МИНОБРНАУКИ РОССИИ

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ

ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)

Кафедра МО ЭВМ

ОТЧЕТ

по лабораторной работе №7

по дисциплине «Искуственные нейронные сети»

Тема: Классификация обзоров фильмов

Студент гр. 8382 _____ Щемель Д.А.

Преподаватель _____ 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.

Задачи

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

Требования

1. Найти набор оптимальных ИНС для классификации текста
2. Провести ансамблирование моделей
3. Написать функцию/функции, которые позволят загружать текст и получать результат ансамбля сетей
4. Провести тестирование сетей на своих текстах (привести в отчёте)

Ход работы

1. Построение нейронной сети

Во всех моделях используются следующие параметры:

Оптимизатор: `adam`

Функция потерь: `binary_crossentropy`

Метрика: `accuracy`

Архитектуры сетей:

Модель 1:

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, 500, 32)           160000
_________________________________________________________________
conv1d (Conv1D)              (None, 500, 32)           3104
_________________________________________________________________
max_pooling1d (MaxPooling1D) (None, 250, 32)           0
_________________________________________________________________
lstm (LSTM)                  (None, 100)               53200
_________________________________________________________________
dense (Dense)                (None, 1)                 101
=================================================================
```

Точность: 90.55%

Модель 2:

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 500, 32)           160000
_________________________________________________________________
```

```
conv1d_1 (Conv1D)            (None, 500, 32)           3104
_________________________________________________________________
max_pooling1d_1 (MaxPooling1 (None, 250, 32)           0
_________________________________________________________________
simple_rnn (SimpleRNN)       (None, 250, 64)           6208
_________________________________________________________________
simple_rnn_1 (SimpleRNN)     (None, 64)                8256
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 65
=================================================================
```

Точность: 90.55%

Модель 3:

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_2 (Embedding)      (None, 500, 32)           160000
_________________________________________________________________
conv1d_2 (Conv1D)            (None, 500, 32)           3104
_________________________________________________________________
max_pooling1d_2 (MaxPooling1 (None, 250, 32)           0
_________________________________________________________________
gru (GRU)                    (None, 250, 64)           18816
_________________________________________________________________
lstm_1 (LSTM)                (None, 250, 32)           12416
_________________________________________________________________
gru_1 (GRU)                  (None, 16)                2400
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 17
=================================================================
```

Точность: 90.74%

Точность ансамбля моделей: 90.82%

Таким образом, точность ансамбля из трёх моделей оказалась ненамного, но всё же выше точности каждой модели в отдельности. При этом точность второй модели имеет большую дисперсию по сравнению с другими моделями.

Результаты:

Тексты оценок:

Оценка 1:

> I felt it was poorly written plot. I felt as if I had joined a movie mid showing as everything was like already a story line in progress and no one had bothered to introduce said story lines. I left the theater feeling as if all they wanted was to get my money one last time. What they state in the preview is the ENTIRE movie, no need to watch it, it dragged very slowly during much of the film. at one point I left to go to the bathroom, got back and still nothing had changed, no action, no nothing but another slow scene. And to add this was set in the late 2020s? I thought from the last movie there was a teaser set near the same time when Professor X and Magneto came for Logan at an airport. what happened there? is this a prequel? many many more questions arose rather than answer. If this is in fact a prequel, than this fact will make me want to wait till all episodes are in and watch them in entirety then maybe it'll make sense. As it is this wasn't very well written and they writers did little to no homage to the previous X-Men franchise movies. They left out completely what had happened other than a reference or 2 to an "event". I don't know I felt like it wasn't worth a theater show, straight to DVD maybe but not a huge success IMO.

Оценка 2:

> The movie was quite depressing. Logan and Charles have aged, and are not doing well. They are hiding out and keeping Charles drugged so he won't hurt anyone. I just didn't like the violence for the sake of violence in the movie. Most of the super hero movies were rated PG-13, now we have another one that is rated R. There one flash of brief nudity, so we could have that added to the rating of the movie, that did nothing to add to the plot of the movie. The movie seemed to go out of its way to kill anything that brought a moment of pleasure to heroes of the story. They don't give Laura's age in the movie, but the actress is 11 years old. Having her kill people over and over and throwing a severed head to the leader of the villains is just going a bit overboard.

Оценка 4:

This film is absolutely brilliant. The makers of the Harry Potter franchise finally pulled themselves together, and made something awesome. I am a huge fan of the books, but I haven't been a real fan of the films, because they've been constantly disappointing me. I've read all the books at least five times, so I now them almost word for word, so maybe that's why the other films (especially the fourth one) were such disappointments for me. But this one, the eighth film in the series, was everything I could hope for. It was true to the book, the visuals were striking, the acting was over the top, the music was moving and the script was (finally) well-adapted. Maybe all these things are due to the fact that J. K. Rowling was an executive producer on this. I liked that there were some funny moments and some extra action in the movie which weren't in the book, but I missed one thing they left out. I won't tell you what it is, because I don't want to spoil the movie for those who haven't seen it yet (shame on you! go see it). I'm glad John Williams didn't come back for this film to compose, Alexandre Desplat's music is so full of emotions and has great tunes. I actually cried at one scene (because of the music and Alan Rickman's great acting), although films don't usually have that effect on me anymore. All in all the film was great fun, in some way better than the book, so for a Potter fan it's a must see.

Оценка 8:

'He is preaching.punishing'. Serial killing with clues from killer is not a new thing to Hollywood but David Fincher's Se7en exceptional. Its evil genius. Se7en opens with a homicide where a fatso is killed in a very strange manner. We meet Detective Somerset(Freeman) who is about to take off and detective David Mills(Pitt) has just fought to get assigned there. Somerset is calm, mature and has achieved mastery over his job. He probably seen so much that now he badly wants to go far away from this bloodshed. David appears cocky first but turns out amiable character of the film. Tracey(Palthow) is David's wife who hasn't got used to with guns even after so many years of their marriage. It begins. Each killing is shockingly repellent. Like an artist's signature every victim holds a note, a sin highlighted with lines from Dante Alighieri' s 'Divine Comedy', Shakespeare's 'Merchant of Venice'. And his name is John Doe(Kevin Spacey). Those lines are screaming that this not just some insanity. He wants to make a point. Each murder is done by reaching horrific extremes of those deadly sins. A fat man is fed until he burst. Then there is an infamous lawyer got gun stuck on his head, handed a knife -a weighing scale is in front of him and he has to cut 1 pound of flesh from his body

by himself. A druggist tied to bed for a whole year!! A famous model, her nose is cut and bandaged again and gave a choice- phone in one hand to call help n sleeping pills in other one to die rather than live deformed. A whore is punished ruthlessly that you should see on screen only. John says, “The world is so shitty and we get used to them”. This Script is sumptuous. It mirrors today’s society skillfully and boldly. Nothing is hidden by the camera. All we watch is naked truth that we hate to admit and most of the time that we hardly care. The spirituality woven in has its impact all over. Most of all ending is a blow. It is stunning. There are number of memorable sequences in Seven. One especially is breath-taking when Somerset and Mills reaches John doe’s apartment and Doe directly starts shooting at them and chase sequence followed is absolute fabulous. On other hand dinner party at Mills house we feel wine mixed in the air. Tracey and Somerset’s meeting at coffee shop is the finest ones I ever seen. The intensity of that conversation can’t be described. This is a very dark film. It is all time raining. All environment surrounding Se7en is so grim and ghastly it clouds viewers with the same. Music refuse to leave us even when it is over. I must mention titles and credits rolled, one of the best I ever seen. Se7en comes out with strong performances. Morgan Freeman is truly outstanding. Brad Pitt made me his fan from this movie. The way he says “Ladies and Gentleman we have a homicide here” and same David when taunted by Somerset “You are saying you care for these people?” says with assuring impulse “Sure I do .!!!”. Gwyneth Paltrow ’s Tracey doesn’t have much screen presence but when she is there, we just can not stop love her character. In final half an hour we are introduced to Kevin Spacey’s John Doe. The preacher. What can I say about this chilling work? Spacey gave such a performance that stings. We hate to admit with this man but somewhere we know he is right. Somerset asks him, “So you are saying some higher power tell you to do this?” Just watch John Doe reply “Lord works in mysterious ways..”. Interesting thing pointed out by one of the fan is Somerset and John Doe both observed the same facts about today’s society but responds in different ways, Somerset want to run away and Doe takes action. No matter how many thrillers I shall see but this will always one of my favorites. It affects deeply. That’s the power of Se7en. Se7en is a terrific thriller and a modern classic.

Оценка 10:

This movie is one of the greatest of all time. It is adapted from a book by Chuck Palahniuk. This movie has very interesting themes like emasculation, violence, chaos, societal breakdown, isolation, the threat of death and consumerism. The direction is

sublime. Perfect cinematography, pacing and editing. The twists and nuances of the book are captured perfectly. Also they did a good job with the inter-cuts. Brad Pitt and Edward Norton were the perfect choice to lead this movie. When you are watching the movie you just are glued to your seat, that is just how good it is.

Оценка на IMDb	Полученная оценка
1	0.06
2	0.38
4	0.99
8	0.99
10	0.99

Несовпадение результатов для оценки 4 и 8 можно объяснить содержимым самих оценок.

Вывод

В ходе лабораторной работы была написана программа, классифицирующая текст с помощью ансамбля моделей.

```
from typing import Dict, List, Tuple

import numpy as np
from keras.datasets import imdb
from keras.layers import GRU, LSTM, Conv1D, Dense, MaxPooling1D, SimpleRNN
from keras.layers.embeddings import Embedding
from keras.models import Model, Sequential
from keras.preprocessing import sequence

INDEX = imdb.get_word_index()


def load_data(
    num_words=5000, max_review_length=500
) -> Tuple[Tuple[np.array, np.array], Tuple[np.array, np.array]]:
    (training_data, training_targets), (testing_data, testing_targets) = imdb.load_
        num_words=num_words
    )
    data = np.concatenate((training_data, testing_data), axis=0)
    targets = np.concatenate((training_targets, testing_targets), axis=0)

    train_length = (data.shape[0] // 10) * 8
    train_data = data[:train_length]
    test_data = data[train_length:]
    train_targets = targets[:train_length]
    test_targets = targets[train_length:]

    train_data = sequence.pad_sequences(train_data, maxlen=max_review_length)
    test_data = sequence.pad_sequences(test_data, maxlen=max_review_length)
    return (train_data, train_targets), (test_data, test_targets)
```

```
def convert_text(text: str, max_size=5000) -> np.array:
    result = []

    for word in text.split():
        index = INDEX.get(word.lower())
        if index is None or index + 3 > max_size:
            continue
        result.append(index + 3)

    reverse_index = dict([(value, key) for (key, value) in INDEX.items()])
    decoded = " ".join([reverse_index.get(i - 3, "#") for i in result])
    print(decoded)

    return result


def ensemble_predictions(models: List[Model], data: np.array) -> float:
    predictions = [model.predict(data) for model in models]
    return np.mean(np.asarray(predictions), 0).flatten()


def evaluate_ensemble(models: List[Model], data: np.array, targets: np.array) -> fl
    predictions = np.round(ensemble_predictions(models, data))
    return np.count_nonzero(predictions == targets) / targets.shape[0]


def create_lstm_model(top_words=5000, embedding_vecor_length=32, max_review_length=
    model = Sequential()
    model.add(Embedding(top_words, embedding_vecor_length, input_length=max_review_
    model.add(Conv1D(filters=32, kernel_size=3, padding="same", activation="relu"))
    model.add(MaxPooling1D(pool_size=2))
    model.add(LSTM(100, dropout=0.2))
    model.add(Dense(1, activation="sigmoid"))
    model.compile(loss="binary_crossentropy", optimizer="adam", metrics=["accuracy"
```

```
    return model


def create_simple_rnn_model(
    top_words=5000, embedding_vecor_length=32, max_review_length=500
) -> Model:
    model = Sequential()
    model.add(Embedding(top_words, embedding_vecor_length, input_length=max_review_
    model.add(Conv1D(filters=32, kernel_size=3, padding="same", activation="relu"))
    model.add(MaxPooling1D(pool_size=2))
    model.add(SimpleRNN(64, return_sequences=True, dropout=0.2))
    model.add(SimpleRNN(64))
    model.add(Dense(1, activation="sigmoid"))
    model.compile(loss="binary_crossentropy", optimizer="adam", metrics=["accuracy"

    return model


def create_gru_model(top_words=5000, embedding_vecor_length=32, max_review_length=5
    model = Sequential()
    model.add(Embedding(top_words, embedding_vecor_length, input_length=max_review_
    model.add(Conv1D(filters=32, kernel_size=3, padding="same", activation="relu"))
    model.add(MaxPooling1D(pool_size=2))
    model.add(GRU(64, return_sequences=True))
    model.add(LSTM(32, return_sequences=True, dropout=0.2))
    model.add(GRU(16))
    model.add(Dense(1, activation="sigmoid"))
    model.compile(loss="binary_crossentropy", optimizer="adam", metrics=["accuracy"

    return model


def main(review: str):
```

```
    (train_data, train_targets), (test_data, test_targets) = load_data()
    models = [create_lstm_model(), create_simple_rnn_model(), create_gru_model()]
    for index, model in enumerate(models):
        print(f"Model {index + 1}")
        print(model.summary())
        model.fit(
            train_data,
            train_targets,
            validation_data=(test_data, test_targets),
            epochs=3,
            batch_size=64,
        )
        scores = model.evaluate(test_data, test_targets)
        print("Accuracy: %.2f%%" % (scores[1] * 100))

    print(f"Ensemble accuracy: {evaluate_ensemble(models, test_data, test_targets)}")

    review_vector = convert_text(review)
    review_vector = sequence.pad_sequences([review_vector], maxlen=500)
    prediction = ensemble_predictions(models, review_vector)
    print(f"Prediction for review is: {prediction}")


if __name__ == "__main__":
    text = input()
    main(text)
```