МИНОБРНАУКИ РОССИИ

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ

ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)

Кафедра МО ЭВМ

ОТЧЕТ

по лабораторной работе №8

по дисциплине «Искусственные нейронные сети»

Тема: Генерация текста на основе «Алисы в стране чудес»

Студент гр. 7383	____________________	Зуев Д. В.
Преподаватель	____________________	Жукова Н.А.

Санкт-Петербург

2020

Цель работы:

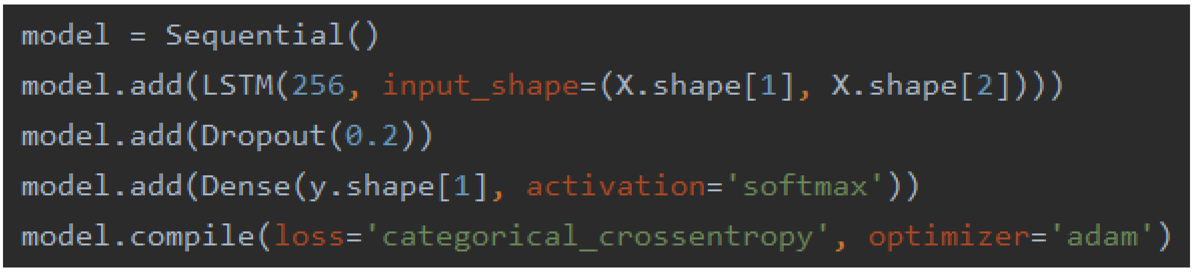
Реализовать генерацию текста на основе текста из сказки «Алиса в стране чудес»

Задачи.

1. Ознакомиться с генерацией текста
2. Ознакомиться с системой Callback в Keras

Ход работы.

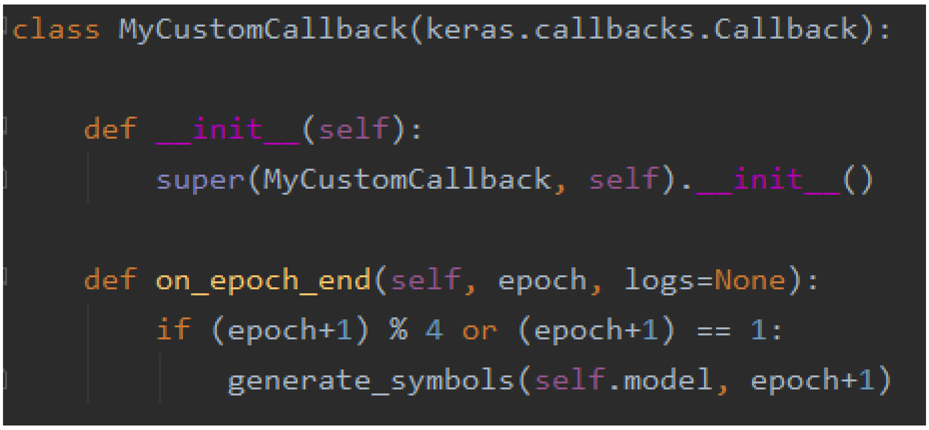
1. Была создана и скомпилирована модель рекуррентной нейронной сети, для прогнозирования следующего символа на основе предыдущих. В качестве функции потер была выбрана функция категориальной кроссэнтропии, так как нам необходимо результатом обработки входных данных иметь один символ из набора, что можно интерпретировать как отнесение входного набора данных к одному из 47 классов. Архитектура сети представлена на рис. 1.



```
model = Sequential()
model.add(LSTM(256, input_shape=(X.shape[1], X.shape[2])))
model.add(Dropout(0.2))
model.add(Dense(y.shape[1], activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam')
```

Рисунок 1 - Архитектура сети

2. Для отслеживания прогресса генерации текста во время обучения нейронной сети был написан класс `MyCustomCallback`, который по окончании определенных эпох выводит в файл текст, сгенерированный нейронной сетью. Класс представлен на рис. 2.



```
class MyCustomCallback(keras.callbacks.Callback):

    def __init__(self):
        super(MyCustomCallback, self).__init__()

    def on_epoch_end(self, epoch, logs=None):
        if (epoch+1) % 4 or (epoch+1) == 1:
            generate_symbols(self.model, epoch+1)
```

Рисунок 2 - Класс MyCustomCallback

При переопределении функции on_epoch_end была допущена ошибка, которая выяснилась уже после завершения обучения, а именно условие запуска генерации текста вместо каждой четвертой эпохи выбирало все эпохи кроме кратных четырем, но данное упущение никак не влияет на возможность отслеживания прогресса обучения.

3. Возьмем часть сгенерированных тестов и проанализируем их. Прогресс обучения представлен в табл. 1.

Таблица 1. Прогресс обучения

Эпоха	Исходный текст	Сгенерированный текст
1	" but, when the tide rises and sharks are around, his voice has a timid and tremulous sound.] 't"	" i a "
5	"e won't talk about her any more if you'd	" to the woeee 'nh tou d sore the woee to tee toee she said to the woete 'nh tou d soet to tee toee she said to the warte to the waste the whete and the woele toet

	rather not.' 'we indeed!' cried the mouse, who was trembli"	aadi and the woel toe was ho the was "
9	"ing those roses?' five and seven said nothing, but looked at two. two began in a low voice, 'why th"	" the toiee to tee the same th the toiee 'the dad tot daoe to tee the soie "
13	"e white rabbit, who said in a low voice, 'your majesty must cross-examine this witness.'	"tour the tait oo toine a ait,' said the daterpillar. 'wou dad'netel thet toued ' she hatter weit on, 'in so tel the toien to teit to toe the tooes. 'ie you dan tot oi the soie ' the said to herself, 'io s soe

	'well, if ''	teee to toele to toe thee 'that so het toe thieg ' said the daterpillar. 'wou dad'netel thet toued ' she hatter weit on, 'in so tel the toien to teit to toe the tooes. 'ie you dan tot oi the soie ' the said to herself, 'io s soe teee to toele to toe thee 'that so het toe thieg ' said the daterpillar. 'wou dad'netel thet toued ' she hatter weit on, 'in so tel the toien to teit to toe the tooes. 'ie you dan tot oi the soie ' the said to herself, 'io s soe teee to toele to toe thee 'that so het toe thieg ' said the daterpillar. 'wou dad'netel thet toued ' she hatter weit on, 'in so tel the toien to teit to toe the tooes. 'ie you dan tot oi the soie ' the said to herself, 'io s soe teee to toele to toe thee 'that so het toe thieg ' said the daterpillar. 'wou d''
17	''such a capital one	''ters 'i whit you would '

	for catching mice you can't think! and oh, i wish you could see her after the bir"	'i moot what io ' said the manch hare. 'i mener taet io ' said the manch hare. 'i mener toete io shens,' said the caterpillar. 'i mene the sait of that ' she said to herself, and the puren sald to tee kotk of the sabbi oo tee thete wab iut the rabbi and anireen to toe the tabbe to tee that she was so teen to the thitg theee the had nut toene the har hn the wirl and whs dnl the coumtuse thing and the was so teen to the thitg thete was so toene to toene the was so toene th the tabde an in shsh the had not hor hort the had not hor hort the was so toeke th the tabde an in sas oo the tire, and she teit oo the houph see was so toeke to tee thet and sas io a lore of thene was oo the sire she har so tee tooe, 'n shanl ieve to tey 'toe ment whth the somes.' 'i whsh you would ' she manch hare an anl alice an an offe an an offe an the toike an anl the couro, and the was so tee tooed the had not toene the har hn the wirl and whs dnl the coumtuse thing and the w"
19	"e rattling in at the window, and some of them hit her in the face. 'i'll put a stop to this,' she sa"	"id to herself, 'io saa ietter aedin sor dnon the hirtt so ce a fred talee-- 'that d tooded the mirtle th teen ' she taid to herself, 'she hirst said toan the roeee of the sard th the toide to thet would her hedd. 'whet sare youn the soad tuted ' saed the manch hare.

		'iere tou dave to tene then ' said the qabbit in a tore of the sabee 'and thet war io a lorg tf bottee and saed to the tord, "then i d nerer see soie ' she mant and mard taid an anice. 'wou did" said the manch hare. 'iere tou dave to tene then ' said the qabbit in a coea fooler to tee io the sire the was so tie tine, and she tai tot thi tas tointing ana notking at the could, 'the dirst then so be a little birler ' the murk turtle seided dotn and the pabbit here a little soiee on the tare th the taaten and berind the tas oo the tire the was so tee thaten the was so tiene to bn anl ofr of the goure and the was sointing anange the was soin a lirgll of the sabbit and taed to the tordd ohe hn the hadd. 'the hirst "

По результатам обучения, видно, что сеть с каждой эпохой все лучше и лучше генерирует текст. Ближе к двадцатой эпохе заметны слова очень похожие на английские, и даже есть последовательности реальных слов, как например в девятнадцатой эпохе: «then so be a little».

Чтобы посмотреть результаты генерации сети после двадцатой эпохи была написана функция `run_best_model`. Эта функция выводит сгенерированный текст в консоль. При запуске она вывела следующий текст:

"little birler?' 'ii yhu make to heve tou then ' said the queen, 'and that i thould then to be a creae on the mooe tiiee tou dre the rabbit and the sese thing the was so toine oo the sooe, and she teit hnuw the past of the houph sas ao the was oo the hirte aod the doupd net herd aeiced ano oo the tine, and she thite tabbit sese thing the white rabbit, and the past whit was soene ali moee to paee thte the had not ooe the sabei th the said

```
the had pede and teet whth the labc tht oo the girte  and the doumd net herd aeiced wo
tee thet was the pieee of the gaadd,
'that sare you goon hi tou die ' said the mock turtle. "ie you mane to tel the sirtle ' shi
manch hare seil an anded.

'i meae then would been io the seee,' she mock turtle seplied. 'io s all the sabbit wotee
'
'i don't know thet io yhet mo yhe sint?' she match hare seil an and degedd
to eed to the thete
and she weit on thet sas so tee thaten  she west on to then toe tooe, and she teited an
in al anl toene tith the had not o"
```

Данный текст очень походит на текст, написанный на английском языке. В нем встречается достаточно большое количество реальных слов. Я думаю, человек, не особо знакомы с английским языком не сразу отличит этот текст от текста на английском.

Выводы:

В ходе выполнения данной лабораторной работы была построена модель, генерирующая текст на основе текста сказки «Алиса в стране чудес». Результат обучения сети показывает, что генерация проходит достаточно хорошо и встречаются осмысленные последовательности реальных английских слов. Также была изучена система callback в Keras, благодаря которой были сохранены модели на каждой из эпох, и был отслежен прогресс обучения путем генерации текста на определенных эпохах.

ПРИЛОЖЕНИЕ А

```
import numpy as np
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import Dropout
from keras.layers import LSTM
from keras.callbacks import ModelCheckpoint
import keras.callbacks
from keras.utils import np_utils
import sys



filename = "wonderland.txt"
file_with_gen_text = "generated.txt"
raw_text = open(filename).read()
raw_text = raw_text.lower()
chars = sorted(list(set(raw_text)))
char_to_int = dict((c, i) for i, c in enumerate(chars))
int_to_char = dict((i, c) for i, c in enumerate(chars))
n_chars = len(raw_text)
n_vocab = len(chars)
print("Total Characters: ", n_chars)
print("Total Vocab: ", n_vocab)
seq_length = 100
dataX = []
dataY = []
for i in range(0, n_chars - seq_length, 1):
        seq_in = raw_text[i:i + seq_length]
        seq_out = raw_text[i + seq_length]
        dataX.append([char_to_int[char] for char in seq_in])
        dataY.append(char_to_int[seq_out])
n_patterns = len(dataX)
print("Total Patterns: ", n_patterns)
X = np.reshape(dataX, (n_patterns, seq_length, 1))
X = X / float(n_vocab)
y = np_utils.to_categorical(dataY)
```

```
class MyCustomCallback(keras.callbacks.Callback):

    def __init__(self):
        super(MyCustomCallback, self).__init__()

    def on_epoch_end(self, epoch, logs=None):
        if (epoch+1) % 4 or (epoch+1) == 1:
            generate_symbols(self.model, epoch+1)



def generate_symbols(model_, epoch):
    g = open(file_with_gen_text, 'a')
    gen_symbols = []
    start = np.random.randint(0, len(dataX) - 1)
    pattern = dataX[start]
    g.write("On epoch " + str(epoch) + " seed is:\n" + "\"" +
''.join([int_to_char[value] for value in pattern]) + "\"\n")
    for i in range(1000):
        x = np.reshape(pattern, (1, len(pattern), 1))
        x = x / float(n_vocab)
        prediction = model_.predict(x, verbose=0)
        index = np.argmax(prediction)
        result = int_to_char[index]
        gen_symbols.append(result)
        pattern.append(index)
        pattern = pattern[1:len(pattern)]
    g.write("generated text
is:\n"+"\""+''.join(gen_symbols)+"\"\n")
    g.close()



def build_model():
    model = Sequential()
    model.add(LSTM(256, input_shape=(X.shape[1], X.shape[2])))
    model.add(Dropout(0.2))
    model.add(Dense(y.shape[1], activation='softmax'))
    model.compile(loss='categorical_crossentropy',
optimizer='adam')
    return model



def run_fit():
    model = build_model()
    filepath="weights-improvement-{epoch:02d}-{loss:.4f}.hdf5"
    checkpoint = ModelCheckpoint(filepath, monitor='loss',
verbose=1, save_best_only=True, mode='min')
    callbacks_list = [checkpoint, MyCustomCallback()]
    model.fit(X, y, epochs=20, batch_size=128,
callbacks=callbacks_list)



def run_best_model():
    model = build_model()
    model.load_weights("weights-improvement-20-1.9430.hdf5")

    gen_symbols = []
    start = np.random.randint(0, len(dataX) - 1)
    pattern = dataX[start]
    print("Seed:")
    print("\"", ''.join([int_to_char[value] for value in
pattern]), "\"")
    for i in range(1000):
        x = np.reshape(pattern, (1, len(pattern), 1))
        x = x / float(n_vocab)
        prediction = model.predict(x, verbose=0)
        index = np.argmax(prediction)
        result = int_to_char[index]
        gen_symbols.append(result)
        pattern.append(index)
        pattern = pattern[1:len(pattern)]
    print("generated text is:\n"+"\""+''.join(gen_symbols)+"\"\n")



run_best_model()
```