

Automated Benevolent Text Evaluation

By:

Bandi Dinesh Kumar

Mrinal Pande

Pushkar Nagpal

Ronak Jain

Saurav Singh

Mentor:

Aritra Saha

(Assistant Professor)

Previous Results

Recap

- First Approach
 - The word was traversed letter by letter and correctness was awarded on substitution and transposition errors.
 - The person was able set the acceptability of the score.
- Second Approach
 - The word was converted into encodings.
 - Then the encodings were traversed number by number and correctness was awarded on substitution and transposition errors. Using naive bayes.

Objective

Targeting one word text.

The objective of the approach is to target one word text converted to encoding in a set of stored dataset.

Why the New Approach?

What is updated in the previous approach and how it is more efficient?

Introduction

What's changed in our new approach.

- Introduction of data set depending on word length.
- More efficient use of Encoding Technique.
- Better method for making the data much accurate via storing.

Since our problem had a very large data set.

Encoding Helps us in making the algorithm much better by reducing the complexity.

Q. How?

A. Complexity of 26^l for words but reduced to 4^l when we use encodings

Where l = length of the word.

Data Set

- The dataset is split into csv files depending on the length of the word.
- This figure consists the words of size 5 into the csv files.
- Storing data set into csv files.

	Standard	Standard	Standard
1	acceptance	encoding	word
2	1	00000	chair
3	1	22000	hcair
4	1	00220	chiar
5	1	00300	cheir
6	1	00330	cheer
7	1	00220	chiar
8	1	00330	chier
9	1	00030	chaer
10	1	00330	chher
11	1	00001	chaie
12	1	00330	choer
13	0	03333	cnwkf

Working

- Question Dataset Creation.
- User Input.
- Encoding generation.
- Using the Training Data for Text Evaluation.
- Acceptability.
- Score generator.

Results

Acceptable
encodings for the
word:

“chair”

Acceptability, Encoding, Word

1,00000,chair

1,22000,hcair

1,00220,chiar

1,00300,cheir

1,00330,cheer

1,00220,chiar

1,00330,chier

1,00030,chaer

1,00330,chher

1,00001,chaie

Unacceptable
Encodings for
the word:

“chair”

Acceptability, Encoding, Word

0,03333,cnwkf

0,11333,vjzjd

0,33333,dnxlg

0,33333,oefms

0,13303,vldig

0,33333,odfmo

0,33133,moslx

0,33333,eorsp

0,33333,mvlsk

0,11333,xjwkf

Scoring for “Chair” Acceptability

Dataset word	Acceptability	Marks Awarded
chair	1	100
hcair	1	89.9
chiar	1	90
cheir	1	80
cheer	1	60
chiar	1	90
chier	1	60
chaer	1	80
chher	1	60
chaie	1	90

Scoring for “Chair” Unacceptability

Dataset word	Acceptability	Marks Awarded
cnwkf	0	20
vjzjd	0	20
dnxlg	0	0
oefms	0	0
vldig	0	30
odfmo	0	0
moslx	0	10
eorsp	0	0
vhprt	0	40
xjwkf	0	20

Demo

- The Q&A.csv file.
- Text evaluation of answers via GNB using the generated training data sets(Csv files).
- Answer Evaluation Scoring.

To view the demo of how the code works please follow the link provided :

<https://asciinema.org/a/108408>

Conclusion and Future Development

Conclusion:

Current Approach helps us in making correct decisions with encodings thus reducing the word complexities.

Using the satisfactory results of the current approach we are aiming to extend this approach to target more number of words for examples a paragraph.

Future Development:

- Aim to find correctness in a passage of text
- Preprocess the file for the keywords using LSA.
- Running our approach on keywords would reduce paragraph complexity.
- Also marks awarded would be from a data bank taking into account the spelling mistakes.

Thank You
