

Гордеев А.С. КЭ-401

Ссылка на Google Colab: <https://colab.research.google.com/drive/1vGBvHb-29a-Zqlr2GVjR2ID6iRwIUt5V?usp=sharing>

Подготовка данных:

1. Нейросеть для классификации отзывов фильма;
2. Построение модели нейронной сети и ее задачи: данная модель нейронной сети предназначена для анализа отзывов в наборе данных IMDb. Это задача обработки естественного языка (Natural Language Processing или же NLP), цель которой — определить выражает ли отзыв положительные или отрицательные настроения;
3. Первоначальный объем данных содержит:
 - Объем данных: набор данных IMDb представляет собой набор из 50 000 обзоров фильмов из базы данных фильмов;
 - Общий размер набора данных: около 80 МБ;
 - Ссылка на ресурс: <https://storage.googleapis.com/tensorflow/tf-keras-datasets/imdb.npz>.
4. Формирование обучающей выборки:
 - Объем данных для обучения модели: 25 000 обзоров (около 40 МБ) фильмов каждый из которых помечен определенным настроением (положительным или отрицательным);
 - Объем данных для тестирования модели: 25 000 обзоров (около 40 МБ) фильмов с соответствующими метками отдельно от данных обучения;
 - Другие действия предварительной обработки и подготовки:
 - Токенизация: набор данных уже токенизирован, каждое слово было заменено целочисленным индексом;
 - Заполнение последовательности: последовательности дополняются, чтобы гарантировать, что все входные последовательности в модель имеют одинаковую длину;

- Разделение: набор данных предварительно разделяется на обучающий и тестовый наборы.

Обучение модели:

1. Нейросеть для классификации отзывов фильма.
2. Описание:
 - а. Оптимизатор: *Adam*
 - б. Функция потерь: для данной задачи двоичной классификации, общей функцией потерь является *бинарная кроссэнтропия*, которая используется, когда модель выводит одну оценку вероятности в качестве прогноза.
 - с. Метрики: для двоичной классификации *точность* является общей метрикой.
3. Параметры обучения:
 - а. Размер батча: 64
 - б. Количество эпох: 3
 - с. Время обучения: 30 минут

Номер этапа изменения параметров нейронной сети	Оптимизатор (optimizer)	Функция потери (lossfunction)	Метрики (metrics)	Размер батчей (batch)	Количество эпох обучения (epoch)	Другие параметры (при наличии указать)	Точность обучения сети (%)
1	Adam	Binary crossentropy	Accuracy	64	3	-	86.78
2	Adam	Binary crossentropy	Accuracy	64	5	-	85,73
3	Adam	Binary crossentropy	Accuracy	32	10	-	85,58