Title
A Comparison of Student Evaluation Algorithms in AutoTutor

Permalink
https://escholarship.org/uc/item/3zg9p7ff

Authors
Chipman, Patrick
Francreschetti, Donald R.

Peer reviewed

A Comparison of Student Evaluation Algorithms in AutoTutor

Patrick Chipman (pchipman@memphis.edu)
Department of Psychology, University of Memphis, 3206 Northwood Drive
Memphis, TN 38111 USA

Donald R. Franceschetti (dfrncsch@memphis.edu)
Department of Physics, University of Memphis, 328 Manning Hall
Memphis, TN 38152 USA

Keywords: latent semantic analysis; student models; intelligent tutoring systems

Introduction

Intelligent tutoring systems (ITSs) require adept student modeling mechanisms in order to adapt their pedagogical strategies to their users. One key element of any student model is its evaluation algorithm for student responses.

AutoTutor, a natural language intelligent tutoring system developed at the University of Memphis (Graesser, Chipman et al., 2005), uses a pair of evaluation algorithms based on the statistics of language. AutoTutor combines Latent Semantic Analysis (LSA) (Landauer, Foltz, & Laham, 1998) with a string matching algorithm in which each word is weighted relative to its inverse word frequency (IWFO). This hybrid approach aims to provide a good evaluation of student input without requiring considerable knowledge engineering. Past analyses have suggested this hybrid algorithm is likely to have good agreement with expert ratings (Graesser et al., 2007). We believe a hybrid model will outperform either LSA or IWFO alone.

Method

330 student essay answers to questions in conceptual physics were collected in a prior study (VanLehn et al., 2005). Each ideal answer to those questions was decomposed into two or more expectations (components of the answer). Each student answer was paired with the expectations for its associated question. A professor of physics rated each student answer on a three point scale. Higher ratings were given to answers that more explicitly stated the expectation. This process produced 1597 expectation-student answer-expert rating trigrams.

LSA and IWFO match scores were then computed. The scores ranged from 0.0 to 1.0; higher values represented a more "correct" student answer. Three algorithms were tested: LSA alone, IWFO alone, and a hybrid algorithm where LSA is weighted 33% and IWFO is weighted 67%.

Results

A 4 x 4 Pearson correlation matrix was computed to determine how well the three algorithms compared to expert ratings. Because of the size of the testing set, all correlations were statistically significant. Table 1 shows the correlations of the three algorithms with the expert ratings.

Table 1: Correlations, expert and algorithm ratings.

Algorithm	r
LSA	.293
IWFO	.430
Hybrid	.419

Discussion

The results of this analysis show that the hybrid model, combining both LSA and IWFO, agrees less with expert ratings of student answers than IWFO alone in the domain of conceptual physics. Physics professors emphasize specific keywords instead of synonyms because of the imprecise nature of language, and so IWFO better represents their grading technique. LSA's poorer representation of this technique reduces the efficacy of the hybrid model, at least in this knowledge domain.

Acknowledgments

References

Graesser, A. C., Chipman, P., Haynes, B. & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed initiative dialogue. *IEEE Transactions on Education, 48(4)*, 612-618.

Graesser, A. C., Penumatsa, P., Ventura, M., Cai, Z., & Hu, X. (2007). Using LSA in AutoTutor: Learning through mixed initiative dialogue in natural language. In T. Landauer, D. McNamara, S. Dennis, and W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Erlbaum.

Landauer, T., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259-284.

VanLehn, K., Graesser, A.C., Jackson, G.T., Jordan, P., Olney, A., & Rose, C.P. (2005). When is reading just as effective as one-on-one interactive tutoring? In B. Bara (Ed.), Proceedings of the 27th Annual Meetings of the Cognitive Science Society. Mahwah, NJ: Erlbaum.