plagirism-checker

August 27, 2023

```
[1]: !pip install -U nltk
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages
(3.8.1)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages
(from nltk) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages
(from nltk) (1.3.2)
Requirement already satisfied: regex>=2021.8.3 in
/usr/local/lib/python3.10/dist-packages (from nltk) (2023.6.3)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages
(from nltk) (4.66.1)
```

```
[2]: import re
     from nltk.util import ngrams, pad_sequence, everygrams
     from nltk.tokenize import word_tokenize
     from nltk.lm import MLE, WittenBellInterpolated
     import numpy as np
     import plotly.graph_objects as go
     from scipy.ndimage import gaussian_filter
```

```
[3]: f1 = open('train.txt', 'w+')
     f2 = open('test.txt', 'w+')
```

```
[7]: f1.write("""
     "Wings of Fire: An Autobiography of Abdul Kalam"  is written by Dr. APJ Abdul␣
     ↪Kalam and Arun Tiwari. This book is an autobiography of Dr. APJ Abdul Kalam,␣
     ↪former president of India. This story sheds light on the journey of a young␣
     ↪boy from Rameswaram to become a renowned scientist. It reflects how simple␣
     ↪living, dedication, strong will and hard work led to success. It also shows␣
     ↪how cultural unity impacts the lives of individuals.
```

```
The story begins with Kalam as a boy  around 1930 in the village of Rameswaram␣
 ↪where he lived with his parents and four brothers and one sister. In this␣
 ↪book, the author tells us how the role of family, relatives, friends, and␣
 ↪society contribute to an individual's life. Kalam describes his village␣
 ↪where everyone lives in harmony. Religious discrimination was unknown at␣
 ↪that time. A passage that shows the closeness  they had: "The high priest of␣
 ↪Rameswaram temple, Pakshi Lakshmana Sastry, was a very close friend of my␣
 ↪father's. One of the most vivid memories of my early childhood is of the two␣
 ↪men, each in his traditional attire, discussing spiritual matters."
""")
f2.write("""
"Wings of fire an autobiography of abdul kalam is a book by dr.APJ Abdul Kalam␣
 ↪and Arun Tiwari.APJ Abdul Kalam, former president of India.This story sheds␣
 ↪light on the journey of a boy from rameswaram to become a famous scientist.␣
 ↪It reflects how simple living, dedication, strong will and hard work led to␣
 ↪success.It also shows how cultural unity influences the lives of individual␣
 ↪individuals.The story begins with kalam as a boy in the village of␣
 ↪rameswaram around 1930 where he lived with his parents four brothers and one␣
 ↪sister.The author tells us in this book how the role of family friends and␣
 ↪society contribute to an individuals life.Kalam describes his village where␣
 ↪everybody lives in harmony.Discrimination was never known at that time and␣
 ↪was unknown today.A passage that shows the closeness they had the priest of␣
 ↪the temple of rameswaram pakshi lakshmana sastry was a very close friend of.␣
 ↪One of the most vivid memories of my early childhood is the conversation␣
 ↪between two males each wearing his traditional clothes about religious␣
 ↪matters."\n\nWhen he was very young, Kalam wanted to join the Indian␣
 ↪Airforce but he was not able to clear the interview.Far from being␣
 ↪demotivated, he focused more on studies, started his career in Aeronautical␣
 ↪Development Establishment (ADE), and was involved in the design of a␣
 ↪hovercraft.He went to nasa on a training programme.He moved to indian space␣
 ↪research and met vikram sarabhai there.He also met many great minds who␣
 ↪inspired him.The book also describes leadership and teamwork, while making␣
 ↪the projects.It shows how a man becomes a leader whenever there is a failed␣
 ↪or a successful project.APJ Kalam, which I find distracting from the story␣
 ↪of his life.Some readers might find these detailed descriptions boring and␣
 ↪unappealing.This publication highlights the successful execution of agni and␣
 ↪other projects.The book tells the story of india's technological␣
 ↪achievements at a time when few nations can boast of its technological␣
 ↪breakthroughs.It talks about some brilliant minds who worked behind the␣
 ↪indian space research such as vikram sarabhai and vikram.Brahm Prakash and␣
 ↪their contribution towards ISRO.As a reader, this book initially is very␣
 ↪interesting particularly the portions from his younger days to a college␣
 ↪student like how his family encouraged him.Later the writing becomes a␣
 ↪little monotonous as he presents minute illustrations of the projects.He␣
 ↪teaches the younger generation not to get demotivated and to live a positive␣
 ↪life.This book demonstrates how dr.Apj abdul kalam's simple lifestyle and␣
 ↪positive thinking became an inspiration to the indian youth"
```

```
""")

f1.close()
f2.close()
```

[8]:
```
# Training data file
train_data_file = "train.txt"

# read training data
with open(train_data_file) as f:
    train_text = f.read().lower()

# apply preprocessing (remove text inside square and curly brackets and rem␣
 ↪punc)
train_text = re.sub(r"\[.*\]|\{.*\}", "", train_text)
train_text = re.sub(r'[^\w\s]', "", train_text)
```

[12]:
```
import nltk
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data…
[nltk_data]    Unzipping tokenizers/punkt.zip.
```

[12]: True

[13]:
```
# set ngram number
n = 4

# pad the text and tokenize
training_data = list(pad_sequence(word_tokenize(train_text), n,  pad_left=True,␣
 ↪left_pad_symbol="<s>"))

# generate ngrams
ngrams = list(everygrams(training_data, max_len=n))
print("Number of ngrams:", len(ngrams))

# build ngram language models
model = WittenBellInterpolated(n)
model.fit([ngrams], vocabulary_text=training_data)
print(model.vocab)
```

```
Number of ngrams: 754
<Vocabulary with cutoff=1 unk_label='<UNK>' and 120 items>
```

[14]:
```
# testing data file
test_data_file = "test.txt"
```

```python
# Read testing data
with open(test_data_file) as f:
    test_text = f.read().lower()
test_text = re.sub(r'[^\w\s]', "", test_text)

# Tokenize and pad the text
testing_data = list(pad_sequence(word_tokenize(test_text), n,
                                 pad_left=True,
                                 left_pad_symbol="<s>"))
print("Length of test data:", len(testing_data))
```

```
Length of test data: 417
```

[15]:
```python
# assign scores
scores = []
for i, item in enumerate(testing_data[n-1:]):
    s = model.score(item, testing_data[i:i+n-1])
    scores.append(s)

scores_np = np.array(scores)

# set width and height
width = 8
height = np.ceil(len(testing_data)/width).astype("int32")
print("Width, Height:", width, ",", height)

# copy scores to rectangular blank array
a = np.zeros(width*height)
a[:len(scores_np)] = scores_np
diff = len(a) - len(scores_np)

# apply gaussian smoothing for aesthetics
a = gaussian_filter(a, sigma=1.0)

# reshape to fit rectangle
a = a.reshape(-1, width)

# format labels
labels = [" ".join(testing_data[i:i+width]) for i in range(n-1,
  ↪len(testing_data), width)]
labels_individual = [x.split() for x in labels]
labels_individual[-1] += [""]*diff
labels = [f"{x:60.60}" for x in labels]
```

```
Width, Height: 8 , 53
```

```
[17]:  # create heatmap
       fig = go.Figure(data=go.Heatmap(
                       z=a, x0=0, dx=1,
                       y=labels, zmin=0, zmax=1,
                       customdata=labels_individual,
                       hovertemplate='%{customdata} <br><b>Score:%{z:.3f}<extra></
         ↪extra>',
                       colorscale="burg"))
       fig.update_layout({"height":height*28, "width":1000, "font":{"family":"Courier␣
         ↪New"}})
       fig['layout']['yaxis']['autorange'] = "reversed"
       fig.show()

[ ]:
```