An evaluation study of new word generation

[1]Vivek Chandrakar, [2]Virendra Kumar Swarnkar

[1]MTech Scholler, [2]Assistent Professor
Computer Science & Engineering
Bharti College of Engineering & technology, Durg

Abstract:* *Researchers usually assess word prediction victimization keystroke savings, however, this live isn't easy. we have a tendency to gift many complications in computing keystroke savings which can have an effect on interpretation and comparison of results. we have a tendency to address this drawback by developing 2 gold standards as a frame for interpretation. These gold standards live the most keystroke savings un-der 2 totally different approximations of a perfect language model. The gold standards addition-ally slim the scope of deficiencies in a very word prediction system.

Index Terms - word prediction, next word prediction, word generation

---

I. Introduction

Word prediction is AN application of language modeling to rushing up text entry, particularly to coming into utterances to be spoken by AN Augmentative and different Communication (AAC) device. AAC de-vices look for to handle the twin drawback of speech and motor impairment by trying to optimize text in-put. Even still, communication rates with AAC de-vices square measure usually below ten words per minute (Newell et al., 1998), compared to the common 130-200 words per minute speech rate of speaking individuals. Word prediction addresses these problems by reducing the quantity of keystrokes needed to provide a message, that has been shown to boost communication rate (Trnka et al., 2007). The reduction in keystrokes conjointly interprets into a lower degree of fatigue from typewriting all day (Carlberger et al., 1997).

Word prediction systems gift multiple completions of the present word to the user. Systems generate a listing of W predictions on the idea of the word being typewritten and a language model. The vocabulary is filtered to match the prefix of the present word and also the language model ranks the words ac-cording to their probability. within the case that no letters of the present word are entered, the language model is that the sole think about generating predictions. Systems usually use a touchscreen or function/number keys to pick any of the anticipated words.

Because the goal of word prediction systems is to scale back the quantity of keystrokes, the first analysis for word prediction is keystroke savings (Garay-Vitoria and Abascal, 2006; Newell et al., 1998; Li and Hirst, 2005; Trnka and McCoy, 2007; Carlberger et al., 1997). Keystroke savings (KS) measures the proportion reduction in keys ironed compared to letter-by-letter text entry.

$$KS = \frac{keys_{normal} - keys_{with\,prediction}}{keys_{normal}} \times 100\%$$

word prediction system that provides higher savings can profit a user a lot of in follow.

However, the equation for keystroke savings has 2 major deficiencies. Firstly, the equation alone isn't enough to reckon keystroke savings — actually computing keystroke savings needs an explicit definition of a keystroke and conjointly needs a technique for determinative what percentage keystrokes square measure used once predictions square measure out there, mentioned in Section a pair of. Be-yond merely computing keystroke savings, the equation alone doesn't give a lot of within the approach of interpretation — is her keystroke savings good? will we have a tendency to do higher? Section three can gift 2 gold standards to permit better interpretation of keystroke savings.

Computing Keystroke Savings

We should have how to work out what percentage keystrokes a user would take underneath each letter-by-letter entry and word prediction to reckon keystroke savings. The common trend in analysis is to simulate a “perfect” user which will ne'er build typewriting mistakes and can choose a word from the pre-dictions as shortly because it seems.

Implementation of excellent utilization of the pre-dictions isn't perpetually easy. For example, contemplate the prognostic interface in Microsoft WordTM: one prediction is obtainable as AN inline completion. If the prediction is chosen, the user could type and edit the word. However, this freedom makes finding the minimum sequence of keys harder — currently the user could choose a prediction with the inaccurate suffix and proper the suffix because the optimum action. we have a tendency to feel that a lot of intuitive interface would permit a user to undo the pre-diction choice by pressing type, AN interface that doesn't support backspace-editing. In addition to backspacing, future analysis in multi-word prediction can face an analogous drawback, analogous to the garden-path drawback in parsing, wherever a greedy approach doesn't perpetually offer the optimum result.

The keystrokes used for coaching and testing word prediction systems will have an effect on the results. we have a tendency to at-tempt to gauge word prediction as realistically as attainable. Firstly, several corpora have punctuation marks, however AN AAC user in a very colloquial setting is unlikely to use punctuation thanks to the high price of every key press. Therefore, we have a tendency to take away punctuation on the surface of words, like commas and periods, however leave word-internal punctuation intact. Also, we have a tendency to treat capital letters as one key press, reflective the trend of the many AAC users to avoid capitalization. Another drawback happens for a newline or “speak key”, that the user would press once completing AN vocalization. In pilot studies, together with the simulation of a speak key lowered keystroke savings by zero.8–1.0% for window sizes 1–10, as a result of new-lines don't seem to be able to be expected within the system. How-ever, we have a tendency to feel that the simulation of a speak key can manufacture AN analysis metric that's nearer to the actual user's expertise, so we have a tendency to embody a speak key in our evaluations.

An analysis of word prediction should address these problems, if solely implicitly. The result of those doubtless implicit choices on keystroke savings will build comparison of results tough. However, if results square measure bestowed in relation to a gold standard underneath an equivalent assumption, we will draw a lot of reliable conclusions from results.

TOWARDS A GOLD STANDARD

In attempting to boost the state of word prediction, many researchers have noted that it appears extremely tough to boost keystroke savings be-yond a precise purpose. Copestake (1997) mentioned the entropy of English to conclude that 50–60% keystroke savings is also the foremost we will expect in follow. Lesher et al. (2002) replaced the language model in a very word prediction system with a person's to undertake and estimate the limit of keystroke savings. They found that humans may accomplish fifty-nine keystroke savings with access to their advanced language model which their advanced language model alone achieved fifty-four keystroke savings. They noted that one subject achieved nearly seventieth keystroke savings on one explicit text, and concluded that any enhancements on current methods square measure attainable. Garay-Vitoria and Abascal (2006) survey several prediction systems, showing a good spectrum of savings, however no system offers quite seventieth keystroke savings.

We investigated the matter of the constraints of keystroke savings initial from a theoretical perspective, seeking a clearly outlined higher boundary. Keystroke savings will ne'er reach 100% — it might mean that the system divined the complete text they in-tended while not one key.

Theoretical keystroke savings limit

The minimum quantity of input needed corresponds to an ideal system — one that predicts each word as shortly as attainable. in a very word completion system, the predictions square measure delayed till once the primary character of the word is entered. In such a system, the minimum quantity of input employing an excellent language model is 2 keystrokes per word — one for the primary letter and one to pick the prediction. The system would conjointly need one keystroke per sentence. in a very word prediction system, the predictions square measure out there like a shot, therefore the minimal input for an ideal system is one keystroke per word (to choose the prediction) and one keystroke per sentence. we have a tendency to accessorial the flexibility to live the minimum range of keystrokes and most savings to our simulation software package, that we have a tendency to decision the theoretical keystroke savings limit.

We evaluated a baseline written word model underneath 2 conditions with totally different keystroke needs on the plugboard corpus. The simulation software package was changed to output the theoretical limit in addition to actual keystroke savings at numerous window sizes. To demonstrate the result of the theoretical keystroke savings limit on actual savings, we have a tendency to evaluated the written word model underneath conditions with 2 totally different limits — word prediction and word completion. The analysis of the written word model victimisation word completion is shown in Figure one. the particular keystroke savings is graphed by window size in reference to the theoretical limit. As noted by different re-searchers, keystroke savings will increase with window size, however with decreasing returns (this is that the result of putting the foremost probable words first). One of

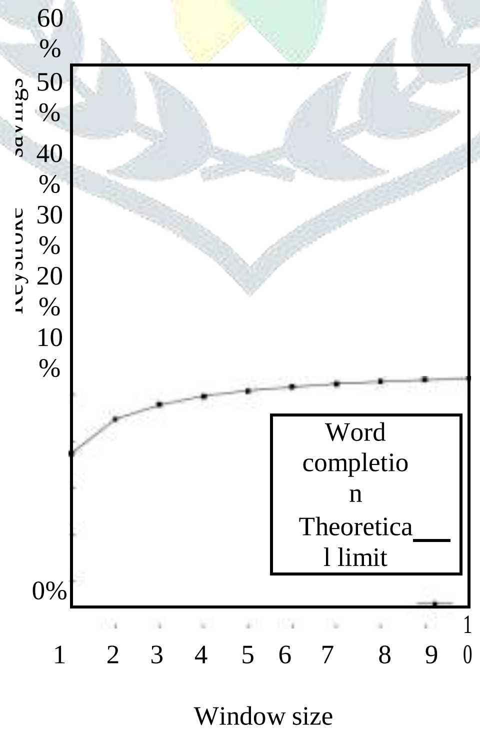

Figure 1: Keystroke savings and the limit vs. window size for word completion.

the problems with word completion is that the theoretical limit is therefore on the point of actual performance — around fifty eight.5% keystroke savings compared to fifty.8% keystroke savings with 5 predictions. At solely 5 predictions, the system has already completed eighty-seven of the attainable keystroke savings. underneath these circum-stances, it might take a forceful modification within the language model to impact keystroke savings.

We recurrent this analysis for word prediction, shown in Figure a pair of aboard word completion. Word prediction is far more than completion, each theoretically (the limit) and in actual keystroke savings.

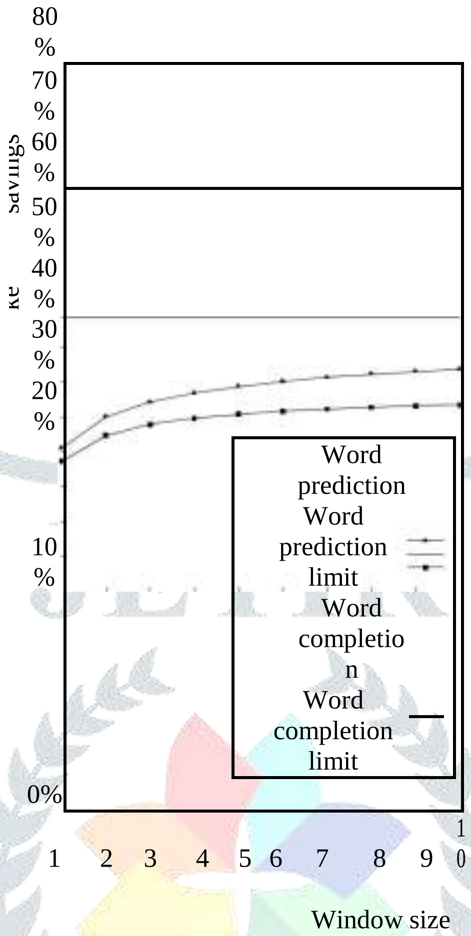

Figure 2: Keystroke savings and the limit vs. window size for word prediction compared to word completion.

Word prediction offers far more headroom in terms of enhancements in keystroke savings. There-fore our current analysis can concentrate on word pre-diction over word completion.

This analysis demonstrates a limit to keystroke savings, however this limit is slightly totally different than Copestake (1997) and Lesher et al. (2002) look for to explain — on the far side the constraints of the user interface, there appears to be a limitation on the predictability of English. Ideally, we'd prefer to have a gold normal that's a better estimate of a perfect language model.

Vocabulary limit

We can derive a lot of sensible limit by simulating word prediction employing a excellent model of all words that occur within the coaching knowledge. This gold normal can predict the proper word like a shot see you later because it happens within the coaching corpus. Words that ne'er occurred in coaching need letter-by-letter entry. we have a tendency to decision this live the vocabulary limit and apply it to gauge whether or not the distinction between coaching and testing vocabulary is critical. Previous analysis has centered on the proportion of out-of-vocabulary (OOV) terms to elucidate changes in keystroke savings (Trnka and McCoy, 2007; Wandmacher and Antoine, 2006). In distinction, the vocabulary limit offers a lot of steering for analysis by translating the matter of OOVs into keystroke savings.

Expanding the results from the theoretical limit, the vocabulary limit is seventy-seven.6% savings, compared to seventy-eight.4% savings for the theoretical limit and fifty eight.7% actual keystroke savings with five predictions. the sensible limit is incredibly on the point of the theoretical limit in the case of plugboard. Therefore, the remaining gap between the sensible limit and actual performance should ensue to different variations between testing and coaching knowledge, limitations of the model, and limitations of language modeling.

Application to corpus studies

We applied the gold standards to our corpus study, during which a written word model was severally trained and tested on many totally different corpora (Trnka and Mc-Coy, 2007). In distinction to the particular written word model

Corpus	Trigram	Vocab. limit	Theor. limit
AAC Email	48.92%	61.94%	84.83%
Callhome	43.76%	54.62%	81.38%
Charlotte	48.30%	65.69%	83.74%
SBCSAE	42.30%	60.81%	79.86%
Micase	49.00%	69.18%	84.08%
Switchboard	60.35%	80.33%	82.57%
Slate	53.13%	81.61%	85.88%

Table 1: A trigram model compared to the limits.

performance, the theoretical limits all fall at intervals a comparatively slim vary, suggesting that the achievable keystroke savings is also similar even across totally different domains. The a lot of technical and formal corpora (Micase, Slate, AAC) show higher limits, because the theoretical limit is predicated on the length of words and sentences in every corpus. the sensible limit exhibits a lot of larger variation. in contrast to the Switch-board analysis, several different corpora have a substantial gap between the theoretical and sensible limits. though the sensible live appears to match the particular savings equally to OOVs testing with cross-validation (Trnka and McCoy, 2007), this live a lot of concretely illustrates the result of OOVs on actual keystroke savings — hr keystroke savings once coaching and testing on AAC Email would be extraordinary.

CONCLUSIONS

Although keystroke savings is that the predominant evaluation for word prediction, this analysis isn't easy, aggravating the matter of interpreting and scrutiny results. we've bestowed a completely unique resolution — deciphering results aboard gold standards that capture the problem of the analysis. These gold standards also are applicable to drive future analysis — if actual performance is incredibly on the point of the theoretical limit, then restful the minimum keystroke needs ought to be the foremost useful (e.g., multi-word prediction). Similarly, if actual performance is incredibly on the point of the vocabulary limit, then the vocabulary of the language model should be improved (e.g., cache modeling, adding general coaching data). within the case that keystroke savings is much from either limit, then analysis into up the language model is probably going to be the foremost useful.

REFERENCES

[1] Alice Carlberger, John Carlberger, Tina Magnuson, M. Sharon Hunnicutt, Sira Palazuelos-Cagigas, and Santiago Aguilera Navarro. 1997. Profet, a new gen-eration of word prediction: An evaluation study. In ACL-97 workshop on Natural Language Processing for Communication Aids.

[2] Ann Copestake. 1997. Augmented and alternative NLP techniques for augmentative and alternative commu-nication. In ACL-97 workshop on Natural Language Processing for Communication Aids, pages 37–42.

[3] Nestor Garay-Vitoria and Julio Abascal. 2006. Text pre-diction systems: a survey. Univ Access Inf Soc, 4:183– 203.

[4] Gregory W. Lesher, Bryan J. Moulton, D Jeffery Higgin-botham, and Brenna Alsofrom. 2002. Limits of hu-man word prediction performance. In CSUN.

[5] Jianhua Li and Graeme Hirst. 2005. Semantic knowl-edge in word completion. In ASSETS, pages 121–128.

[6] .Alan Newell, Stefan Langer, and Marianne Hickey. 1998.The roleˆ of natural language processing in alternative and augmentative communication. Natural Language Engineering, 4(1):1–16.

[7] Keith Trnka and Kathleen F. McCoy. 2007. Corpus Stud-ies in Word Prediction. In ASSETS, pages 195–202.

[8] Keith Trnka, Debra Yarrington, John McCaw, Kathleen F.McCoy, and Christopher Pennington. 2007. The Ef-fects of Word Prediction on Communication Rate for AAC. In NAACL-HLT; Companion Volume: Short Pa-pers, pages 173–176.

[9] Tonio Wandmacher and Jean-Yves Antoine. 2006. Training Language Models without Appropriate Lan-guage Resources: Experiments with an AAC System for Disabled People. In Eurospeech.