Subject: Final Reports for 09-11 LRDC grants
Date: Thursday, June 16, 2011 11:04 AM
From: Mary Kay Stein <mkstein@pitt.edu>
To: "Litman, Diane J" <dlitman@pitt.edu>, "Schunn, Christian Dieter" <schunn+@pitt.edu>, "Ashley, Kevin" <ashley@pitt.edu>, Jenn Russell <jrussel@pitt.edu>, Louis Gomez <lgomez+@pitt.edu>, James Greeno <jimgrno@pitt.edu>, "Gomez, Kimberly" <kgomez@pitt.edu>, "Crowley, Kevin J" <crowleyk@pitt.edu>
Cc: "Perfetti, Charles A" <perfetti@pitt.edu>, "Prine, Deanna Weber" <dwp17@pitt.edu>

Dear LRDC Internal Grant Holders:

We are nearing the closure of the 2009-2011 grant cycle for the LRDC Internal Grants Program. I am writing to request a final report on the goals, activities, and accomplishments associated with these grants. These final reports are important because they provide a way for us to track the challenges and successes of this program.

Please submit an approximately 2-page written report that addresses: Goals, Activities, and Accomplishments. Within the accomplishment section, please identify (1) pending or awarded external funding for which the LRDC monies served as a seed grant; (2) the names of graduate students supported by the grant.

Our records indicate the following 2009-11 grants:

IMPROVING LEARNING FROM PEER REVIEW WITH NLP AND ITS TECHNIQUES
Litman, Schunn, Ashley

ROLES, TOOLS AND PRACTICES OF TEACHERS WITHIN INCLUSIVE SCHOOLS
Russell, L. Gomez, Greeno

SCHOOL/COMMUNITY PARTNERSHIPS IN THE 21ST CENTURY: HOW DIGITAL TECHNOLOGIES CAN BUILD A CULTURE OF LEARNING
L. Gomez, K. Gomez, Crowley

Please send the final reports to me by August 1, 2011. Thanks!

Mary Kay

Final Report: Improving Learning from Peer Review with NLP and ITS Techniques

Diane Litman, Christian Schunn and Kevin Ashley.

July 2009 - June 2011

Goals

Prior research with SWoRD, a web-based system to support peer reviewing in a wide variety of disciplinary classroom settings, generated an enormous database of written materials. This corpus was ripe for analysis and exploitation in support of research on natural language processing (NLP), intelligent tutoring systems (ITS), cognitive science, educational data mining, and improving learning from peer review. The research goal of our project was to use existing SWoRD-generated data to develop new types of components for inclusion in SWoRD, with the ultimate goal of improving the quality of student reviewing and writing performance. Our software development goal was to implement an improved version of SWoRD for use in further experimentation. In the process of achieving these goals, we created a very productive new collaboration in a novel LRDC research direction, and a thriving interdisciplinary research group with regular meetings of faculty, graduate students and postdocs (including LRDC members who have not been financially supported by the grant).

Activities and Accomplishments

Adding Artificial Intelligence to SWoRD

Current peer-review software lacks intelligence for responding to students' reviewing performance. Through the use of Artificial Intelligence (AI techniques), we developed components for automatically detecting key features (identified in prior learning sciences research) related to increased quality in both the reviews and papers managed by SWoRD. This will help ensure that peer reviewers state their reviews effectively and that reviews and papers focus more frequently on important paper aspects. We hypothesize that future intelligent scaffolding of such features will enable reviewers to generate better feedback, and authors to implement feedback more effectively, resulting ultimately in improved writing quality overall.

First, we developed systems for assessing reviewers' skills in producing feedback with three types of desirable features: problem localization, solution identification, and helpfulness. Taking a data mining approach, we used machine learning to automatically build classifiers based on attributes extracted from free text peer-review data via Natural Language Processing techniques. Our experiments demonstrated that our learned classifiers could predict whether or not feedback was problem-localized significantly better than a baseline algorithm [1]. We also examined the impact of annotation difficulty for detecting problem localization [2]. Since our goal is to generate assessment of reviewing performance for each reviewer, we added an aggregation step, in which we made a binary decision on whether the reviewer provided enough problem localization in their reviews in general. Our system achieved comparatively high accuracy and precision for identifying reviewers who generally failed to provide localization information in their reviews [3]. We generalized our approach by obtaining similar results for detecting solution identification [4]. Finally, we demonstrated that prediction methods for analyzing the helpfulness of product reviews (e.g. reviews of movies, books, appliances) applied to our context of peer reviews, although

prediction accuracy could be further improved through the use of peer-review specific knowledge [5]. In addition, we showed that low-level linguistic features were more useful for predicting student ratings of helpfulness, while high-level cognitive constructs were more effective in modeling expert ratings [6].

Second, with support from this grant, Collin Lynch (Ashley's student) has been studying argument diagrams and their relationship to written argumentation and peer-review. In the SWoRD system students engage in producing, sharing and discussing writing assignments and responding to feedback in a variety of argumentation-oriented domains including cognitive science, physics, history and philosophy. Lynch's dissertation proposal [14] focuses on students' ability to compose their own written arguments, the diagnosticity of diagrammatic models of argument, and the practical application of diagrams to argument writing. When diagrammatically annotating an existing argument or producing a novel one, it is assumed, students demonstrate their ability to recognize, reify, or instantiate individual argument schema and to connect them into a cohesive whole. This diagnosticity of argument diagrams is also assumed to be domain-general, that is, student-produced argument diagrams will be diagnostic across domains although the type of diagram production tasks may vary across domains. For a given argumentation task (e.g., diagramming introductory material for a research proposal), it is further assumed, there exists a set of argument schema that are pedagogically relevant, will be exhibited by a properly-designed argument diagram structure, and can be connected to a students' written argument. In a variety of classroom settings, Lynch has collected students' argument diagrams, prepared with visual realizations of these schema, as well as written arguments and instructor-assigned scores of the accompanying students' written arguments; he will soon collect more detailed scores for various argumentation features in both diagrams and written arguments using instructor-approved scoring instruments. Lynch will analyze the data to determine the extent to which the argument diagrams really are diagnostic, that is, if they reveal pedagogically relevant information about a students' understanding of argumentation and their ability to make novel arguments. Based in part on this work, we plan to investigate the uses of students' argument diagrams in peer reviewing as a precursor to writing arguments.

In related work, we developed a prototype thesis detection system that identified the presence or absence of clear thesis statements in student papers. When forming an effective essay it is essential for an author to have a clear and coherent thesis statement that is well-integrated into his work. This is true both of scientific papers, which must have a clear hypothesis statement with support, and persuasive essays in the humanities such as history. In this summer project he will explore the feasibility of developing a thesis detection system that identifies the presence or absence of clear thesis statements in students' writing. The presence of a robust thesis classifier would help us to identify whether authors include a clear statement of a thesis and peer reviewers identify and comment upon thesis statements in their peers' writings, and thus to provide targeted feedback to authors and reviewers in the SWoRD system as well as to identify other key argument structures in the data such as citing examples and drawing analogies.

SWoRD Software Development

We developed a new version of the SWoRD program from scratch that 1) is a more robust platform for continued deployment in actual classrooms (the old system was breaking too often), 2) included many options that enabled it to be used flexibly across more kinds of classrooms (varying numbers of rubrics, varying numbers of drafts, varying grading options), 3) included features that made it easier to use for research purposes (pre-segmenting of comments, easy export of anonymous data including only students who agreed to the participate for research, changeable back-evaluation prompt), and 4) was easily dividable into multiple ongoing variations to enable the testing of advanced AI features within a few classrooms. The base system was built in the summer / early Fall of 2009. It tested in one trial classroom in late Fall of 2009. It was then made available for Spring 2010, and has been used since then in over 4,000 students in 100 classrooms distributed across 8 countries, many disciplines, and many levels (from middle school through graduate courses). A visiting scholar from Beihang University in China ran a successful pilot test of SWoRD for use in L2 Writing instruction in China that may lead to many thousands of users next year.

Related Developments

In two related developments:

1. With support from a University of Pittsburgh Innovations in Education Award, Ashley's ISP Ph.D. candidate, Ilya Goldin, developed a structured peer review approach and system that enabled comparing the effects of conceptually-oriented, problem-specific review criteria and more general domain-writing-related criteria. The former scaffold the reviewer on how well the author addressed in his writing particular substantive issues raised by the problem. The latter focus on more general features of effective written legal argumentation. Ashley and Goldin applied the approach in two successive intellectual property courses, once with Comrade, a program Goldin developed, and once with the new SWoRD program. He found that the both kinds of criteria were valid in that they were correlated with the instructor's independently-assigned scores, and that the correlation of the problem-specific criteria was stronger, a result of potential significance in law school pedagogy [7; 8; 13]. He also found that reviewers applied the problem-specific criteria more independently of the other problem-specific criteria than with respect to the domain-writing-related criteria. By treating the peer-review process as a social network and by applying hierarchical statistical analysis, Goldin has tested various models for predicting the instructor's scores based on students' in- and outbound ratings [9]. They model student authors' and reviewers' (i.e., the entire class's) understanding of the relevant issues in the assignment and can provide feedback to the instructor, based on the ratings that students gave each other. Goldin has reported on all of the above experiments and findings in his successful Ph.D. defense [10]. Goldin has been a frequent participant in our Internal Grant project meetings.

2. In addition, LRDC's Internal Grant support has materially contributed to the successful ITS 2010 Workshop entitled, Computer-Supported Peer Review in Education (CSPRED) http://www.cspred.org/, organized by Schunn, Ashley, Goldin, et al. and held on Monday, June 14, in Pittsburgh [11]. As far as known, this was the first ITS workshop to focus on computer-supported peer review, and, according to the ITS 2010 Workshops/Tutorials Chairpersons, our workshop received the highest number of registrants of any submitted to the Conference. In addition, we gratefully acknowledge LRDC's generous support to underwrite a Workshop lunch and keynote address by Prof. Judy Kay, University of Sydney. In turn, the Workshop led to the development of a special issue of the *Journal of Writing Research* devoted to computer-supported peer review [12].

Funding applications for which the LRDC monies served as a seed grant

- DIP: Teaching Writing and Argumentation with AI-supported Diagramming and Peer Review (K. Ashley (PI), D. Litman and C. Schunn (Co-PIs)), *National Science Foundation*, 9/01/11-8/31/15, $1,349,986, recommended for funding.

- Intelligent Scaffolding for Peer Reviews of Writing (D. Litman (PI), K. Ashley and C. Schunn (Co-PIs)), *Institute of Education Sciences*, 7/01/11-6/30 2014, $1,484,222, declined (revision to be submitted September 2011).

- Keeping Instructors Well-informed in Computer-Supported Peer Review (K. Ashley (PI), D. Litman, C. Schunn and J. Wang (Co-PIs)), *LRDC Internal Grant*, 7/11-6/13, $131,100, funded.

Graduate students supported by the grant

- Michael Lipschultz, Computer Science
- Collin Lynch, Intelligent Systems
- Melissa Patchan, Psychology
- Wenting Xiong, Computer Science

Publications

1. Wenting Xiong and Diane Litman. Identifying Problem Localization in Peer-Review Feedback. *Proceedings 10th International Conference on Intelligent Tutoring Systems,* Pittsburgh, PA, June 2010 (poster paper).

2. Wenting Xiong, Diane Litman and Christian Schunn. Impact of Annotation Difficulty on Automatically Detecting Program Localization of Peer-Review Feedback. *Proceedings of Computer-Supported Peer Review in Education (ITS Workshop),* Pittsburgh, PA, June 2010.

3. Wenting Xiong, Diane Litman and Christian Schunn. Assessing Reviewers' Performance Based on Mining Problem Localization in Peer-Review Data. *Proceedings Third International Conference on Educational Data Mining,* Pittsburgh, PA, June 2010.

4. Wenting Xiong, Diane Litman, and Christian Schunn, Toward Improving the Quality of Peer Feedback through Natural Language Processing, *Journal of Writing Research*, under review.

5. Wenting Xiong and Diane Litman, Automatically Predicting Peer-Review Helpfulness, *Proceedings 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, OR, June 2011.

6. Wenting Xiong and Diane Litman, Understanding Differences in Perceived Peer-Review Helpfulness using Natural Language Processing, *Proceedings 6th Workshop on Innovative Use of NLP for Building Educational Applications (ACL-HLT Workshop)*, Portland, OR, June 2011.

7. Goldin, Ilya M., Kevin D. Ashley. Learning by Reviewing through Peer Feedback Refinement. *Proceedings of Computer-Supported Peer Review in Education (ITS Workshop),* Pittsburgh, PA, June 2010.

8. Goldin, I., Ashley, K. "Eliciting Informative Feedback in Peer Review: Importance of Problem-Specific Scaffolding". *Proceedings of the 10th International Conference on Intelligent Tutoring Systems*. Lecture Notes in Computer Science 6094. pp. 95-104. Pittsburgh, USA. June 2010.

9. Goldin, Ilya M., Kevin D. Ashley. "Peering Inside Peer Review with Bayesian Models." *Proceedings of the 15th International Conference on Artificial Intelligence in Education.* Lecture Notes in Computer Science, 6738, pp. 90-97. Auckland, New Zealand. July 2011.

10. Goldin, I., "A Focus on Content: The Use of Rubrics in Peer Review to Guide Students and Instructors". Ph.D. Dissertation. University of Pittsburgh Intelligent Systems Program. (submitted).

11. Goldin, I., Brusilovsky, P., Schunn, C., Ashley, K., Hsiao, I-H. (Eds.) *Proceedings of the Workshop on Computer-Supported Peer Review in Education (CSPRED-2010) at the Tenth International Conference on Intelligent Tutoring Systems.* Pittsburgh, PA. June 2010. http://www.cspred.org/proceedings

12. Schunn, C., Ashley, K. and Goldin, I. (Eds.) *Journal of Writing Research,* Special issue on Redesigning Peer Review Interactions Using Computer Tools. (revised and resubmitted).

13. Goldin, Ilya M. and Ashley, K. D. Conceptually focusing peer feedback using rating dimensions, *Journal of Writing Research,* Special issue on Redesigning Peer Review Interactions Using Computer Tools. 32 pages. (revised and resubmitted)

14. Lynch, C. "The Diagnosticity of Argument Diagrams". Ph.D. Dissertation Proposal. University of Pittsburgh Intelligent Systems Program. 2011