

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

ОТЧЕТ
по лабораторной работе №7
"Классификация обзоров фильмов"
по дисциплине «Искусственные нейронные сети»

Студентка гр. 8383

Преподаватель

Ишанина Л.Н.

Жангиров Т.Р.

Санкт-Петербург

2021

Цель.

Классификация последовательностей – это проблема прогнозирующего моделирования, когда у вас есть некоторая последовательность входных данных в пространстве или времени, и задача состоит в том, чтобы предсказать категорию для последовательности.

Проблема усложняется тем, что последовательности могут различаться по длине, состоять из очень большого словарного запаса входных символов и могут потребовать от модели изучения долгосрочного контекста или зависимостей между символами во входной последовательности.

В данной лабораторной работе также будет использоваться датасет IMDb, однако обучение будет проводиться с помощью рекуррентной нейронной сети.

Задание.

- Ознакомиться с рекуррентными нейронными сетями
- Изучить способы классификации текста
- Ознакомиться с ансамблированием сетей
- Построить ансамбль сетей, который позволит получать точность не менее 97%

Найти набор оптимальных ИНС для классификации текста

Провести ансамблирование моделей

Написать функцию/функции, которые позволят загружать текст и получать результат ансамбля сетей

Провести тестирование сетей на своих текстах (привести в отчете)

Выполнение работы.

Были подключены необходимые библиотеки:

```
import numpy
from keras.datasets import imdb
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from keras.layers.embeddings import Embedding
from keras.preprocessing import sequence
```

Далее был загружен датасет IMDB, который уже встроен в Keras и объединены данные обучения и тестирования:

```
from keras.datasets import imdb
(training_data, training_targets), (testing_data,
testing_targets) = imdb.load_data(num_words=10000)
data = np.concatenate((training_data, testing_data), axis=0)
targets = np.concatenate((training_targets, testing_targets),
axis=0)
```

Затем были обрезаны и дополнены входные последовательности так, чтобы они были одинаковой длины для моделирования:

```
max_review_length = 500
X_train = sequence.pad_sequences(X_train,
maxlen=max_review_length)
X_test = sequence.pad_sequences(X_test, maxlen=max_review_length)
```

Затем была создана модель, представленная в листинге ниже:

```
models = []
    model1 = Sequential()
    model1.add(Embedding(top_words, embedding_vector_length,
input_length=max_review_length))
    model1.add(LSTM(100))
    model1.add(Dropout(0.4))
    model1.add(Dense(64, activation='relu'))
    model1.add(Dropout(0.2))
    model1.add(Dense(1, activation='sigmoid'))
    models.append(model1)

    model2 = Sequential()
```

```

model2.add(Embedding(top_words, embedding_vector_length,
input_length=max_review_length))
model2.add(Conv1D(filters=32, kernel_size=3, padding='same',
activation='relu'))
model2.add(MaxPooling1D(pool_size=2))
model2.add(Dropout(0.3))
model2.add(LSTM(150))
model2.add(Dropout(0.3))
model2.add(Dense(1, activation='sigmoid'))
models.append(model2)
return models

```

Результаты работы

Графики точности и потерь рекуррентной модели представлены на рис. 1-2.

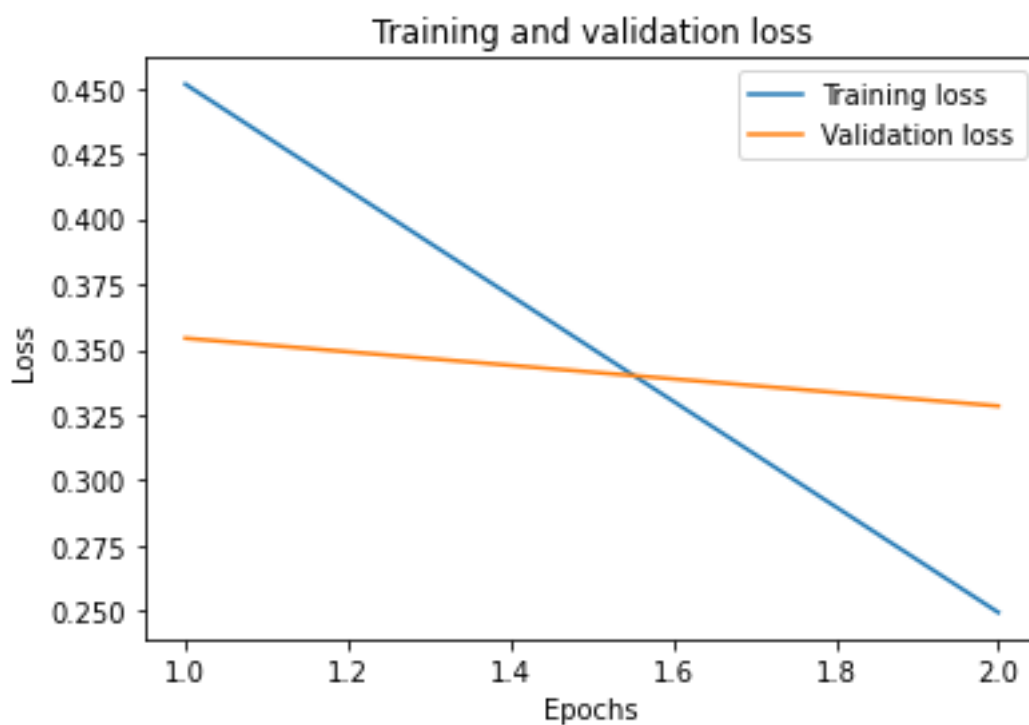


Рисунок 1 – графики потерь

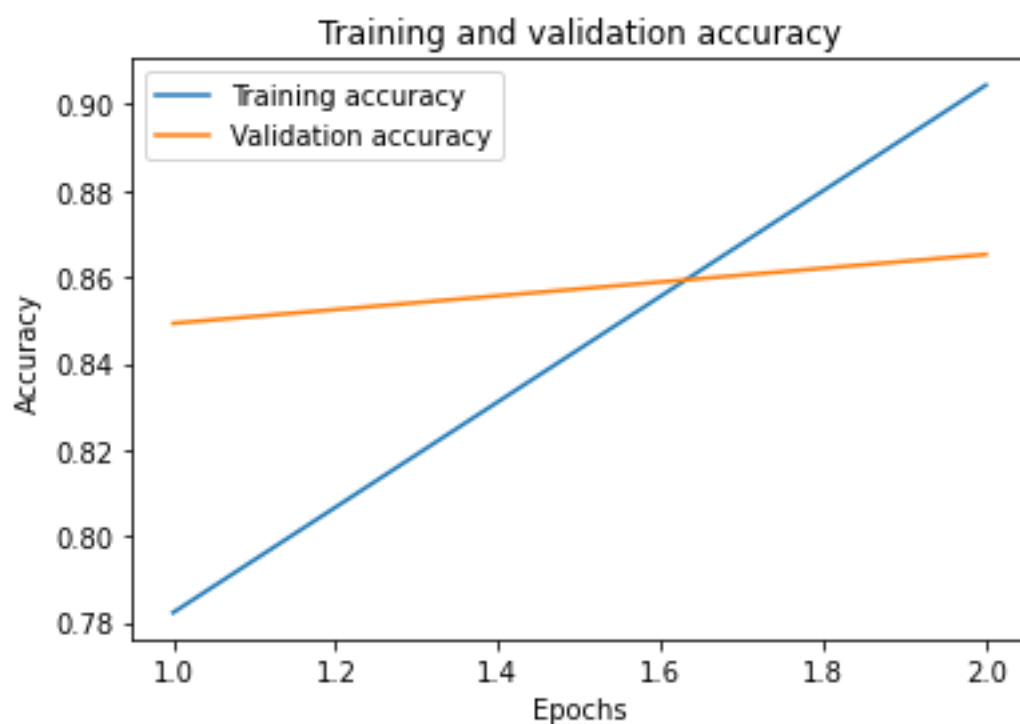


Рисунок 2 – графики точности

Значение точности было достигнуто:

accuracy value: 86.52%

Графики точности и потерь рекуррентной сверточной модели представлены на рис. 3-4.

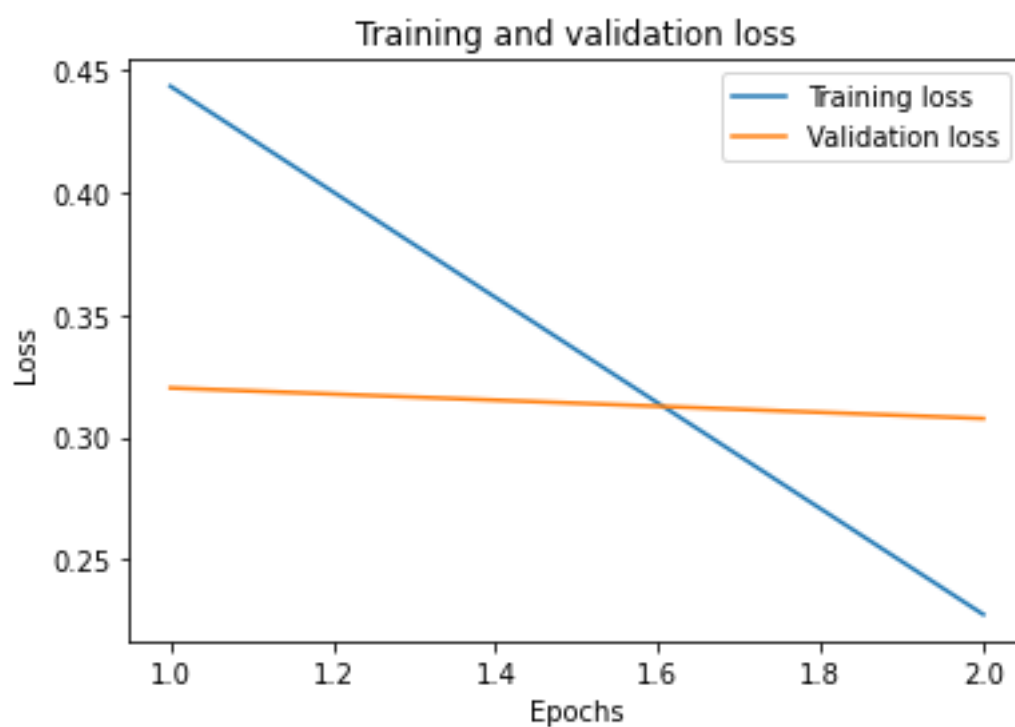


Рисунок 3 – графики потерь

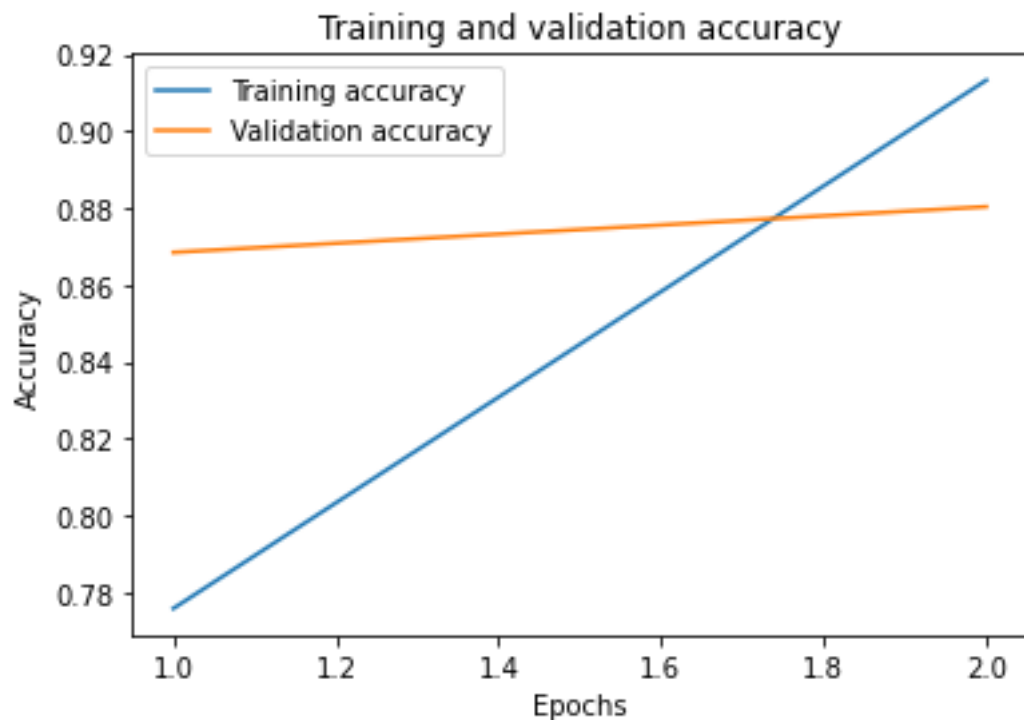


Рисунок 4 – графики точности

Значение точности было достигнуто:

accuracy value: 88.03%

Далее был протестирован следующий отзыв:

A bad movie, I think that actors are ugly, storyline is boring.

Результат работы программы представлен ниже:

```
Accuracy of ensembling 0.885
['a', 'bad', 'movie', 'i', 'think', 'that', 'actors', 'are',
'ugly', 'storyline', 'is', 'boring']
[array([[0.03024552]], dtype=float32), array([[0.02059536]],
dtype=float32)]
[[0.02542044]]
```

Таким образом, точность при ансамблировании сетей лучше, чем у каждой модели по отдельности.

Выводы.

В ходе выполнения лабораторной работы было проведено ознакомление со рекуррентными нейронными сетями. Были изучены способы классификации текста и также проведено ознакомление с ансамблированием сетей.