DOMAIN-SPECIFIC WORD PREDICTION FOR AUGMENTATIVE COMMUNICATION

Gregory W. Lesher, Ph.D. and Gerard J. Rinkus, Ph.D.
Enkidu Research, Inc.
247 Pine Hill Road
Spencerport, NY 14559

ABSTRACT

Many augmentative communication systems employ word prediction to help minimize the number of user actions needed to construct messages. Statistical prediction techniques rely upon a database (model) of word frequencies and inter-word correlations derived from a large text corpus. One potential means to improve prediction is to create a set of models derived from domain-specific corpora, dynamically switching to the model most appropriate for the current conversation. Using telephone transcripts to generate prediction models for 20 different topic domains, we have observed a clear benefit to including domain-specific models in an overall prediction scheme.

BACKGROUND

Statistical word prediction systems for augmentative communication commonly utilize both word frequencies and inter-word correlations (word contexts). An *ngram* prediction model utilizes the past *n-1* words to predict the *n*th (current) word. Easily derived from large samples of text, ngram models can provide impressive prediction performance – Lesher (1) reports on a trigram (*n*=3) model derived from a 3 million word corpus that yielded keystroke savings in excess of 54%.

There have been numerous techniques suggested for enhancing traditional ngram word prediction, including recency, syntactic analysis, and syntax-based ngrams. One technique that has not been fully explored is the use of domain-specific ngram models – models derived from text samples that are focused on distinct subjects or genres. In theory, these ngram models could be dynamically swapped in and out of use to match the direction of an ongoing conversation.

The text used to train a word prediction system should match as closely as possible the kind of messages produced by the augmented communicator. Although core vocabulary stays fairly constant (2), fringe vocabulary may change substantially through the course of a day as different topics and settings are encountered. The same is likely to hold true for inter-word correlations. We know of no studies that have attempted to quantify the effect of domain shifts on word prediction efficacy. As a precursor to developing a system that can automatically shift between appropriate domain-specific models, we undertook to find the keystroke savings possible in such a system.

Utilizing transcripts from the Switchboard Corpus, a series of 2,400 telephone conversations organized into approximately 60 topic domains (for example, recycling, food/cooking), we have quantified the performance gains associated with utilizing domain-specific ngram models. Although this corpus does not involve augmented communicators, it is conversational, large, and organized into specific topical domains – by far the most suitable large corpus currently available.

RESEARCH QUESTION

The question we addressed is: Can the use of domain-specific ngram models appreciably enhance word prediction performance in the context of augmentative communication? While our early studies indicated that database domain specificity did not play a significant role in system performance, recent pilot studies indicated that this question merited a more focused investigation.

METHODS

We chose to study the 20 most frequently occurring topic domains in the Switchboard Corpus. The testing texts for each of these 20 target domains was generated by concatenating all conversations of that domain from the first 12.5% of the corpus. The remainder of the corpus was used to generate the domain-specific training texts. We generated two other training texts: 1) 'Small', consisting of 5% of the training text for each of the 20 target domains, resulting in a text approximately the same size as the average of the 20 domain-specific training texts; and 2) 'Big', comprised of the entire training text. Trigram models were created for each of the 22 training texts.

The experiments were carried out using our IMPACT augmentative communication software. Running in emulation mode, this system can simulate a human using its interface to produce a message. The testing interface consisted of a standard QWERTY keyboard augmented by a dynamic 6-word prediction list. Keystroke savings (KS) were used as the performance measure.

RESULTS

We measured the performance of four prediction model configurations on each of the 20 domain-specific testing texts. The four configurations were: 1) 'Small' only; 2) 'Auto', meaning that the prediction model was derived from the same domain as the testing text; 3) 'Big' only; and 4) 'Big+Auto', a blending of two ngram models. Figure 1 shows performance on 10 representative domains. Table 1 shows average performance over all 20 domains for the four configurations.

Not surprisingly, the 'Big+Auto' configuration, with its equally weighted general and specific components yielded the best results, followed in turn by 'Big', 'Auto', and 'Small'. The almost 2% advantage of 'Big' over 'Auto' is also reasonable given its much larger training text size. However this effect is also due to the fact that the conversants were generally not experts in these domains. This boosts the relative importance of those testing text statistics correlated with *conversation* in general and the 'Big' training text constitutes a fairly large and therefore reliable sample of such general conversational statistics.

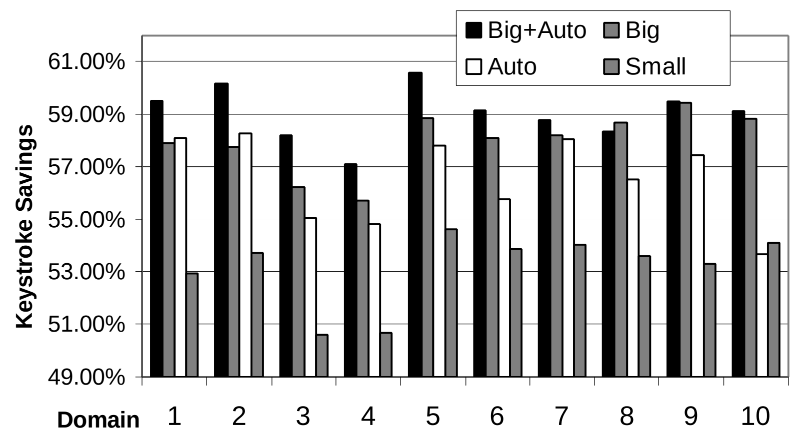

Figure 1: Comparison of Predictive Models

The 'Small' training text is only about 1/40th the size of the 'Big' training text, yet it covers a significant fraction of the domains that 'Big' does, thus rendering it a far less reliable sample of general conversational statistics. This accounts for our most interesting result which is the nearly 3% advantage of the 'Auto' configuration over the comparably-sized 'Small' configuration. Thus, for a given model size, using a model derived from text of the same domain as the testing text yields better prediction than using a model derived from a more general pool of text. As noted earlier, this is because the 'Auto' models provide a better match between the training and testing text word usage patterns.

Table 1: Average KS for four configurations

Config.	Ave KS
Big+Auto	58.74%
Big	57.81%
Auto	55.91%
Small	53.00%

Finally, we emphasize that the 'Big+Auto' configuration exceeds the 'Big' configuration by nearly 1%, despite the fact that it is only very slightly larger than the 'Big' configuration. This reinforces our main finding of the benefit of using domain-specific prediction models and suggests that as we consider larger and larger total model storage capacities, we expect the greatest incremental improvements to result from additions of domain-specific training text rather than of general text. This is the subject of ongoing studies.

DISCUSSION

Since it appears that domain-specific databases can provide substantial improvements in word prediction, where can appropriate databases be found? Existing corpora such as the Brown, Switchboard, and British National Corpora consist of text categorized roughly along various domain boundaries – topic, genre, sophistication, etc. By dividing these corpora along these categories, a series of baseline domain-specific models could be derived.

Our research team is also investigating the feasibility of culling appropriate databases from the internet using an autonomous "web crawler" (3). While the web offers a wealth of text – perhaps as much as a trillion words – this text varies widely in content, style, and sophistication. We have developed a prototype web crawler capable of searching out and retrieving specific genres of text. Such a system opens up exciting new possibilities for domain-specific word prediction since it can potentially produce very large databases – an important determiner of word prediction accuracy (1).

This paper has focused on prediction databases specific to a particular *topic* domain. The model can clearly be extended to other domain classification schemes such as style, formalness, or genre. For example, at different points during a day, a student might be working on an essay for class, a work of fiction, and a letter to a friend. By switching domains between appropriate text genres (essay, narrative, and correspondence), a word prediction system could take advantage of the word usage and syntactic peculiarities of each genre to offer more appropriate predictions.

The performance enhancements described above assume that the appropriate ngram model is always being applied to a conversation. Certainly this could be true if the augmented communicator manually switched databases as needed. Of more interest, however, is a system that *automatically* switches databases. We are currently developing a system that utilizes local conversational context to determine the current domain. Preliminary results show that on a limited set of topic domains, 80% domain switching accuracy is possible.

REFERENCES

1. Lesher, G.W. (1998). *Advanced Prediction Techniques for Augmentative Communication*. Final Report, Department of Education SBIR Phase I Contract RW97076002.

2. Beukelman, D., Yorkston, K., Pobleto, M., & Naranjo, C. (1984). Frequency of word occurrence in communication samples produced by adult communication aid users. *Journal of Speech and Hearing Disorders*, 49, 360-367.

3. Lesher, G.W. & Sanelli, C. (2000). A web-based system for autonomous text corpus generation. *ISAAC 2000*, Washington, DC.

ACKNOWLEDGMENTS

This study was supported in part by a National Institute of Deafness and Communication Disorders (NIDCD) Small Business Innovation Research grant (1 R43 DC04383-01).

Gregory W. Lesher
Enkidu Research, Inc., Spencerport, NY 14559
716-352-0507, 716-352-0508 (fax) , lesher@enkidu.net