

Word Prediction using Bi-gram Model

Here's implementation that focuses on interactive word prediction using Bi-gram model, filtered to exclude stopwords using the NLTK stopwords corpus:

1. Initial Setup:

- The user is prompted to enter a word. If the word is not found in the corpus, the user can choose to enter another word or quit.
- The word, once validated, is used as the basis for predictions.

2. Language Model Construction:

- Using the Brown corpus, I build a 2-gram model that excludes all stopwords.
- The model calculates the conditional probability of each word following the initial word, considering only words that appear in the corpus.

3. Interactive Prediction:

- A menu displays the top 3 most likely words to follow the initial word, based on their probability scores.
- The user can continue building a sentence by selecting words from the menu or quit the interaction.

4. Handling User Input:

- The program handles invalid selections by defaulting to the first option.
- The process continues until the user decides to quit or no valid predictions can be made.

5. Final Output:

- Once the user quits, the constructed sentence is displayed.

Result_1:

Please enter a word : Divya

The word 'divya' is not in the corpus.

Choose an option ('a' or 'b'):

- a. Again - To choose a word
- b. Quit - To Quit

Case1:

a

Please enter a word : GOOD

The word 'good' is in the corpus and so proceeding further...

good ...

Which word should follow : 'good'

- 1) . $P(\text{good } .) = 0.07444$
- 2) , $P(\text{good } ,) = 0.05459$
- 3) deal $P(\text{good deal}) = 0.03350$
- 4) QUIT

Select an option (1-4): 2

, ...

Which word should follow : ','

- 1) `` $P(, ``) = 0.02273$
- 2) , $P(, ,) = 0.01301$
- 3) said $P(, \text{said}) = 0.01181$
- 4) QUIT

Select an option (1-4): 3

said ...

Which word should follow : 'said'

1) , $P(\text{said} ,) = 0.27180$

2) . $P(\text{said} .) = 0.22591$

3) : $P(\text{said} :) = 0.02958$

4) QUIT

Select an option (1-4): 3

: ...

Which word should follow : ':'

1) `` $P(: ``) = 0.24568$

2) ($P(: () = 0.02340$

3) 1 $P(: 1) = 0.02340$

4) QUIT

Select an option (1-4): 4

Quitting...

Constructed sentence: good , said :

Case2: if entered 'b'

Quitting the program.

Result_2:

Please enter a word : DIVYA

The word 'divya' is not in the corpus.

Choose an option ('a' or 'b'):

- a. Again - To choose a word
- b. Quit - To Quit

A

Please enter a word : BAD

The word 'bad' is in the corpus and so proceeding further...

bad ...

Which word should follow : 'bad'

- 1) . $P(\text{bad } .) = 0.09859$
- 2) , $P(\text{bad } ,) = 0.04930$
- 3) luck $P(\text{bad luck}) = 0.03521$
- 4) QUIT

Select an option (1-4): 2

, ...

Which word should follow :','

- 1) `` P(, ``) = 0.02273
- 2) , P(, ,) = 0.01301
- 3) said P(, said) = 0.01181
- 4) QUIT

Select an option (1-4): 5

Invalid choice, defaulting to the first option.

`` ...

Which word should follow :'``'

- 1) " P(`` ") = 0.02150
- 2) , P(`` ,) = 0.01403
- 3) know P(`` know) = 0.01222
- 4) QUIT

Select an option (1-4): 3

know ...

Which word should follow :'know'

- 1) , P(know ,) = 0.10395
- 2) . P(know .) = 0.09956
- 3) " P(know ") = 0.07321
- 4) QUIT

Select an option (1-4): 2

No more likely words to follow '.', Do you want to enter another word and continue... or Quit.

Choose an option ('a' or 'b'):

a. Again - To choose a word

b. Quit - To Quit

B

Constructed sentence: bad , `` know .