

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

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

Студент гр. 8382

Янкин Д.О.

Преподаватель

Жангиров Т.Р.

Санкт-Петербург

2021

Цель работы.

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.

Постановка задачи.

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

Выполнение работы.

Для классификации используется датасет imdb с 10000 самых популярных слов. Так как длина отзывов, подаваемая на вход сети, должна быть одинаковой, они будут обрезаны или дополнены нулями до 500 слов.

Была создана следующая модель сети:

```
model = Sequential()  
model.add(Embedding(top_words, embedding_vector_length,  
input_length=max_review_length))  
model.add(LSTM(100))  
model.add(Dense(1, activation='sigmoid'))
```

После ее обучения была показана точность на тестовой выборке 87.91%

Затем в модель были добавлены слои свертки и максимального пулинга:

```
model = Sequential()
model.add(Embedding(top_words, embedding_vector_length,
input_length=max_review_length))
model.add(Conv1D(filters=32, kernel_size=3, padding='same',
activation='relu'))
model.add(MaxPooling1D(pool_size=2))
model.add(LSTM(100))
model.add(Dense(1, activation='sigmoid'))
```

Обучение этой модели дало точность 89.49% и заняло немного меньше времени.

В модель добавлен Dropout:

```
model = Sequential()
model.add(Embedding(top_words, embedding_vector_length,
input_length=max_review_length))
model.add(Conv1D(filters=32, kernel_size=3, padding='same',
activation='relu'))
model.add(MaxPooling1D(pool_size=2))
model.add(Dropout(0.2))
model.add(LSTM(100))
model.add(Dropout(0.2))
model.add(Dense(1, activation='sigmoid'))
model.compile(loss='binary_crossentropy', optimizer='adam',
metrics=['accuracy'])
```

Точность 89.57%

Был создан ансамбль из четырех штук третьей модели. Их точность:

```
Model 1 accuracy: 87.26%
Model 1 accuracy: 86.88%
Model 1 accuracy: 86.49%
Model 1 accuracy: 86.32%
Ensemble Accuracy: 89.42%
```

Совокупная точность ансамбля сетей несколько выше, чем у каждой сети в отдельности.

Была написана функция для классификации настроений в отзывах с загрузкой текста из файла. Результаты тестирования ансамбля на собственных отзывах приведены в таблице 1.

ОТЗЫВ	Предсказание	Реальность
You'll end up like me, binge watching it until 5 AM (and I don't even know anything about chess). Very well written, directed and acted.	0.61462814 Positive	Positive
This TV Mini-series is excellent! It is wonderful to see Anya Taylor-Joy in a well crafted show. She shines brilliantly as Beth Harmon a chess playing prodigy. The story takes us from when she is a small girl, orphaned, continuing through her teenage angst years. I highly recommend this show. Great cinematography and directing and of course to get back to the acting - its top notch! I only wish there was more than 7 episodes.	0.9992835 Positive	Positive
I know there is a lot of hype about this show, but I just couldn't get into it. I thought it was ridiculous that a little kid was taking so much pain medication....totally unrealistic. And, doesn't say much for the character if that is the only way she can win. I just didn't care enough about the characters to keep watching.	0.03226701 Negative	Negative
Started off with promise, but ended up being a not-very-interesting soap opera. I can only watch a pretty girl get drunk and dance around so many times in seven episodes before I start to notice all of the huge, gaping, ridiculous holes in the plot, like the fact that the main character experiences at least 5 lives worth of tragedy in about 7 years. This way, she's can still look gorgeous through the whole thing is the point, I guess. Episode 6 in particular is just bad.	0.42771667 Negative	Negative

I was expecting a masterpiece and oscar worthy film from all the hype. I actually didn't enjoy watching this movie. there was no message, it was slow and the joker turned out just to be the son of a schizophrenic and went crazy over a slight amount of bullying. I don't see how anyone could relate to this character and the laughing is lame acting way overdone. I'm just stunned that this has a 9/10 rating right now, how the hell	0.32242432 Negative	Negative
--	------------------------	----------

Ансамбль сетей дал правильный ответ во всех пяти случаях.

Выводы.

В ходе лабораторной работы на датасете IMDB была обучена нейронная сеть для классификации окраса настроения в тексте. Было проведено ансамблирование моделей, совокупная точность ансамбля превышает точность каждой модели в отдельности. Была написана функция для классификации пользовательского текста, все отзывы были классифицированы правильно.