Performance Analysis of N-gram Language Models

FirstName Surname†  
 Department Name  
 Institution/University Name  
 City State Country  
 email@email.com

FirstName Surname  
 Department Name  
 Institution/University Name  
 City State Country  
 [email@email.com](mailto:email@email.com)

FirstName Surname  
 Department Name  
 Institution/University Name  
 City State Country  
 email@email.com

ABSTRACT

This study aims to measure the effectiveness of n-gram language models at predicting the likelihood of a document having valid sentences and phrases. We use a probabilistic model trained on a sample corpus to study different sizes of n-grams, with Laplace smoothing to account for n-grams not present in the training data. INSERT FINDINGS HERE

INTROCUTION

Language modeling

ngram model

applications

DATASET

For this study, we used training and test corpuses from the DUC 2005 Conference with permission.

RESULT ANALYSIS

Maybe a sentence or 2 about results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| Average log-likelihood |  |  |  |  |
| Perplexity |  |  |  |  |
|  |  |  |  |  |

DISCUSSION

comments and remarks about study

REFERENCES

[1] Dan Jurafsky and James H. Martin. 2021. Speech and Language Processing (3rd ed.). From https://web.stanford.edu/~jurafsky/slp3/

[2] DUC 2005. Training and Testing Datasets. NIST. From https://duc.nist.gov/duc2005/tasks.html

Conference Name:ACM Woodstock conference

Conference Short Name:WOODSTOCK’18

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00