

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

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

Студент гр. 8383

Шишкин И.В.

Преподаватель

Жангиров Т.Р.

Санкт-Петербург

2021

Цель работы.

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

Задачи.

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

Выполнение работы.

Построена нейронная сеть, представленная в листинге 1.

Листинг 1 – Модель

```
model = Sequential()
model.add(layers.Dense(50, activation="relu", input_shape=(10000,)))
model.add(layers.Dropout(0.3, noise_shape=None, seed=None))
model.add(layers.Dense(50, activation="relu"))
model.add(layers.Dropout(0.2, noise_shape=None, seed=None))
model.add(layers.Dense(50, activation="relu"))
model.add(layers.Dense(1, activation="sigmoid"))
NumEpochs = 10
BatchSize = 512
model.compile(optimizer="adam", loss="binary_crossentropy",
metrics=["accuracy"])
```

Схема модели представлена на рис. 1.

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 50)	500050
dropout (Dropout)	(None, 50)	0
dense_1 (Dense)	(None, 50)	2550
dropout_1 (Dropout)	(None, 50)	0
dense_2 (Dense)	(None, 50)	2550
dense_3 (Dense)	(None, 1)	51

```

Total params: 505,201
Trainable params: 505,201
Non-trainable params: 0

```

Рисунок 1 – Схема модели

Результат работы модели с 10 эпохами представлен на рис. 2.

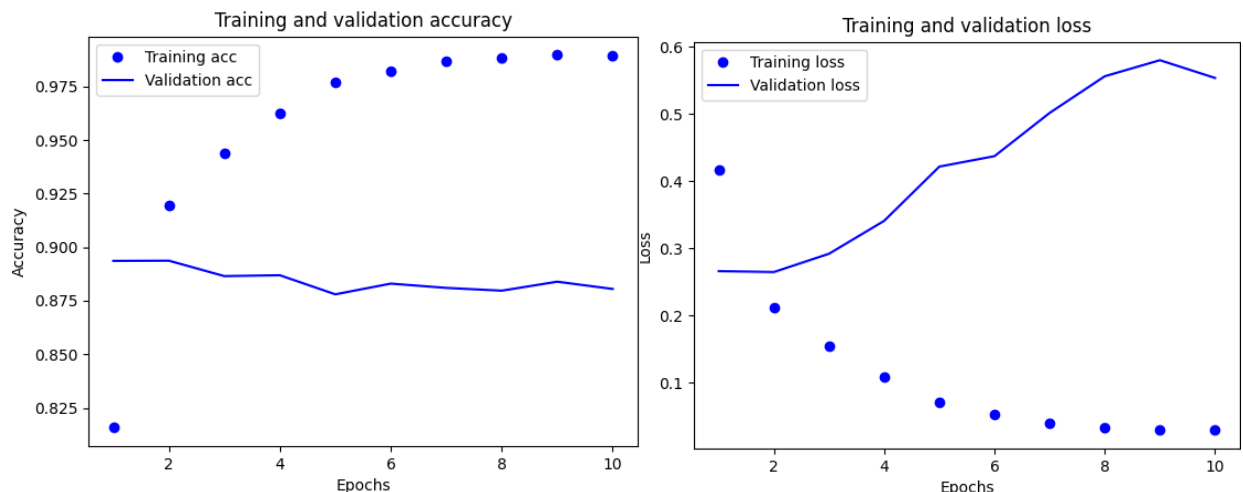


Рисунок 2 – Результат работы модели с 10 эпохами

Как видно из рис. 2., точность на тренировочных растет, в то время как на тестовых данных – падает. Это является типичным признаком переобучения. Причем можно увидеть, что переобучение начинается со 2-й эпохи, поэтому понизим количество эпох до 2.

Размер вектора представления текста на данный момент равен 10000, и точность на тренировочных данных – 92%, а на тестовых – 90%, потери на тренировочных – 0.22, на тестовых – 0.26.

Уменьшим размер вектора представления текста до 8000. Результаты остались такими же. Если уменьшить размер до 5000, то точность на

тренировочных данных упадет на 1%, потери поднимутся до 0.25, точность на тестовых данных тоже упадет на 1%, потери увеличатся на 0.01. Изменения небольшие, но они есть. Теперь уменьшим размер до 1000. Теперь точность на тренировочных данных упала еще на 3%, потери поднялись на 0.05, на тестовых данных точность упала до 86%, потери поднялись до 0.31. Можно сделать вывод, что размер от 8000 до 1000 является наилучшим.

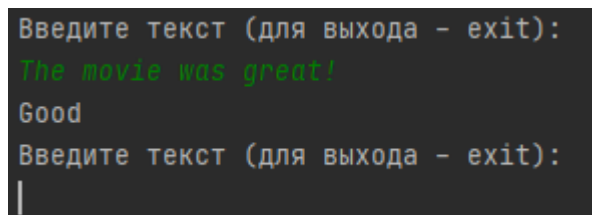
Была добавлена возможность ввода пользовательского текста. Код представлен в листинге 2.

Листинг 2 – Ввод текста

```
while True:
    print("Введите текст (для выхода - exit):")
    text = str(input())
    if text == "exit":
        break
    else:
        words = text_to_word_sequence(text)
        vocabulary = imdb.get_word_index()
        x_predict = []
        for i in range(len(words)):
            if words[i] not in vocabulary:
                continue
            if vocabulary[words[i]]+3 < 10000:
                x_predict.append(vocabulary[words[i]]+3)
        x_predict = vectorize_sequences(np.asarray([x_predict]))
        prediction = (model.predict(x_predict) > 0.5).astype("int32")
        print("Good" if prediction[0] == 1 else "Bad")
```

Результат работы программы для текста «The movie was great!» на рис.

3.



```
Введите текст (для выхода - exit):
The movie was great!
Good
Введите текст (для выхода - exit):
|
```

Рисунок 3 – Результат для первого текста

Результат работы программы для текста «The Green Mile is a masterwork. This is film as art, at it's very best. The depth of the cast is extraordinary, with all of the players delivering excellent performances. There is a clear sense here that all involved in the production knew that this was something special, and gave it their all. See this film if you truly enjoy actors giving everything to their craft. Watch for the countless subtleties of expression, and the great power that the cast creates with silence. This is evident in the opening sequence and remains throughout. Above all, Michael Duncan as John Coffey is exceptional. He brings gripping emotion to a unique, fascinating character» на рис. 4.

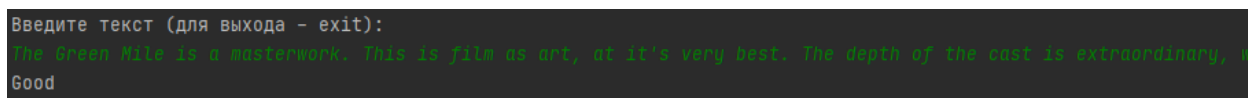


Рисунок 4 – Результат для второго текста

Результат работы программы для текста «I can't believe that an actor with the credentials Tom Hanks would star in this film. It was so slow, my family & I were praying for it to end. The plot didn't go anywhere, and the ending was so weak I threw my hands in the air and just asked, "why?" How could they drag the movie out that long for such a stock plot and predictable ending? And the execution scenes seemed to be there for shock value only, and were quite gruesome and gratuitous» на рис. 5.

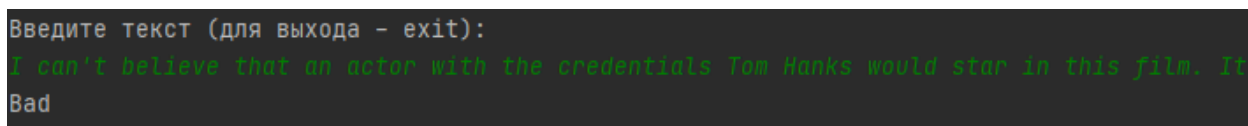


Рисунок 5 – Результат для третьего текста

Выводы.

Была построена и обучена нейронная сеть для обработки текста. Исследованы результаты при различном размере вектора представления текста. Написана функция, которая позволяет ввести пользовательский текст.