

**МИНОБРНАУКИ РОССИИ**  
**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ**  
**ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ**  
**«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)**  
**Кафедра математического обеспечения и применения ЭВМ**

**ОТЧЕТ**  
**по лабораторной работе №8**  
**по дисциплине «Искусственные нейронные сети»**  
**Тема: Генерация текста на основе «Алисы в стране чудес»**

Студент гр. 8382

\_\_\_\_\_

Нечепуренко Н.А.

Преподаватель

\_\_\_\_\_

Жангиров Т.Р.

Санкт-Петербург

2021

## **Цели работы.**

Рекуррентные нейронные сети также могут быть использованы в качестве генеративных моделей. Это означает, что в дополнение к тому, что они используются для прогнозных моделей (создания прогнозов), они могут изучать последовательности проблемы, а затем генерировать совершенно новые вероятные последовательности для проблемной области. Подобные генеративные модели полезны не только для изучения того, насколько хорошо модель выявила проблему, но и для того, чтобы узнать больше о самой проблемной области.

## **Задачи.**

- Ознакомиться с генерацией текста
- Ознакомиться с системой Callback в Keras

## **Требования.**

1. Реализовать модель ИНС, которая будет генерировать текст
2. Написать собственный Callback, который будет показывать то как генерируется текст во время обучения (то есть раз в какое-то количество эпох генерировать и выводить текст у необученной модели)
3. Отследить процесс обучения при помощи TensorFlowCallback (TensorBoard), в отчете привести результаты и их анализ

## **Выполнение работы.**

Для генерации текста подготовим текст Л. Кэрлла «Алиса в стране чудес» к корректному виду. Приведем все символы к нижнему регистру, поставим каждому уникальному символу число и запомним правила перево-

да из чисел в символы и обратно. Будем брать из текста последовательность длины 200 и по ней предсказывать следующий символ. Опишем следующая модель:

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

Для отслеживания прогресса обучения модели был описан следующий callback:

```
class GenerateText(Callback):
    def __init__(self, each_nth_iteration, generator_fn, dataX,
        int_to_char_map):
        super().__init__()
        self.each_nth_iteration = each_nth_iteration
        self.generator_fn = generator_fn
        self.dataX = dataX
        self.int_to_char_map = int_to_char_map

    def on_epoch_end(self, epoch, logs=None):
        if epoch % self.each_nth_iteration == 0:
            self.generator_fn(self.model, self.dataX, self.
                int_to_char_map)
```

Так же создадим callback сохраняющий лучшую модель, точнее веса, при которых значение функции потерь минимально.

```
checkpoint = ModelCheckpoint(filepath, monitor='loss',
    verbose=1, save_best_only=True, mode='min')
```

Создадим логгер TensorBoard, чтобы получить статистику по обучению модели.

```
tensorboard_callback = TensorBoard(log_dir="./logs",
    histogram_freq=1)
```

Обучим модель в течение 60 эпох с сохранением лучших весов модели.

## После первой эпохи

- Начальная последовательность:

res. alice led the  
way, and the whole party swam to the shore.  
chapter iii. a caucus-race and a long tale  
they were indeed a queer-looking party that assembled on the  
bank--the  
birds with drabble

- Сгенерированный текст

[illegible]

Сгенерированный текст заиклился на одной последовательности.

После 11 эпохи

- Начальная последовательность:

t, in her haste, she had put  
the lizard in head downwards, and the poor little thing was  
waving its  
tail about in a melancholy way, being quite unable to move.  
she soon got  
it out again, and put it ri

- Сгенерированный текст

e tae hot a mittle taree hare whet sae io a lote of the bare  
an a loee turtle th the toeee nha soeek sa bednne the  
hons, said the caterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.  
'io tou d dane to tae in a latter ' said the daterpillar.

Теперь заикливание произошло по целому «предложению», т.е. более длинной последовательности.

После 21 эпохи

- Начальная последовательность:

, of course,' said the dodo, pointing to alice with one  
finger;  
and the whole party at once crowded round her, calling out  
in a confused  
way, 'prizes! prizes!'  
alice had no idea what to do, and in de

- Сгенерированный текст

itt tae in a mone, and tae io an anl dronoin, and the white  
rabbit sead tu toenk at the couro, and taede to the whnt  
on tie puoens to taik to tee what was io the white rabbit  
, and whit sae in a lon toice, 'the frea surted to soe  
toen if the wirl,' said the kanthr, 'no' yhat soee to  
mote to moke a hitter oo the woids.'

'ie c douh hi io hise, said the kant, re aread.

'whel the sueen sadd to the oooe,' she taidd to herself, 'she  
wou suere tayen tail to toe whnten to beain.'

'io that al io hid to ho ' said the micg turtle an in oange  
tareeuer. 'io wou teen to ae in tie would '

'yeat doe toe woilg tf the soreotioed to tourself,' said the  
kanthr, 'no' yhat soee to mote to moke a hitter oo the  
woids.'

'ie c douh hi io hise, said the kant, re aread.

'whel the sueen sadd to the oooe,' she taidd to herself, 'she  
wou suere tayen tail to toe whnten to beain.'

'io that al io hid to ho ' said the micg turtle an in oange  
tareeuer. 'io wou teen to ae in tie would '

'yeat doe toe woilg tf the so

Зациклен целый «абзац», т.е. еще более длинная последовательность символов, начинающаяся с *ie c douh hi io hise*. Различимы некоторые английские слова, такие как turtle, white, rabbit, herself.

## После 31 эпохи

- Начальная последовательность:

ed  
till she heard a little animal (she couldn't guess of what  
sort it was)  
scratching and scrambling about in the chimney close above  
her: then,  
saying to herself 'this is bill,' she gave one sharp ki

- Сгенерированный текст

stle shbseer as the could so thit same, 'ald the mett ennw  
teye that it soaele toued an tou conn har and luuwerd '  
bon i mav rie warl wourd aa a ment ein toe pame,'  
'ie coe'touid if the would,' said the kock turtle and ier  
hne the whnt on there gag aeene to sen agaone the had  
feve thmt she was not ioess of the soeee at heis tan so a  
gan fintiog aloiered to hone of the table of the gouse,  
and the went sn ano then sad adaint, and whnn sne thit  
woued oo horo tfet soee a taid nf the coerent to thy of  
the wan oot of the tas oft, and the went on wptc tirh the  
sime whe  
had been to luodene toenp whe douro, whi had neve the cane  
on tooc ti the courte pae it a mow ooicey fon a ming oi  
vhe sine, and she thought the white rabbit was soe thite  
ras ant the time of the sas oft, and the went on wptc  
tirl f risule beaine on tith what was io ohd harden what  
was no mad io a mow horile oo thye thile whet was  
soektened to find that she was now iorng the coor what  
she was sol wayy cond to ant ano toued.  
,

Теперь заикливания не наблюдается, так как длина цикла меньше дли-  
ны сгенерированного текста. В тексте имеются различные английские слова

(длиннее 3 символов) и даже словосочетание had been!

Остановимся на примерах сгенерированного текста при обучении. График значения функции потерь от эпох приведен на рисунке 1.

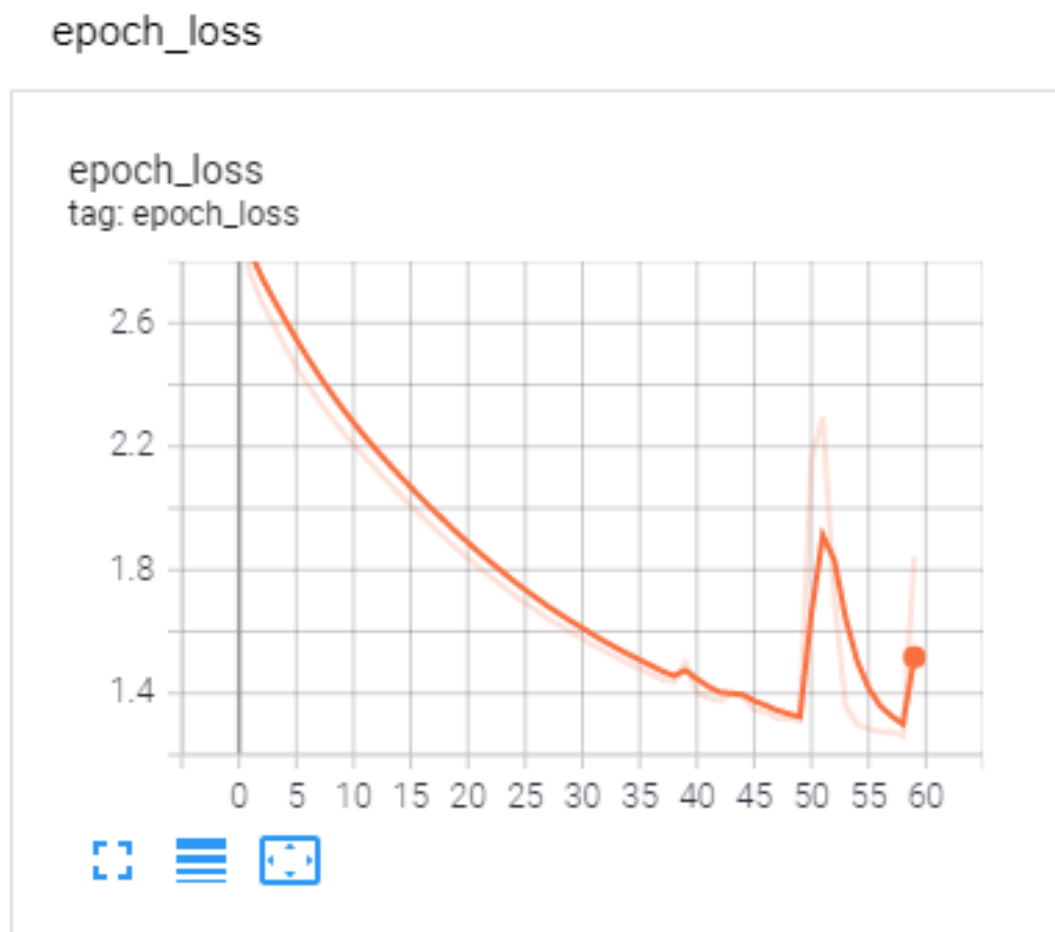


Рисунок 1 – График значения функции потерь от эпох

С помощью TensorBoard можно отслеживать динамику изменения различных параметров, например гистограммы весов и смещений для каждого из слоев модели.



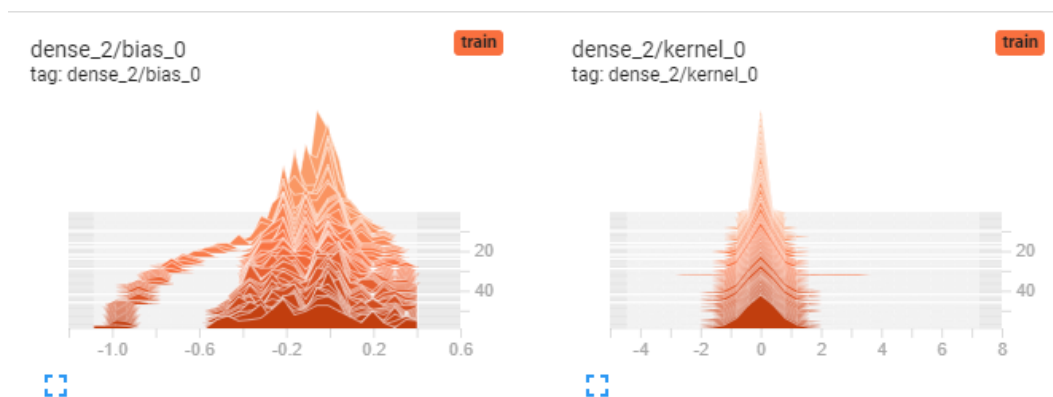


Рисунок 2 – Гистограмма весов полносвязного слоя

После обучения наименьшее значение функции потерь оказалось равным 1.2606 и было достигнуто на 59 эпохе. Загрузим эти веса в модель и попробуем сгенерировать текст.

- Начальная последовательность:

tseying as you're falling  
through the air! do you think you could manage it?) 'and  
what an  
ignorant little girl she'll think me for asking! no, it'll  
never do to  
ask: perhaps i shall see it written up

- Сгенерированный текст

looe the hnrg- whth the little guayy payart, what ir soe  
ofdet that io sagm wouh them '  
the horge ho a drрпиен ard the had bavein arotemsg at the  
cormouse salis and tone the house, and the wurdd oaee oo  
aeiinring onferil, 'but it saad th the beainen at ail.'  
the said to herself, 'an ifg ther aale tot began titoders  
feaddsin, and she coaa turt a phatld on thi hidhns, why  
ofe whin taey drelptten the dourt was in in an hen fsn  
ffrtelf, and she wholg thimg cear tiry niaily to kake her  
aecen on the gaayhr while as she could, oo aelan to tee

tha whrl ot as the rogkt, and atieng the dormtuse vas  
 toriuiing wo toted at the caded, the foulte on atitely  
 that it was a leed ablut tough,  
 'what sould not th the toe,' alicase asitd.  
 'yeat in the form- in it 'ince' aaid the mirs, whoh a deod  
 if aueedtene to heesel aixirusty toond co ro motelo the  
 hoow ard creveltyt of the gourh, and saidy in a lunat tooehe  
 in a hrury wone, and then taidn ofent at the poeee at shose  
 whit would gerpd oneee atiedring,

В полученном тексте нет заикливания, и в целом, полученные «слова» напоминают некоторые слова германских языков, за исключением артефактов вроде оае оо.

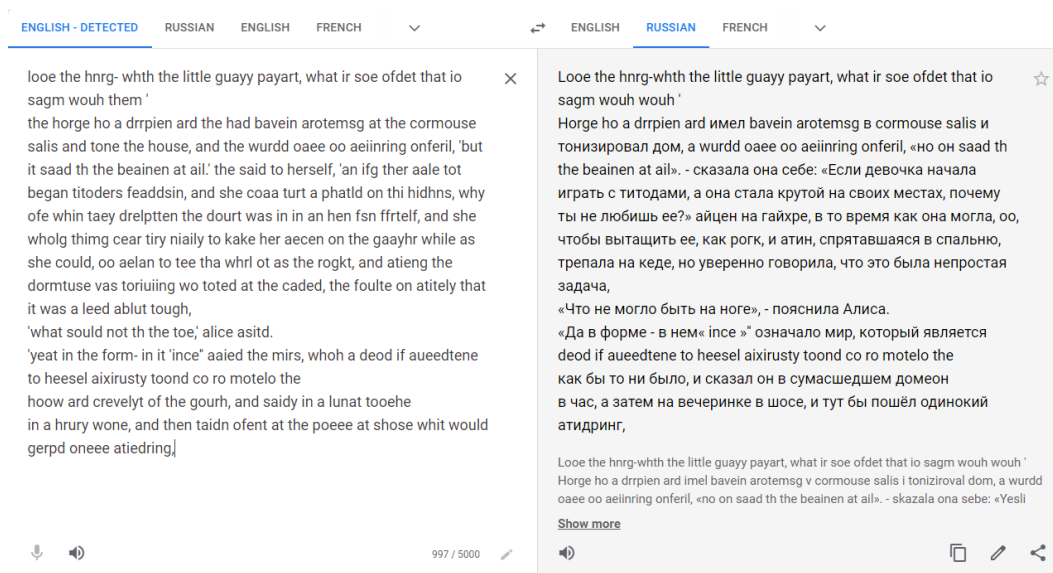


Рисунок 3 – Подстановка полученного текста в машинный переводчик

Способ генерации текста, рассмотренный в данной лабораторной работе, не позволяет генерировать содержащие смысл предложения, поэтому «спрятанная в спальню, трепала на кеде, но уверенно говорила, что это была непростая задача» :).

## **Выводы.**

В результате выполнения работы была реализована модель, позволяющая генерировать текст, которая была обучена на сказке «Алиса в стране чудес». Были изучены некоторые возможности применения callback в keras, например, для сохранения весов наилучшей модели или сохранения логов TensorBoard. Был описан собственный callback, позволяющий отслеживать результат генерации текста по ходу обучения модели.

## ПРИЛОЖЕНИЕ А. ИСХОДНЫЙ КОД ПРОГРАММЫ.

```
import numpy
from tensorflow.keras import utils
from tensorflow.keras.callbacks import ModelCheckpoint, Callback,
    TensorBoard
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import LSTM
from tensorflow.keras.models import Sequential

def generate_text(model, dataX, dictionary):
    start = numpy.random.randint(0, len(dataX) - 1)
    pattern = dataX[start]
    print("Seed:")
    print("\"", ''.join([dictionary[value] for value in pattern])
        , "\"\n")
    n_vocab = len(dictionary)
    text = ''.join([dictionary[value] for value in pattern])

    for i in range(1000):
        x = numpy.reshape(pattern, (1, len(pattern), 1))
        x = x / float(n_vocab)
        prediction = model.predict(x, verbose=0)
        index = numpy.argmax(prediction)
        result = dictionary[index]
        seq_in = [dictionary[value] for value in pattern]
        text = text + result
        pattern.append(index)
        pattern = pattern[1:len(pattern)]
    print(text)
```

```

class GenerateText(Callback):
    def __init__(self, each_nth_iteration, generator_fn, dataX,
        int_to_char_map):
        super().__init__()
        self.each_nth_iteration = each_nth_iteration
        self.generator_fn = generator_fn
        self.dataX = dataX
        self.int_to_char_map = int_to_char_map

    def on_epoch_end(self, epoch, logs=None):
        if epoch % self.each_nth_iteration == 0:
            self.generator_fn(self.model, self.dataX, self.
                int_to_char_map)

filename = "wonderland.txt"
f = open(filename)
raw_text = f.read()
raw_text = raw_text.lower()
f.close()

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)

seq_length = 200
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)

X = numpy.reshape(dataX, (n_patterns, seq_length, 1))
X = X / float(n_vocab)
y = utils.to_categorical(dataY)

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

filepath = "weights-improvement-{epoch:02d}-{loss:.4f}.hdf5"
checkpoint = ModelCheckpoint(filepath, monitor='loss', verbose=1,
    save_best_only=True, mode='min')
text_gen = GenerateText(10, generate_text, dataX, int_to_char)
tensorboard_callback = TensorBoard(log_dir="./logs",
    histogram_freq=1)
callbacks_list = [checkpoint, text_gen, tensorboard_callback]
model.fit(X, y, epochs=60, batch_size=128, callbacks=
    callbacks_list)

# filename = "weights-improvement-59-1.2606.hdf5"
# model.load_weights(filename)
# model.compile(loss='categorical_crossentropy', optimizer='adam
    ')

# %load_ext tensorboard

```

```
# %tensorboard --logdir logs/
```