

IT 302

Assignment 4

Submitted By

Mohit Awachar
191IT231

First we import the necessary libraries required. IN the next step we input all the data from the text files.

We have considered 2 text file

1. EnglishText.txt : Contains data from a novel in English
2. MarathiText.txt : Conatins data from a article in Marathi

Next we convert all the data in the files to words and make an array out of it.

For all the unique words among the list we make a matrix out of that where dimension are $n \times n$, Where n is the no of unique words.

For all word2 encountered after word1 we increase matrix[word1][word1] by 1.

Finally we calculate probability in the matrix by dividing with appropriate no(sum of col).

Now to predict we use each word as a state and order contains how many words we want to predict. According to markov's chain the future results only depend on the current word. We use this fact to predict the next word with the help of transition matrix created by us. We generate a random word according to the probability. And then update prevWord as the newly generated word.

For each sentence, we run our code 1000 times to get the most probable result.

Following are results for Marathi :

- The Word arrays 1st words

```
['lovely',  
'ilonka',  
'there',  
'was',  
'once',  
'a',  
'king's',  
'son',  
'who',  
'told',  
'his',  
'father',  
'that',  
'he',  
'wish  
ed',  
'to',  
'marry.',  
'no',  
'no!']
```

The matrix and it's description

	whispered	breast	easily	make	smallest	orders.	lilona;	crown	somewhere	destruction;	...	fears
count	2222.000000	2222.000000	2222.000000	2222.000000	2222.000000	2222.000000	2222.000000	2222.000000	2222.000000	2222.000000	...	2222.000000
mean	0.000450	0.000450	0.000450	0.002700	0.000900	0.000450	0.000450	0.000900	0.000450	0.000450	...	0.000450
std	0.021214	0.021214	0.021214	0.059956	0.029995	0.021214	0.021214	0.042429	0.021214	0.021214	...	0.021214
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
75%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
max	1.000000	1.000000	1.000000	2.000000	1.000000	1.000000	1.000000	2.000000	1.000000	1.000000	...	1.000000

8 rows × 2222 columns

	whispered	breast	easily	make	smallest	orders.	ilonka;	crown	somewhere	destruction;	...	fears	feet	before;	gentle	'why	dreams.	bulrus
whispered	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
breast	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
easily	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
make	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
smallest	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
...	
dreams.	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
bulrushes?'	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
bulrush	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
guess	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	
ten	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	

- The final results in English

```
Enter the sentence
You should
Enter how many words to add
3
```

Last Word is should

Most probable result occurred 10 times

The best result occurred 10 among 1000 attempts :
You should like each other

```
1 . See all the results
2 . Continue
2
1 : Enter another sentence.
2 : Stop
1
```

```
Enter the sentence
His life is
Enter how many words to add
3
```

Last Word is is

Most probable result occurred 16 times

The best result occurred 16 among 1000 attempts :
His life is mere smith's work.

```
1 . See all the results
2 . Continue
2
1 : Enter another sentence.
2 : Stop
1
```

```
Enter the sentence
My laptop
Enter how many words to add
2
```

```
Last Word is  laptop
No matching words found
```

In the example shown below we are showing all results that occurred in the 1000 tries we tried

Enter the sentence
He disappeared quickly
Enter how many words to add
3

Last Word is quickly

Most probable result occurred 48 times

The best result occurred 48 among 1000 attempts :
He disappeared quickly enough, and the

1 . See all the results
2 . Continue
1

Printing all the results :

Freq = 2	enough, and were
Freq = 2	enough, and you
Freq = 11	enough, and in
Freq = 4	enough, and there
Freq = 2	up of busan,
Freq = 5	enough, and ordered
Freq = 15	enough, and said:
Freq = 1	enough, and lovely
Freq = 1	up his mind
Freq = 3	enough, and sad
Freq = 3	enough, and dashed
Freq = 25	enough, and he
Freq = 8	up there was
Freq = 1	up the cloak
Freq = 10	enough, and with
Freq = 17	up inside the
Freq = 2	enough, and remained
Freq = 2	enough, and drinking
Freq = 3	up it was
Freq = 22	up into the
Freq = 2	enough, and embraced
Freq = 9	enough, and was
Freq = 7	enough, and if
Freq = 1	up the crows
Freq = 3	up which he
Freq = 3	enough, and been
Freq = 4	enough, and putting
Freq = 4	enough, and died.
Freq = 9	up into a
Freq = 48	enough, and the
Freq = 9	enough, and to
Freq = 1	up the mill

Now let's look at the results for Marathi..

The 1st 20 words

```
: words = formWords(data)

# words is an array containing all word
print( len(words) )
words[:20]
```

60033

```
: ['लोकसभा',
  'निवडणूक',
  '२०१९',
  'मध्ये',
  'चे',
  'निष्ठावान',
  'नेते',
  'आणि',
  'माझे',
  'स्नेही',
  'श्री.',
  'जी',
  'रायगड',
  'लोकसभा',
  'मतदारसंघातून',
  'विजयी',
  'झाल्याबद्दल',
  'त्यांच',
  'मनापासून',
  'अभिनंदन']
```

Now let's see the matrix

```
len(matrix)
matrix.describe()
```

	अचूक	माझे	जहाल	केले?	पाठिराख्यांच्या	रूपये	चव्हाण	निधीही	शासन	बांधवांसाठी	...	महिन्याचे
count	4901.000000	4901.000000	4901.000000	4901.000000	4901.000000	4901.000000	4901.000000	4901.000000	4901.000000	4901.000000	...	4901.000000
mean	0.000204	0.002040	0.000204	0.000204	0.000204	0.000408	0.002244	0.000204	0.000408	0.000612	...	0.000204
std	0.014284	0.049445	0.014284	0.014284	0.014284	0.020199	0.068475	0.014284	0.020199	0.024736	...	0.014284
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
75%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000
max	1.000000	2.000000	1.000000	1.000000	1.000000	1.000000	4.000000	1.000000	1.000000	1.000000	...	1.000000

8 rows × 4901 columns

4

matrix

	अचूक	माझे	जहाल	केले?	पाठिराख्यांच्या	रूपये	चव्हाण	निधीही	शासन	बांधवांसाठी	...	महिन्याचे	२१	दूर	आले	सडक	शुभेच्छा	भेट	हिंदू	अजून	स्वीकारता
अचूक	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
माझे	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
जहाल	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
केले?	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
पाठिराख्यांच्या	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
...
शुभेच्छा	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
भेट	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
हिंदू	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
अजून	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
स्वीकारता	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

4901 rows × 4901 columns

4

The final results

1. EX 1

Enter the sentence
भारताचे पहिले राष्ट्रपती डॉ. राजेंद्र प्रसाद
Enter how many words to add
3

Last Word is प्रसाद

Most probable result occurred 101 times

The best result occurred 101 among 1000 attempts :
भारताचे पहिले राष्ट्रपती डॉ. राजेंद्र प्रसाद यांना वाढदिवसाच्या हार्दिक

- 1 . See all the results
 - 2 . Continue
- 2

2. EX 2

Enter the sentence
हिंदू राष्ट्रवादाची मांडणी करणारे तत्वज्ञ विज्ञानिष्ठ ज्वलंत साहित्यिक
Enter how many words to add
5

Last Word is साहित्यिक

Most probable result occurred 68 times

The best result occurred 68 among 1000 attempts :
हिंदू राष्ट्रवादाची मांडणी करणारे तत्वज्ञ विज्ञानिष्ठ ज्वलंत साहित्यिक क्रियाशील समाजसुधारक गोपाळ गणेश आगरकर

- 1 . See all the results
 - 2 . Continue
- 1

Printing all the results :

Freq = 13	क्रियाशील समाजसुधारक शैक्षणिक संस्थांमार्फत अखंड
Freq = 5	क्रियाशील समाजसुधारक शैक्षणिक क्रांतीचे जनक
Freq = 3	क्रियाशील समाजसुधारक जयोस्तुते' हे खऱ्या
Freq = 21	क्रियाशील समाजसुधारक ग्रामसुधारणेचे जनक महाराष्ट्राचे
Freq = 17	क्रियाशील समाजसुधारक कर्मयोगी महात्मा जोतिबा
Freq = 15	क्रियाशील समाजसुधारक पद्मभूषण बाबा आढाव!
Freq = 53	क्रियाशील समाजसुधारक लोकशाहीर अण्णाभाऊ साठे
Freq = 1	क्रियाशील समाजसुधारक आणि या मनापासून
Freq = 6	क्रियाशील समाजसुधारक जयोस्तुते' हे पूर्णब्रह्म!
Freq = 8	क्रियाशील समाजसुधारक श्री उद्धव ठाकरेजी
Freq = 15	क्रियाशील समाजसुधारक स्वातंत्र्यवीर विनायक दामोदर
Freq = 1	क्रियाशील समाजसुधारक श्री यशवंतराव मोहिते
Freq = 4	क्रियाशील समाजसुधारक जयोस्तुते' हे दिले
Freq = 68	क्रियाशील समाजसुधारक गोपाळ गणेश आगरकर
Freq = 1	क्रियाशील समाजसुधारक श्री तुकाराम बीज
Freq = 8	क्रियाशील समाजसुधारक पद्मभूषण बाबा आढाव
Freq = 2	क्रियाशील समाजसुधारक आणि प्रकरण दाबण्याचा
Freq = 6	क्रियाशील समाजसुधारक स्त्री शिक्षणाच्या प्रणेत्या
Freq = 10	क्रियाशील समाजसुधारक ग्रामसुधारणेचे जनक दूरसंचार
Freq = 36	क्रियाशील समाजसुधारक पद्मभूषण कर्मवीर भाऊराव
Freq = 3	क्रियाशील समाजसुधारक आणि 'धनत्रयोदशी'च्या हार्दिक
Freq = 3	क्रियाशील समाजसुधारक शैक्षणिक कर्ज फेडता
Freq = 3	क्रियाशील समाजसुधारक जयोस्तुते' हे दिव्यांग
Freq = 5	क्रियाशील समाजसुधारक लोकशाहीर अण्णा हवे
Freq = 6	क्रियाशील समाजसुधारक जयोस्तुते' हे जीवन
Freq = 11	क्रियाशील समाजसुधारक जयोस्तुते' हे जीवन

Freq = 2	क्रियाशील समाजसुधारक जयास्तुते' हे सरकारच्या
Freq = 1	क्रियाशील समाजसुधारक आणि सर्वाना विनंती
Freq = 1	क्रियाशील समाजसुधारक जयोस्तुते' हे सर्वच
Freq = 2	क्रियाशील समाजसुधारक स्त्री शिक्षण झालं
Freq = 4	क्रियाशील समाजसुधारक श्री २०२० च्या
Freq = 1	क्रियाशील समाजसुधारक आणि पक्षाचे म्हणजेच
Freq = 1	क्रियाशील समाजसुधारक जयोस्तुते' हे कार्य
Freq = 2	क्रियाशील समाजसुधारक आपल्या सर्वांचं दैवत
Freq = 1	क्रियाशील समाजसुधारक श्री संत शिरोमणी
Freq = 1	क्रियाशील समाजसुधारक आणि आरोग्य विभाग
Freq = 1	क्रियाशील समाजसुधारक आणि प्रत्यक्ष कृती
Freq = 3	क्रियाशील समाजसुधारक जयोस्तुते' हे शक्य
Freq = 1	क्रियाशील समाजसुधारक आपल्या अभंगांच्या माध्यमातून
Freq = 1	क्रियाशील समाजसुधारक श्री संत गाडगेबाबा
Freq = 2	क्रियाशील समाजसुधारक जयोस्तुते' हे राष्ट्रवादीच्या
Freq = 2	क्रियाशील समाजसुधारक आणि कृषि विद्यापीठे
Freq = 1	क्रियाशील समाजसुधारक आणि आयोजन हा
Freq = 1	क्रियाशील समाजसुधारक आणि प्रतिमेला मनःपूर्वक
Freq = 2	क्रियाशील समाजसुधारक आणि प्रसारमाध्यम संस्था
Freq = 3	क्रियाशील समाजसुधारक आणि कोरेनाच्या संकटकाळात
Freq = 1	क्रियाशील समाजसुधारक आणि प्रशिक्षक मिळणे
Freq = 1	क्रियाशील समाजसुधारक आणि यापुढेही स्वावलंबी
Freq = 3	क्रियाशील समाजसुधारक जयोस्तुते' हे पक्षाचे
Freq = 1	क्रियाशील समाजसुधारक आणि मरणोत्तर भारतरत्न
Freq = 2	क्रियाशील समाजसुधारक आणि सर्वाना हार्दिक
Freq = 1	क्रियाशील समाजसुधारक आणि आपल्याला दिला
Freq = 1	क्रियाशील समाजसुधारक आपल्या जीवनात रंगांचे
Freq = 1	क्रियाशील समाजसुधारक आणि देशाचं संरक्षण
Freq = 1	क्रियाशील समाजसुधारक आणि कोळी बांधवांसाठी
Freq = 3	क्रियाशील समाजसुधारक श्री रामांच्या चरणी
Freq = 1	क्रियाशील समाजसुधारक आणि महाराष्ट्र आणि
Freq = 1	क्रियाशील समाजसुधारक आणि प्रतिमेला कुठेही
Freq = 2	क्रियाशील समाजसुधारक आपल्या कृतीतून व
Freq = 2	क्रियाशील समाजसुधारक आणि थेट मदत
Freq = 1	क्रियाशील समाजसुधारक आपल्या भाजपाच्या चंद्रातील
Freq = 1	क्रियाशील समाजसुधारक आणि पर्यावरण विभागाने

3. EX 3

Enter the sentence

माझे दैवत आदरणीय साहेबांच्या हस्ते एवढ्या लहान वयात

Enter how many words to add

1

Last Word is वयात

Most probable result occurred 1000 times

The best result occurred 1000 among 1000 attempts :

माझे दैवत आदरणीय साहेबांच्या हस्ते एवढ्या लहान वयात कार्यक्षम

4. EX 4

Enter the sentence

सांगोला तालुक्यातील इमडेवाडी लक्ष्मीदहीवडी माण नदीवर

Enter how many words to add

2

Last Word is नदीवर

No matching words found

THANK YOU!!