

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

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

Студент гр. 8383

Колмыков В.Д.

Преподаватель

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.

Ход работы.

Была реализована нейросеть со следующими параметрами архитектуры и обучения:

```
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(128, return_sequences=True))
    model.add(Dropout(0.5))
    model.add(LSTM(64))
    model.add(Dropout(0.5))
    model.add(Dense(1, activation='sigmoid'))

    model.compile(loss='binary_crossentropy',      optimizer='adam',
metrics=['accuracy'])
    model.fit(train_x,  train_y,  validation_data=(test_x,  test_y),
epochs=3,  batch_size=64)
```

Для достижения более точных результатов данная модель была создана и обучена 5 раз на разных совокупностях тренировочных и валидационных данных и применено ансамблирование моделей для достижения более высокой точности.

Для организации ввода пользовательского текста была написана следующая функция:

```
def interface(model, vec_len):
```

```

index = imdb.get_word_index()
while (True):
    print('Input text (or "stop"):')
    text = input()
    text = clear_text(text)
    if text == 'stop':
        break

    ind_arr = get_indexes_from_text(text, index, vec_len)
    vec = vectorize(np.asarray([ind_arr]))
    res = model.predict(vec)
    if res >= 0.5:
        print('Positive')
    else:
        print('Negative')

```

Принимает на вход модель и длину входного вектора, обрабатывает входную строку и выводит результат.

Для обработки строки написана следующая функция:

```

def clear_text(text):
    text = re.sub(r'([^\A-z ]|[\[\]])', '', text).lower()
    text = text.strip()
    text = re.sub(r' +', ' ', text)
    return text

```

Убирает из строки все лишнее (знаки препинания, лишние пробелы и тд).

Для преобразования строки в массив индексов используется следующая функция:

```

def get_indexes_from_text(text, index, vec_len):
    text_arr = text.split(" ")
    ind_arr = []
    for word in text_arr:
        num = index.get(word)
        if num is not None and num < vec_len:
            ind_arr.append(num + 3)
    return ind_arr

```

Используя индекс переводит слова в индексы.

Пример работы пользовательского ввода:

Input text (or "stop"):

This film is amazing. It is really interesting and cool.

Positive

Input text (or "stop"):

This film is the worst film i have ever seen. It is too boring and stupid.

Negative

Интерес представляет тест на отзыве, в котором будут как положительные, так и отрицательные предложения. Например, следующий текст: «My mom said that this film is awesome and very interesting. But i think it is too bad and boring.» содержит как положительные, так и отрицательные слова, но не является положительным. В результате, ансамбль нейросетей выдаст сообщение о том, что в целом отзыв негативный. Интересно обратить внимание на результат работы каждой нейросети:

[0.14150757 0.40707037 0.10020601 0.60462677 0.92401487]

Как видно, две из пяти нейросети оценили данный текст как скорее положительный. Из результата можно сделать вывод, что ансамблирование помогло разработать более точную систему для синтентального анализа.

Выводы.

В ходе выполнения работы была реализована нейросеть для синтентального анализа текста отзыва фильма. Было проведено ознакомление с ансамблированием моделей, способами представления текста для передачи в нейросеть. Была написана программа для предсказания результата пользовательского ввода.