

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

Студент гр. 8383

Колмыков В.Д.

Преподаватель

Санкт-Петербург

2021

Цель работы.

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

Ход работы.

Была реализована нейросеть из методических указаний. Ее точность на тестовых данных составила 89.4%.

Размер вектора входных данных был изменен на 11 000. Точность на тестовых данных не сильно изменилась – 89.38%.

Размер вектора входных данных был изменен на 15 000. Точность упала до 89.33%.

Видно, что при избыточном размере вектора результаты обучения немного, но ухудшаются. Следовательно, нет смысла брать вектор размера большего, чем необходимо.

Для организации ввода пользовательского текста была написана следующая функция:

```
def interface(model, vec_len):
    index = imdb.get_word_index()
    while (True):
        print('Input text (or "stop"):')
        text = input()
        text = clear_text(text)
        if text == 'stop':
            break

        ind_arr = get_indexes_from_text(text, index, vec_len)
        vec = vectorize(np.asarray([ind_arr]))
        res = model.predict(vec)
        if res >= 0.5:
            print('Positive')
        else:
            print('Negative')
```

Принимает на вход модель и длину входного вектора, обрабатывает входную строку и выводит результат.

Для обработки строки написана следующая функция:

```
def clear_text(text):
```

```

text = re.sub(r'([^\A-z ]|[\[\]])', '', text).lower()
text = text.strip()
text = re.sub(r'  +', ' ', text)
return text

```

Убирает из строки все лишнее (знаки препинания, лишние пробелы и тд).

Для преобразования строки в массив индексов используется следующая функция:

```

def get_indexes_from_text(text, index, vec_len):
    text_arr = text.split(" ")
    ind_arr = []
    for word in text_arr:
        num = index.get(word)
        if num is not None and num < vec_len:
            ind_arr.append(num + 3)
    return ind_arr

```

Используя индекс переводит слова в индексы.

Пример работы пользовательского ввода:

Input text (or "stop"):

This film is amazing. It is really interesting and cool.

Positive

Input text (or "stop"):

This film is the worst film i have ever seen. It is too boring and stupid.

Negative

Выводы.

В ходе выполнения работы был реализована нейросеть для прогноза успеха фильмов по обзорам. Было проведено ознакомление с задачей регрессии, способами представления текста для передачи в нейросеть. Была написана программа для предсказания результата пользовательского ввода.