CNN_text_classification_word_embeddings

April 22, 2021

```python
from google.colab import drive
drive.mount('/content/gdrive')
```

Drive already mounted at /content/gdrive; to attempt to forcibly remount, call
drive.mount("/content/gdrive", force_remount=True).

```python
import copy
import nltk
from sklearn import preprocessing
from keras.utils import plot_model
import tensorflow as tf
import pandas as pd
import numpy as np
from tensorflow import keras
from itertools import islice
import os
import regex as re
import nltk
from nltk import word_tokenize, pos_tag, ne_chunk
from nltk.chunk import tree2conlltags
from nltk.tokenize.treebank import TreebankWordDetokenizer
from sklearn.model_selection import train_test_split
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras import Input
from keras.layers import concatenate, Input, Lambda, Dense, Flatten, Conv1D,␣
 ↪Conv3D, Conv2D, Activation, MaxPooling2D, Dropout, MaxPool1D, MaxPool2D,␣
 ↪Embedding
from keras import optimizers
from keras.models import Model
import matplotlib.pyplot as plt
import tensorflow.keras.backend as K
from termcolor import colored
from tensorflow.keras.callbacks import ReduceLROnPlateau
import datetime
import tensorboard
import os
from sklearn.preprocessing import OneHotEncoder
```

```
%load_ext tensorboard
!rm -rf ./logs/
```

The tensorboard extension is already loaded. To reload it, use:
  %reload_ext tensorboard

```
import nltk
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
nltk.download('maxent_ne_chunker')
nltk.download('words')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data…
[nltk_data]    Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]        /root/nltk_data…
[nltk_data]    Package averaged_perceptron_tagger is already up-to-
[nltk_data]          date!
[nltk_data] Downloading package maxent_ne_chunker to
[nltk_data]        /root/nltk_data…
[nltk_data]    Package maxent_ne_chunker is already up-to-date!
[nltk_data] Downloading package words to /root/nltk_data…
[nltk_data]    Package words is already up-to-date!
```

[ ]: True

1 Email

[ ]:

```
def extract_emailid(text):
    email_per_line_list = []
    emails = re.findall("([a-zA-Z0-9_.+-]+@[a-zA-Z0-9-]+\.[a-zA-Z0-9-.]+)",
    ↪text)
    for mail in emails:
        email_per_line_list.append(mail)

    return email_per_line_list
```

```
def get_emails_from_file(file):
    # file = open(os.path.join(path, file))
    # file_lines = file.readlines()
    with open(path+file, encoding="utf8", errors='ignore') as f:
      file_lines = f.readlines()
```

```
    emails_in_file = []
    emails_preprocessed_list = []
    emails_preprocessed_string = ""

    for line in file_lines:
        email_per_line_list = []
        emails = re.findall("([a-zA-Z0-9_.+-]+@[a-zA-Z0-9-]+\.[a-zA-Z0-9-.]+)",␣
↪line)
        for mail in emails:
            emails_in_file.append(mail)



    for i, email in enumerate(emails_in_file):
        email = email.split("@", 1)[1]
        email = email.split('.')
        for c in email:
            if c=="com":
                email.remove(c)
            else:
                emails_preprocessed_list.append(c)

    emails_preprocessed_list = list(set(emails_preprocessed_list))

    for c in emails_preprocessed_list:
        c.lower()
        if c.find('-') != -1:
            c = c.replace("-","")
        if len(c) > 2:
            if len(emails_preprocessed_string) == 0:
                emails_preprocessed_string = emails_preprocessed_string + c
            else:
                emails_preprocessed_string = emails_preprocessed_string + " " +␣
↪c



    return emails_preprocessed_string, emails_in_file, emails_preprocessed_list
```

2 Subject

```
[ ]: bad_chars = [';', ':', '!', "*", "\n", "\t"]

def get_subject_from_file(file):

    with open(path+file, encoding="utf8", errors='ignore') as f:
```

```python
        file_lines = f.readlines()

    # file = open(os.path.join(path, file))
    # file_lines = file.readlines()
    subject_in_file = []
    subject_preprocessed_list = []
    subjects_preprocessed_strings = ""


#     line = re.sub(' Re: ',"", line)
#            line = line.rstrip('\n')
#            line = line.rstrip('\t')
#            line = re.sub(r'[^\w\s]', '', line)
#            line = ''.join(i for i in line if not i in bad_chars)

    for line in file_lines:
        if (line.find('Subject: ')!=-1):
            line = line.split('Subject: ')[1]
            line = line.rstrip('\n')
            line = line.rstrip('\t')
            line = re.sub(r'[^\w\s]', ' ', line)
            line = re.sub(' Re ',"", line)
            line = re.sub("  ", "", line)
            line = ''.join(i for i in line if not i in bad_chars)
            line = line.lstrip()
            subject_in_file.append(line)

    if len(subject_in_file) == 1:
        for i, c in enumerate(subject_in_file):
            if i==0 and c==" ":
                continue
            elif (i==1 and c==" "):
                continue
            else:
                subjects_preprocessed_strings = subjects_preprocessed_strings +␣
→c

    if len(subject_in_file) > 1:
        for item in subject_in_file:
            for i, c in enumerate(item):
                if i==0 and c==" ":
                    continue
                elif (i==1 and c==" "):
                    continue
                else:
                    subjects_preprocessed_strings =␣
→subjects_preprocessed_strings + c
```

```
    subjects_preprocessed_strings = re.sub('  ',"",␣
 ↪subjects_preprocessed_strings)


    return file, subjects_preprocessed_strings
```

3   Text

```
[ ]: bad_chars = [';', ':', '!', "*", "\n", "\t", '\\','/', "-", "@", "#", "$", "%",␣
     ↪"^", "&", '"', '.', ',', ')', '(', '[', ']', '{', '}']



     # Remove words of len<=2 and >=15
     def removeWord_with_len(line):
         wordList = word_tokenize(line)

         for i, word in enumerate(wordList):
             if len(word)<=2:
                 wordList[i] = ""
             if len(word)>=15:
                 wordList[i] = ""
         new_line = TreebankWordDetokenizer().detokenize(wordList)
         return new_line



     # Remove _word, word_, _word_ and keep word2 in word1_word2
     def remove_word_type(line):

         wordlist = word_tokenize(line)

         for i, word in enumerate(wordlist):
             if ((re.search(r"^_\w.*_", word)) or (re.search(r"^_\w.*", word)) or␣
     ↪(re.search(r"^\w.*_", word)) ):
                 wordlist.pop(i)
             elif (re.search(r"^\w.*_\w.*", word)):

                 subword_list = word.split("_")
                 new_word = subword_list[1]
                 index_of_word = wordlist.index(word)
                 wordlist[index_of_word] = new_word
```

```python
    line = TreebankWordDetokenizer().detokenize(wordlist)
    return line




def remove_person_name(sentence):

    personNames = []
    wordsList = word_tokenize(sentence)
    for item in tree2conlltags(ne_chunk(pos_tag(wordsList))):
        if 'PERSON' in item[2]:
            personNames.append(item[0])

    return personNames




def get_text_from_file(file):
    text_in_file = []
    text_preprocessed_list = []
    text_preprocessed_strings = ""

    with open(path+file, encoding="utf8", errors='ignore') as f:
      file_lines = f.readlines()

    # file = open(os.path.join(path, file))
    # file_lines = file.readlines()

    for line in file_lines:
        # regex pattern --> word:
        if re.search(r"^\w.*:", line):
            continue
        # Remove [words]
        pattern = r"(\[(?:\[??[^\[]*?\]))"
        line = re.sub(pattern, "", line)

        # Remove (words)
        pattern = r".*?\((.*?)\)"
        line = re.sub(pattern, "", line)

        # Remove <words>
        pattern = r".*?\<(.*?)\>"
        line = re.sub(pattern, "", line)

        line = ''.join(i for i in line if not i in bad_chars)
```

```python
        line = line.rstrip("\n")
        line = line.rstrip("\t")
        line = line.rstrip("--")
        line = line.lstrip()

        # specific
        line = re.sub(r"won't", "will not", line)
        line = re.sub(r"can\'t", "can not", line)

        # general
        line = re.sub(r"n\'t", " not", line)
        line = re.sub(r"\'re", " are", line)
        line = re.sub(r"\'s", " is", line)
        line = re.sub(r"\'d", " would", line)
        line = re.sub(r"\'ll", " will", line)
        line = re.sub(r"\'t", " not", line)
        line = re.sub(r"\'ve", " have", line)
        line = re.sub(r"\'m", " am", line)

        for c in line:
            line = line.replace('\\', '')
            line = line.replace('/', '')


        # remove person names
        personNames = remove_person_name(line)
        for name in personNames:
            line = line.rstrip(name)

        # remove digits
        line = ''.join([i for i in line if not i.isdigit()])

        # remove words like _word, word_, _word_ and [get word2 from pattern
→word1_word2]
        line = remove_word_type(line)


        # convert to lower case
        line = line.lower()

        # remove words that are <=2 and >=15
        line = removeWord_with_len(line)

        # remove all the words except (a-zA-Z and space)
        line = re.sub(r'[^a-zA-Z ]+', '', line)

        if line.strip() == "":
```

```
            continue

        text_in_file.append(line)

    if len(text_in_file) == 1:
        for i, c in enumerate(text_in_file):
            if i==0 and c==" ":
                continue
            elif (i==1 and c==" "):
                continue
            else:
                text_preprocessed_strings = text_preprocessed_strings + c

    if len(text_in_file) > 1:
        for item in text_in_file:
            for i, c in enumerate(item):
                if i==0 and c==" ":
                    continue
                elif (i==1 and c==" "):
                    continue
                else:
                    text_preprocessed_strings = text_preprocessed_strings + c


    text_preprocessed_strings = re.sub('  ',"", text_preprocessed_strings)

    return text_preprocessed_strings
```

4 Raw Text

```
[ ]: def get_raw_text_from_file(file):
    text_in_file = []
    raw_text_string = ""

    with open(path+file, encoding="utf8", errors='ignore') as f:
      file_lines = f.readlines()

    for line in file_lines:
        text_in_file.append(line)

    if len(text_in_file) == 1:
        for i, c in enumerate(text_in_file):
            raw_text_string = raw_text_string + c

    if len(text_in_file) > 1:
```

```
        for item in text_in_file:
            for i, c in enumerate(item):
                raw_text_string = raw_text_string + c

    return raw_text_string
```

5 Extracting emails, subject and texts from all files

```python
def pre_process_text_file(file_list):
    preprocessed_email_list = []
    preprocessed_subject_list = []
    preprocessed_text_list = []
    raw_text_list = []

    for i, file in enumerate(file_list):

        preprocessed_subject_list.append(get_subject_from_file(file)[1])
        preprocessed_email_list.append(get_emails_from_file(file)[0])
        preprocessed_text_list.append(get_text_from_file(file))
        raw_text_list.append(get_raw_text_from_file(file))

    return preprocessed_subject_list, preprocessed_email_list,␣
 ↪preprocessed_text_list, raw_text_list
```

6 Extracting text from alt.atheism_49960.txt

```python
def pre_process_text_file_specific(file_list):
    preprocessed_email_list = []
    preprocessed_subject_list = []
    preprocessed_text_list = []
    raw_text_list = []
    file_name = []

    for file in file_list:
        if file == "alt.atheism_49960.txt":
            preprocessed_subject_list.append(get_subject_from_file(file)[1])
            preprocessed_email_list.append(get_emails_from_file(file)[0])
            preprocessed_text_list.append(get_text_from_file(file))
            raw_text_list.append(get_raw_text_from_file(file))
```

```
            file_name.append(file)

            break

    return preprocessed_subject_list, preprocessed_email_list,␣
↪preprocessed_text_list, raw_text_list, file_name
```

7   List all files and extract email, subject, text and raw text

```
[ ]: path = '/content/gdrive/My Drive/Colab Notebooks/CNN on Text data/documents/
     ↪documents/'

     file_list = os.listdir(path)

     file_list
```

```
[ ]: ['talk.politics.misc_178714.txt',
      'talk.politics.misc_179113.txt',
      'talk.politics.misc_178837.txt',
      'talk.politics.misc_179067.txt',
      'talk.politics.misc_178648.txt',
      'talk.politics.misc_178977.txt',
      'talk.politics.misc_176850.txt',
      'talk.politics.misc_178788.txt',
      'talk.politics.misc_178431.txt',
      'talk.politics.misc_178663.txt',
      'talk.politics.misc_178770.txt',
      'talk.politics.misc_178842.txt',
      'talk.politics.misc_178643.txt',
      'talk.politics.misc_178545.txt',
      'talk.politics.misc_178870.txt',
      'talk.politics.misc_178348.txt',
      'talk.politics.misc_179081.txt',
      'talk.politics.misc_176904.txt',
      'talk.politics.misc_178329.txt',
      'talk.politics.misc_178626.txt',
      'talk.politics.misc_178629.txt',
      'talk.politics.misc_178673.txt',
      'talk.politics.misc_178849.txt',
      'talk.politics.misc_178672.txt',
      'talk.politics.misc_179025.txt',
      'talk.politics.misc_179102.txt',
      'talk.politics.misc_178772.txt',
      'talk.politics.misc_179048.txt',
      'talk.politics.misc_178333.txt',
```

```
'talk.politics.misc_178435.txt',
'talk.politics.misc_176852.txt',
'talk.politics.misc_176980.txt',
'talk.politics.misc_178411.txt',
'talk.politics.misc_179091.txt',
'talk.politics.misc_178989.txt',
'talk.politics.misc_178876.txt',
'talk.politics.misc_178860.txt',
'talk.politics.misc_176891.txt',
'talk.politics.misc_178543.txt',
'talk.politics.misc_178310.txt',
'talk.politics.misc_178627.txt',
'talk.politics.misc_178895.txt',
'talk.politics.misc_178556.txt',
'talk.politics.misc_176887.txt',
'talk.politics.misc_178781.txt',
'talk.politics.misc_178802.txt',
'talk.politics.misc_178302.txt',
'talk.politics.misc_179088.txt',
'talk.politics.misc_178410.txt',
'talk.politics.misc_178725.txt',
'talk.politics.misc_176931.txt',
'talk.politics.misc_178438.txt',
'talk.politics.misc_178605.txt',
'talk.politics.misc_179022.txt',
'talk.politics.misc_177010.txt',
'talk.politics.misc_176942.txt',
'talk.politics.misc_178855.txt',
'talk.politics.misc_177011.txt',
'talk.politics.misc_178952.txt',
'talk.politics.misc_176903.txt',
'talk.politics.misc_178835.txt',
'talk.politics.misc_178836.txt',
'talk.politics.misc_178510.txt',
'talk.politics.misc_178596.txt',
'talk.politics.misc_178967.txt',
'talk.politics.misc_177007.txt',
'talk.politics.misc_176847.txt',
'talk.politics.misc_178394.txt',
'talk.politics.misc_178866.txt',
'talk.politics.misc_178520.txt',
'talk.politics.misc_178901.txt',
'talk.politics.misc_178555.txt',
'talk.politics.misc_178576.txt',
'talk.politics.misc_178417.txt',
'talk.politics.misc_178982.txt',
'talk.politics.misc_178299.txt',
```

```
'talk.politics.misc_179063.txt',
'talk.politics.misc_178906.txt',
'talk.politics.misc_178664.txt',
'talk.politics.misc_178569.txt',
'talk.politics.misc_178296.txt',
'talk.politics.misc_177022.txt',
'talk.politics.misc_178671.txt',
'talk.politics.misc_178752.txt',
'talk.politics.misc_176902.txt',
'talk.politics.misc_176861.txt',
'talk.politics.misc_178676.txt',
'talk.politics.misc_178767.txt',
'talk.politics.misc_178468.txt',
'talk.politics.misc_178871.txt',
'talk.politics.misc_179006.txt',
'talk.politics.misc_176976.txt',
'talk.politics.misc_178768.txt',
'talk.politics.misc_178448.txt',
'talk.politics.misc_179089.txt',
'talk.politics.misc_176935.txt',
'talk.politics.misc_178407.txt',
'talk.politics.misc_178987.txt',
'talk.politics.misc_178887.txt',
'talk.politics.misc_178903.txt',
'talk.politics.misc_178929.txt',
'talk.politics.misc_178979.txt',
'talk.politics.misc_178891.txt',
'talk.politics.misc_178753.txt',
'talk.politics.misc_178326.txt',
'talk.politics.misc_177008.txt',
'talk.politics.misc_178682.txt',
'talk.politics.misc_179116.txt',
'talk.politics.misc_178868.txt',
'talk.politics.misc_176878.txt',
'talk.politics.misc_176969.txt',
'talk.politics.misc_176906.txt',
'talk.politics.misc_178668.txt',
'talk.politics.misc_178487.txt',
'talk.politics.misc_176936.txt',
'talk.politics.misc_178557.txt',
'talk.politics.misc_178869.txt',
'talk.politics.misc_178380.txt',
'talk.politics.misc_178669.txt',
'talk.politics.misc_178759.txt',
'talk.politics.misc_178486.txt',
'talk.politics.misc_176860.txt',
'talk.politics.misc_178433.txt',
```

```
'talk.politics.misc_178694.txt',
'talk.politics.misc_178401.txt',
'talk.politics.misc_178378.txt',
'talk.politics.misc_178732.txt',
'talk.politics.misc_176897.txt',
'talk.politics.misc_178784.txt',
'talk.politics.misc_176890.txt',
'talk.politics.misc_178392.txt',
'talk.politics.misc_178763.txt',
'talk.politics.misc_176959.txt',
'talk.politics.misc_178928.txt',
'talk.politics.misc_177020.txt',
'talk.politics.misc_179008.txt',
'talk.politics.misc_178986.txt',
'talk.politics.misc_176885.txt',
'talk.politics.misc_178446.txt',
'talk.politics.misc_178355.txt',
'talk.politics.misc_178840.txt',
'talk.politics.misc_179007.txt',
'talk.politics.misc_178956.txt',
'talk.politics.misc_178735.txt',
'talk.politics.misc_178973.txt',
'talk.politics.misc_178563.txt',
'talk.politics.misc_178716.txt',
'talk.politics.misc_179015.txt',
'talk.politics.misc_178665.txt',
'talk.politics.misc_178872.txt',
'talk.politics.misc_176958.txt',
'talk.politics.misc_178483.txt',
'talk.politics.misc_178531.txt',
'talk.politics.misc_176929.txt',
'talk.politics.misc_178619.txt',
'talk.politics.misc_178300.txt',
'talk.politics.misc_178595.txt',
'talk.politics.misc_178526.txt',
'talk.politics.misc_178419.txt',
'talk.politics.misc_178935.txt',
'talk.politics.misc_178308.txt',
'talk.politics.misc_178608.txt',
'talk.politics.misc_178353.txt',
'talk.politics.misc_178790.txt',
'talk.politics.misc_178970.txt',
'talk.politics.misc_179046.txt',
'talk.politics.misc_178771.txt',
'talk.politics.misc_179075.txt',
'talk.politics.misc_178640.txt',
'talk.politics.misc_176970.txt',
```

```
'talk.politics.misc_178983.txt',
'talk.politics.misc_178447.txt',
'talk.politics.misc_178533.txt',
'talk.politics.misc_178942.txt',
'talk.politics.misc_177016.txt',
'talk.politics.misc_178882.txt',
'talk.politics.misc_178624.txt',
'talk.politics.misc_178375.txt',
'talk.politics.misc_176962.txt',
'talk.politics.misc_178383.txt',
'talk.politics.misc_176952.txt',
'talk.politics.misc_178453.txt',
'talk.politics.misc_178518.txt',
'talk.politics.misc_178884.txt',
'talk.politics.misc_176896.txt',
'talk.politics.misc_178910.txt',
'talk.politics.misc_178832.txt',
'talk.politics.misc_176868.txt',
'talk.politics.misc_178776.txt',
'talk.politics.misc_178420.txt',
'talk.politics.misc_178361.txt',
'talk.politics.misc_176864.txt',
'talk.politics.misc_176922.txt',
'talk.politics.misc_176876.txt',
'talk.politics.misc_176926.txt',
'talk.politics.misc_178723.txt',
'talk.politics.misc_179035.txt',
'talk.politics.misc_178731.txt',
'talk.politics.misc_178541.txt',
'talk.politics.misc_178654.txt',
'talk.politics.misc_178583.txt',
'talk.politics.misc_177024.txt',
'talk.politics.misc_179098.txt',
'talk.politics.misc_178606.txt',
'talk.politics.misc_178854.txt',
'talk.politics.misc_176944.txt',
'talk.politics.misc_178602.txt',
'talk.politics.misc_178430.txt',
'talk.politics.misc_178319.txt',
'talk.politics.misc_178502.txt',
'talk.politics.misc_178564.txt',
'talk.politics.misc_178641.txt',
'talk.politics.misc_178722.txt',
'talk.politics.misc_178425.txt',
'talk.politics.misc_178416.txt',
'talk.politics.misc_179000.txt',
'talk.politics.misc_178707.txt',
```

```
'talk.politics.misc_178342.txt',
'talk.politics.misc_178844.txt',
'talk.politics.misc_178712.txt',
'talk.politics.misc_176898.txt',
'talk.politics.misc_178631.txt',
'talk.politics.misc_178890.txt',
'talk.politics.misc_178764.txt',
'talk.politics.misc_179037.txt',
'talk.politics.misc_178314.txt',
'talk.politics.misc_178742.txt',
'talk.politics.misc_179001.txt',
'talk.politics.misc_178649.txt',
'talk.politics.misc_176856.txt',
'talk.politics.misc_178907.txt',
'talk.politics.misc_176996.txt',
'talk.politics.misc_178458.txt',
'talk.politics.misc_176859.txt',
'talk.politics.misc_176930.txt',
'talk.politics.misc_179020.txt',
'talk.politics.misc_176947.txt',
'talk.politics.misc_178693.txt',
'talk.politics.misc_176862.txt',
'talk.politics.misc_178560.txt',
'talk.politics.misc_178653.txt',
'talk.politics.misc_176858.txt',
'talk.politics.misc_178316.txt',
'talk.politics.misc_178374.txt',
'talk.politics.misc_176895.txt',
'talk.politics.misc_176989.txt',
'talk.politics.misc_178702.txt',
'talk.politics.misc_178325.txt',
'talk.politics.misc_178939.txt',
'talk.politics.misc_176886.txt',
'talk.politics.misc_176909.txt',
'talk.politics.misc_178897.txt',
'talk.politics.misc_177021.txt',
'talk.politics.misc_178867.txt',
'talk.politics.misc_176849.txt',
'talk.politics.misc_178994.txt',
'talk.politics.misc_176981.txt',
'talk.politics.misc_178584.txt',
'talk.politics.misc_178794.txt',
'talk.politics.misc_178508.txt',
'talk.politics.misc_176923.txt',
'talk.politics.misc_178358.txt',
'talk.politics.misc_178498.txt',
'talk.politics.misc_178346.txt',
```

```
'talk.politics.misc_176917.txt',
'talk.politics.misc_178588.txt',
'talk.politics.misc_178724.txt',
'talk.politics.misc_178636.txt',
'talk.politics.misc_178912.txt',
'talk.politics.misc_178382.txt',
'talk.politics.misc_178542.txt',
'talk.politics.misc_176973.txt',
'talk.politics.misc_178611.txt',
'talk.politics.misc_178370.txt',
'talk.politics.misc_178681.txt',
'talk.politics.misc_178638.txt',
'talk.politics.misc_178445.txt',
'talk.politics.misc_178409.txt',
'talk.politics.misc_178423.txt',
'talk.politics.misc_178490.txt',
'talk.politics.misc_178941.txt',
'talk.politics.misc_176993.txt',
'talk.politics.misc_177018.txt',
'talk.politics.misc_178857.txt',
'talk.politics.misc_178343.txt',
'talk.politics.misc_176948.txt',
'talk.politics.misc_179051.txt',
'talk.politics.misc_176979.txt',
'talk.politics.misc_178743.txt',
'talk.politics.misc_178511.txt',
'talk.politics.misc_179033.txt',
'talk.politics.misc_178992.txt',
'talk.politics.misc_178452.txt',
'talk.politics.misc_176870.txt',
'talk.politics.misc_178457.txt',
'talk.politics.misc_178387.txt',
'talk.politics.misc_178485.txt',
'talk.politics.misc_179009.txt',
'talk.politics.misc_178940.txt',
'talk.politics.misc_176889.txt',
'talk.politics.misc_177012.txt',
'talk.politics.misc_176877.txt',
'talk.politics.misc_178838.txt',
'talk.politics.misc_178718.txt',
'talk.politics.misc_178313.txt',
'talk.politics.misc_176913.txt',
'talk.politics.misc_176865.txt',
'talk.politics.misc_178848.txt',
'talk.religion.misc_83669.txt',
'talk.politics.misc_178981.txt',
'talk.politics.misc_178499.txt',
```

```
'talk.politics.misc_178880.txt',
'talk.politics.misc_179060.txt',
'talk.politics.misc_178708.txt',
'talk.politics.misc_178709.txt',
'talk.politics.misc_178544.txt',
'talk.politics.misc_178749.txt',
'talk.politics.misc_176964.txt',
'talk.politics.misc_178883.txt',
'talk.politics.misc_179047.txt',
'talk.politics.misc_178515.txt',
'talk.politics.misc_176946.txt',
'talk.politics.misc_178859.txt',
'talk.politics.misc_179031.txt',
'talk.politics.misc_178893.txt',
'talk.politics.misc_178697.txt',
'talk.politics.misc_176938.txt',
'talk.politics.misc_178311.txt',
'talk.politics.misc_178943.txt',
'talk.politics.misc_178466.txt',
'talk.politics.misc_179112.txt',
'talk.politics.misc_178628.txt',
'talk.politics.misc_178926.txt',
'talk.politics.misc_178578.txt',
'talk.politics.misc_179111.txt',
'talk.politics.misc_176921.txt',
'talk.politics.misc_178703.txt',
'talk.politics.misc_178932.txt',
'talk.politics.misc_178389.txt',
'talk.religion.misc_84288.txt',
'talk.politics.misc_178911.txt',
'talk.politics.misc_178962.txt',
'talk.politics.misc_176888.txt',
'talk.politics.misc_178782.txt',
'talk.politics.misc_178908.txt',
'talk.politics.misc_178309.txt',
'talk.politics.misc_178597.txt',
'talk.politics.misc_178408.txt',
'talk.politics.misc_178393.txt',
'talk.politics.misc_178514.txt',
'talk.politics.misc_179099.txt',
'talk.politics.misc_178757.txt',
'talk.politics.misc_178523.txt',
'talk.politics.misc_179071.txt',
'talk.politics.misc_177004.txt',
'talk.politics.misc_178930.txt',
'talk.politics.misc_176971.txt',
'talk.politics.misc_178761.txt',
```

```
'talk.politics.misc_176995.txt',
'talk.politics.misc_179027.txt',
'talk.politics.misc_178946.txt',
'talk.politics.misc_178696.txt',
'talk.politics.misc_176880.txt',
'talk.politics.misc_178902.txt',
'talk.politics.misc_178760.txt',
'talk.politics.misc_176965.txt',
'talk.politics.misc_178550.txt',
'talk.politics.misc_178372.txt',
'talk.politics.misc_178475.txt',
'talk.politics.misc_179095.txt',
'talk.politics.misc_179003.txt',
'talk.politics.misc_178958.txt',
'talk.politics.misc_178336.txt',
'talk.politics.misc_178795.txt',
'talk.religion.misc_84276.txt',
'talk.religion.misc_84564.txt',
'talk.religion.misc_84153.txt',
'talk.religion.misc_83972.txt',
'talk.religion.misc_84286.txt',
'talk.religion.misc_84140.txt',
'talk.religion.misc_83823.txt',
'talk.religion.misc_83807.txt',
'talk.religion.misc_83978.txt',
'talk.religion.misc_83478.txt',
'talk.religion.misc_83622.txt',
'talk.religion.misc_83469.txt',
'talk.religion.misc_83483.txt',
'talk.religion.misc_84018.txt',
'talk.religion.misc_84346.txt',
'talk.religion.misc_84084.txt',
'talk.religion.misc_84294.txt',
'talk.religion.misc_84182.txt',
'talk.religion.misc_83471.txt',
'talk.religion.misc_83915.txt',
'talk.religion.misc_84445.txt',
'talk.religion.misc_82757.txt',
'talk.religion.misc_83472.txt',
'talk.religion.misc_84006.txt',
'talk.religion.misc_82766.txt',
'talk.religion.misc_83497.txt',
'talk.religion.misc_84282.txt',
'talk.religion.misc_84248.txt',
'talk.religion.misc_83907.txt',
'talk.religion.misc_83817.txt',
'talk.religion.misc_83621.txt',
```

```
'talk.religion.misc_83856.txt',
'talk.religion.misc_83728.txt',
'talk.religion.misc_83688.txt',
'talk.religion.misc_84148.txt',
'talk.religion.misc_84135.txt',
'talk.religion.misc_84195.txt',
'talk.religion.misc_82818.txt',
'talk.religion.misc_83449.txt',
'talk.religion.misc_83593.txt',
'talk.religion.misc_84033.txt',
'talk.religion.misc_84350.txt',
'talk.religion.misc_83738.txt',
'talk.religion.misc_83812.txt',
'talk.religion.misc_83662.txt',
'talk.religion.misc_83674.txt',
'talk.religion.misc_82799.txt',
'talk.religion.misc_82796.txt',
'talk.religion.misc_83801.txt',
'talk.religion.misc_83986.txt',
'talk.religion.misc_84304.txt',
'talk.religion.misc_84207.txt',
'talk.religion.misc_84329.txt',
'talk.religion.misc_84212.txt',
'talk.religion.misc_84007.txt',
'talk.religion.misc_83985.txt',
'talk.religion.misc_84139.txt',
'talk.religion.misc_84023.txt',
'talk.religion.misc_83441.txt',
'talk.religion.misc_82774.txt',
'talk.religion.misc_83968.txt',
'talk.religion.misc_82819.txt',
'talk.religion.misc_83983.txt',
'talk.religion.misc_83725.txt',
'talk.religion.misc_83594.txt',
'talk.religion.misc_82813.txt',
'talk.religion.misc_84348.txt',
'talk.religion.misc_84134.txt',
'talk.religion.misc_82801.txt',
'talk.religion.misc_82797.txt',
'talk.religion.misc_84065.txt',
'talk.religion.misc_84144.txt',
'talk.religion.misc_83437.txt',
'talk.religion.misc_83467.txt',
'talk.religion.misc_84312.txt',
'talk.religion.misc_84341.txt',
'talk.religion.misc_84062.txt',
'talk.religion.misc_83672.txt',
```

```
'talk.religion.misc_83689.txt',
'talk.religion.misc_84339.txt',
'talk.religion.misc_84185.txt',
'talk.religion.misc_84218.txt',
'talk.religion.misc_83855.txt',
'talk.religion.misc_84096.txt',
'talk.religion.misc_84215.txt',
'talk.religion.misc_83708.txt',
'talk.religion.misc_83673.txt',
'talk.religion.misc_84236.txt',
'talk.religion.misc_84308.txt',
'talk.religion.misc_84054.txt',
'talk.religion.misc_83835.txt',
'talk.religion.misc_84249.txt',
'talk.religion.misc_84170.txt',
'talk.religion.misc_84272.txt',
'talk.religion.misc_84310.txt',
'talk.religion.misc_84245.txt',
'talk.religion.misc_83513.txt',
'talk.religion.misc_82800.txt',
'talk.religion.misc_83526.txt',
'talk.religion.misc_84057.txt',
'talk.religion.misc_83453.txt',
'talk.religion.misc_84232.txt',
'talk.religion.misc_84209.txt',
'talk.religion.misc_84437.txt',
'talk.religion.misc_83910.txt',
'talk.religion.misc_84291.txt',
'talk.religion.misc_83897.txt',
'talk.religion.misc_84133.txt',
'talk.religion.misc_83479.txt',
'talk.religion.misc_83781.txt',
'talk.religion.misc_83451.txt',
'talk.religion.misc_83506.txt',
'talk.religion.misc_83979.txt',
'talk.religion.misc_83525.txt',
'talk.religion.misc_83668.txt',
'talk.religion.misc_83504.txt',
'talk.religion.misc_84334.txt',
'talk.religion.misc_84358.txt',
'talk.religion.misc_83498.txt',
'talk.religion.misc_83852.txt',
'talk.religion.misc_84538.txt',
'talk.religion.misc_82759.txt',
'talk.religion.misc_83439.txt',
'talk.religion.misc_83974.txt',
'talk.religion.misc_83445.txt',
```

```
'talk.religion.misc_84557.txt',
'talk.religion.misc_84446.txt',
'talk.religion.misc_84289.txt',
'talk.religion.misc_84120.txt',
'talk.religion.misc_83592.txt',
'talk.religion.misc_84240.txt',
'talk.religion.misc_82763.txt',
'talk.religion.misc_83496.txt',
'talk.religion.misc_84048.txt',
'talk.religion.misc_83900.txt',
'talk.religion.misc_84241.txt',
'talk.religion.misc_83488.txt',
'talk.religion.misc_84428.txt',
'talk.religion.misc_84199.txt',
'talk.religion.misc_83845.txt',
'talk.religion.misc_84174.txt',
'talk.religion.misc_84184.txt',
'talk.religion.misc_83442.txt',
'talk.religion.misc_84229.txt',
'talk.religion.misc_84295.txt',
'talk.religion.misc_84251.txt',
'talk.religion.misc_84011.txt',
'talk.religion.misc_84290.txt',
'talk.religion.misc_84401.txt',
'talk.religion.misc_84146.txt',
'talk.religion.misc_84128.txt',
'talk.religion.misc_83586.txt',
'talk.religion.misc_83816.txt',
'talk.religion.misc_84328.txt',
'talk.religion.misc_84204.txt',
'talk.religion.misc_82785.txt',
'talk.religion.misc_83789.txt',
'talk.religion.misc_84278.txt',
'talk.religion.misc_82778.txt',
'talk.religion.misc_82795.txt',
'talk.religion.misc_83454.txt',
'talk.religion.misc_83876.txt',
'talk.religion.misc_83487.txt',
'talk.religion.misc_83994.txt',
'talk.religion.misc_84353.txt',
'talk.religion.misc_84156.txt',
'talk.religion.misc_83491.txt',
'talk.religion.misc_84508.txt',
'talk.religion.misc_83799.txt',
'talk.religion.misc_84444.txt',
'talk.religion.misc_84298.txt',
'talk.religion.misc_82788.txt',
```

```
'talk.religion.misc_83967.txt',
'talk.religion.misc_84274.txt',
'talk.religion.misc_82777.txt',
'talk.religion.misc_83971.txt',
'talk.religion.misc_84558.txt',
'talk.religion.misc_83843.txt',
'talk.religion.misc_84252.txt',
'talk.religion.misc_84222.txt',
'talk.religion.misc_83671.txt',
'talk.religion.misc_84265.txt',
'talk.religion.misc_84554.txt',
'talk.religion.misc_84143.txt',
'talk.religion.misc_84297.txt',
'talk.religion.misc_84193.txt',
'talk.religion.misc_84301.txt',
'talk.religion.misc_84203.txt',
'talk.religion.misc_84263.txt',
'talk.religion.misc_84147.txt',
'talk.religion.misc_83818.txt',
'talk.religion.misc_84105.txt',
'talk.religion.misc_83828.txt',
'talk.religion.misc_83495.txt',
'talk.religion.misc_83786.txt',
'talk.religion.misc_84101.txt',
'talk.religion.misc_84553.txt',
'talk.religion.misc_84275.txt',
'talk.religion.misc_83500.txt',
'talk.religion.misc_84149.txt',
'talk.religion.misc_84552.txt',
'talk.religion.misc_83744.txt',
'talk.religion.misc_83932.txt',
'talk.religion.misc_83501.txt',
'talk.religion.misc_83867.txt',
'talk.religion.misc_84004.txt',
'talk.religion.misc_84022.txt',
'talk.religion.misc_83520.txt',
'talk.religion.misc_84061.txt',
'talk.religion.misc_83514.txt',
'talk.religion.misc_84281.txt',
'talk.religion.misc_84257.txt',
'talk.religion.misc_84562.txt',
'talk.religion.misc_83788.txt',
'talk.religion.misc_84239.txt',
'talk.religion.misc_84198.txt',
'talk.religion.misc_83808.txt',
'talk.religion.misc_84091.txt',
'talk.religion.misc_84422.txt',
```

```
'talk.religion.misc_83572.txt',
'talk.religion.misc_84438.txt',
'talk.religion.misc_84079.txt',
'talk.religion.misc_83508.txt',
'talk.religion.misc_84507.txt',
'talk.religion.misc_82782.txt',
'talk.religion.misc_83510.txt',
'talk.religion.misc_83535.txt',
'talk.religion.misc_83846.txt',
'talk.religion.misc_83860.txt',
'talk.religion.misc_84395.txt',
'talk.religion.misc_83998.txt',
'talk.religion.misc_83934.txt',
'talk.religion.misc_84436.txt',
'talk.religion.misc_83977.txt',
'talk.religion.misc_84165.txt',
'talk.religion.misc_84043.txt',
'talk.religion.misc_83455.txt',
'talk.religion.misc_84098.txt',
'talk.religion.misc_82772.txt',
'talk.religion.misc_83485.txt',
'talk.religion.misc_82812.txt',
'talk.religion.misc_82807.txt',
'talk.religion.misc_84083.txt',
'talk.religion.misc_83790.txt',
'talk.religion.misc_84442.txt',
'talk.religion.misc_83957.txt',
'talk.religion.misc_82793.txt',
'talk.religion.misc_84261.txt',
'talk.religion.misc_83899.txt',
'talk.religion.misc_82775.txt',
'talk.religion.misc_83627.txt',
'talk.religion.misc_84223.txt',
'talk.religion.misc_84569.txt',
'talk.religion.misc_84344.txt',
'talk.religion.misc_83581.txt',
'talk.religion.misc_83624.txt',
'talk.religion.misc_84127.txt',
'talk.religion.misc_84210.txt',
'talk.religion.misc_83524.txt',
'talk.religion.misc_84219.txt',
'talk.religion.misc_84064.txt',
'talk.religion.misc_83519.txt',
'talk.religion.misc_83798.txt',
'talk.religion.misc_84077.txt',
'talk.religion.misc_84075.txt',
'talk.religion.misc_84279.txt',
```

```
'talk.religion.misc_83913.txt',
'talk.religion.misc_84292.txt',
'talk.religion.misc_84217.txt',
'talk.religion.misc_83705.txt',
'talk.religion.misc_84202.txt',
'talk.religion.misc_83842.txt',
'talk.religion.misc_83444.txt',
'talk.religion.misc_83505.txt',
'talk.religion.misc_84087.txt',
'talk.religion.misc_84142.txt',
'talk.religion.misc_84244.txt',
'talk.religion.misc_84506.txt',
'talk.religion.misc_84243.txt',
'talk.religion.misc_83830.txt',
'talk.religion.misc_84078.txt',
'talk.religion.misc_83680.txt',
'talk.religion.misc_83811.txt',
'talk.religion.misc_84277.txt',
'talk.religion.misc_83944.txt',
'talk.religion.misc_84254.txt',
'talk.religion.misc_84267.txt',
'talk.religion.misc_84269.txt',
'talk.religion.misc_84060.txt',
'talk.religion.misc_84280.txt',
'talk.religion.misc_84285.txt',
'talk.religion.misc_84213.txt',
'talk.religion.misc_84338.txt',
'talk.religion.misc_84021.txt',
'talk.religion.misc_84399.txt',
'talk.religion.misc_83805.txt',
'talk.religion.misc_83561.txt',
'talk.religion.misc_84396.txt',
'talk.religion.misc_84264.txt',
'talk.religion.misc_84345.txt',
'talk.religion.misc_84080.txt',
'talk.religion.misc_84226.txt',
'talk.religion.misc_84268.txt',
'talk.religion.misc_84189.txt',
'talk.religion.misc_83509.txt',
'talk.religion.misc_83844.txt',
'talk.religion.misc_83670.txt',
'talk.religion.misc_84397.txt',
'talk.religion.misc_83729.txt',
'talk.religion.misc_83568.txt',
'talk.religion.misc_84380.txt',
'talk.religion.misc_84273.txt',
'talk.religion.misc_84343.txt',
```

```
'talk.religion.misc_84072.txt',
'talk.religion.misc_83885.txt',
'talk.religion.misc_82815.txt',
'talk.religion.misc_83841.txt',
'talk.religion.misc_84178.txt',
'talk.religion.misc_84086.txt',
'talk.religion.misc_84164.txt',
'talk.religion.misc_83922.txt',
'talk.religion.misc_84429.txt',
'talk.religion.misc_84042.txt',
'talk.religion.misc_83706.txt',
'talk.religion.misc_83880.txt',
'talk.religion.misc_83937.txt',
'talk.religion.misc_83447.txt',
'talk.religion.misc_83661.txt',
'talk.religion.misc_84227.txt',
'talk.religion.misc_84319.txt',
'talk.religion.misc_84235.txt',
'talk.religion.misc_84145.txt',
'talk.religion.misc_83529.txt',
'talk.religion.misc_84020.txt',
'talk.religion.misc_83936.txt',
'talk.religion.misc_84009.txt',
'talk.religion.misc_84019.txt',
'talk.religion.misc_84430.txt',
'talk.religion.misc_83690.txt',
'talk.religion.misc_84224.txt',
'talk.religion.misc_83486.txt',
'talk.religion.misc_84056.txt',
'talk.religion.misc_84313.txt',
'talk.religion.misc_83992.txt',
'talk.religion.misc_84284.txt',
'talk.religion.misc_84138.txt',
'talk.religion.misc_83931.txt',
'talk.religion.misc_83468.txt',
'talk.religion.misc_83527.txt',
'talk.religion.misc_84121.txt',
'talk.religion.misc_84187.txt',
'talk.religion.misc_82776.txt',
'talk.religion.misc_84201.txt',
'talk.religion.misc_84250.txt',
'talk.religion.misc_84307.txt',
'talk.religion.misc_84230.txt',
'talk.religion.misc_84305.txt',
'talk.religion.misc_83544.txt',
'talk.religion.misc_82798.txt',
'talk.religion.misc_84247.txt',
```

```
'talk.religion.misc_84129.txt',
'talk.religion.misc_84159.txt',
'talk.religion.misc_84306.txt',
'talk.religion.misc_84565.txt',
'talk.religion.misc_84082.txt',
'talk.religion.misc_82802.txt',
'talk.religion.misc_84270.txt',
'talk.religion.misc_84150.txt',
'talk.religion.misc_83882.txt',
'talk.religion.misc_84063.txt',
'talk.religion.misc_83849.txt',
'talk.religion.misc_84205.txt',
'talk.religion.misc_83745.txt',
'talk.religion.misc_83659.txt',
'talk.religion.misc_84302.txt',
'talk.religion.misc_84266.txt',
'talk.religion.misc_83943.txt',
'talk.religion.misc_83895.txt',
'talk.religion.misc_84398.txt',
'talk.religion.misc_83456.txt',
'talk.religion.misc_84296.txt',
'talk.religion.misc_84242.txt',
'talk.religion.misc_83892.txt',
'talk.religion.misc_82787.txt',
'talk.religion.misc_83658.txt',
'talk.religion.misc_84106.txt',
'talk.religion.misc_83736.txt',
'talk.religion.misc_84124.txt',
'talk.religion.misc_84220.txt',
'talk.religion.misc_83528.txt',
'talk.religion.misc_84197.txt',
'talk.religion.misc_83512.txt',
'talk.religion.misc_84559.txt',
'talk.religion.misc_83523.txt',
'talk.religion.misc_83476.txt',
'talk.religion.misc_84260.txt',
'talk.religion.misc_84081.txt',
'talk.religion.misc_84208.txt',
'talk.religion.misc_84423.txt',
'talk.religion.misc_83457.txt',
'talk.religion.misc_83461.txt',
'talk.religion.misc_83450.txt',
'talk.religion.misc_84190.txt',
'talk.religion.misc_83701.txt',
'talk.religion.misc_84076.txt',
'talk.religion.misc_84067.txt',
'talk.religion.misc_83976.txt',
```

```
'talk.religion.misc_84433.txt',
'talk.religion.misc_83741.txt',
'talk.religion.misc_83803.txt',
'talk.religion.misc_84136.txt',
'talk.religion.misc_84448.txt',
'talk.religion.misc_83918.txt',
'talk.religion.misc_84300.txt',
'talk.religion.misc_83651.txt',
'talk.religion.misc_84015.txt',
'talk.religion.misc_83601.txt',
'talk.religion.misc_83975.txt',
'talk.religion.misc_82806.txt',
'talk.religion.misc_83714.txt',
'talk.religion.misc_84052.txt',
'talk.religion.misc_84435.txt',
'talk.religion.misc_84192.txt',
'talk.religion.misc_83602.txt',
'talk.religion.misc_84347.txt',
'talk.religion.misc_84439.txt',
'talk.religion.misc_84258.txt',
'talk.religion.misc_82794.txt',
'talk.religion.misc_84431.txt',
'talk.religion.misc_84238.txt',
'talk.religion.misc_83582.txt',
'talk.religion.misc_83494.txt',
'talk.religion.misc_84175.txt',
'talk.religion.misc_84055.txt',
'talk.religion.misc_83608.txt',
'talk.religion.misc_82786.txt',
'talk.religion.misc_83481.txt',
'talk.religion.misc_84510.txt',
'talk.religion.misc_84211.txt',
'talk.religion.misc_84158.txt',
'talk.religion.misc_82781.txt',
'talk.religion.misc_84151.txt',
'talk.religion.misc_83562.txt',
'talk.religion.misc_84315.txt',
'talk.religion.misc_83492.txt',
'talk.religion.misc_84303.txt',
'talk.religion.misc_82816.txt',
'talk.religion.misc_84336.txt',
'talk.religion.misc_84196.txt',
'talk.religion.misc_83610.txt',
'talk.religion.misc_83629.txt',
'talk.religion.misc_84008.txt',
'talk.religion.misc_82783.txt',
'talk.religion.misc_84332.txt',
```

```
'talk.religion.misc_83477.txt',
'talk.religion.misc_83919.txt',
'talk.religion.misc_82767.txt',
'talk.religion.misc_84194.txt',
'talk.religion.misc_83547.txt',
'talk.religion.misc_83800.txt',
'talk.religion.misc_84169.txt',
'talk.religion.misc_84100.txt',
'talk.religion.misc_83511.txt',
'talk.religion.misc_84200.txt',
'talk.religion.misc_83473.txt',
'talk.religion.misc_84357.txt',
'talk.religion.misc_83620.txt',
'talk.religion.misc_84321.txt',
'talk.religion.misc_84231.txt',
'talk.religion.misc_82804.txt',
'talk.religion.misc_83911.txt',
'talk.religion.misc_83795.txt',
'talk.religion.misc_84099.txt',
'talk.religion.misc_84441.txt',
'talk.religion.misc_84188.txt',
'talk.religion.misc_83438.txt',
'talk.religion.misc_83587.txt',
'talk.religion.misc_84183.txt',
'talk.religion.misc_84059.txt',
'talk.religion.misc_83623.txt',
'talk.religion.misc_84331.txt',
'talk.religion.misc_83929.txt',
'talk.religion.misc_83482.txt',
'talk.religion.misc_82784.txt',
'talk.religion.misc_84234.txt',
'talk.religion.misc_83686.txt',
'talk.religion.misc_84066.txt',
'talk.religion.misc_84237.txt',
'talk.religion.misc_83650.txt',
'talk.religion.misc_84309.txt',
'talk.religion.misc_84255.txt',
'talk.religion.misc_82810.txt',
'talk.religion.misc_84259.txt',
'talk.religion.misc_83980.txt',
'talk.religion.misc_82792.txt',
'talk.religion.misc_84349.txt',
'talk.religion.misc_83902.txt',
'talk.religion.misc_84228.txt',
'talk.religion.misc_83999.txt',
'talk.religion.misc_84221.txt',
'talk.religion.misc_83905.txt',
```

```
'talk.religion.misc_84326.txt',
'talk.religion.misc_83829.txt',
'talk.religion.misc_84452.txt',
'talk.religion.misc_84414.txt',
'talk.religion.misc_84253.txt',
'talk.religion.misc_83516.txt',
'talk.religion.misc_84123.txt',
'talk.religion.misc_83503.txt',
'talk.religion.misc_84570.txt',
'talk.religion.misc_84069.txt',
'talk.religion.misc_84354.txt',
'talk.religion.misc_83995.txt',
'talk.religion.misc_84157.txt',
'talk.religion.misc_84447.txt',
'talk.religion.misc_83727.txt',
'talk.religion.misc_84555.txt',
'talk.religion.misc_84186.txt',
'talk.religion.misc_83687.txt',
'talk.religion.misc_83499.txt',
'talk.religion.misc_84316.txt',
'talk.religion.misc_83713.txt',
'talk.religion.misc_82760.txt',
'talk.religion.misc_84311.txt',
'talk.religion.misc_84434.txt',
'talk.religion.misc_84450.txt',
'talk.religion.misc_84352.txt',
'talk.religion.misc_84071.txt',
'talk.religion.misc_83630.txt',
'talk.religion.misc_83777.txt',
'talk.religion.misc_83719.txt',
'talk.religion.misc_84342.txt',
'talk.religion.misc_84025.txt',
'talk.religion.misc_84132.txt',
'talk.religion.misc_84074.txt',
'talk.religion.misc_83776.txt',
'talk.religion.misc_83711.txt',
'talk.religion.misc_83691.txt',
'talk.religion.misc_83891.txt',
'talk.religion.misc_84283.txt',
'talk.religion.misc_83740.txt',
'talk.religion.misc_83591.txt',
'talk.religion.misc_83515.txt',
'talk.religion.misc_82771.txt',
'talk.religion.misc_83614.txt',
'talk.religion.misc_83847.txt',
'talk.religion.misc_84152.txt',
'talk.religion.misc_83470.txt',
```

```
'talk.religion.misc_83484.txt',
'talk.religion.misc_84293.txt',
'talk.religion.misc_84014.txt',
'talk.religion.misc_83517.txt',
'talk.religion.misc_84125.txt',
'talk.religion.misc_83646.txt',
'talk.religion.misc_84568.txt',
'talk.religion.misc_83681.txt',
'talk.religion.misc_84299.txt',
'talk.religion.misc_82808.txt',
'talk.religion.misc_84131.txt',
'talk.religion.misc_84024.txt',
'talk.religion.misc_84340.txt',
'talk.religion.misc_83463.txt',
'talk.religion.misc_84325.txt',
'talk.religion.misc_83863.txt',
'talk.religion.misc_83901.txt',
'talk.religion.misc_84233.txt',
'talk.religion.misc_83704.txt',
'talk.religion.misc_84070.txt',
'talk.religion.misc_83827.txt',
'talk.religion.misc_84113.txt',
'talk.religion.misc_84246.txt',
'talk.religion.misc_83643.txt',
'talk.religion.misc_84314.txt',
'talk.religion.misc_82758.txt',
'talk.religion.misc_82814.txt',
'talk.religion.misc_84137.txt',
'talk.religion.misc_84560.txt',
'talk.religion.misc_83518.txt',
'talk.religion.misc_83660.txt',
'talk.religion.misc_82779.txt',
'talk.religion.misc_83440.txt',
'talk.religion.misc_84351.txt',
'talk.religion.misc_83783.txt',
'talk.religion.misc_84256.txt',
'talk.religion.misc_83558.txt',
'talk.religion.misc_83460.txt',
'talk.religion.misc_83490.txt',
'talk.religion.misc_84511.txt',
'talk.religion.misc_84085.txt',
'talk.religion.misc_84509.txt',
'talk.religion.misc_84271.txt',
'talk.religion.misc_84216.txt',
'talk.religion.misc_83609.txt',
'talk.religion.misc_84451.txt',
'talk.religion.misc_83677.txt',
```

```
       'talk.religion.misc_83585.txt',
       'talk.religion.misc_84262.txt',
       'talk.religion.misc_83848.txt',
       'talk.religion.misc_84323.txt',
       'talk.religion.misc_84122.txt',
       'talk.religion.misc_84092.txt',
       'talk.religion.misc_84225.txt',
       'talk.religion.misc_83732.txt',
       'talk.religion.misc_84563.txt',
       'talk.religion.misc_83884.txt',
       'talk.religion.misc_84443.txt',
       'talk.religion.misc_84327.txt',
       'talk.religion.misc_84360.txt',
       'talk.religion.misc_84206.txt',
       'talk.religion.misc_83642.txt',
       'talk.religion.misc_83617.txt',
       'talk.religion.misc_84053.txt',
       'talk.religion.misc_84097.txt',
       'talk.religion.misc_84141.txt',
       'talk.religion.misc_84214.txt',
       'talk.religion.misc_83507.txt',
       'talk.religion.misc_84324.txt',
       'talk.religion.misc_84449.txt',
       'talk.religion.misc_84130.txt',
       'talk.religion.misc_84073.txt',
       'talk.religion.misc_83722.txt',
       'talk.religion.misc_83522.txt',
       'talk.religion.misc_84440.txt',
       'talk.religion.misc_84058.txt',
       'talk.religion.misc_84320.txt',
       'talk.religion.misc_84047.txt',
       …]
```

```python
# Extracting text, email and subject from ---> alt.atheism_49960.txt
preprocessed_subject_list, preprocessed_email_list, preprocessed_text_list,
 ↪raw_text_list, name = pre_process_text_file_specific(file_list)


print(name)
print()
print(preprocessed_subject_list)
print()
print(preprocessed_email_list)
print()
print(preprocessed_text_list)
print()
```

['alt.atheism_49960.txt']

['Alt Atheism FAQAtheist Resources']

['netcom mantis']

['atheist resourcesaddressesatheist organizationsusafreedom from religion foundationdarwin fish bumper stickers and assorted other atheist paraphernalia areavailable from the freedom from religion foundationtheevolution designsevolution designs sell the darwin fishfish symbol like the oneschristians sticktheir cars but with feet and the word darwin writteinside the deluxe mouldedplastic fishpostpaidthepeoplethe san francisco bay area can get darwin fish from lynn goldfor net people who lynn directly thepriceper fishamerican atheist pressaap publish various atheist books critiquesthe bible liststhe bible handbook ball andfoote american atheist pressisbnedition bible contradictionscontradicts itself aap basedthe king james versionthe bibcameron road austinprometheus booksprometheus books glenn drive buffalofor humanismorganization promoting black secular humanism and uncovering the historyblack freethought they publishquarterly newsletter aah examinerbuffalounited kingdomrationalist press association national secular societyislington high street holloway roadlondon londonbritish humanist association south place ethical societylambconduit passage conwaylondon wcrred lion squarelondon wcrfaxthe national secular society publish the freethinkermonthly magazinefoundedgermanyibkabund derundpostfachberlin germanymiz materialien und informationen zur zeit politischesjournal derund atheisten hrsg ibkamizvertrieb postfachberlin germanyibdkbucherdienst derpostfachhannover germanybooks fictionthomasdischthe santa claus compromiseshort story the ultimate proof that santa exists all characters andevents are fictitious any similaritylivingdead godswellwaltermillercanticle for leibowitzone gemthis post atomic doomsday novelthe monks who spent their livescopying blueprints from saint leibowitz filling the sheetspaper withink and leaving white lines and lettersedgar pangborndavypost atomic doomsday novel setclerical states the church for exampleforbids that anyone produce describeuse any substance containingatomsphilipdickphilipdick dick wrote many philosophical andshortstories and novels his stories are bizarretimes but very approachablewrote mainlybutwrote about people truth and religion rather thantechnology althoughoften believed thathad met some sortgodgalactic pothealerfallible alien deity summonsgroupearth craftsmen and womenremote planetraisegiant cathedral from beneath the oceans when thedeity beginsdemand faith from the earthers pothealer joe fernwrightunablecomplypolished ironic and amusing novelmazedeathnoteworthy for its descriptionreligionvalisthe schizophrenic hero searches for the hidden mysteriesgnosticchristianity after realityfired into his brain pink laser beamunknown but possibly divine origin accompaniedhis dogmatic anddismissively atheist friend and assorted other odd charactersthe divine invasiongod invades earthmakingyoung woman pregnantshe returns fromanother star system unfortunately sheterminally ill and mustassisted dead man whose brainwiredhour easy listening musicmargaret atwoodthe handmaidstory basedthe premise that thecongressmysteriouslyassassinated

andquickly take chargethe nationsetright again the bookthe diary womanlifeshe triesliveunder the new christian theocracy womenrightown propertyrevokedand their bank accounts are closed sinful luxuries are outlawed and theradioonly used for readings from the bible crimes are punishedhunted down and hanged atwoodwriting styledifficultget usedfirst but the tale grows more and more chilling goesvarious authorsthe biblethis somewhat dull and rambling work has often been criticized howeverprobably worth readingonlythat you will know what all the fussaboutexistsmany different versionsmake sure you get the onetrue versionbooks nonfictionpeterrosavicarschrist bantam pressalthoughrosa seems christianeven catholic this veryenlighting historypapal immoralities adulteries fallacies etcgerman translation gottes erste diener die dunkle seite desdroemerknaurmichael martinatheismphilosophical justification temple university pressphiladelphia usadetailed and scholarly justificationatheism containsoutstandingtendentiousarea argues both for negative atheismthe nonbelieftheand also for positive atheism the belieftheincludes great refutationsthe mostchallenging arguments for god particular attentionpaidrefutingcontempory theists suchplatinga and swinburnethe case against christianity temple university presscomprehensive critiquechristianitywhichconsidersdemonstrates that they are unsupportable andor incoherentpages isbnjames turnerwithout god without creed the johns hopkins university press baltimoreusasubtitled the originsunbeliefamerica examines the waywhichbecamemainstream alternativeworldview focussesthe period and while considering franceand britain the emphasis american and particularly new englanddevelopments neitherreligious historysecularizationatheismwithout god without creedrather the intellectual historythe fatsingle idea the belief that god existsthe great thoughts ballantine books new york usadictionaryquotations different kind concentratingstatementsand writings which explicitlyimplicitly present the personphilosophyopinions from manypeople for some popular observations traces the waywhich variouspeople expressed and twisted the idea over the centuries quitenumberthe quotations are derived from cardiffwhat great men thinkreligioand noyesviewsreligionrichard swinburneclarendon paperbacks oxfthis bookthe second volume trilogy that began with the coherencethiswork swinburne attemptsconstructseriesinductive arguments for theexistencegod his arguments which are somewhat tendentious and relyupon the imputationlatecentury western christian values andaesthetics god whichsupposedlysimplecanconceived weredecisively rejectedmackiethe miracletheismthe revisededitionthe existencegod swinburne includesappendixwhichmakessomewhat incoherent attemptrebutmackiethe miracletheism oxfordvolume containscomprehensive reviewthe principalarguments for and against the existencegodranges from the classicalphilosophical positionsdescartes anselm berkeley hume throughthe moral argumentsnewman kant and sidgwickthe recent restatementsthe classical thesesplantinga and swinburnealso addresses thosepositions which push the conceptgod beyond the realmthe rationalsuchthosekierkegaard kung and philipswellreplacements forgod suchlelieaxiarchism the book delightread lessformalistic and better written than martinworks and refreshingly direcwhen compared with the handwavingswinburnejameshaughtholy horrorsillustrated historyreligious murder and madnessprometheus bookslooksreligious persecution from ancient timesthe present day andnot onlychristianslibrarycongress catalog card

numbernormallenafrican american humanismanthologysee the listing for african
americans for humanism abovegordon steinanthologyatheism and rationalism
prometheusanthology coveringwide rangesubjects including the devil eviland
moralityand the historyfreethoughtcomprehensive bibliographyedmundcohenthe
mindthe biblebeliever prometheus booksstudywhy people become christianand what
effecthasthemnet resourcesthere small mailbased archive servermantiscouk which
carriesarchivesoldarticles and assorted other files formore information send
mail sayinghelpsend atheismindexandwill mail backreplymathew']

```python
preprocessed_subject_list, preprocessed_email_list, preprocessed_text_list,␣
 ↪raw_text_list = pre_process_text_file(file_list)
```

```python
fileNames = []

for i, file in enumerate(file_list):
    fileNames.append(file.split("_")[0])
```

```python
# Data Frame
```

```python
len(preprocessed_email_list), len(preprocessed_subject_list),␣
 ↪len(preprocessed_text_list), len(raw_text_list), "fileNames"
```

```
(18828, 18828, 18828, 18828, 'fileNames')
```

```python
data = {"pre_emails":preprocessed_email_list, "pre_subject":
 ↪preprocessed_subject_list, "pre_text":preprocessed_text_list,"text":
 ↪raw_text_list, "class":fileNames}
df_1 = pd.DataFrame(data)

len(preprocessed_text_list), len(preprocessed_email_list)␣
 ↪,len(preprocessed_subject_list), len(fileNames)


df_1
```

```
                                      pre_emails  …
class
0                aston uhura apexgrp terminus halcyon  …
talk.politics.misc
1                  uhura optilink COM galileo rochester edu  …
talk.politics.misc
2        psuvm INS Edu usenet Freenet psu CWRU edu clev…  …
talk.politics.misc
3        Campaign92 alchemy Org mit life edu utoronto chem  …
talk.politics.misc
4                            NCR netcom COM ncratl AtlantaGA  …
```

```
talk.politics.misc
…                                                              …  …
…
18823                      fido asd solntze buphy wpd sgi edu  …
alt.atheism
18824                            buphy dbstu1 tubs edu  …
alt.atheism
18825                        galki toppoint alice att  …
comp.graphics
18826      isc rit cco CWRU gap edu mantis ultb caltech  …
alt.atheism
18827          solntze cco wpd CWRU gap sgi edu caltech  …
alt.atheism

[18828 rows x 5 columns]
```

8 Combining 3 columns to a single column

```python
df_2 = df_1.assign(x = df_1.pre_emails.astype(str) +" " + df_1.pre_subject.
 ↪astype(str)+ " " + df_1.pre_text.astype(str))
df_2 = df_2.drop(["pre_emails", "pre_subject", "pre_text", 'text'], axis=1)
```

```python
df_2
```

```
                      class                                                    x
0         talk.politics.misc  aston uhura apexgrp terminus halcyon ReNew Stu…
1         talk.politics.misc  uhura optilink COM galileo rochester edu ReNew…
2         talk.politics.misc  psuvm INS Edu usenet Freenet psu CWRU edu clev…
3         talk.politics.misc  Campaign92 alchemy Org mit life edu utoronto c…
4         talk.politics.misc  NCR netcom COM ncratl AtlantaGA ReTemper tantr…
…                         …                                                  …
18823           alt.atheism  fido asd solntze buphy wpd sgi edu ReYet more …
18824           alt.atheism  buphy dbstu1 tubs edu ReThe Inimitable Rushdie…
18825         comp.graphics  galki toppoint alice att ReTIFFphilosophical s…
18826           alt.atheism  isc rit cco CWRU gap edu mantis ultb caltech R…
18827           alt.atheism  solntze cco wpd CWRU gap sgi edu caltech Re Cr…

[18828 rows x 2 columns]
```

```python
le = preprocessing.LabelEncoder()
le.fit(df_2['class'])
df_2['class'] = le.transform(df_2['class'])
```

```python
df_2
```

```
[ ]:        class                                                    x
     0          18   aston uhura apexgrp terminus halcyon ReNew Stu…
     1          18   uhura optilink COM galileo rochester edu ReNew…
     2          18   psuvm INS Edu usenet Freenet psu CWRU edu clev…
     3          18   Campaign92 alchemy Org mit life edu utoronto c…
     4          18   NCR netcom COM ncratl AtlantaGA ReTemper tantr…
     …           …                                                  …
     18823       0   fido asd solntze buphy wpd sgi edu ReYet more …
     18824       0   buphy dbstu1 tubs edu ReThe Inimitable Rushdie…
     18825       1   galki toppoint alice att ReTIFFphilosophical s…
     18826       0   isc rit cco CWRU gap edu mantis ultb caltech R…
     18827       0   solntze cco wpd CWRU gap sgi edu caltech Re Cr…

     [18828 rows x 2 columns]
```

```
[ ]: df_2["class"].nunique()
```

```
[ ]: 20
```

9 Traint test split of sentences

```
[ ]: X = df_2['x']
     y = df_2['class']
```

```
[ ]: # stop = stopwords.words('english')
     # df_2['x'].apply(lambda x: len([x for x in x.split() if x in stop]))
```

```
[ ]: sentences_train, sentences_test, y_train, y_test = train_test_split(X, y,␣
      ↪stratify=y, test_size=0.25)
```

```
[ ]: sentences_train.shape, sentences_test.shape, y_train.shape, y_test.shape
```

```
[ ]: ((14121,), (4707,), (14121,), (4707,))
```

```
[ ]: l = []
     for s in sentences_train:
         l.append(len(s))
```

```
[ ]: np.mean(l)
```

```
[ ]: 1068.7188584377877
```

10 Text to Sequence

```python
sentences = df_2['x'].to_numpy()

tokenizer   = Tokenizer( num_words=None,
    filters='!"#$%&()*+,-./:;<=>?@[\\]^`{|}~\t\n',
    lower=True,
    split=' ',
    char_level=False,
    oov_token=None,
    document_count=0)

tokenizer.fit_on_texts(sentences_train)

X_train = tokenizer.texts_to_sequences(sentences_train)
X_test = tokenizer.texts_to_sequences(sentences_test)

vocab_size = len(tokenizer.word_index) + 1
maxlen = 909

X_train = pad_sequences(X_train, padding='post', maxlen=maxlen)
X_test = pad_sequences(X_test, padding='post', maxlen=maxlen)

print("train shape {} \ntest shape {}".format(X_train.shape, X_test.shape))
```

```
train shape (14121, 909)
test shape (4707, 909)
```

```python
y_train = y_train.to_numpy().reshape(-1, 1)
y_test = y_test.to_numpy().reshape(-1, 1)
```

```python
ohe = OneHotEncoder()
y_train = ohe.fit_transform(y_train).toarray()
y_test = ohe.transform(y_test).toarray()
```

11 dictionary of all words and it's vector of 300 dimension

```python
embeddings_index = dict()

word_embedding_50_path = '/content/gdrive/My Drive/Colab Notebooks/CNN on Text␣
 ↪data/glove.6B.50d.txt'
word_embedding_100_path = '/content/gdrive/My Drive/Colab Notebooks/CNN on Text␣
 ↪data/glove.6B.100d.txt'
word_embedding_200_path = '/content/gdrive/My Drive/Colab Notebooks/CNN on Text␣
 ↪data/glove.6B.200d.txt'
```

```
word_embedding_300_path = '/content/gdrive/My Drive/Colab Notebooks/CNN on Text␣
 ↪data/glove.6B.300d.txt'

f = open(word_embedding_300_path, encoding="utf8")

for line in f:
    values = line.split()
    word = values[0]
    coefs = np.asarray(values[1:], dtype='float32')
    embeddings_index[word] = coefs
f.close()
```

12 Embedded matrix

```
[ ]: embedding_matrix = np.zeros((vocab_size, 300))
```

```
[ ]: vocab_size
```

```
[ ]: 392549
```

```
[ ]: for word, i in tokenizer.word_index.items():
        embedding_vector = embeddings_index.get(word)
        if embedding_vector is not None:
            embedding_matrix[i] = embedding_vector
```

```
[ ]: embedding_matrix.shape
```

```
[ ]: (392549, 300)
```

```
[ ]: embedding_dim = 300
```

```
[ ]: X_train[0].shape
```

```
[ ]: (909,)
```

13 Custom Callback

```
[ ]: from sklearn.metrics import f1_score


    class LossHistory(tf.keras.callbacks.Callback):

        def __init__(self, model_no):
            self.val_acc_list = []
            self.init_lr = 0.01
```

```python
    def on_train_begin(self, logs={}):
        self.history={'loss': [],'val_loss': []}


    def on_epoch_end(self, epoch, logs={}):
        ## on end of each epoch, we will get logs and update the self.history
↪dict
        y_pred_train_f1 = []
        y_pred_test_f1 = []
        y_pred_train_auc = []
        y_pred_test_auc = []

        # ===================== LOSS RECORD=============================
        self.history['val_loss'].append(logs.get('val_loss'))
        self.history['loss'].append(logs.get('loss'))

        if epoch > 2 :
          loss_curr = self.history['val_loss'][epoch]
          loss_prev = self.history['val_loss'][epoch-1]

          if loss_curr > loss_prev:
            self.init_lr = self.init_lr * 0.53
            K.set_value(self.model.optimizer.learning_rate, self.init_lr)


        # ========================= PREDICTION

        y_train_pred = self.model.predict(X_train)
        y_test_pred = self.model.predict(X_test)


        y_train_pred_l = np.argmax(y_train_pred, axis=-1)
        y_test_pred_l = np.argmax(y_test_pred, axis=-1)

        y_train_l = np.argmax(y_train, axis=-1)
        y_test_l = np.argmax(y_test, axis=-1)


        # ========================== MICRO-F1 SCORE

        train_f1_score = f1_score(y_train_l, y_train_pred_l, average='micro')
        print("train f1 score {}".format(np.round(train_f1_score, 5)))
        test_f1_score = f1_score(y_test_l, y_test_pred_l, average='micro')
        print("test f1 score {}".format(np.round(test_f1_score, 5)))
```

14 Creating Model

```python
tf.keras.backend.clear_session()

input_layer = Input(shape=(909,))

# Embedding layer
embedding_layer = Embedding(vocab_size, embedding_dim,
 weights=[embedding_matrix], input_length=maxlen,
 trainable=False)(input_layer)

# conv1D - kernel_size = 2
conv1D_1 = Conv1D(filters=16, kernel_size=2, activation='relu',
 kernel_initializer='he_normal') (embedding_layer)

# conv1D - kernel_size = 3
conv1D_2 = Conv1D(filters=16, kernel_size=3, activation='relu',
 kernel_initializer='he_normal') (embedding_layer)

# conv1D - kernel_size = 4
conv1D_3 = Conv1D(filters=16, kernel_size=4, activation='relu',
 kernel_initializer='he_normal') (embedding_layer)

# concatination
concat_1  = concatenate([conv1D_1, conv1D_2, conv1D_3], axis=1)

# maxpool 1D
Max_pool_1 = MaxPool1D(pool_size=2, strides=1 ,padding='valid',
 name='Pool1')(concat_1)


# conv1D - kernel_size = 5
conv1D_4 = Conv1D(filters=8, kernel_size=5, activation='relu',
 kernel_initializer='he_normal') (Max_pool_1)

# conv1D - kernel_size = 6
conv1D_5 = Conv1D(filters=8, kernel_size=6, activation='relu',
 kernel_initializer='he_normal') (Max_pool_1)
```

```python
# conv1D - kernel_size = 7
conv1D_6 = Conv1D(filters=8, kernel_size=7, activation='relu',
 ↪kernel_initializer='he_normal') (Max_pool_1)


# concatination
concat_2  = concatenate([conv1D_4, conv1D_5, conv1D_6], axis=1)

# maxpool 1D
Max_pool_2 = MaxPool1D(pool_size=2, strides=1 ,padding='valid',
 ↪name='Pool2')(concat_2)

# conv1D - p = 11
conv1D_7 = Conv1D(filters=4, kernel_size=9, activation='relu',
 ↪kernel_initializer='he_normal') (Max_pool_2)


# flatten
flatten_1 = Flatten()(conv1D_7)

# drop out
dropout1 = Dropout(0.2) (flatten_1)

# dense
Dense_1 = Dense(4, activation='relu', name='Dense_1',
 ↪kernel_initializer='he_normal')(dropout1)


# output
output_layer = Dense(20, activation='softmax', name="Output")(Dense_1)

model = Model(inputs=input_layer, outputs=output_layer)


callback = LossHistory(1)

stop_early = tf.keras.callbacks.EarlyStopping(monitor='val_loss', mode='min',
 ↪patience=5,)

log_dir="/content/gdrive/My Drive/Colab Notebooks/CNN on Text data/logs/fit/
 ↪model_1__" + datetime.datetime.now().strftime("%Y%m%d-%H%M%S")
tensorboard_callback = tf.keras.callbacks.TensorBoard(log_dir=log_dir,
 ↪histogram_freq=1, write_graph=True, write_grads=True)



optim = tf.keras.optimizers.Adam(learning_rate=0.01)

model.compile(optimizer=optim,
              loss='categorical_crossentropy',
```

```
            metrics=[None])
```

WARNING:tensorflow:`write_grads` will be ignored in TensorFlow 2.0 for the `TensorBoard` Callback.

```
[ ]: model.summary()
```

```
Model: "model"
_____
_____
Layer (type)                   Output Shape          Param #     Connected to
============================================================================
==================
input_1 (InputLayer)           [(None, 909)]         0
_____
_____
embedding (Embedding)          (None, 909, 300)      117764700   input_1[0][0]
_____
_____
conv1d (Conv1D)                (None, 908, 16)       9616        embedding[0][0]
_____
_____
conv1d_1 (Conv1D)              (None, 907, 16)       14416       embedding[0][0]
_____
_____
conv1d_2 (Conv1D)              (None, 906, 16)       19216       embedding[0][0]
_____
_____
concatenate (Concatenate)      (None, 2721, 16)      0           conv1d[0][0]
                                                                 conv1d_1[0][0]
                                                                 conv1d_2[0][0]
_____
_____
Pool1 (MaxPooling1D)           (None, 2720, 16)      0
concatenate[0][0]
_____
_____
conv1d_3 (Conv1D)              (None, 2716, 8)       648         Pool1[0][0]
_____
_____
conv1d_4 (Conv1D)              (None, 2715, 8)       776         Pool1[0][0]
_____
_____
conv1d_5 (Conv1D)              (None, 2714, 8)       904         Pool1[0][0]
_____
_____
concatenate_1 (Concatenate)    (None, 8145, 8)       0           conv1d_3[0][0]
                                                                 conv1d_4[0][0]
```

```
                                                          conv1d_5[0][0]
_____
_____
Pool2 (MaxPooling1D)             (None, 8144, 8)      0
concatenate_1[0][0]

_____
_____
conv1d_6 (Conv1D)                (None, 8136, 4)      292         Pool2[0][0]
_____
_____
flatten (Flatten)                (None, 32544)        0           conv1d_6[0][0]
_____
_____
dropout (Dropout)                (None, 32544)        0           flatten[0][0]
_____
_____
Dense_1 (Dense)                  (None, 4)            130180      dropout[0][0]
_____
_____
Output (Dense)                   (None, 20)           100         Dense_1[0][0]
==============================================================================
==================
Total params: 117,940,848
Trainable params: 176,148
Non-trainable params: 117,764,700

_____
_____
```
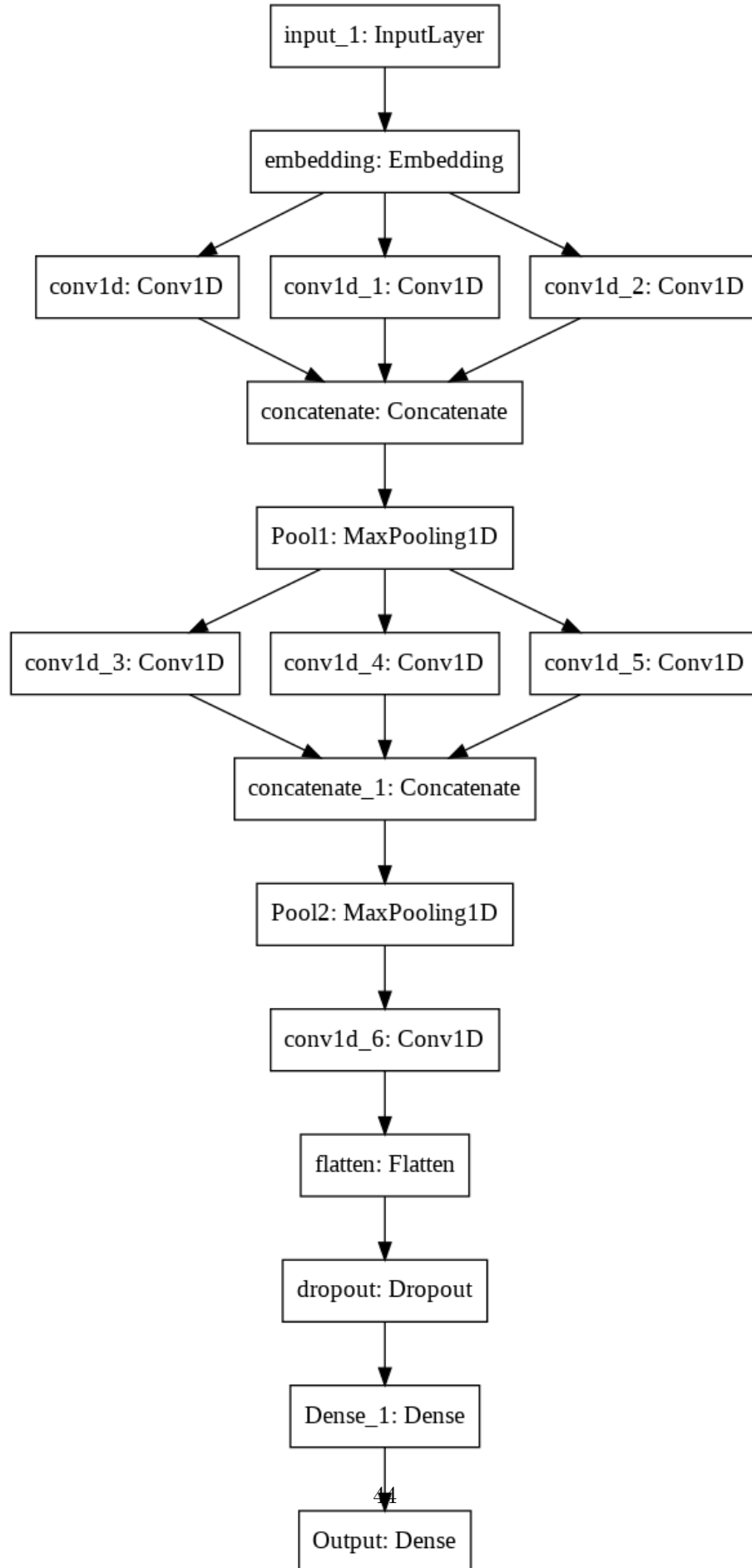
```python
plot_model(model, to_file='/content/gdrive/My Drive/Colab Notebooks/CNN on Text
   data/convolutional_neural_network_word_embedding.png')
```

[ ]:

```
                    ┌──────────────────────┐
                    │  input_1: InputLayer │
                    └──────────────────────┘
                               │
                               ▼
                    ┌──────────────────────┐
                    │ embedding: Embedding │
                    └──────────────────────┘
                    ╱          │          ╲
                   ▼           ▼           ▼
      ┌───────────────┐ ┌─────────────────┐ ┌─────────────────┐
      │ conv1d: Conv1D│ │conv1d_1: Conv1D │ │conv1d_2: Conv1D │
      └───────────────┘ └─────────────────┘ └─────────────────┘
                    ╲          │          ╱
                     ▼         ▼         ▼
                    ┌──────────────────────┐
                    │concatenate: Concatenate│
                    └──────────────────────┘
                               │
                               ▼
                    ┌──────────────────────┐
                    │ Pool1: MaxPooling1D  │
                    └──────────────────────┘
                    ╱          │          ╲
                   ▼           ▼           ▼
    ┌─────────────────┐ ┌─────────────────┐ ┌─────────────────┐
    │conv1d_3: Conv1D │ │conv1d_4: Conv1D │ │conv1d_5: Conv1D │
    └─────────────────┘ └─────────────────┘ └─────────────────┘
                    ╲          │          ╱
                     ▼         ▼         ▼
                 ┌──────────────────────────┐
                 │concatenate_1: Concatenate│
                 └──────────────────────────┘
                               │
                               ▼
                    ┌──────────────────────┐
                    │ Pool2: MaxPooling1D  │
                    └──────────────────────┘
                               │
                               ▼
                    ┌──────────────────────┐
                    │ conv1d_6: Conv1D     │
                    └──────────────────────┘
                               │
                               ▼
                    ┌──────────────────────┐
                    │ flatten: Flatten     │
                    └──────────────────────┘
                               │
                               ▼
                    ┌──────────────────────┐
                    │ dropout: Dropout     │
                    └──────────────────────┘
                               │
                               ▼
                    ┌──────────────────────┐
                    │ Dense_1: Dense       │
                    └──────────────────────┘
                               │
                               ▼
                    ┌──────────────────────┐
                    │ Output: Dense        │
                    └──────────────────────┘
```

44

```python
history = model.fit(X_train, y_train,
                    epochs=50,
                    verbose=True,
                    validation_data=(X_test, y_test),
                    batch_size=200, callbacks=[callback, tensorboard_callback,
 ↪stop_early])
```

```
Epoch 1/50
 6/71 [=>…] - ETA: 17s - loss:
3.3428WARNING:tensorflow:Callback method `on_train_batch_end` is slow compared
to the batch time (batch time: 0.0759s vs `on_train_batch_end` time: 0.1055s).
Check your callbacks.
71/71 [==============================] - 12s 150ms/step - loss: 3.0704 -
val_loss: 2.9409
train f1 score 0.08675
test f1 score 0.0854
Epoch 2/50
71/71 [==============================] - 10s 137ms/step - loss: 2.8786 -
val_loss: 2.5453
train f1 score 0.11692
test f1 score 0.11557
Epoch 3/50
71/71 [==============================] - 10s 136ms/step - loss: 2.4706 -
val_loss: 2.3421
train f1 score 0.16868
test f1 score 0.15764
Epoch 4/50
71/71 [==============================] - 10s 135ms/step - loss: 2.2671 -
val_loss: 2.2470
train f1 score 0.22045
test f1 score 0.20076
Epoch 5/50
71/71 [==============================] - 10s 135ms/step - loss: 2.1293 -
val_loss: 2.2545
train f1 score 0.25579
test f1 score 0.22881
Epoch 6/50
71/71 [==============================] - 10s 136ms/step - loss: 2.0373 -
val_loss: 2.1672
train f1 score 0.29113
test f1 score 0.25366
Epoch 7/50
71/71 [==============================] - 10s 134ms/step - loss: 1.9414 -
val_loss: 2.1602
train f1 score 0.30593
```
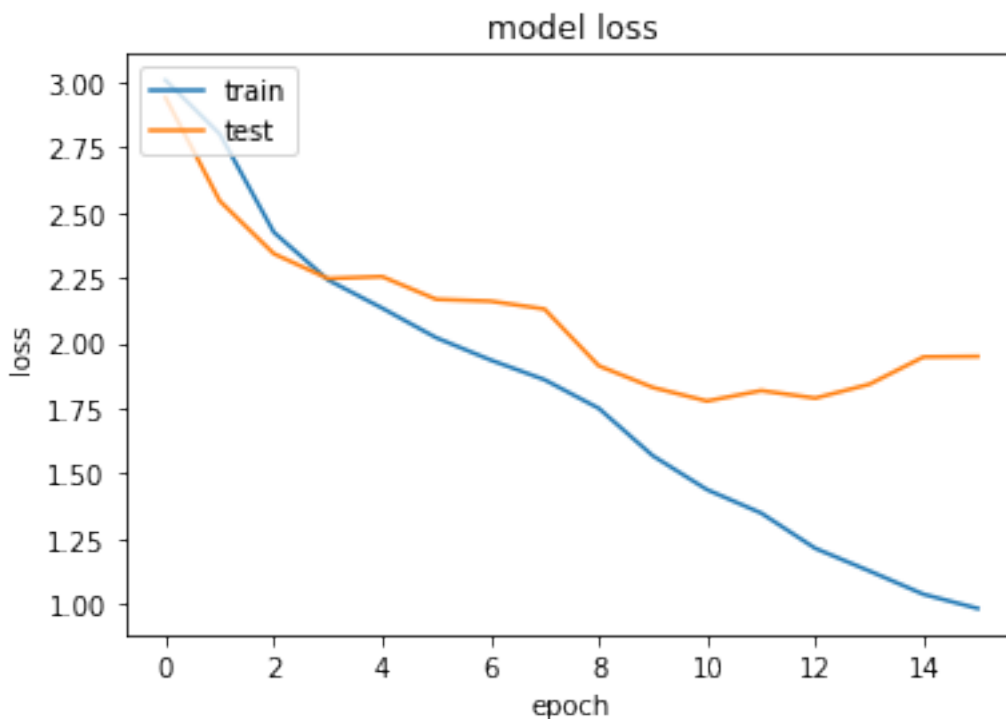
```
test f1 score 0.26068
Epoch 8/50
71/71 [==============================] - 10s 135ms/step - loss: 1.8638 -
val_loss: 2.1300
train f1 score 0.32264
test f1 score 0.27279
Epoch 9/50
71/71 [==============================] - 10s 135ms/step - loss: 1.7737 -
val_loss: 1.9124
train f1 score 0.40564
test f1 score 0.34077
Epoch 10/50
71/71 [==============================] - 10s 135ms/step - loss: 1.5793 -
val_loss: 1.8291
train f1 score 0.47907
test f1 score 0.39558
Epoch 11/50
71/71 [==============================] - 10s 135ms/step - loss: 1.4207 -
val_loss: 1.7771
train f1 score 0.50237
test f1 score 0.40493
Epoch 12/50
71/71 [==============================] - 10s 135ms/step - loss: 1.3563 -
val_loss: 1.8170
train f1 score 0.5438
test f1 score 0.41151
Epoch 13/50
71/71 [==============================] - 10s 135ms/step - loss: 1.2106 -
val_loss: 1.7882
train f1 score 0.60711
test f1 score 0.43616
Epoch 14/50
71/71 [==============================] - 10s 135ms/step - loss: 1.1272 -
val_loss: 1.8413
train f1 score 0.64478
test f1 score 0.46272
Epoch 15/50
71/71 [==============================] - 10s 135ms/step - loss: 1.0323 -
val_loss: 1.9464
train f1 score 0.6598
test f1 score 0.45847
Epoch 16/50
71/71 [==============================] - 10s 136ms/step - loss: 0.9856 -
val_loss: 1.9482
train f1 score 0.68451
test f1 score 0.46866
```

```
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('model loss')
plt.ylabel('loss')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.show()
```



```
%tensorboard --logdir /content/gdrive/My Drive/Colab Notebooks/CNN on Text data/
 ↪logs/fit
```

```
ERROR: Failed to launch TensorBoard (exited with 2).
Contents of stderr:
2021-04-21 14:08:28.740676: I tensorflow/stream_executor/platform/default/
 ↪dso_loader.cc:49] Successfully opened dynamic library libcudart.so.11.0
usage: tensorboard [-h] [--helpfull] [--logdir PATH] [--logdir_spec PATH_SPEC]
                   [--host ADDR] [--bind_all] [--port PORT]
                   [--purge_orphaned_data BOOL] [--db URI] [--db_import]
                   [--inspect] [--version_tb] [--tag TAG] [--event_file PATH]
                   [--path_prefix PATH] [--window_title TEXT]
                   [--max_reload_threads COUNT] [--reload_interval SECONDS]
                   [--reload_task TYPE] [--reload_multifile BOOL]
                   [--reload_multifile_inactive_secs SECONDS]
                   [--generic_data TYPE]
```

```
                    [--samples_per_plugin SAMPLES_PER_PLUGIN]
                    [--whatif-use-unsafe-custom-prediction␣
 ↪YOUR_CUSTOM_PREDICT_FUNCTION.py]
                    [--whatif-data-dir PATH]
                    {serve,dev} …
tensorboard: error: invalid choice: 'Drive/Colab' (choose from 'serve', 'dev')
```