UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Automated Detection of Strategies in Free Text Responses

Permalink
https://escholarship.org/uc/item/36m8h815

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 24(24)

ISSN
1069-7977

Authors
Harrison, Anthony
Saner, Lelyn
Cookson, Celestine
et al.

Publication Date
2002

Peer reviewed

Automated Detection of Strategies in Free Text Responses

Anthony Harrison (anh23@pitt.edu)
Lelyn Saner (les53@pitt.edu)
Celestine Cookson (clcst70@pitt.edu)
Darcie Kunder (dakst67@pitt.edu)
Christian D. Schunn (schunn@pitt.edu)
Learning Research and Development Center, University of Pittsburgh
3939 O'Hara St., Pittsburgh, PA 15260 USA

When solving problems, people often use a wide array of different strategies. Effective teaching often requires isolating what strategies students are using (or not using) in order to more effectively structure the instructional intervention. Nowhere is this truer than in the realm of intelligent adaptive tutors. The classification of strategy use in complex domains presents an interesting challenge to intelligent tutors. This is made even greater if the strategies are to be extracted from free text responses given by the students.

To this end, we have been using Latent Semantic Analysis (LSA) as an automatic strategy classification tool. LSA is a computational tool that extracts the co-occurrence of words in a corpus. Through high-dimensional matrix decomposition, LSA is able to produce a "semantic-space" allowing all experienced words, phrases, and sentences to be represented as vectors within that space. The more similar the vectors are to each other, the more similar their meanings. As LSA has matured, some have suggested that it may be a psychologically plausible theory of semantic learning. We remain noncommittal in this regard, choosing instead to rely upon LSA in its original capacity as a fast and efficient text-processing tool.

Strategy Classification

Our current endeavor is to use LSA to intelligently classify strategy use in day-to-day military operations. The hope is that by accurately classifying young officers' strategy uses, we can develop tutoring systems to broaden their range of strategies as well as train them to more appropriately apply the strategies.

The strategy classification system relies upon a series of key steps. First the LSA semantic space was generated based on a set of military handbooks, training documents, and pedagogical examples. Free text responses to military scenarios were collected from officers in training as well as experienced military officers. These were then human coded into different strategy categories. The responses were then fed into LSA to generate their vector representations in the semantic space. These two sources yielded two databases of semantically coded (vectors in semantic space) strategies. The novice database (officers in training) is used as a descriptive reference, while the expert database (experienced officers) provides the normative references. The final steps are to take the free text responses of other novices to the same vignettes, transform them into vectors in semantic space, and then each of these vectors is compared against those in the databases. Since they are vectors, the cosine between the two serves as a simple similarity score. As similarity increases, the cosine value will approach one. This process yields a ranking of similarities to the descriptive database, where the classified strategy is merely the most similar. Additionally, since we have a sample of strategy exemplars, we can also look at the distribution of similarity scores across strategies. This yields a simple measure of *confidence*: the greater the number of high similarity matches within a given strategy gets, the more confident we can be that it is representative of that strategy.

At this early stage in the development of the system, we were pleased to see that LSA was classifying strategies about as well as our human coders, with almost equivalent inter-rater reliabilities. This is a significant accomplishment given how limited our semantic space is currently (only 100,000+ words, in comparison to the millions of most other LSA corpora), and the limited scale of our descriptive database (10 strategies, approx. 16 exemplars each).

Future Directions

Aside from increasing the scale of both the semantic space and the reference databases, we hope to begin working on the tutoring system proper. This will mean developing a training system that adapts to the strategy use of the individual to provide sufficient scaffolding to enable them to explore alternative strategies, as well as to learn how to appropriately apply them. Then, as the student progresses through the tutor, the normative database (provided by experienced military officers) will come into greater play.

References

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104(2)*. 211-240.

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