

QUANTITATIVE EVALUATION OF A PORTUGUESE WORD PREDICTOR

Luis Garcia,

*School of Technology and Management of Beja, Rua Afonso III, 1, BEJA
7800, PORTUGAL*

Luís de Oliveira,
INESC-ID/IST

Keywords

Word Prediction, Rate Enhancement, Quantitative Evaluation

1 Abstract

This paper reports the results of a first evaluation of a Portuguese statistical word predictor. For a quantitative evaluation we developed a simulator that acts as an ideal user of the system. Simulations using texts from three distinct sources presented similar values of efficacy (e.g. 46%, 47% and 50% keystroke saving rate). We also studied the effect of changing the weight of each statistical factor. In general the introduction of each factor improved the system efficacy, however the variation of each factor's coefficient did not produced significant changes to that efficacy.

2 Introduction

We developed a complete augmentative and alternative communication system for the European Portuguese, we called "Eugénio – O Génio das Palavras" (<http://www.l2f.inesc.pt/~lco/eugenio>). In order to reduce the user's physical effort and to improve the writing speed it has a statistical word predictor. Statistical predictors combine different statistical factors (e.g. word probability) to find a set of high probable words that are further presented to the user. Table 1 shows the statistical factors used by "Eugénio". The efficacy of these systems under the context of the Portuguese language has not yet been measured. With long-term studies in progress we decided to develop a simulator that acts as an ideal user of the system to measure our word predictor efficacy and the influence of each factor used in the prediction.

3 Evaluation

Simulations were carried out with three compilations of texts from three distinct sources: (1) school texts from a young student with special needs; (2) email texts from an adult; (3) articles from the "Diário do Minho" newspaper. We started studying the influence of each statistical factor in the system efficacy. In these first simulations we introduced each factor abruptly, setting its coefficient from 0% to 100%. It was used a list with 5 words. Table 1 shows the different sets of factors used along the simulations. Graph of the Figure 1 shows keystroke savings for these simulations.

	SF1	SF2	SF3	SF4	SF5
$p(w)$	•	•	•	•	•
$p(w_n/w_{n-1})$		•	•	•	•
$p(pos_n/pos_{n-1})$			•		
$p_i(w)$				•	•
$p_i(w_n/w_{n-1})$					•

Table 1 - Sets of factors used along the simulations.

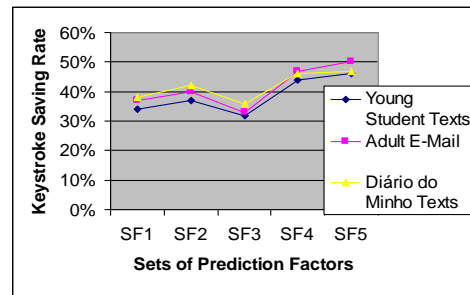


Figure 1 - Keystroke savings along the simulations where each factor was introduced abruptly (from 0% to 100%).

We can verify that apart the part-of-speech conditional probability the introduction of all other factors brought improvements to the keystroke saving rate. We think this results form not having assigned part-of-speech tags to all the words. With sets of factors SF1, SF2 and SF3 the keystroke saving rate was always higher with “Diário do Minho” texts. This is a consequence of the similarity between these texts and the ones used in the lexicon construction. However, with the adoption of user adaptation factors (set of factors SF4 and SF5) the second compilation of texts got higher keystroke saving rate. This result shows the word predictor successfully adapted to the user writing style. For the three texts and set of factors SF5 the simulator reported 1.5, 1.5 and 1.7 characters needed in the successful predictions. Under the same conditions the rate of successful predictions was respectively 31%, 32% and 28%. With SF5 we got the best keystroke saving rates, 46%, 50% and 47%, respectively. Along all the simulations the average position of the successfully predicted words was around 1.5. These values are in agreement with values reported about systems for other languages (Higginbotham, 1992).

To study the gradual introduction of each factor in the prediction process we developed new simulations changing smoothly the importance of each factor from 0% up to 100%. After studying each factor, it remained present in the prediction process but with a weight coefficient of 50%. We verified that this variation did not introduced substantial changes in the simulator performance measures. These results show that the variation of the factors does not introduce substantial changes in the system efficacy.

4 Conclusions and Future Work

Our word predictor presented performance measures similar to systems of other languages (e.g. keystroke savings of 50%). We verified that the introduction of some factors in the prediction process has positively influenced the system efficacy, however changing the weight of each factor did not introduced significant changes to the performance measures. We think factors weighting and its optimization deserves further study.

5 References

HIGGINBOTHAM, D. Jeffery. Evaluation of Keystroke Savings across Five Assistive Communication Technologies. AAC Augmentative Alternative Communication. 1992.