

Application of Artificial Intelligence Methods in a Word-Prediction Aid

Nestor Garay-Vitoria and Julio González-Abascal

Laboratory of Human-Computer Interaction for Special Needs
Informatika Fakultatea, Euskal Herriko Unibertsitatea
649 postakutxa; 20080 Donostia. SPAIN
E-mail: acbgavin@si.ehu.es; julio@si.ehu.es
Telephone: +34 43 21 91 81; Fax: +34 43 21 93 06

Abstract. Word-prediction appears to be a good aid to enhance message-composition rate for people with physical disabilities. Usually word-prediction is based on statistical information (mainly on word frequencies). Hit rate can be enhanced by trying to imitate the behaviour of a human interlocutor (who uses syntactic and semantic information). In this paper some new approaches based on Artificial Intelligence methods are presented. Advantages of syntactic and semantic analysis in relation to bare statistical methods are studied. Furthermore, the integration with human-computer interfaces for disabled users is also described.

1 Introduction

In the last few years several devices to assist people with severe speech and motor disabilities in personal communication have been developed in different countries. Even though most of them have supposed an enhancement of disabled people communication possibilities, the speed achieved remains far lower than that of a normal dialogue. The rate of an oral conversation can be estimated at 150-200 words/minute, whilst the writing rate of a user who has some motor problems tends to be much slower (in the order of 2-10 words/minute) [2].

Word prediction is an interesting choice to speed up writing composition. Basically, prediction or anticipation lies in trying to guess what a user is going to 'say' next, just as a human interlocutor normally does. The feasibility of obtaining good results in word prediction is due to the excellent balance existing between the average word size and the reachable hit ratio. This is based, to a great extent, on the large amount of redundancy and the high quantity of information (statistical, morphological, syntactic, semantic...) accompanying words [3, 4].

Depending on the kind of information used to make predictions, different types of predictors can be distinguished. Predictors by frequencies use purely statistical information, whereas syntactical predictors add syntactic information, and semantic predictors operate with semantic categories.

In the next section some general notions about how our predictors operate and the methods we apply for our study are presented. Later, some Natural Language processing techniques and their application in word prediction will be discussed.

2 Revision of the Notions

We have implemented a system which emulates the behaviour of a user to evaluate the performance of distinct type of anticipators: a process which reads words from a 'trial text' and sends them to the predictor character by character emulating the behaviour of a user. The predictor receives what is sent by the user and stores the information relevant for prediction. Then, depending on the strategy, it sends the user the most likely words or, where appropriate, informs the user that it is impossible to make a prediction. The user, depending on the case, will accept or reject the proposals given by the predictor. If a proposal is accepted, the user computes the improvements so far, and the predictor updates the dictionary. The predictor can also update the lexicon if the user produces words that were not included. The functioning of these predictors is shown in Fig. 1

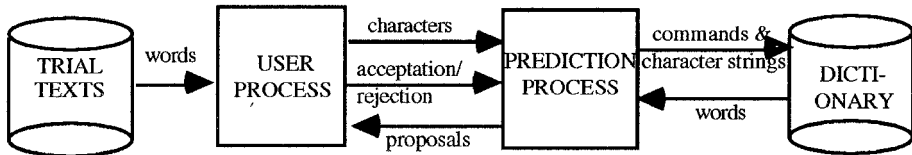


Fig. 1 Predictors functioning

This way data about the performance of each proposed predictor can be obtained and some comparisons working under the same conditions can be made.

In a previous project, two strategies of word-prediction using purely statistical information have been developed. The first method uses word frequencies to make its proposals, taking into account only the current word. Each time the user inputs a character, the predictor proposes the most frequent word (or words) which begins with the last characters typed. The system has to memorise the words last proposed but not accepted in order to avoid repetition. Basically, the entries of the dictionary are composed of the words and their associated frequencies. In order to properly adapt the system to the user the frequencies associated with the used words are updated. This first approach has a great computational speed, but results are not very satisfactory because the context is not taken into account (words are seen as isolated entities within the conversation).

The second approach tries to avoid this problem by placing words into a wider unit: the current sentence. Each word plays a role (specified by the syntactic category) within the sentence. A table was formed with information about the probabilities of each syntactic category following any other category, and also the likelihood of a category starting a sentence. The behaviour of this approach still based on statistical criteria: the category of the preceding word is stored, and the predictor looks into the table for the most probable categories to follow the current one. Using the beginning of the current word, the words stored in the dictionary with related information to their category and frequency are obtained. The predictor makes the proposals, combining adequately the values of the entries of the table and the frequencies of the lexicon. Basically, an entry of the lexicon is composed of the word, its frequency and

its syntactic category. The personalisation occurs by changing the word frequencies in the dictionary and the entries of the table. This approach maintains a satisfying speed of operation. More details about these prediction systems can be found in [4]. Both predictors have been implemented in C language and are part of a human-computer interaction system for disabled people developed in our Laboratory. They are very useful to make comparisons with the newly proposed methods.

3 The Chart Bottom-up Technique

Charts are used in a bottom-up parsing method to analyse sentences in Natural Language [1, 8]. This technique works with a Natural Language grammar defined by rules in the manner:

LEFT \rightarrow [RIGHT]⁺

That is, the left compound constituent is decomposed into one or several constituents which appear on the right of the rule. The bottom-up parser uses the rule to take the sequence of symbols and to match it to the right-hand side of the rule. Then, it identifies them as the left symbol. Matches are always considered from the point of view of one symbol, called the key. To find rules that match a string involving the key, the parser looks for rules that start with the key, or for rules that have already been started by earlier keys and require the present key either to complete the rule or to extend it.

A record of the state of a bottom-up parse is kept in a structure called a *chart*. This structure is a record of the relative positions of the words in the sentence and the new structures derived from the sentence. The chart also stores the rules that have matched previously but are not complete: they constitute the active arcs on the chart.

This method is very adequate for our goals because it is not complex, it adds low computational load and it allows a reasonably good natural language analysis.

4 Syntactic Word-Prediction Using Charts

This technique for syntactic analysis of sentences can also be applied in word-prediction. The goal is to make use of the information provided by the syntactical structure of the sentence. This information allows a more accurate selection of the possible words to be proposed.

The procedure is as follows: each grammatical rule in the chart has an associated weight, which depends on its frequency of use. To weigh each rule a lexicon that contains words associated with their frequency, syntactical category and other morphological marks, if existing, such as gender and number, was previously built.

To completely describe any natural language a large amount of rules must be defined as there are many different kinds of structures and sentences (e.g., active, passive, interrogative, affirmative, negative, imperative, ...). Keeping in mind that our purpose is to obtain a high hit rate within a short time, a complete grammar would be too complex from the computational point of view, and hence too slow [5]. Moreover, this approach seems to be excessive, because some rules add very little information due to their low frequency. For this reason, a partial grammar containing the most representative rules (from the statistical point of view) was defined.

This type of anticipator takes into account syntactical and statistical information. The syntactic information is accessed by using the active arcs on the chart (that is,

the partially completed rules) and the position of the current word in the sentence. Looking at the active arcs pointing to a word, the most probable syntactic categories are obtained, according to the previously defined grammar. The probability of each category depends on the weights of the active arcs and their associated rules.

While the user is writing a word "W", the predictor proposes the most frequent word (or words) belonging to the most probable syntactical category of the current position in the chart and having the same beginning as "W". As more than one arc can be active simultaneously, the statistical weights of the different arcs are crucial to determine the most probable category. Due to the syntactic knowledge obtained, proposals are given with the most appropriate gender and number, depending on the considered rules.

The operational way with the first word in the sentence must be different, because there is not an active arc. In this particular case, an array with the probabilities of the syntactical categories beginning a sentence is used. A combination of the entries in this array (that is, the weights of the categories beginning a sentence) and word frequencies is used to give proposals.

Tailoring of the anticipation method to the individual user is done by updating the word-frequencies in the dictionary and the weights of the rules (or the entries of the array with the probabilities of the syntactic categories in the case of the beginning of a sentence). To include new words in the lexicon all the information needed must be provided either when the new word appears or later. In this last case, the new word is set aside into a special category until in a particular session someone adds the complete information. When new words are accepted in the lexicon some problems arise. The most important one is the appearance of syntactic ambiguities. In our case, this problem can be solved by creating special categories for the ambiguous cases involving more than one single category.

5 Semantic Prediction Using Grammars.

5.1 Antecedents

An interesting approach to semantic prediction is mentioned by Hunnicutt in [7]. A semantic classification of the words is made. This classification can have several levels and which may include some generalisations and inheritances.

A possible implementation of this approach is to build a lexicon as a knowledge base in which each word is interconnected with some others depending on their meaning. This approach allows access to the base by meanings or by clues, such as the number of syllables, intermediate letters, etc. These characteristics are very useful if the user suffers from amnesia, aphasia or anomie [6]. However, the increase of the complexity of this approach can represent a problem. This is a direct consequence of the dictionary's structure and the related working procedures. For this reason the access time can experience a great increase. There is another problem if new words appear, because the words in the base are interrelated. The special sessions to categorise new words will be longer and very complex, because a great portion of the data base will have to be changed.

5.2 Our Approach

We tried a different approach to the semantic prediction, similar to the one designed for syntactic prediction. The difference is that now semantic categories of the words are defined, and a semantic grammar is built. The procedure is the following: with regard to the categories, semantic information related to the theme of the conversation is linked to each word. In relation to the grammar, the syntactic approach is enriched with semantic categories. That procedure narrows the field of feasible words, which should enhance the hit ratio. Thus, we can model the conversation taking into consideration which themes are treated, and in which context they are suitable. The defined grammar covers the proposed modelling of the conversation. Entries of the dictionary are similar to the preceding approach, adding to the syntactic category the semantic one. User adaptation is done in the same way as in the previous approach.

Results for people with limited conversation possibilities may be better than the ones obtained with statistical prediction [4], but they will be close to the results of the syntactic approach, because they present the same behaviour. With this approach a new problem arises. It is a direct consequence of the modelling of the conversation: when new topics are treated, the hit ratio is expected to decrease, because of the lack of semantic information.

A promising possibility for prediction taking into account semantic information is the use of neural networks. In a neural network approach, neurones can reflect the stage of the conversation, and the connections between neurones reflect the weights of the transitions at this stage. A transition happens if the identified word appears. This seems to be the most "natural" approach, because it is similar to the way people make predictions. An advantage of this approach is that, whilst the previous semantic approaches may be constrain-limited, this one can manage non-constrained conversations, and can learn from the experience.

6 Integration of the Predictor in Human-Computer Interfaces

All these predictors are intended to be integrated in some different interfaces designed in our laboratory for severely motor and speech impaired people. These interfaces have a scan based input: matrixes containing different options are scanned sequentially and the user selects the element (character, word, sentence...) pressing one or more buttons when this element is highlighted. Inside these matrixes a place is reserved for the words offered by the anticipator. An important parameter to be tuned up is the number of words that the predictors offered, because the time (and sometimes the number of keystrokes) needed to accept or reject a proposal, depend on the number of proposals received. When this number increases, the hit ratio also increases, but more time is needed to make a selection. So, to obtain the best results a balance between the number of proposals issued each time and the increment in the effort to make a selection should be reached.

If only one proposal is made a new possible procedure is possible, namely explicit rejection. That is, the user must explicitly accept or reject the proposed word. When the hit ratio is very high, this method considerably enhances communication speed, because the selection effort is minimised. Otherwise, continuous rejections can make this procedure very boring. Explicit rejection is not possible when more than one

proposal is made, because the selection of one of the words implicitly supposes its acceptance.

In our study four cases have been tested: predictors which offer one, five and ten proposals with implicit rejection and one proposal with explicit rejection (ER).

7 Comparison of the Results

Two standard Spanish texts, as seen in figure 2, have been used to compare the results of the implemented anticipatory methods. Four prediction methods (word frequencies, syntactic with automaton, syntactic chart and semantic-based chart) are compared. In figure 3 text A and in figure 4 text B are used.

Text	Size (in characters)	origin
A	994	colloquial language
B	2444	colloquial language

Fig. 2 Characteristics of the trial texts

To test the behaviour of the predictors saving percentage is taken as a figure of merit. Saving percentage is the savings in the number of keystrokes obtained when the predictor is used, in comparison to the production of the same text without any prediction method.

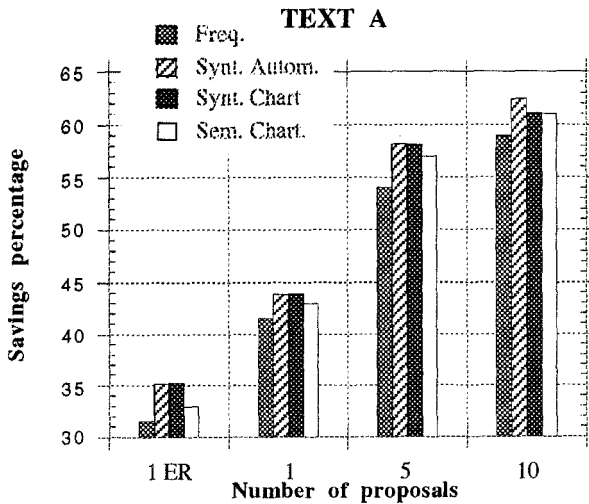


Fig. 3 Results with text A

To make comparisons, the method of prediction by frequencies is taken as a model. This method gives low results, but the needed computational effort is minimal. According to its results the semantic prediction method using grammars follows. Nevertheless, results are worse than expected because the number of the sentences

which do not fit the defined semantic grammar is too large in relation to the number of sentences fitting it.

The other two approaches are very similar, considering only the results. But the needed computational effort is greater in the syntactic prediction using grammars.

In both texts, as can be seen in figs. 3 and 4, the savings percentage increases with the number of proposals. But, when the number of proposals increases, the acceptance protocol has to be more complex and therefore slower. Only in some applications a large set of proposals is useful.

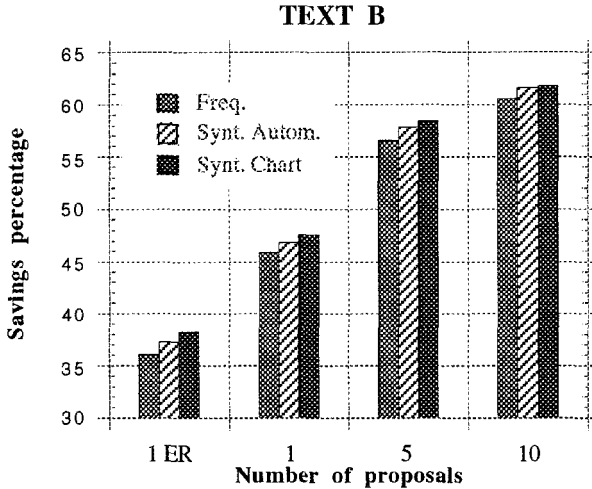


Fig. 4 Results with text B

8 Conclusions

In this paper we have shown two new methods of anticipation based on techniques of Artificial Intelligence, such as the *chart* bottom-up technique (used in the Natural Language Analysis). On the one hand, syntactic approaches appear to offer the best results, compared to the prediction using frequencies that need a low computational effort but obtains worse results. On the other hand, the implemented semantic approach produce results between those of the prediction using frequencies and the syntactic predictions, but is constrained to a narrow model of the conversation. If this model changes, the performance of this method decreases quickly.

Our proposal is to use syntactic prediction using grammars, since this approach works with a reasonable computational cost and produces a high hit ratio, and therefore a reasonable enhancement of the communication speed. Furthermore, some possible solutions for problems detected in this approach have been proposed.

Acknowledgements

This work has received financial support from the Department of Economy of the local government "Gipuzkoako Foru Aldundia". Furthermore, Nestor Garay-Vitoria holds a doctoral grant from the Department of Education, Universities and Research of the Basque Government, to whom the authors would like to acknowledge its support.

References

1. James Allen: Natural Language Understanding. Benjamin Cummings Publishing Company. (1987)
2. Norman Alm, John L. Arnott and Alan F. Newell: Prediction and Conversational Momentum in an Augmentative Communication System. Communications of the ACM. Volume 35, Number 5 (May 1992). Pages 46-57.
3. Arnott, J. L., Pickering, J. A., Swiffin, A. L. and Battison M.: An Adaptive and Predictive Communication Aid for the Disabled that Exploits the Redundancy in Natural Language. Proceedings of the 2nd International Conference on Rehabilitation Engineering (Ottawa, Canada. 1984) pp. 349-350.
4. Nestor Garay and Julio González-Abascal: Using Statistical and Syntactic Information in Word Prediction for Input Speed Enhancement. Information Systems Design and Hypermedia. Proceedings of the BIWIT-94. Biarritz, France (February 1994), pp. 223-230.
5. Craig W. Heckathorne and Dudley S. Childress: Applying Anticipatory Text Selection in a Writing Aid for People with Severe Motor Impairment. IEEE MICRO, June 1983. Pages 17-23.
6. Sheri Hunnicutt: Input and Output Alternatives in Word Prediction. STL-QPSR 2-3/1987. Pages 15-29.
7. Sheri Hunnicutt: Using Syntactic and Semantic Information in a Word Prediction Aid. Proc. Europ. Conf. Speech Commun. Paris, France. September 1989, vol. 1, pp. 191-193.
8. Kepa Sarasola: Natural Language Processing (in Spanish). Notes of lectures Academic Course 1989-1990.