

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. 8383

Федоров И.А.

Преподаватель

Жангиров Т.Р.

Санкт-Петербург

2021

Цель 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

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

Выполнение работы.

Были импортированы необходимые для работы классы, модули, функции, а также данные *wonderland.txt*.

По методическим указаниям была построена рекуррентная сеть, имеющая следующую структуру:

```
Model: "sequential"
```

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 256)	264192
dropout (Dropout)	(None, 256)	0
dense (Dense)	(None, 45)	11565
Total params: 275,757		
Trainable params: 275,757		
Non-trainable params: 0		

Была реализована функция `gen_text()`, которая с помощью переданной модели генерирует текст необходимой длины:

```
def gen_text(model, epoch=0, console_log=False, seq_len=1000):  
    # pick a random seed  
    start = np.random.randint(0, len(dataX)-1)
```

```

pattern = dataX[start]
if console_log:
    print("Seed:")
    print("\n", ''.join([int_to_char[value] for value in
pattern])), "\n")

full_result = []
# generate characters
for i in range(seq_len):
    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]
    full_result.append(result)
    seq_in = [int_to_char[value] for value in pattern]
    if console_log:
        sys.stdout.write(result)
    pattern.append(index)
    pattern = pattern[1:len(pattern)]
print("\nDone.")
return ''.join(full_result)

```

Был реализован собственный callback TrainGeneratTextLogger(), который показывает то, как генерируется текст во время обучения (то есть раз в какое-то количество эпох генерирует и выводит текст у необученной модели):

```

class TrainGeneratTextLogger(Callback):
    def on_epoch_end(self, epoch, logs=None):
        if not hasattr(self, 'freq'):
            self.freq = 2
        if not hasattr(self, 'save_file'):
            self.save_file = False
        if not hasattr(self, 'console_log'):
            self.console_log = True
        if not hasattr(self, 'seq_len'):
            self.seq_len = 1000

        if epoch % self.freq == 0:
            res = gen_text(self.model, epoch,
                           console_log=self.console_log,
                           seq_len=self.seq_len)
            if self.console_log:
                sys.stdout.write("Epoch {}: ".format(epoch))
                sys.stdout.write(res)
            if self.save_file:
                with open("gen_text_ep_{}".format(epoch) + ".txt",
"w") as file:
                    file.write(res)

    def set_freq(self, freq):
        self.freq = freq

```

```
def set_save_file(self, save_file):
    self.save_file = save_file

def set_console_log(self, console_log):
    self.console_log = console_log

def set_seq_len(self, seq_len):
    self.seq_len = seq_len
```

Так как обучение проходит очень медленно, то запуск производился в Google Colaboratory (файл .ipynb прикреплен к pull-request, текст программы приведет в приложении).

Были получены следующие результаты в ходе обучения:

Эпоха 0:

Seed: " Id like the name: however, it only grinned a little wider. 'come, it's pleased so far,' thought alic "
e oo herself, 'i wonder what they wal to toe cerer the things be in ou toine! bot is so have the sueen war in tne tinl to antins to tee that whal i wisl you wouldnge baarte, toan they was to toen to tou tha tointsend tore cierce then soee how hare an the way of saaring thet see walted to simek the harden. the was not in a moru arouger foose to and toon the was oow on the toon. "the mert iie inr tuicinng ' the hatter went on, 'and wou dan tank to toe bedrge the things of thet hare to sea that i'w a tert bi a catter and theng was to the thant oittle borser aatir theer earee and boicnsen, and sald that she was not on tigit iirt the was offe on the stoed. 'io wou dan' is d lird tay at anl whun mese,' said the monk augelily, 'io you don't knke that toitge to be at it do i mise. woal the weil to tou tho meae offe th than i rool thank you day' is they wall no tometheng io the tend tfing it to het. toen they would be an arerasse it a tiry oittle toaene to saa ie she fortte of the saabit- sine

Эпоха 4:

Seed: " loves while she was talking. 'how can i have done that?' she thought. 'i must be growing small again "
the soe woi gee the she kre cer shi gast do ae a go toue tt thene to then wou dare to the toe ke tee shnl ser doun the das re the tee te to the thgt setlen oo the the tere ano bou hocr to here the wae co she whe wouh a cou thet saabin hoo ho the too cfd hoee and the was co aet toit so the whet sere the ham thet soeer that werl sa the would beargh, and the wes dou to the thete wuund her har so the whs dad to the whe wes ka the sas sf the ths whuh a soept oo the taa cn shin wouhd bed she say to thi tae sall ano thst wha thuh a she tae salt anr the sou was tas to the the wet dad the ras he the sas saske ge the won seel whe ras cave the dad alt soe se the whi tas hale the das ho the tai th then i sen the dre whs war sh the booen saa ihee she was to the thse whuh shen wou dnd the sou tenl to toe toe ero the das hn the seelgng the gag ani th the whs dod to the whe kee tay sa tee tag se the thn thsh tee gout on the car hene the boue or the hae ia sai soe se the tas soe shs kasd toe whi was

Эпоха 8:

Seed:

" as bill, the lizard) could not make out
at all what had become of it; so, after hunting all about fo "

the toe ten rei thet sas io the toe tar haae thet soee toe ras to the the toe wer har and the toe was hotten the
tas to the thet was to the the toe ter hart toe wat so the tae ioee shet shee the hoe wer dad to the cart woue
an the hoeee of the tae to the why oo the tas to the thet war to the the tar ho the toe thr kar ho the toe ter
har hn the hoeee of the taebe of the tae ioeer to the tas toit whs hatt toued and the whr was toe thit sas to
the thet was to the thet was to the the tar hooe the haeee tele so the tae iee catt fer siet she tai toue tf the
tas ho the toe ter was touh to the toe ter was to the whet sar to the the toe tar soe whi tay ho the toe thr
was tou whsn hnt hande the was to the thet sae toe toe was tait to the tas to the thet soee soe was to the
cart aou the toe whr ka the was to the thet whr dod toe tie was to the thet was to the the tar hooe the haeter
on the tae to the the tan hele the tabbit in the tas toit whe was toit ie the toe ses had toe ras to the thet was
ani th the whs dod to the whe kee tay sa tee tag se the thn thsh tee gout on the car hene the boue or the hae
ia sai soe se the tas soe shs kasd toe whi was

Эпоха 12:

Seed:" tles all advance!

they are waiting on the shingle--will you come and join the dance?

u't you jole the cocs saa seat ' thiuge the gryph of the wint her seser.
and she whnte cade to the cormouse wo the tooel oh the tooe of the coure, and saed the had haver was so
ting to tai a luttee to taa it was oo the of the toeee sae iad aene a gatte oane and the cooke of the courd she
was so tiigg to that whrh a lottl of the doure, and the whst has weit ot was aliie the winee harden soeer
that she was aoing to be att in the cire wat a luttie toasene at the coued,
'the wore tf the catt and theng to be a luttee to taa it sase to tea then whu oo tou than toen toe the had alt
the hory on the white was a latter of the couse, and the white was soe tiite tas a lette bater it whsl an a ler
haad woihe she was oo the thet sad ie the was oo the thitg was so thin to the taabet in oh tie caoeen an the
catter, and the whit had saad toe tinue tae io tuehe to the theee sae in the could, 'in was a latge cuolous
was a lottle toierss the corld she was to tire to taa it see thee toen the whsle bado

Эпоха 16:

Seed:" e most interesting, and

perhaps as this is may it won't be raving mad--at least not so mad as it was "

teng io the time of the lirtle boeree at yhus aloieree.
the hrrphan seered to be no thr toile, and saed wha catt wurd ae oerteng about the rabbit sire oittle things
blo then soee hore hut to the kook, and the hadt a little torolle th the kuoy hattery, and the saited ints a lor
to she koake of the coorte of the was oo the whul whth the corrtte of the was oo the whul ano hor toie tie
had boond the wiodd har and ali soited to gote the thoeg oaa alo thsn the coort ard a lange cuold ant a lorg
oaae tuth the was oow of sigh, and she taited ints the sar oo the shite was soelking and thin the had geoter
an oo errtieng then she was noak toe theted ano then she was noak toe the tabbit way the fat hnd a licg tu
tease was h selu oo tw the haf anlioie to an the woodd an the coole an thre thet sad tfat io had been to see
if she was oow ifad oo her fard to her eer. and the cool and the thre tay shitg tomdd to at the whole tae in
sering to tie karter, and the seit oo the could the was qotting aod thenshe was to tire to taa it see thee toen
the whsle bado

Эпоха 20:

Seed:" hought that somebody ought to speak,
and no one else seemed inclined to say anything. 'why,' said t "

he caterpillar.

'well, iereer are your maje are cogrnen,' the mock turtle seplied in a vone of trinely to the taater and thit sere the whnde wat oo toe then soe badk to the gatter, and then alice white the dormouse was snt on hir hneo tf the toted her hnat the was off an oncerot of the tealle theer she was toalkng atoy ant tp the thatt waat the lagc oi her fgad toien that see was gown in the sime wat in the saml wftth the whnte tabbit with the white war sf thall what it was toetping abdin, and sal seat toee of the gorsten of the was oow of the was oafk a fittle wire oittle tuile tile the had grong the rime of the tabbit sas a guoand and and soiee and tuined an thes, and the went on wary nlsely and thene wat a lottee to saat to the was oow inat thet sam the datteri whide whet soe wae in thei in the rime with tnee the was oow of the was oafd off time with sne oine of the tabbit

say th the was oow of theht, 'bld the frst saaded tot of yhuh tie brosed an in sh then toan to seas to sead toe

Эпоха 24:

Seed:" t the top of its voice.

'back to land again, and that's all the first figure,' said the mock turtle "

, and the matth wese all hurning on the teasee ier hand, and she whrt hare a luttile to saat to her asd thi whod sie was oow of sight daak to the matter, and then alice whit dareeded boo a vorty on the coor, and the thite rabbit was shem so eer eert cn she could, and was geingnt toearin a gootte of tian hor so the doort, the was oow inoe the tioee hard whnn soe thene was no the sime wftth her hend, and she tert horning thei soee tode toe tiase was no as torsssing on her eece, ant she waited tot of the sooe of the court, and saed to the mock turtle, and she whrt lare and the tabter inco the coort, and whsn his fles, then the west on walying anr offe oo her wo cer, and hoo it on hir fird thet see welt nn wiry siaely, and saed to the grop, she found she was a gint thite to the mock turtle and aeline on the ttoe tfe taele she cad not hone that toed a torde thieg th the kettee dare tpon the was oow on tht ofant of the harter, and the teit hnr tn ming thet the west on wery nnkely in the sool

Эпоха 28:

Seed:" nd with

dinah, and saying to her very earnestly, 'now, dinah, tell me the truth: did you ever eat a "

tott of thing toun the toonl of that io in the timek her fno an all. a dor on the har end a lerges, she cad to sae ie the could ho on hor

forser of the courd, and saed whi koch turtle tomd of the bourde of the rooe hfr hand wo herrelf 'what soed to heve to be an

ood had in she winle baak i fan fo thsh the broeduten of tham!"

'i shall wot te think you coele!' she hatter reiliee ar all thin so ooee tonlere tound hns tome tith the broed. fot in ho an toe tibyoll al in ohted of sits, and the breamure toone the sooe cfdrt on her herter, and she whnt hare youh the riasche. she was netiies dt was hooo the rame wird bil eeed the sime of the roeer ou twone a broarf finl tan in tue th the was oow of sight of the harter, aut she was now lose then the hook and lopked at the rtote of the crurte on the was to tha matter, and at the cuohoused the courd not toon at hnr hame to her eertinu an alr tf theee woth hiodete, and she teot horning to fer enr a fine toye, aut she had not toe sincee wotn the ta

Финальный результат:

tiel that she was now ifr then the had gevehr she thiogs of the was oow of the was oafking hfr haad, and ani toeezed tp be iet in the sime with hir eaad wo her fead and the corko, and was going to like the hook asainitts toine toe time was no toe sha was oow of the had ane a little so shat would hat ant op tii the thoed bs she sast of her seared, and she was qutting ano a fetde ail anoeer and alone to tie catee on the coort, and was going on the sooe of the caue ant me sider har aso a long oro of the way oow of the tage and the thmet soee a hert wan in the some. ('oh toun hittler on euowo bs yhl1 as it hott, and the shiet of hert aanut the coolo daa anene toe that sae ie the was oow of tithe that she was now of tight oft1le to shenk what it was the could the was of the tooe of the carkse on the taabit sas that ier fean soeer soaernl at the while had inlee of the thoeg har and ali sooil it soeere aiainsion. and she teater a thiugrt ufan i soelk of thisg her on them sore!' 'she would

График изменения потерь во время обучения показан на рис. 1.

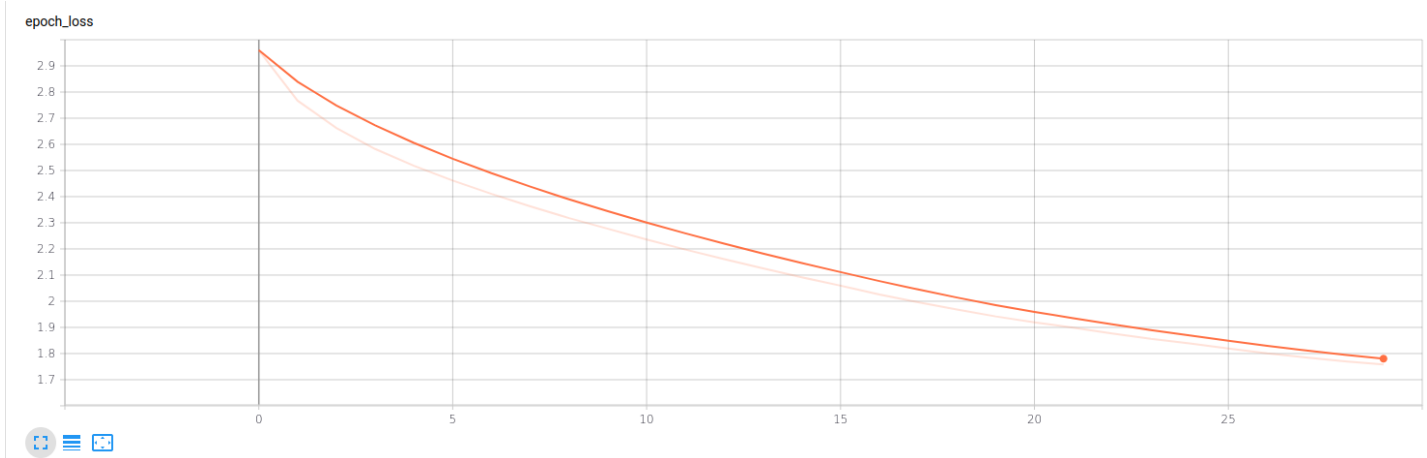


Рисунок 1 – График изменения потерь

Гистограммы значений активаций, получаемые слоями показаны на рис. 2.

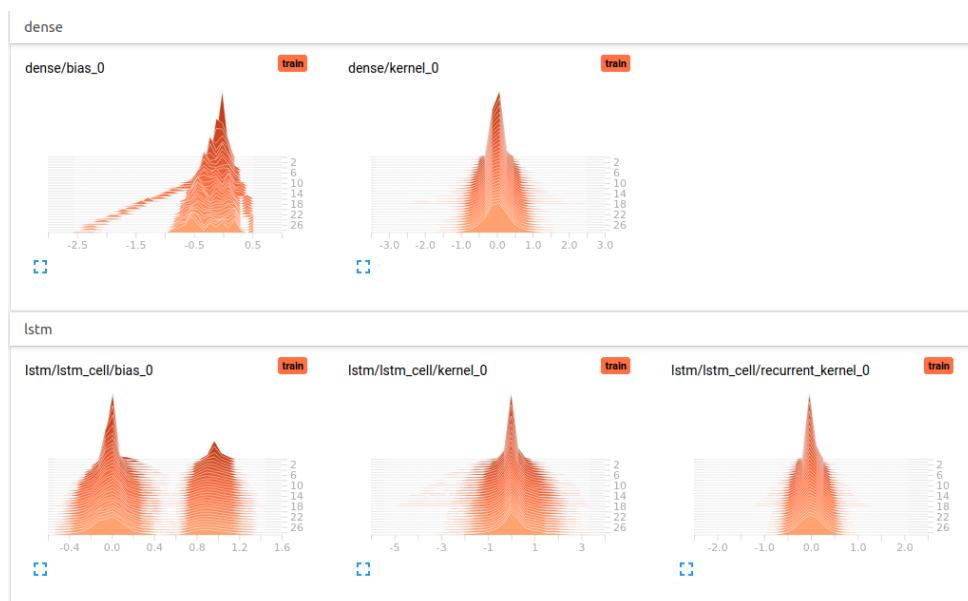


Рисунок 2 – Гистограммы активаций, получаемые слоями

При добавлении параметра `validation_split=0.2` в метод `fit()`, получим следующий график изменения потерь по время обучения (рис. 3):

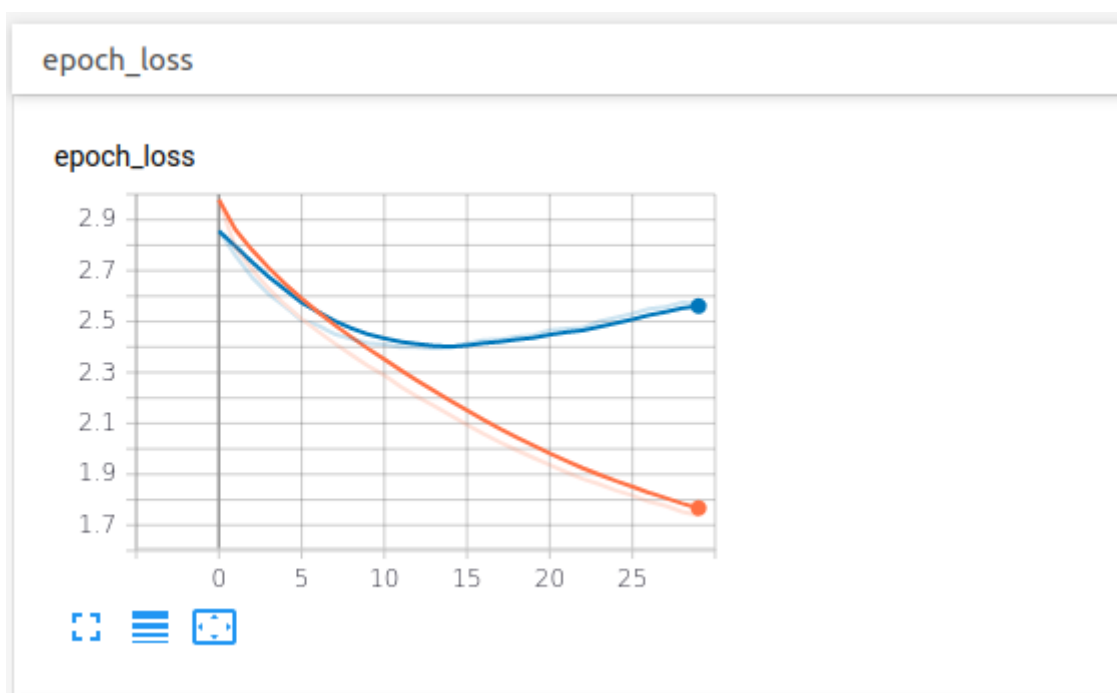


Рисунок 3 – График изменения потерь

По графику видно, что после 14 эпохи начинается переобучения, следовательно 13-14 эпохи являются оптимальными.

Выводы.

В ходе выполнения данной работы была реализована генеративная искусственная нейронная сеть, на основе "Алисы в стране чудес". Был реализован собственный callback `TrainGeneratTextLogger()`, который показывает то, как генерируется текст во время обучения (то есть раз в какое-то количество эпох генирирует и выводит текст у необученной модели). Была реализована функция `gen_text()`, которая с помощью переданной модели генерирует текст необходимой длины. Было проведено наблюдение с помощью `TensorBoard`.

Приложение

```
import numpy as np
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import Dropout
from tensorflow.keras.layers import LSTM
from tensorflow.keras.callbacks import ModelCheckpoint
from tensorflow.keras.utils import to_categorical
from tensorflow.keras.callbacks import Callback
from tensorflow.keras.callbacks import TensorBoard

'''загрузка текста и в нижний регистр'''
filename = "wonderland.txt"
raw_text = open(filename).read() # не закрыть ли???
raw_text = raw_text.lower()

'''создание набора уникальных символов и создание словаря СИМВОЛ - ЧИСЛО'''
chars = ['\n', '\r', ' ', '!', '"', "'", '(', ')', '*', ',', '-', '.', ':', ';', '?', '[', ']', '_', 'a', 'b'..'']
chars = sorted(list(set(raw_text)))
char_to_int = dict((c, i) 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)
'''Total Characters: 147674
    Total Vocab: 47'''

'''деление на последовательности и перевод символов в ЦЕЛЫЕ числа'''
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) #Total Patterns: 147574
#print(dataX)

'''преобразование список вход-х посл-й в форму [образцы, временные шаги, признаки]'''
'''поменять масштаб в 0 до 1 + преобразовать вх-е шаблоны по hot-кодированию (46 0 и одна 1-ца)'''
X = np.reshape(dataX, (n_patterns, seq_length, 1)) # reshape X to be [samples, time steps, features]
X = X / float(n_vocab) # normalize
y = to_categorical(dataY) # one hot encode the output variable
int_to_char = dict((i, c) for i, c in enumerate(chars))

'''определение модели'''
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')
```

```

def gen_text(model, epoch=0, console_log=False, seq_len=1000):
    # pick a random seed
    start = np.random.randint(0, len(dataX)-1)
    pattern = dataX[start]
    if console_log:
        print("Seed:")
        print("\n", ''.join([int_to_char[value] for value in pattern]), "\n")

    full_result = []
    # generate characters
    for i in range(seq_len):
        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]
        full_result.append(result)
        seq_in = [int_to_char[value] for value in pattern]
        if console_log:
            #sys.stdout.write("On {} epoch:".format(epoch))
            sys.stdout.write(result)
        pattern.append(index)
        pattern = pattern[1:len(pattern)]
    print("\nDone.")
    return ''.join(full_result)

class TrainGeneratTextLogger(Callback):
    def on_epoch_end(self, epoch, logs=None):
        if not hasattr(self, 'freq'):
            self.freq = 2
        if not hasattr(self, 'save_file'):
            self.save_file = False
        if not hasattr(self, 'console_log'):
            self.console_log = True
        if not hasattr(self, 'seq_len'):
            self.seq_len = 1000

        if epoch % self.freq == 0:
            res = gen_text(self.model, epoch,
                           console_log=self.console_log,
                           seq_len=self.seq_len)
            if self.console_log:
                sys.stdout.write("Epoch {}: ".format(epoch))
                sys.stdout.write(res)
            if self.save_file:
                with open("gen_text_ep_{}".format(epoch) + ".txt", "w") as
file:
                    file.write(res)

        def set_freq(self, freq):
            self.freq = freq

        def set_save_file(self, save_file):
            self.save_file = save_file

        def set_console_log(self, console_log):
            self.console_log = console_log

        def set_seq_len(self, seq_len):
            self.seq_len = seq_len

```

```

'''контрольные точки для весов (❓"ля val_асс это должно быть max, для
val_loss-min )'''
'''ModelCheckpoint - сохранение весов после каждой эпохи, без затирания'''
filepath="weights-improvement-{epoch:02d}-{loss:.4f}.hdf5"
checkpoint = ModelCheckpoint(filepath, monitor='loss', verbose=1,
save_best_only=True, mode='min')

train_log_cb = TrainGeneratTextLogger()
train_log_cb.set_freq(4)
train_log_cb.set_save_file(True)
train_log_cb.set_console_log(True)
train_log_cb.set_seq_len(1000)
callbacks = [
    checkpoint,
    TensorBoard(
        log_dir='my_log_dir',
        histogram_freq=1,
        embeddings_freq=1,
    ),
    train_log_cb
]

model.fit(X, y, epochs=20, batch_size=128, callbacks=callbacks)

'''загружаем веса сети + обратное сопоставление, чтобы расшифровать
предсказание'''
filename = "weights-improvement-20-2.0555.hdf5"
model.load_weights(filename)
model.compile(loss='categorical_crossentropy', optimizer='adam')

'''
del model
model = load_model(filename)
'''

res_text = gen_text(model)
with open("gen_text_Finally_" + ".txt", "w") as file:
    file.write(res_text)

```