

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. Н.Э. Баумана

Кафедра «Системы обработки информации и управления»

ОТЧЕТ

Лабораторная работа №7

по курсу «Проектирование интеллектуальных систем»

Тема: «Использование нейронных сетей для анализа текста»

ИСПОЛНИТЕЛЬ:

группа ИУ5-22

Петрова Ирина

ФИО

подпись

"25" 05 2020 г.

ПРЕПОДАВАТЕЛЬ:

ФИО

подпись

" " _____ 2020 г.

Москва - 2020

Задание

Итоговый код для обучения нейросети и оценки ее точности содержится в приложении. Необходимо увеличить количество скрытых слоев до 3-ех, а количество нейронов в этих слоях так, чтобы обеспечить точность работы нейросети не менее 75%. Темы текстов необходимо изменить в соответствии с вариантом:

Вариант	Темы текстов
10	comp.windows.x, rec.motorcycles, sci.crypt, sci.space

Выполнение

```
# Параметры обучения
learning_rate = 0.01
training_epochs = 10
batch_size = 150
display_step = 1

# Network Parameters
n_hidden_1 = 10 # скрытый слой
n_hidden_2 = 10 # скрытый слой
n_hidden_3 = 5 # скрытый слой
n_input = total_words # количество уникальных слов в наших текстах
n_classes = 4 # 4 класса
input_tensor = tf.placeholder(tf.float32,[None, n_input],name="input")
output_tensor = tf.placeholder(tf.float32,[None, n_classes],name="output")

def multilayer_perceptron(input_tensor, weights, biases):
    # скрытый слой
    layer_1_multiplication = tf.matmul(input_tensor, weights['h1'])
    layer_1_addition = tf.add(layer_1_multiplication, biases['b1'])
    layer_1 = tf.nn.relu(layer_1_addition)

    # скрытый слой
    layer_2_multiplication = tf.matmul(layer_1, weights['h2'])
    layer_2_addition = tf.add(layer_2_multiplication, biases['b2'])
    layer_2 = tf.nn.relu(layer_2_addition)

    layer_3_multiplication = tf.matmul(layer_2, weights['h3'])
    layer_3_addition = tf.add(layer_3_multiplication, biases['b3'])
    layer_3 = tf.nn.relu(layer_3_addition)

    # выходной слой
    out_layer_multiplication = tf.matmul(layer_3, weights['out'])
    out_layer_addition = out_layer_multiplication + biases['out']
    return out_layer_addition
```

```
Эпоха: 0001 loss= 188.6810341278712144
Эпоха: 0002 loss= 24.6067061821619646
Эпоха: 0003 loss= 1.7154735962549845
Эпоха: 0004 loss= 1.5524506131807962
Эпоха: 0005 loss= 1.5262028694152832
Эпоха: 0006 loss= 1.4997873028119408
Эпоха: 0007 loss= 1.4774873534838362
Эпоха: 0008 loss= 1.4622506340344745
Эпоха: 0009 loss= 1.4442233363787336
Эпоха: 0010 loss= 1.4331837773323057
Обучение завершено!
Точность: 0.9974732
```

Выводы

В ходе лабораторной работы были получены навыки предобработки текстовых данных для использования в нейросетях, а также использования нейронной сети для классификации текстов.

Контрольные вопросы

1. Какие вы знаете задачи обработки текстов, в чем они заключаются?

Классификация текста – определение его тематики, кластеризация – разделение набора текстов на группы и машинный перевод текстов.

2. Зачем нужна предобработка текста для машинного обучения?

Чтобы повысить точность работы нейронной сети

3. Какие виды предобработки текста вы знаете?

Стемминг, токенизация, удаление знаков препинания, удаление шумовых слов, нерелевантных слов и символов, приведение текста к нижнему регистру.

4. Что такое стемминг?

Замена различных форм слова одной, также замена синонимичных слов одним словом.

5. Что такое 20 Newsgroups?

Набор, который состоит из около 20 тысяч постов на 20 различных тем

6. Чему должно равняться число входных и выходных нейронов в задаче классификации текстов?

Число входных нейронов – количеству уникальных слов в текстах, выходных – количеству тем для классификации.

Список литературы

- [1] Документация по tensorflow. <https://www.tensorflow.org>.
- [2] Описание 20 Newsgroups. <http://qwone.com/~jason/20Newsgroups/>
- [3] Глубокое обучение для NLP. <https://nlp.stanford.edu/courses/NAACL2013/>