
Практическая работа №6

курс Алгоритмы глубокого обучения - 2022

Задание 1 (30%)

Выполните задания, по каждому варианту напишите точность распознавания на проверочной выборке и сделайте выводы. (База Тексты писателей:

https://drive.google.com/drive/folders/1WgX2XkSe8k7vpl4m2KFfcPdBQ_GAPXuA?usp=sharing)

1. Запустите нейронку с bag of words (01) при разных maxWordsCount
 - a. 100
 - b. 1000
 - c. 10000
 - d. 50000
2. Запустите нейронку с bag of words (01) при maxWordsCount = 20000 и разных архитектурах
 - a. Поменяйте количество нейронов в слоях
 - b. Поменяйте количество слоев
 - c. Поменяйте активационные функции слоев
3. Запустите нейронку с Embedding при maxWordsCount = 50000, поменяйте размер Embedding пространства
 - a. 10
 - b. 50
 - c. 200

Задание 2 (10%)

Добейтесь точности распознавания 97% и верно распознанных всех писателей с помощью любой нейронной сети без фильтрации данных.

Задание 3 (40%)

1. Используйте разбиение текстов на фрагменты на базе обращения граждан в органы власти, как при распознавании писателей, например, размер блока - 20 слов, шаг - 1 слово.
2. Сравните точность с базовым ноутбуком, проверьте 5 различных архитектур сети.
3. Напишите функцию распознавания нужного обращения разделением его на фрагменты и распознаванием фрагментов.
4. Попробуйте разные параметры фильтрации выборки. Добавьте какой-нибудь свой критерий для фильтрации.
5. Запишите в конце ноутбука выводы о полученных исследованиях.

База Обращения граждан:

https://drive.google.com/file/d/1Sp-O0yp_hxTxGW6B3AJlmsZZqKTUQta/view?usp=sharing