Topic Modeling in Fringe Word Prediction for AAC

Keith Trnka, Debra Yarrington, Kathleen McCoy
Computer Science Department
University of Delaware Newark, DE 19716
{trnka,yarringt,mccoy}@cis.udel.edu

Christopher Pennington
AgoraNet, Inc.
penningt@agora-net.com

ABSTRACT
Some AAC devices make a distinction between core vocabulary (the relatively small set of frequently used words) and fringe vocabulary (the much larger set of infrequently used words). In this poster, we concentrate on using word prediction to reduce the number of keystrokes necessary for accessing fringe vocabulary words. We will introduce the notion of topic modeling to improve predictions and present evaluations of keystroke savings.

Keywords: Communication, Research, Technology, Theoretical models and issues

INTRODUCTION
Communication is an essential component of day-to-day life. Individuals with communication difficulties therefore face significant challenges throughout their lives. Our research focuses on improving the state of the art for user input on VOCA devices. Specifically, we seek to improve language processing in fringe word prediction.

Currently, many VOCA devices use a static user interface for entry of core words. For example, many of the devices of Prentke Romich Company use sequences of icons to allow efficient access to core words. However, ordinary conversation often requires the use of some non-core words. To accommodate this, devices require the user to spell out the word and provide a dynamic list of words the user is likely to be entering.

Word prediction uses a language model to statistically predict the most likely next word on the basis of what has been typed so far. We attempt to have the language model influenced by the current topic of the conversation so that it more accurately predicts the next word. Many fringe words are content words, so we feel that topic modeling is well suited for fringe word prediction.

LANGUAGE MODELING
Early work on word prediction used a dictionary of words and information on word frequency and recency to predict the most likely next word given the first couple of letters typed by the user.

Statistical natural language processing (Manning and Schütze, 2000) is an area of research that tries to incorporate the context of the preceding words into account when predicting the word to come next. For instance, even though “may” is a high frequency word, it is not likely to be what is intended given that the user has typed “Let's go to the m”. Statistical NLP uses the notion of ngrams – developing language models that take the previous (n-1) words into account when deciding what is most likely to come next. Taking frequency into account is the unigram case where a single word's probability is taken into account. Bigrams (2-grams) conditions the likely-hood of a given word given the previous word. In this instance, “the” would be the word taken into account. So, rather than just considering the frequency of the word “may” we would consider the frequency of the string “the may”. This probability is fairly low since "the" is usually followed by a noun or adjective. Trigrams take the previous two words into account, "to the" in this case. And so on. Ngram methods require increasingly larger text samples as the size of the context is increased. For this reason, most researchers do not consider contexts any larger than the two previous words.

Ngram methods have worked very well for the amount of effort they require to implement. However, ngram methods suffer the same sort of problem as unigram methods: predictions are sometimes inappropriate because a portion of the context is ignored. In particular, ngrams ignore the topic of conversation (e.g., sports, food, weather). Some systems have tried to incorporate topic by having the user select a particular dictionary to use (e.g., Lesher and Rinkus, 2002). However, this puts a lot of work on the user and is unable to handle many normal conversations (e.g., talking about food at a baseball game). In our topic modeling approach, we do not place any burden on the user for selecting a specific dictionary and we allow the topics to change during a conversation.

Our approach uses what the user has said so far in the conversation to identify the relevance of each of several predefined topics. The relevance is then converted into percents (e.g., this conversation is 67% food, 20% sports, and 13% other). Each topic has its own set of ngram statistics. The ngram statistics to use in word prediction are determined by combining the individual ngram statistics of each topic using the percentages. Then the relevance of each prestored topic to the current conversation is recomputed every few sentences.

The effect of this approach is that the occurrence of one content word will tend to boost the position of similar words up in the prediction list. For example, if one utterance is “What's for dinner?”, words dealing with food will be higher in the prediction list than they would be otherwise. Because our method allows the topic of conversation to shift over time, if the user starts to discuss a topic such as local news, words related to local news will be boosted higher in the prediction lists while words related to food will fall back to a normal position.

EVALUATION

We evaluate the impact of topic modeling on fringe word prediction by measuring keystroke savings on a designated part of the Switchboard corpus. The Switchboard corpus is a collection of transcribed telephone conversations. Although a collection of AAC user text would be more suitable for evaluation, there is no widely available large scale collection of AAC text. The keystroke savings (KS) of a conversation is

$$KS = \frac{keys_{normal} - keys_{prediction}}{keys_{normal}}$$

$keys_{prediction}$ includes keys required to select a word from the list of predictions, as well as a key at the end of each turn. $keys_{normal}$ includes all the keys requires to enter a word normally, as well as spaces pressed after a word and the key at the end of each turn.

The results of fringe word prediction with a trigram method and with topic modeling are shown below. Because the number of words to predict affects the results dramatically and the ideal window size is largely determined by user preference, we have shown the evaluation of each method at various window sizes.

Window size	Trigram method	Topic modeling	Improvement
1	42.3%	43.1%	0.8%
3	55.1%	56.4%	1.3%
5	58.8%	60.2%	1.4%
6	60.0%	61.4%	1.4%

Topic modeling shows a clear improvement over a trigram baseline method. Our poster presentation will provide more details regarding our implementation of topic modeling as well as further evaluations. We will also show several side-by-side examples of what word prediction would look like with and without topic modeling.

RELATED WORK

Language modeling

The topic modeling approach that we have designed is somewhat similar to research in Computational Linguistics and Speech Recognition. Our work in topic modeling was originally inspired by the work of Mahajan et. al. (1999), although we found it necessary to define topics explicitly for practical reasons. The interested reader might also see Florian and Yarowsky (1999) for another similar approach, but with a hierarchical organization of topics. Bellegarda (2000) was the primary motivation for our representation of the current conversation as well as an optimized implementation of topic modeling. However, Bellegarda additionally uses Latent Semantic Analysis with word-based clustering to improve speech recognition.

Word prediction

There have been attempts at topic modeling within the word prediction community. Lesher and Rinkus (2002) explored topic modeling using the Switchboard corpus. However, Lesher and Rinkus' work assumed that the topic of conversation was known ahead of time in testing. The most similar work to our own is that of Li and Hirst (2005), which explicitly models the notion that the occurrence of a word increases the likelihood of seeing related words. Their work shows that topic modeling can yield an improvement in keystroke savings, but because it was evaluated on different material, we are unable to make a direct comparison.

REFERENCES

J. Bellegarda. Large vocabulary speech recognition with multispan language models. IEEE Trans. On Speech and Audio Processing, 8(1), 2000.

R. Florian and D. Yarowsky. Dynamic nonlocal language modeling via hierarchical topic-based adaptation. In Proceedings of ACL'99, pages 167–174, 1999.

G. Lesher and G. Rinkus. Domain-specific word prediction for augmentative communication. In Proceedings of the RESNA '02 Annual Conference, 2002.

J. Li and G. Hirst. Semantic knowledge in word completion. In The Seventh International ACM SIGACCESS Conference on Computers and Accessibility, pages 121–128, 2005.

M. Mahajan, D. Beeferman, and X. D. Huang. Improved topic-dependent language modeling using information retrieval techniques. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, 1999.

C. Manning and H. Schütze. Foundations of Statistical Natural Language Processing. MIT Press, 2000.