2019SIY7580

AMAN BHARDWAJ

Assignment1 Part 2

Report: Cryptanalysis of Hill Cipher

User Inputs:

- **Cipher Text:** Read from **encrypted_text.txt** file
- **Key Length: N^2** for key matrix (N*N). **(Eg. for 2x2 user enters 4)** //Could be 1, 4, 9.

Pre-defined Most Frequent N-Grams lists in English:

- **Monograms =** ["e","t","a","o","i","n","s","h","r","d"] //Used for Key of Size (1*1) N = 1
- **Bi-Grams =** ["th","he","in","er","an","re","on","at","en","nd","ti","es","or","te","of","ed","is", "it","al","ar","st"] //Used for Key of Size (2*2) N = 2
- **Tri-Grams =** ["the","and","tha","ent","ent","ing","ion","tio","for","nde","has"] //for Key (3*3) N = 3

Algorithm Implementation: (**Note**: To read algorithm quickly just read the **highlighted headings (Bold)** of following steps)

1. **Read Cipher Text:** Remove spaces, special chars, numbers, "\n" if any.
2. **Validation:** If encrypted_text.txt file empty or key length is not a perfect square then validate and throw error. If everything is fine, the let it proceed further.
3. **Extract top N-Grams from Cipher Text:** For N*N size key split Cipher Text into N-Grams and find the frequency of each distinct N-gram and get a list of top 5 N-Grams based on number of occurrences in the cipher text. **Eg**. **Cipher Text** = "abcdef". Corresponding **Tri-grams** = ["abc", "bcd", "cde", "def"]
4. **Set Counter for number of trials trails_counter = 0**
5. **IF (**trails_counter <= Max_Counter**):**

   Generate P and C matrix of size (N*N):

   These will be used in formula **Key = (C * P^{-1}) mod 26** //Dot Product of C and P $^{-1}$

   P = Plain Text Matrix from top English N-Grams (Pre-defined lists)

   C = Cipher Text Matrix from top Cipher Text N-Gram (Calculated in step 3)

   For above N*N Matrices each column is from the respective list of top N Grams

   How to choose columns for above matrices?

   - From respective lists of top N-grams, get N random elements. (Note that for key length of N*N the length of each element of the lists will also ne = N). Therefore, we have selected here N elements of length N each.
   - From each element split the gram into chars and those chars will occupy a column. Place characters from each element into one column of the matrix. And convert each character to its respective integer. **Eg. a = 0, b = 1, c =2, … , z = 25**
   - **Eg**. For Key of (2*2) let list of top bigrams = ["ab, "bc", "cd", "df", "fg"]. Choose any two elements at random. Let those two elements = ["bc", "fg"].

     Matrix = $\begin{bmatrix} b & f \\ c & g \end{bmatrix} = \begin{bmatrix} 1 & 5 \\ 2 & 6 \end{bmatrix}$.
   - Compute both **C and P** matrix as above.
6. **Check if P invertible:** P should be an invertible matrix

   If TRUE: Proceed to step 7;

   If False: trails_counter ++; Return to step 5;
7. **Calculate Key:**

   Key = (C * P^{-1}) mod 26

   Check if Key is invertible: (Key should be invertible to decrypt the cipher text**)**

   If TRUE: Proceed to step 8

   If False: trails_counter ++; Return to step 5;
8. **Decrypt Cipher Text:** Operate the key, calculated in the step 7 and decrypt cipher and extract plain text from the same. Formula used **Plain Text = (Cipher_Text * (Key)$^{-1}$) mod 26** //Dot Product

9. **Calculate Index of Coincidence (IC):** Calculate IC for the above plain text.

$$IC = \frac{\sum_{i=1}^{n} f_i(f_i - 1)}{N(N-1)}$$

Where f_i is the frequency count of ith letter in the ciphertext of length N.

- IC English = 0.0686, IC Random ≈ 1/26 = 0.038466

```
IF (IC > 0.062 and IC < 0.069): // IC value of English = 0.0686
        // We have got our Deciphered Key of size (N*N) as plain text IC = English
        Output (Key, Plain Text, IC Value)
ELSE:
        trails_counter ++; Return to Step 5;
```

10. I have kept max no. of trials to 5000 attempts, if we do not get the desired output in this trial. Then run the code multiple times. If the input length of key in wrong, then try for some different length. Following alert will be generated.
    Could not find the key combination in this try OR Key size is not what you thought OR Encryption key might not be invertible.

If we increase the size of key matrix (N*N) then there is very less probability of getting repeated N-Grams for higher values of N. Therefore, the cipher becomes stronger, cryptanalysis of hill cipher becomes very difficult. And Human Intervention may be required for cryptanalyzing.

HOW TO RUN THE CODE

In case you do not get the desired output or face any issue in running the code. Please contact me at following: **Phone**: 9882305248 **Email**: aman.bhardwaj.cse.iitd.ac.in

How to run:

0. **Extract** 2019SIY7580_AMAN_BHARDWAJ_HILL_CIPHER_CRYPTANALYSIS.zip

    This contains two files:

    a. **2019SIY7580_AMAN_BHARDWAJ_HILL_CIPHER_CRYPTANALYSIS.py**

    b. **encrypted_text.txt :** To place the encrypted text for cryptanalysis

1. **Open file**

    a. 2019SIY7580_AMAN_BHARDWAJ_HILL_CIPHER_CRYPTANALYSIS.py and copy all code to Jupyter notebook in one cell. And Run it (Recommended)

    b. OR you can directly run through command prompt. python 2019SIY7580_AMAN_BHARDWAJ_HILL_CIPHER_CRYPTANALYSIS.py (python3 in case version >= 3)

2. **Instructions to Run code:**

    a. Copy and paste the cipher text for cryptanalysis in "encrypted_text.txt" in the same folder and Save it. (It should only contain a-z chars)

    Test Cases with desired output could be found in **EXAMPLES_A1_PART2.txt**

b. Now you run the code, you will be asked to input the name of encrypted text file:

write "encrypted_text.txt" then press ENTER.

c. Next you will be asked to enter Cipher Key Length. (Eg. for 2*2 Matrix enter Key length = 4)

Provide Key Length and press ENTER.

d. You should get deciphered text. and Cracked Key Matrix for hill cipher in the console.

e. in case you get the following message.

"ALERT:

Could not find the key combination in this try

or Key size is not what you thought

or Encryption key might not be invertible.

Try again"

Please run the code a few more times for same configurations. This is because I have capped the number of attempts for cryptanalysis to 5000. if key is not found for given attempts this Alert is generated. So please try a few more times.

NOTE: You will have to run the 2019SIY7580_AMAN_BHARDWAJ_HILL_CIPHER_ENCRYPT_DECRYPT.py again for every crypt analysis

****