# Adding to LSA's Bag For Information Retrieval

**Michael Glass**
michael.glass@valpo.edu
Math & CS Dept., Valparaiso University
Valparaiso, IN 46383 USA

## Introduction

Feature LSA (FLSA) is the practice of creating non-word features and adding them, as synthetic words, into the bags of words that LSA models. Here we report the largely unsuccessful results of two experiments to enhance LSA effectiveness, one adding a syntactically-derived feature and one adding a semantic one.

## Experiment 1: Word Collocations Within Noun Chunks

For our task we imitated part of the TREC-4 information retrieval task, applying full-sentence queries to retrieve wire service articles.

Starting from a chunked corpus, we added synthetic words representing collocations. Each collocation came from two words $w_1, w_2$, not necessarily adjacent, that co-occurred within the same noun group. We kept the synthetic words only for pairs that co-occur within a noun group more frequently than chance would predict. These synthetic words thus recognize many word associations that are not phrasal nouns or frozen phrases. For example *rocket, solid*, which occurs 174 times more likely than chance, is derived from *solid fuel booster rocket* and *solid rocket booster motor* as well as a number of other NPs including the shorthand form *solid rocket*.

A 500-dimension LSA semantic space was constructed from the documents in the collection being searched. Both recall and R-Precision, the precision at the moment of the last retrieved relevant document, show a very slight improvement, shown in Table 1. Various versions of this experiment, e.g. different cutoffs for selecting high-probability pairs, synthetic collocations extracted from the verb groups, etc., all yield similarly small, or even negative, changes in R-Precision and recall.

The failure of this method stems from the almost $4\times$ explosion in the vocabulary from 110,000 to 416,000 words. We can readily decrease performance by adding extra vocabulary words while holding the number of dimensions constant. That we saw a small performance increase could indicate success of the word pairs, masked by the vocabulary increase effect. There being no principled way to increase the semantic space size as a function of vocabulary size, there is no effective way to compare the two conditions using different numbers of dimensions. (Wiemer-Hastings & Zipitria, 2001) achieved greater success in adding structural information to LSA comparisons of sentence-length texts without our problem of increasing the vocabulary.

Table 1: R-Precision with Synthetic Word Pairs

| Exp No. | Rel. Docs Retrieved | R-Precision |
|---|---|---|
| Control | 867 | 0.12 |
| NP word pairs | 879 | 0.14 |

## Experiment 2: Artificial Semantic Tag

(Serafin & Di Eugenio, 2004) achieved excellent results at dialogue act annotation of the CallHome Spanish corpus by encoding structural information as synthetic features along with the annotation tags. By contrast, our task used a simple collection of annotated text, adding the annotation as an artificial word into LSA's bag. Using the hand-coded relevance judgment for each document that comes with the TREC-4 data, we added a synthetic feature word $r (relevant to one of the queries) or $i (irrelevant to any query) to 90% of our documents. All of the $r-tagged documents then have a word–the feature annotation–in common, in addition to whatever semantic similarities existed, making their LSA document vectors slightly more similar to each other. The same holds for the $i-tagged documents.

The experiment injected $i and $r tags alternately into an untagged document, retrieved the most relevant documents from the entire collection under each condition, and compared the retrieval sets. We were unable to use these differences to reliably annotate the query document, the differences induced by a single feature word in 440 words (average) of wire service text appeared to be too slight.

## Acknowledgments

Jessica Warnier provided software support and considerable technical assistance to this project.

This work was supported by the Cognitive Science Program, Office of Naval Research, under grant N00014-03-1-0037 to Valparaiso University. The content does not reflect the position or policy of the government and no official endorsement should be inferred.

## References

Serafin, R. and B. Di Eugenio (2004). FLSA: Extending Latent Semantic Analysis with Features for Dialogue Act Classification. *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, ACL-04, Barcelona*.

Wiemer-Hastings, P. & I. Zipitria (2001). Rules for Syntax, Vectors for Semantics. In *Proceedings of the 23nd Annual Conference of the Cognitive Science Society*.

# Expert and Novice Algebra Tutor Behaviors Compared

**Jung Hee Kim (jungkim@ncat.edu) and Hyeun Me Chae (hchae@ncat.edu)**
Dept. Computer Science, North Carolina A&T State University
Greensboro, NC 27411 USA

**Michael Glass (michael.glass@valpo.edu)**
Dept. Math & Computer Science, Valparaiso University
Valparaiso, IN 46383 USA

## The Experiment

Analyzing computer-mediated, keyboard-to-keyboard algebra tutoring transcripts (Kim & Glass, 2004), we studied quantitative differences in tutoring behaviors between an expert and a novice tutor. This should highlight the kinds of behaviors expert tutors acquire that set them apart from novices. The most pronounced difference is that the expert is far more likely to set procedural goals.

In this study the expert tutor is Dr. Kathy Cousins-Cooper, a professor in the Mathematics Department at NC A&T State University who has taught and tutored basic algebra for many years. The novice is an upper-level undergraduate mathematics student who prior to these sessions had the typical student tutoring experiences of a mathematics major. Students were volunteers from an undergraduate elementary algebra class. The problems that were addressed in a tutoring session were selected by examining student performance on a pre-test, so in all cases we had reason to believe the student could not solve these problems beforehand.

We focused on the two symbolic manipulation problems shown in Figure 1. We have 9 examples of problem 1 taught by the expert and 6 by the novice tutors. For problem 2 we have 8 expert and 4 novice examples. Over all, we have 24 expert sessions and 10 novice, but not every problem was tutored in every session.

| | |
|---|---|
| Problem 1: subtract | $\frac{x}{x^2 - x - 6} - \frac{2}{x^2 - 7x + 12}$ |
| Problem 2: solve for x | $5x = 2x^2 + 1$ |

Figure 1: The two problems studied in this paper.

Phenomena we examined included: learning gains, time to tutor the problem, number of turns needed, frequency of goal-setting acts, and tutor responses to impasses.

## Results and Discussion

The mean learning gains were 0.47 for the expert sessions and 0.24 for the novice, where learning gain is calculated as:

(posttest – pretest) / (1 – pretest)

This difference is only weakly significant, $p<0.1$.

The most marked difference is that the expert engages in significantly more goal-setting episodes, see Table 1.

Table 1: Goal-Setting Episodes

| | Expert | | Novice | | t-test |
|---|---|---|---|---|---|
| | Avg./Prob | n | Avg./Prob | n | |
| Collab. | 3.88 | 66 | 1.00 | 10 | p<.001 |
| Informed | 1.59 | 27 | 0.50 | 5 | p<.001 |
| Total | 5.47 | 93 | 1.50 | 15 | p<.001 |

This finding is consistent with teaching algebra as a procedural skill. In this table, collaborative goal setting episodes involved both parties deciding the next procedural step, while informed episodes consisted of the tutor telling the next procedural step to the student. . Chi-squared shows that despite the difference in numbers of episodes, there is no significant difference between novice's and expert's choice of collaborative vs. informed goal-setting.

Average dialogue turns per problem was significantly ($p<0.01$) higher for the expert at 17.5 turns (divided between both parties) vs. the novice at 12.4 turns. Combined with the goal-setting results, this shows that the bulk of the expert tutor's nine turns were devoted to goal setting.

Consistent with the findings of VanLehn et al (2003), we discovered that for the expert tutor most student impasses (4.5 out of 4.7 per problem tutored) were recognized by tutor intervention. The novice intervened in significantly fewer instances, 1.7 out of 3.9 impasses. The two tutors did not differ in their responses to the impasses.

## Acknowledgements

This work was supported by the Cognitive Science Program, Office of Naval Re-search, under grant N00014-02-1-0164, to North Carolina A&T State University. The content does not reflect the position or policy of the government and no official endorsement should be inferred.

## References

Kim ,Jung Hee and Michael Glass. 2004. Evaluating Dialogue Schmata with the Wizard of Oz Computer-Assisted Algebra Tutor. In Lester, J., R. M. Vicari, and F. Paraguacu, eds., *Intelligent Tutoring Systems: 7th International Conference (ITS 2004), Maceio, Brazil*. Berlin: Springer. Published as LNCS 3220.

VanLehn, Kurt, Stephanie Siler, Charles Murray, Takashi Yamauchi, and William Baggett. 2003. Why do only some events cause Learning During Human Tutoring? Cognition and Instruction 21(3), pp. 209-249.