

МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра МО ЭВМ

ОТЧЕТ
по лабораторной работе №6
по дисциплине «Искусственные нейронные сети»
Тема: Прогноз успеха фильмов по обзорам

Студентка гр. 8383

Гречко В.Д.

Преподаватель

Жангиров Т.Р.

Санкт-Петербург

2021

Цель работы

Прогноз успеха фильмов по обзорам (Predict Sentiment From Movie Reviews).

Задачи

- Изучить способы представления текста для передачи в ИНС
- Достигнуть точность прогноза не менее 95%

Требования

1. Построить и обучить нейронную сеть для обработки текста
2. Исследовать результаты при различном размере вектора представления текста
3. Написать функцию, которая позволяет ввести пользовательский текст (в отчете привести пример работы сети на пользовательском тексте)

Выполнение работы

1. Построение и обучение нейронной сети для обработки текста

На первом шаге была построена сеть, представленная в методических указаниях.

Получили следующие результаты:

```
Epoch 1/2
80/80 [=====] - 10s 90ms/step - loss: 0.5188 -
accuracy: 0.7407 - val_loss: 0.2643 - val_accuracy: 0.8954
Epoch 2/2
80/80 [=====] - 2s 19ms/step - loss: 0.2141 -
accuracy: 0.9196 - val_loss: 0.2730 - val_accuracy: 0.8905
0.8929499983787537
```

Модель обладает точностью 89.29%

2. Исследование результатов при различном размере вектора представления текста

Рассмотрим 3000 слов, наиболее часто встречающихся в обучающем наборе

ОТЗЫВОВ:

Epoch 1/2

```
80/80 [=====] - 2s 20ms/step - loss: 0.5349 - accuracy: 0.7246 - val_loss: 0.2768 - val_accuracy: 0.8858
```

Epoch 2/2

```
80/80 [=====] - 1s 12ms/step - loss: 0.2452 - accuracy: 0.9037 - val_loss: 0.2719 - val_accuracy: 0.8873
```

0.8865500092506409

Модель обладает 88.66% точностью.

Рассмотрим 800 слов:

Epoch 1/2

```
80/80 [=====] - 2s 12ms/step - loss: 0.6206 - accuracy: 0.6327 - val_loss: 0.3613 - val_accuracy: 0.8410
```

Epoch 2/2

```
80/80 [=====] - 0s 4ms/step - loss: 0.3693 - accuracy: 0.8394 - val_loss: 0.3365 - val_accuracy: 0.8546
```

0.8478000164031982

Модель обладает 84.78% точностью.

Рассмотрим 15000 слов:

Epoch 1/2

```
80/80 [=====] - 13s 128ms/step - loss: 0.5164 - accuracy: 0.7343 - val_loss: 0.2541 - val_accuracy: 0.8993
```

Epoch 2/2

```
80/80 [=====] - 2s 31ms/step - loss: 0.1916 - accuracy: 0.9297 - val_loss: 0.2621 - val_accuracy: 0.8963
```

0.8977999985218048

Модель обладает 89.78% точностью.

Рассмотрим 20000 слов:

Epoch 1/2

```
80/80 [=====] - 16s 155ms/step - loss: 0.5140  
- accuracy: 0.7492 - val_loss: 0.2572 - val_accuracy: 0.8970
```

Epoch 2/2

```
80/80 [=====] - 5s 65ms/step - loss: 0.1853 -  
accuracy: 0.9322 - val_loss: 0.2633 - val_accuracy: 0.9005  
0.8987500071525574
```

```
80/80 [=====] - 20s 189ms/step - loss: 0.5016  
- accuracy: 0.7546 - val_loss: 0.2517 - val_accuracy: 0.8995
```

Epoch 2/2

```
80/80 [=====] - 6s 78ms/step - loss: 0.1875 -  
accuracy: 0.9328 - val_loss: 0.2562 - val_accuracy: 0.9005  
0.90000000059604645
```

Модель обладает практически 90% точностью.

После нескольких запусков с 10000 словами и 20000 было выявлено, что максимальная точность достигается при 20 000 слов. Интуитивно понятно, что уменьшение размера вектора представления текста влечет за собой потерю слов, а значит и точность. Однако колебания точности достаточно малы – увеличение размера приводит к сдвигу всего лишь на 0.5% в среднем (дальше 20000 не хватает вычислительных мощностей), в то время как значительное уменьшение размера приводит к сдвигу в приблизительно -5%.

3. Ввод пользовательский текст

Результаты анализа пользовательского текста:

Файл 1.txt содержал положительную рецензию. Вот, что продемонстрировала сеть:

```
test 1
```

```
[[0.8818437]]
```

Файл 2.txt содержал отрицательную рецензию. Вот, что продемонстрировала сеть:

```
test 2
```

```
[[0.2039097]]
```

Файл 3.txt содержал положительную рецензию. Вот, что продемонстрировала сеть:

```
test 3
```

```
[[0.96575123]]
```

Выводы

В ходе выполнения лабораторной работы была создана нейронная сеть для задачи прогнозирования успеха фильмов по обзорам. Также была изучена работа с текстом для представления его в нейронной сети.