

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

ЛАБОРАТОРНАЯ РАБОТА №6
по дисциплине «Искусственные нейронные сети»
Тема: «прогноз успеха фильма по обзорам»

Студент гр. 7381

Ильясов А.В.

Преподаватель

Жукова Н. А.

Санкт-Петербург

2020

Цели.

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%

Требования.

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

Выполнение работы.

При выполнении работы были обучены модели с размерами вектора представления текста 500, 1000, 2500, 5000 и 10000. Результаты обучения представлены ниже.

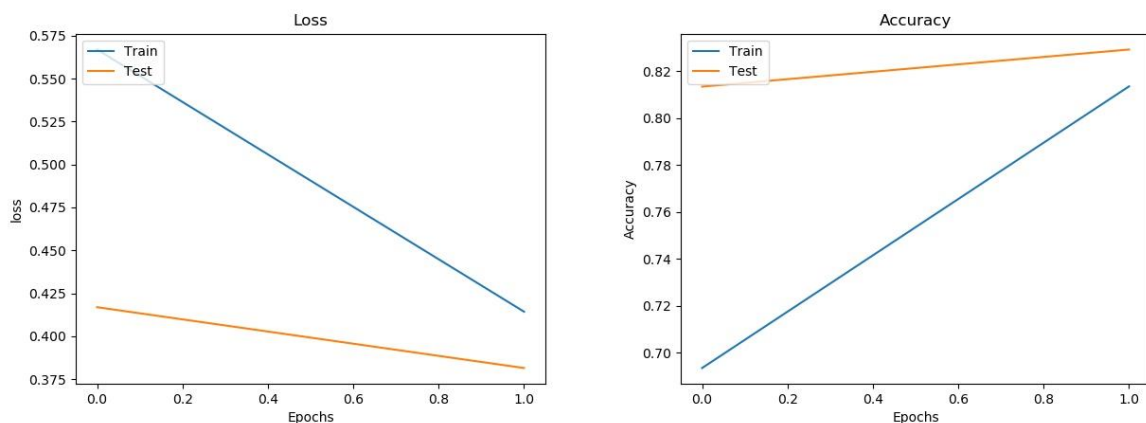


Рисунок 1 – графики точности и функции потерь модели с вектором длины 500

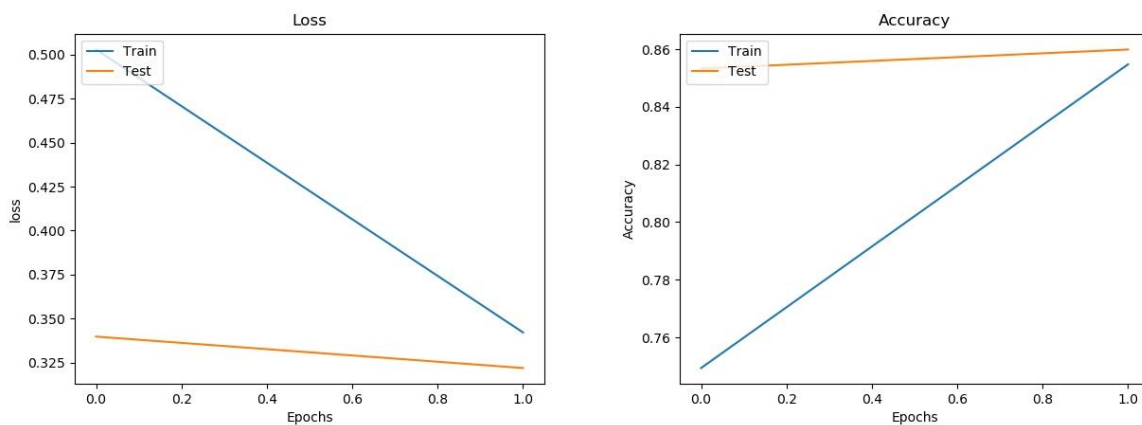


Рисунок 2 – графики точности и функции потерь модели с вектором длины 1000

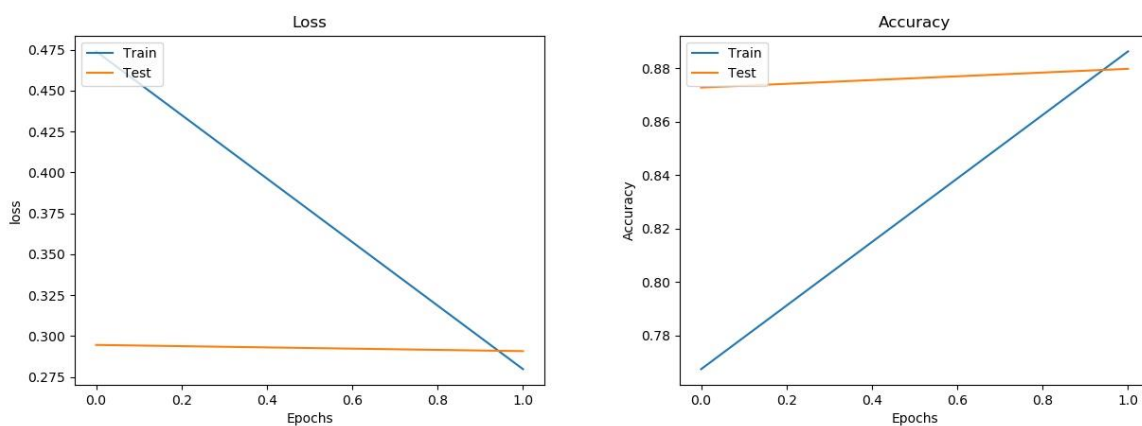


Рисунок 3 – графики точности и функции потерь модели с вектором длины 2500

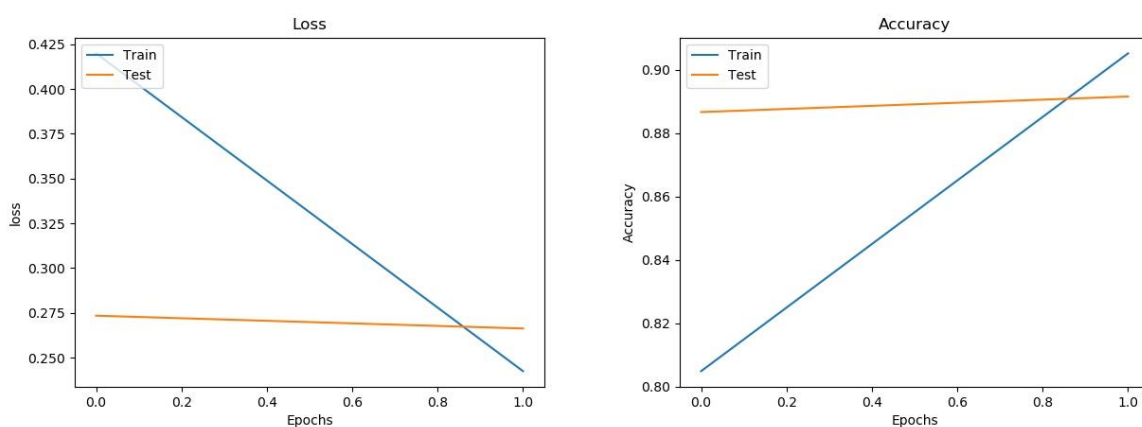


Рисунок 4 – графики точности и функции потерь модели с вектором длины 5000

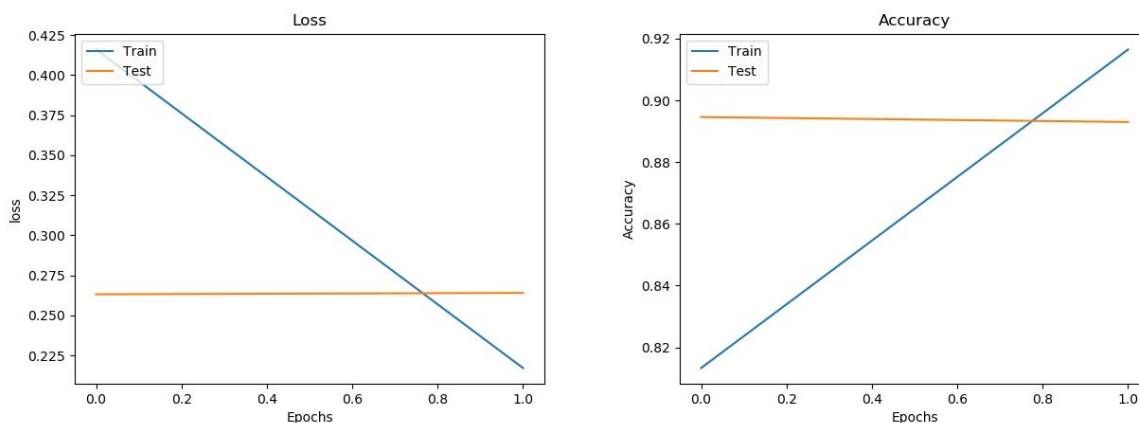


Рисунок 4 – графики точности и функции потерь модели с вектором длины 10000

Можно заметить, что, чем длиннее вектор, тем более точное предсказание выдает сеть.

Для ввода пользовательского текста была написана функция, которая переводит исходный текст, в массив индексов слов датасета `imdb`, который уже можно передавать модели, в функцию предсказания:

```
def load_text(text):
    file = open(text, 'r').read()
    word_index = imdb.get_word_index()
    text = []
    for i in file:
        if i in word_index and word_index[i] < 10000:
            text.append(word_index[i])

    text = vectorize([text])
    return text
```

функция была протестирована на примере отзыва на фильм «Джокер» с оценкой 10/10 «I was a person that saw all the hype and claims of masterpiece as overreacting and overblown excitement for another Joker based film. I thought this looked solid at best and even a bit too pretentious in the trailer, but in here to say I was incredibly wrong. This is a massive achievement of cinema that's extremely rare in a day and age of cgi nonsense and reboots. While this is somewhat of a reboot of sorts, the standalone origin tale is impeccable from start

to finish and echoes resemblance to the best joker origin comics from the past. Joaquin bleeds, sweats, and cries his every drop into this magnificently dedicated performance. Heath Ledger would be proud. This is undoubtedly the greatest acting performance since Heath's joker. The directing and writing is slickly brilliant and the bleak settings and tones are palpable throughout. When this film was over the place was blown away and every audience member was awestruck that they witnessed a film that could still transport them into a character's world and very existence. Believe the hype. This is going to be revered as a transcending masterpiece of cinema.» выдает результат 0.78748045.

Вывод.

В ходе выполнения данной работы была изучена задача прогноза настроения отзывов фильмов на примере датасета imdb.