

Rozpoznawanie języka tekstu na podstawie częstotliwości liter

Krzysztof Kutt, Michał Nowak

Abstrakt—TODO: Tutaj leci abstrakt

Słowa kluczowe—sieć neuronowa, język, rozpoznawanie.

I. WPROWADZENIE

PRZEDSTAWIONA w tym artykule aplikacja powstała w ramach przedmiotu Sztuczne Sieci Neuronowe, semestr letni 2011/2012, prowadzący: mgr inż. Tomasz Orzechowski.

A. Sieci neuronowe

Sztuczne Sieci Neuronowe to modele matematyczne, stworzone na wzór naturalnych połączeń neuronów znajdujących się w ludzkim mózgu. Każdy sztuczny neuron zbiera sygnały wejścia, dokonuje na nich odpowiedniej transformacji i zwraca odpowiednią wartość na wyjściu. Sztuczna sieć neuronowa może składać się z tysięcy takich elementarnych neuronów. W zależności od budowy sieci oraz od wykorzystywanej przez neurony funkcji aktywacji, wyróżniamy różne rodzaje sieci.

B. Rodzaje sieci

Skorzystać z książki Tadeusiewicza [1] - główne rodzaje sieci to rozdziały książki - napisać zdanie-dwa o każdym z nich

C. Zastosowania sieci

Wiki i książka tadeusiewicza.

D. Rozpoznawanie języka

Coś ogólnie o naszym problemie. Do czego to może być potrzebne, kto już walczył z tym problemem i jak (np google translator ma opcję rozpoznawania języka).

Przedstawione opcje to nie musi być tylko częstotliwość liter - pewnie są jakieś inne (choćby sprawdzanie, czy występują znaki specjalne danego języka :P)

E. Określenie języka przy pomocy sieci neuronowej

Czyli wprowadzenie teoretyczne do naszego projektu. Uznaliśmy, że ten problem można rozwiązać przy pomocy sieci neuronowej. Uznaliśmy, że fajnie będzie to zrobić przy pomocy prostej sieci, złożonej z trzech warstw, z neuronami z aktywacją sigmoidalną (i dlaczego akurat taką :P)

Do środkowej warstwy wzięliśmy 10 neuronów - tak z dupy

trochę - może to potestować? (i wtedy w testach informację o tym, że zweryfikowaliśmy wstępne założenia i okazały się dobre / złe - musieliśmy zmienić drastycznie liczbę neuronów, etc.)

W książce [2] są informacje o tym jak wybierać sieć (na tej podstawie ustalić np wstępną liczbę neuronów)

II. IMPLEMENTACJA

Tutaj idzie sekcja o tym jak podzieliliśmy na moduły, jaką sieć wybraliśmy, jaką bibliotekę (i dlaczego)

III. TESTY

Przeprowadzone testy. Wyniki. Informacje o częstotliwości liter, ale też o skuteczności sieci. Wykresy?

Co nam wyszło, dla których języków działa, które okazały się podobne, etc

IV. WNIOSKI

Projekt jest fajny, bądź nie i jak go można dalej rozwijać, albo gdzie go można zastosować

LITERATURA CYTOWANA

- [1] R. Tadeusiewicz, *Sieci neuronowe*, wyd. 2, Warszawa, Polska: Akademicka Oficyna Wydawnicza RM, 1993.
- [2] R. Tadeusiewicz, *Elementarne wprowadzenie do techniki sieci neuronowych z przykładowymi programami*, Warszawa, Polska: Akademicka Oficyna Wydawnicza PLJ, 1998.