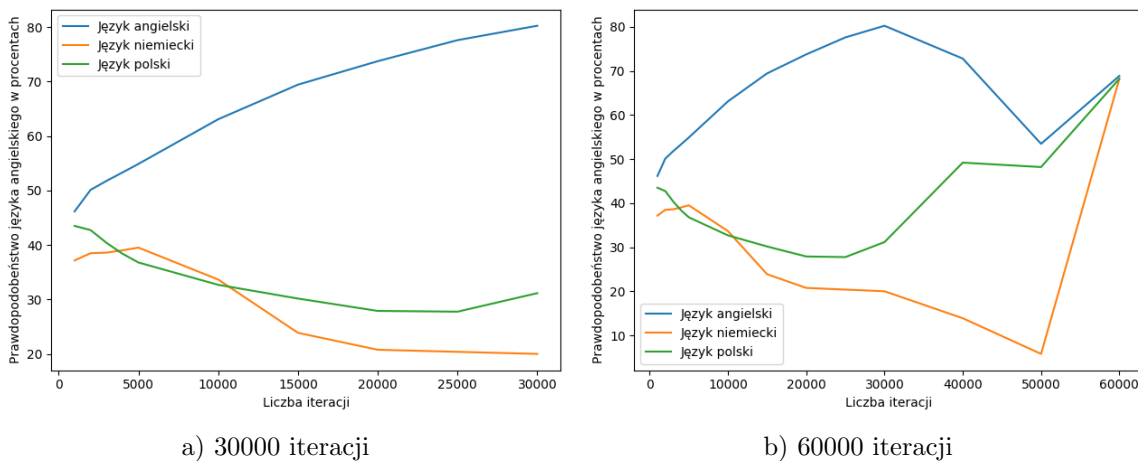
Z konfiguracji wynika że system był uczony do rozpoznawania języka angielskiego. To wynika z wartości podanych po argumentcie **-a**.

Żeby zaobserwować proces uczenia sztucznej sieci neuronowej zacząłem od tysiąca iteracji. Na początku krok był co tysiąc iteracji.

Po pięciu tysiącach iteracji, zwiększyłem krok o pięć tysięcy. Poniższa tabela oraz wykres reprezentuje odpowiedzi systemu na różne pliki wejściowe w zależności od liczby iteracji.

Liczba iteracji	Plik w języku angielskim, %	Plik w języku niemieckim, %	Plik w języku polskim, %
1000	46.18%	37.18%	43.51%
2000	50.12%	38.5%	42.75%
3000	51.78%	38.61%	49.41%
4000	53.32%	39.03%	38.41%
5000	54.89%	39.53%	36.8%
10000	63.09%	33.67%	32.7%
15000	69.44%	23.89%	30.18%
20000	73.75%	20.78%	27.91%
25000	77.61%	20.4%	27.77%
30000	80.25%	20.02%	31.16%
40000	72.81%	13.90%	49.21%
50000	53.47%	5.81%	48.19%
60000	68.87%	68.22%	68.22%

Tabela 1: Zależność odpowiedzi systemu od liczby iteracji



Rysunek 1: Wykres pokazujący odpowiedzi systemu w zależności od liczby iteracji

Z tabeli 1 oraz z rysunku 1 b) łatwo zauważyć, że w pewnym momencie sztuczna sieć neuronowa przestaje się uczyć, wynika to z algorytmu, który został zaimplementowany w systemie. W rzeczywistości ten regres jest skutkiem tego, że współczynnik zadający prędkość uczenia sieci jest ustawiony na sztywno w systemie.

Analizując tabele, można łatwo zauważyć, że po trzydziestu tysiącach iteracji, aby dalej skutecznie uczyć sieć neuronową, trzeba zmniejszyć współczynnik zadający prędkość uczenia sieci. Ten przypadek również został sprawdzony przy testowaniu systemu, ale w rzeczywistości po zmianie współczynnika wynik sieci zaczął zmieniać dopiero w tysięcznych wyniku. Z czego wynika, że ustawiony współczynnik jest prawie dobrze dobrany.

#### 4 Podsumowanie projektu

Finalnie implementacja programu nie odbiega znacząco od koncepcji przedstawionych w specyfikacji implementacyjnej i funkcjonalnej. Bardzo żałuje, że nie udało mi się dopracować algorytmu doboru współczynnika zadającego prędkość uczenia sieci.

Z drugiej strony jestem bardzo ucieszony, że za pomocą tego projektu udało się poznać takie zagadnienia jak sztuczna sieć neuronowa, konwolucyjna sieć neuronowa oraz zdobyć doświadczenie w języku programowania Python.

Również jestem bardzo ucieszony, że udało mi się stworzyć taki system, który później można będzie ponownie użyć i opracować system do rozpoznawania dowolnego języka tekstu.