

Automated Content Assessment Tools for E-Learning Environments

Lynn A. Streeter & Noelle LaVoie
Knowledge Analysis Technologies, 4940 Pearl East Circle, Suite 200, Boulder, CO 80301
Tel: (303) 545-9092, Fax: (303) 545-9113
Email: lstreeter@k-a-t.com

Charles Krupnick (1)
US Army War College, Carlisle, PA

Joseph Psotka (1)
US Army Research Institute, Alexandria, VA

Technology can make Asynchronous Learning Networks more accessible, useful, and effective. Here we examine two new automated tools: (1) embedded assessment of participant contributions, and (2) classification of comments into policy and planning/coordinating categories. Both tools use Latent Semantic Analysis (LSA), a machine learning algorithm for text understanding (see Landauer, Foltz, & Laham, 1998, An introduction to latent semantic analysis, *Discourse Processes*, 25).

The US Army War College supplied a complete electronic copy of an extensive asynchronous threaded simulation of the US national security policy development process on the future of NATO. There were 20 discussion groups with 12 to 15 participants per group, who produced six MB of text consisting of over 1600 individual contributions. Each group wrote both an intermediate and final policy paper. We were able to devise methods for accurately assessing the students' policy papers and individual contributions.

Four experienced human graders assigned grades to the policy papers. We then used the Intelligent Essay Assessor, based on LSA, and several other language measures, to predict the average human grade for each policy paper. The Intelligent Essay Assessor creates the best fitting regression model which then can grade new essays as well as human graders. In this case, the model correlated with the average human scores 0.64. Since it is not usually possible to access skilled human graders for online discussion groups we devised a method of assessing written contributions without human input by comparing every "essay" to every other essay on a series of LSA and language measures to produce the best linear ordering of the set of essays. The intuition behind the algorithm is that a smart human grader, but not necessarily a subject matter expert, could take essays on the same topic, read them all, compare them and in the end form some fairly accurate opinions of which were the good essays and which were the poor essays. The implementation of the no-human-input model was highly correlated (.80) with the model based on human grades. Since this procedure is very general, embedded assessment software agents can be built to score any content-based course, and could assess the relative contributions of groups and individuals quickly in a non-obtrusive and automatic way.

Moderators of discussion groups often pay special attention to groups who are having problems accomplishing their goals. They would like to know which groups are focusing on the task and which ones need some intervention. We devised an automatic method for separating policy from support postings, which often indicate coordination difficulties. From the LSA – based model used to assess the policy papers, we computed a score for each of the approximately 1600 comments. A human then rated each comment as either a policy or planning/coordinating kind of comment, based on whether it was low or high in policy content. A threshold for separating the two categories was selected that produced the greatest number of hits with the fewest number of false alarms. The LSA – based model correlated highly with the human scores (0.72) with only an 8% misclassification rate. Since the LSA – based no-human-input model can be devised for any new course content and runs in real time, it is possible to automate embedded assessment and monitoring of any threaded discussion group. Such tools hold great promise for the automated assistance of human moderators in large, complex e-Learning environments. These tools can improve the quality and consistency of learning, and reduce instructors' workload by focusing their efforts on groups and individuals who need it most.

Endnotes

- (1) With two authors from government organizations, we emphasize that the views expressed in this paper are those of the contributors and do not necessarily reflect the official policy or position of the US Army, Department of Defense, or the US government.