

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»

Студент гр. 8383

Преподаватель

Сахаров В.М.

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.

1. Ознакомиться с рекуррентными нейронными сетями
2. Изучить способы классификации текста
3. Ознакомиться с ансамблированием сетей
4. Построить ансамбль сетей, который позволит получать точность не менее 97%
5. Найти набор оптимальных ИНС для классификации текста
6. Провести ансамблирование моделей
7. Написать функцию/функции, которые позволят загружать текст и получать результат ансамбля сетей
8. Провести тестирование сетей на своих текстах (привести в отчете)

Выполнение работы.

В программе было реализовано три модели. Листинг моделей представлен ниже:

```
model_lstm = Sequential([
```

```

        Embedding(10000, embedding_length, input_length=reviews_count),
        LSTM(128),
        Dropout(0.3),
        Dense(64, activation="relu"),
        Dropout(0.4),
        Dense(1, activation="sigmoid"]])

model_cnn = Sequential([
    Embedding(10000, embedding_length, input_length=reviews_count),
    Conv1D(filters=64, kernel_size=3, padding='same', activation='relu'),
    MaxPool1D(pool_size=2),
    Dropout(0.4),
    LSTM(128),
    Dropout(0.3),
    Dense(1, activation='sigmoid']])

model_cnl = Sequential([
    Embedding(10000, embedding_length, input_length=reviews_count),
    Conv1D(filters=16, kernel_size=3, padding='same', activation='relu'),
    MaxPool1D(pool_size=2),
    Dropout(0.2),
    Conv1D(filters=32, kernel_size=3, padding='same', activation='relu'),
    MaxPool1D(pool_size=2),
    Dropout(0.2),
    Flatten(),
    Dense(64),
    Dropout(0.1),
    Dense(1, activation='sigmoid']])

```

Результаты обучения моделей представлены на рис. 1.

```

Epoch 1/2
188/188 [=====] - 9s 36ms/step
Epoch 2/2
188/188 [=====] - 6s 33ms/step
105/105 - 1s - loss: 0.3545 - accuracy: 0.8611
Accuracy: 86.11%
Epoch 1/2
188/188 [=====] - 7s 27ms/step
Epoch 2/2
188/188 [=====] - 4s 23ms/step
105/105 - 1s - loss: 0.3261 - accuracy: 0.8587
Accuracy: 85.87%
Epoch 1/2
188/188 [=====] - 3s 12ms/step
Epoch 2/2
188/188 [=====] - 2s 10ms/step
105/105 - 0s - loss: 0.3181 - accuracy: 0.8668
Accuracy: 86.68%
Accuracy: 88.04%
Test string:

```

Рисунок 1 – обучение моделей

Точность ансамбля составила 88%, что выше любой отдельной сети

На пользовательских тестах был проверен ансамбль моделей

Текст: Im very prous after watching this movie thank you dear producer

Результат:

[[0.9375119]]

[[0.87352663]]

[[0.5451025]]

Общий: [[1.]]

Текст: this is masterpiece except the style of decorations it is not so good as
other

Результат:

[[0.95161617]]

[[0.869758]]

[[0.8750143]]

Общий: [[1.]]

Текст: bad movie and actors they just cant play

Результат:

[[0.08369088]]

[[0.31129512]]

[[0.15745948]]

Общий: [[0.]]

Предсказания ансамбля корректны.

Выводы.

В ходе выполнения лабораторной работы было изучено решение задачи классификации отзывов на фильмы для определения успешности фильма с помощью рекуррентных сетей. Был создан и протестирован ансамбль из двух моделей.