IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, April 2022
Impact Factor: 6.252

Predictive Keyboard

Harshad Bhale[1], Yash Chaudhari[2], Gaurav Gaikwad[3], Prof. Sujata Kolhe[4]
Students, Department of Information Technology[1,2,3]
Faculty, Department of Information Technology[4]
Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India

Abstract: *Predicting the most probable word for immediate selection is one of the most high-ticket methods for enhancing the communication experience. With the growth in mobile technologies and the vast spread of the internet, socializing has gotten much easier. People around the world spend further and further time on their mobile affection for dispatch, social networking, banking, and a variety of other conditioning. Due to the fast-paced nature of similar exchanges, it's necessary to save as major as time possible while categorizing. Hence a prophetic textbook exercise is necessary for this. Text prediction is one of the most typically used approaches for adding the rate of communication. Still, the speed at which the textbook is prognosticated is also actually important in this case. The motive of this work is to design and apply a new word predictor algorithm that suggests words that are grammatically more applicable, with a lower burden for the system, and significantly reduces the volume of keystrokes required by users. The predictor uses a probabilistic language model grounded on the methodology of the N-Grams for text prediction.*

Keywords: Predictive Keyboard, Natural Language Processing, Corpus, N-Grams Model.

I. Introduction

Predicting the succeeding word has been an important technique for better communication for more than a decade. Traditional systems used word frequency lists to complete the words already spelled out by the user. However, in the last several years more advanced predictive techniques have been emerged based on the preceding word or syntactic rules. More advanced prediction methods can save keys to advanced rates. several experimenters have plant out that the increased cognitive load associated with word prediction may affect with the quick communication, recent discovery have stated that more accurate predictions can compensate more as compared to these loads [1]. The benefits of increased accuracy of prediction accuracy cannot be limited to keystrokes saved by the predictions. An effective word prediction model can enhance the standard alongside quantity of text generated for persons with language impairments, and people with learning disabilities (2). Word prediction approaches also can be utilized in order to separate key pad sequences, correct typing errors and supply more correct scanning interface features forecasts (3). The prediction of the letter sequences was analyzed by Shannone (1951). He found out that written English has high level of repetitions. Based upon this research an clear question was whether users can be supported by systems which forecast the next keystrokes, words, or phrases while writing text [4]. A Unix shell was developed by Matodana Yoshida [5] which predicts the command which a user is possible to use based upon this history of previously entered commands.

II. Literature Review

The preceding n-1 word is employed in estimating the present (nth) term within the N-Grams word prediction methods. This has been increased by Korvemaker and Greiner [6] which would predict the entire command lines. Scanning and selecting skilled authors from various displayed options may, in contrast, slow up (Langlais et al. 2002; Magnuson and Hunnicutt 2002). These tools provide a possible list of words from which the user could select the required or most approximate word. For these users it is usually more efficient to scan and choose from lists of proposed words than to type. Scanning and selecting capable authors from several displayed options may, in contrast, slow up (Langlais et al. 2002; Magnuson and Hunnicutt 2002). A prediction system can make more suitable word choices for the user by exploiting the present sentence context using analytical techniques. The previous n-1 word is employed in predicting the present (nth) term within the N-Grams word prediction technique. In a large corpus, known as the training text, the N-Grams data is collected by counting each single n-word-sequence. In case of increasing communications usage N-Grams techniques were limited to unigram and bigram word prediction, but in many other areas associated with nlp like speech recognition and MT trigram

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, April 2022
Impact Factor: 6.252

and higher N-Grams orders were often used [10]. There are such numbers of linguistically valid n-word sequences for N-Grams-order orders above unigrams, that certain sequences don't appear or occur to supply statistically meaningful data even in extensive training texts.

III. Problem Statement

The main problem addressed is predicting the next appropriate word in a conversation. To achieve this, the model has been trained using various sources of data. The sources of data include various blogs, news and tweets, the detailed description of the same. The challenge is to predict the words with minimum waiting time. An experiment designed to investigate three design issues for predictive keyboards. Do the types of text that users might reasonably input with a predictive keyboard come from the same or different text populations as measured by the frequency with which users would need to touch an "Other" button when inputting sets of test texts? Does displaying eight keys rather than six keys substantially increase the likelihood that the desired next key will be immediately available for typing in other words, reduces the frequency.

IV. Proposed Methodology

Predictive Keyboard provides the capability to autocomplete words and suggests predictions for the succeeding word. This makes typing quick, more intelligent and reduces trouble. The execution involves using a large corpus. The methodologies used by us are as follows:

4.1 N-Grams Model

Probabilistic methodologies are used for computing the probability of an entire sentence or for giving a probabilistic prediction of what the succeeding word will be in a series. This methodologies involves looking at the qualified probability of a term given the preceding words. If we review each word occurring in its correct location as an independent event, we might represent this probability as follows:

$$P(w_1, w_2 \ldots, w_{n-1}, w_n)$$

Chain rule of probability is used to decompose this probability:

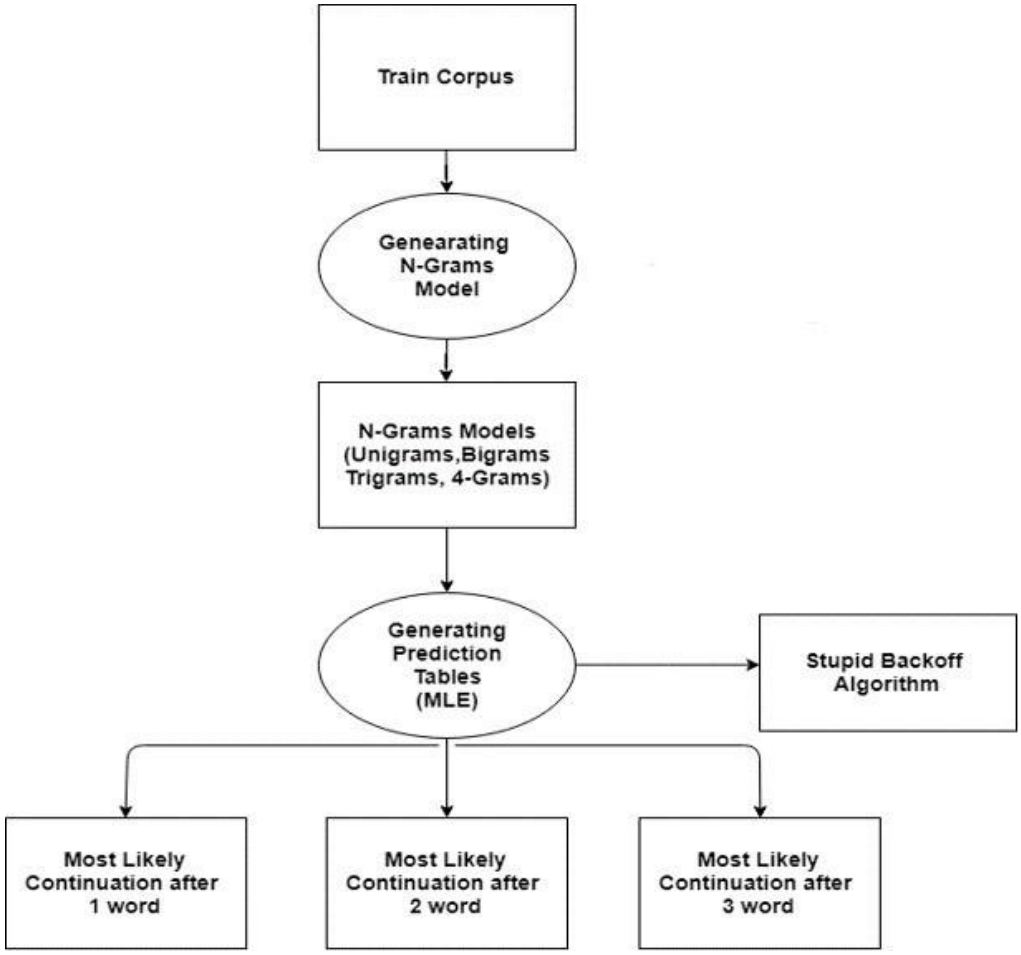
$$\begin{aligned} P(w_1^n) &= P(w_1)P(w_2|w_1)P(w_3|w_1^2)\ldots P(w_n|w_1^{n-1}) \\ &= \prod_{k=1}^{n} P(w_k|w_1^{k-1}) \end{aligned}$$

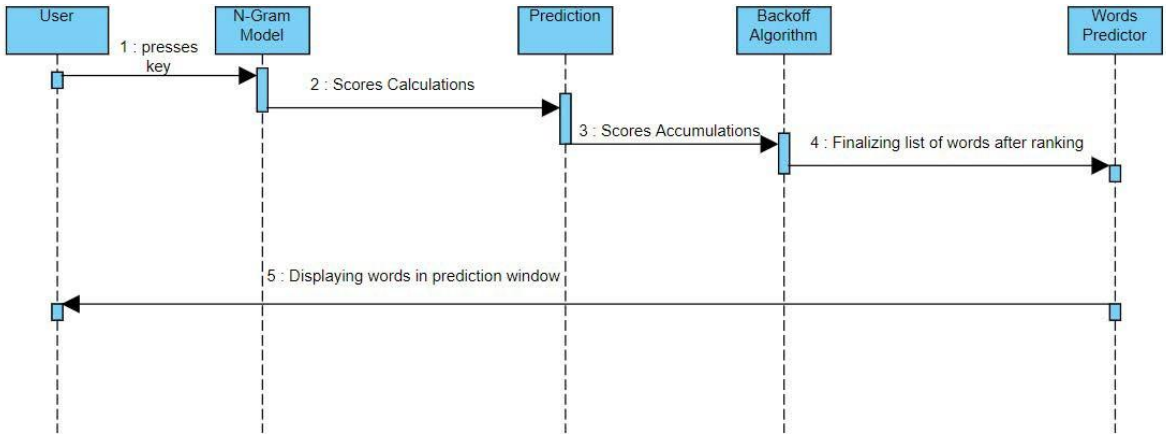
- Bigram Model: bigram model is used to approximate the probability of a word given all the previous words by the conditional probability of the preceding word.
- Trigram: A trigram model looks just the same as a bigram model, except that we condition on the two-previous words.

4.2 Corpus

A corpus is a library of authentic text or audio organized into datasets. Authentic then means text written or audio vocalized by a native of the language or dialect. A corpus can be made up of everything from newspapers, novels, fashions, radio broadcasts to TV shows, pictures, and tweets.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, April 2022
Impact Factor: 6.252





V. Conclusion and Future Scope

This paper shows that N-Grams being a fairly simple language model can prove to be efficient for real time prediction conditions. The perfection of the model can be seen perfecting as the value of N-Grams increases. Consider further improvements setting up a 5-grams model; avoiding pruning when while training the models; or applying more significant smoothing algorithms than 'stupid backoff & can be applied for increase the precision of the model. However, as the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, April 2022
Impact Factor: 6.252

size of model increase the prediction time would also escalate. Hence the tradeoff between precision and speed needs to be considered while enhancing the model.

ACKNOWLEDGEMENT

Our project, which we have submitted is result of our dedicated team effort and also the hard work done by each of us. We express our sincere thanks to respected "Prof. Dr. Sujata Kolhe" for her guidance, patience, insightful comments, helpful information, practical advice and unceasing ideas that have helped us tremendously at all times as well as for sharing her expertise whilst giving us the space to work in our own way.

REFERENCES

[1]. Horstmann Koester, H. & Levine, S. (1996). Effect of a word prediction feature on user performance. Augmentative and Alternative Communication, 12, 155-168.

[2]. Morris, C., Newell, A., Booth, L., Ricketts, I., & Arnott, J. (1992). Syntax PAL: A system to improve the written syntax of languageimpaired users. Assistive Technology, 4, 51-59.

[3]. Nagalavi D, Hanumanthappa M (2016) N-gram Word prediction language models to identify the sequence of article blocks in English e-newspapers. IEEE. https://doi.org/10.1109/CSITSS.2016.7779376

[4]. Kumar P (2018) An Introduction to N-grams. https://blog.xrds.acm.org/2017/10/introduction-n-grams-need. Accessed 26 June 2019

[5]. Hernandez D, Calvo H (2014) CoNLL 2014 shared task: grammatical error correction with a syntactic N-gram language model from a big corpora. In: Proceedings of the eighteenth conference on computational natural language learning: shared task. CoNLL-53-59.