Modeling Peer Review in Example Annotation

I-Han Hsiao, Peter Brusilovsky
School of Information Sciences, University of Pittsburgh, USA
{ihh4, peterb}@pitt.edu

Abstract: The purpose of this study was to utilize the strengths of stronger students, in order to facilitate the example annotation process and ensure the quality of the annotations. The system presented in this study, AnnotEx, offers a focused way to integrate the Peer Review Model in example annotation and can be extended to meet advanced educational purposes. The results confirm that modeling the peer review when annotating examples guarantees the quality and efficiency of the annotations.

Keywords: Peer reviews, example annotating, example-based learning.

Introduction

In traditional e-learning environments, students are passive consumers and teachers are creators of educational content. However, the lack of content has encouraged a number of researchers to explore the idea of using students as creators rather than consumers of the content (Arroyo, 2003; Hsiao & Brusilovsky, 2007). Several studies demonstrated that this approach works really well: not only the students were able to produce content that was useful for their peers, but they also learned a lot in this process. For example, Jonassen and Reeves (1996) demonstrated that students are likely to learn more by constructing hypermedia instructional materials than by studying hypermedia created by others. Similarly, Hundhausen and Douglas (2000) argued that students can learn more by constructing algorithm animations than by watching animations contracted by experts.

As shown by several authors, student collaboration in the process of content authoring is critical for producing quality content. This collaboration can be both direct and indirect (in the form of peer reviewing). For example, Animal Watch Web-based Environment (AWE) engages student authors in content-creation and allows students to collaborate with each other to produce comments, which enhance the content of the system (Arroyo, 2003). The study confirms that students are willing to author the content, while still enjoying the benefits of the ITS as a student. In our previous studies (Hsiao & Brusilovsky, 2007; Hsiao, 2008), we explored the value of peer reviewing for creating quality educational content. We used Example Annotator system (AnnotEx), which supported collaborative authoring and peer-reviewing. The system was designed to increase the likelihood of example self-explanation and to provide collaborative learning opportunity through raising questions and creating the opportunity to collect peers perspectives. Both cases contribute to personal learning experience.

While earlier studies with AnnotEx demonstrated the value of peer reviewing process, the question of how to organize the peer reviewing process so that its benefits can be maximized was still not answered. One of the research questions explored in this study is whether the value of peer review could be increased if we pair students with strong and weak knowledge of the subject. In our past studies we observed that both quality of annotation and level of knowledge of weak students can be dramatically increased when strong students serve as peer reviewers for their annotations. Interesting is that we also observed a reverse trend: feedback provided by weak students in the form of questions and requests to expand annotation caused strong students to improve their annotation as well. In the study presented below, we attempted to put a more solid ground for the idea of pairing strong and weak students.

1. Literature Review

Peer Grader (PG) is a web-based peer grading system (Gehringer, 2000). It adopts simple mapping strategy by setting constraints in assigning reviewers. Crespo et al. (2006) proposed an adaptive strategy to map reviewers and authors based on user profiles and based on fuzzy classification techniques. It advances traditional peer-review matching by blindly assigning random associations. According to (Chi et al., 1989), students can learn a lot when attempting to explain examples. "Self-explanations," formulating the unwritten steps of an example or concept, help students understand examples and problems. Other cognitive science studies have shown that students acquired less shallow procedural knowledge by specifically creating their own explanation (Aleven & Koedinger, 2002). The benefits of generating self-explanations extend to explanations created in response to specific questions (Pressley, 1992).

2. System Description and Study Design

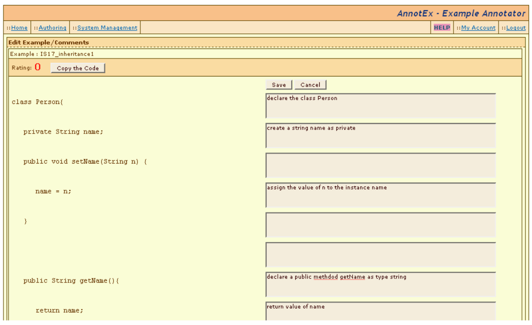


Figure 1 AnnotEx: authoring annotations interface. Figure 2 AnnotEx: reviewing annotations/comments interface

AnnotEx stands for Example Annotator System (http://adapt2.sis.pitt.edu/annotex/). It was developed to support community based authoring. The model is that students work in a group within a community. There are three stages of a complete example annotating process. The first stage is to author the annotation of the example. The second stage is to peer review others' work by providing ratings/comments on the example annotations. The third stage is to re-annotate the example annotations. The tasks appear once the student logs in. Figure 1 shows the interface where students author the annotations. Figure 2 shows the pages where the student can review others' annotations or comments.

There were 21 undergraduate students from an Object-Oriented Programming I class which was offered by School of Information Sciences, University of Pittsburgh in the Spring term of 2008. The study was carried out as an extra credit assignment. The assignment included three phases with three different tasks: Annotating, Peer-reviewing and Re-annotating. Examples covered three topics in java programming language, including

ArrayList, Interface and Inheritance. Each topic contains at least 4 same level examples that have different content. The average example has 45 lines of code.

Based on previous study results (Hsiao & Brusilovsky, 2007; Hsiao, 2008), superior students help improve the quality of annotations. To further investigate the annotation process, we hypothesized that by assigning stronger students to weaker ones the annotation quality would be improved more efficiently. In order to verify this hypothesis with the example-annotating Peer Review model, we divided a group of students with various levels of knowledge into two groups, weaker and stronger students, based on their previous average in-class quiz scores. The average in-class quiz below 50 percentile was defined as stronger students; above 50 percentile was defined as weaker students. We expected to facilitate the example annotation process and see weaker students take advantage of the presence of stronger students to maximize the benefit of Peer Review modeling.

- Phase 1 (Annotating): each student was given three examples to annotate, one topic each.
- Phase 2 (Peer-reviewing): each student had to provide ratings and comments for three annotated examples, covering three different topics. For the topic ArrayList, examples and students were randomly assigned to his/her peers, without regard to their knowledge level. For the topic Interface, examples were randomly assigned, but student knowledge levels were cross-assigned to match stronger students with weaker ones and vice versa. For the topic Inheritance, stronger students were assigned to weaker ones and vice versa, using exactly the same example content, but it was annotated by his/her peers (Table 1).
- Phase 3 (Re-Annotating): the three examples assigned to students in phase 1 were re-assigned back to them to author re-annotations.

Topic	*Same Example Content*	*Cross Assignment by Student Knowledge Level*
ArrayList	No	No
Interface	No	Yes
Inheritance	Yes	Yes

Table 1 Peer-Review Settings

3. Results and Analyses

18 out of 21 students completed the assignment. The annotations collected after phase 1 and re-annotations collected after phase 3 were passed through Expert Review for quality examination. There were a number of effects found in this study; discussion follows. The data summary is provided in Table 2.

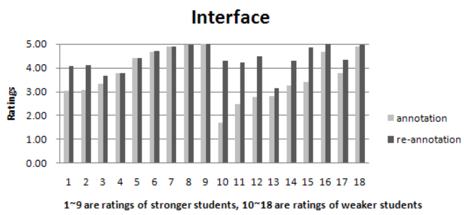
In this study, the ArrayList topic group was randomly assigned during peer-review. The stronger students did not have much growth. On the contrary, re-annotations resulted in a coherent outcome for weaker students. The overall ratings increased from 4.17 to 4.50, which is significant (Table 2). The results of higher quality re-annotations are consistent with our previous studies (Hsiao & Brusilovsky, 2007; Hsiao, 2008). Therefore, we call it the baseline group.

With the cross assignment of peer-reviewed students according to knowledge level, weaker students increased their re-annotation ratings and obtained a significant increase (Table 2). By reviewing the stronger students' annotations, weaker students managed to improve the quality of their annotations. Yet, we do not see this effect on the stronger students who reviewed weaker students' work. The final ratings increased, but not significantly. We also observed that weaker students tended to explain every line, while stronger students tended to annotate the key concepts only. The final increase of stronger students' ratings may be due to the number of annotations completed, so it was a quantity increase, instead of a quality increase, since they were already explaining the essential points and thus didn't have room for improvement. There is a significantly higher increase in overall ratings in the cross-assignment-by-student-knowledge-level group compared to the baseline group (Table 2).

Topic	***Student Knowledge Level***	***Annotation***	***Re-Annotation***	***p-value***
ArrayList	Stronger	4.34	4.41	.033*
	Weaker	3.99	4.58	.023*
	Overall	4.17	4.50	.014*
Interface	Stronger	4.13	4.40	.104
	Weaker	3.31	4.40	.004**
	Overall	3.72	4.40	.001**
Inheritance	Stronger	4.16	4.70	.022*
	Weaker	3.66	4.51	.021*
	Overall	3.91	4.61	.001**
Overall		3.93	4.50	.000**

* p-value <.05, ** p-value <.01

Table 2 Expert ratings before/after peer-review in three topic groups across student knowledge levels



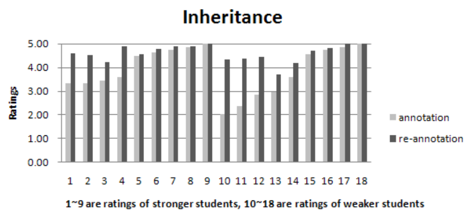


Figure 3 & Figure 4 Sorted expert ratings of experimental group a & b

In order to closely monitor the efficiency of the annotation process, the example content annotated by others was carefully distributed in the cross assignment. Both stronger and weaker students had significant increases in ratings (Table 2). For the stronger students, the final ratings reached 4.7 and became the highest result within this study. Comparing Figures 3 and 4, the ratings of re-annotations are more coherent in Figure 4, the same example content manipulation. The overall ratings were significantly increased too. The results outperformed those from the previous baseline group and experimental group a. The outcome indicated that keeping the example content the same while cross-assigning the students by student knowledge level produced a uniformly good quality of annotations.

A non-mandatory questionnaire was administered at the end of the assignment to collect students' opinions and suggestions. 13 out of 18 students completed the survey. 84.6% of the students were satisfied with the overall annotating and peer-reviewing experience. 100% of the students agreed or strongly agreed about the need for such a tool in general. 84.6% of the students agreed or strongly agreed that reviewing and rating the annotations of others, in phase 2, helped them to improve her/his annotations in phase 3.We also asked students to order the activities based on how much they contributed to their learning. Students, across the board, considered that the flow of the existing annotation process which contributed most to their learning was this one: begin by 1) annotate examples, 2) peer-review others' work, 3) read review comments about their own annotations, and finally 4) re-annotate the initial example.

4. Discussions

In the baseline group setting, we once again proved that the student community is capable of creating valuable content. In order to highlight the stronger students' advantage, we cross-assigned reviewers by student knowledge level during the peer-review phase. Weaker students made use of the stronger students' reviews and obtained a significant growth in the quality of their re-annotations. Stronger students did not obtain a significant growth in re-annotations, but found only a marginal increase in terms of annotation quantity. The accountability of significance in the increase of overall ratings is higher in cross-assignment-by-student-knowledge-level group than baseline group. With the same example content and cross-assignment by student knowledge level, students managed to produce a uniformly good quality of annotations. The results uphold our hypotheses; we can not only harness the students' power to create example annotations, we have also utilized the stronger students' power to assure the annotation quality. It opens up a number of research possibilities for the future, such as utilizing the existing student model and assigning reviewers adaptively or providing simultaneous support during peer-review.

Acknowledgements

This work was supported by National Science Foundation under Grants IIS0447083.

References

[1] Aleven, V. & Koedinger, K., (2002) An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor, Cognitive Science: A Multidisciplinary Journal, 26(2), 147-179

[2] Arroyo, I. & Woolf, B. P. (2003) Students in AWE: changing their role from consumers to producers of ITS content. Advanced Technologies for Mathematics Education Workshop. Supplementary Proceedings of the 11th International Conference on Artificial Intelligence in Education.

[3] Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R., (1989) Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science, 13, 145-182

[4] Crespo, R.M., Pardo,A. & Kloos, C.D (2006) Adaptive Peer Review Based on Student Profiles. Intelligent Tutoring Systems. Lecture Notes in Computer Science. Volumen 4053/2006, páginas 781-783. ISBN: 978-3-540-35159-7. Springer, Boston.

[5] Gehringer, E. F., (2000) Strategies and mechanisms for electronic peer review, 30th ASEE/IEEE Frontiers in Education Conference, October 18 - 21, 2000 Kansas City, MO

[6] Hsiao, I.-H. & Brusilovsky, P. (2007) Collaborative Example Authoring System: The Value of Re-annotation based on Community Feedback. In: T. Bastiaens and S. Carliner (eds.) Proceedings of World Conference on E-Learning, E-Learn 2007, Quebec City, Canada, October 15-19, 2007, AACE, pp. 7122-7131

[7] Hsiao, I. (2008) The Role of Community Feedback in the Example Annotating Process: an Evaluation on AnnotEx. In: YRT, Intelligent Tutoring Systems. (Proceedings of 8th International Conference on Intelligent Tutoring Systems, ITS'2008, Montreal, Canada, June 23-27, 2008)

[8] Hundhausen, C.D., Douglas, S.A. (2000) Using visualizations to learn algorithms: Should students construct their own, or view an expert's? In: Proc. of IEEE Symposium on Visual Languages. IEEE Computer Society Press 21-28

[9] Jonassen, D. H. & Reeves, T. C. (1996). Learning with technology: Using computers as cognitive tools. In Jonassen, D. H. (Ed.), Handbook of Research for Educational Communications and Technology (pp. 693-719). New York, NY: Mc Millan.

[10] Pressley, M., Wood, E., Woloshyn, V.E., Martin, V., King, A. and Menke, D. (1992) Encouraging mindful use of prior knowledge: attempting to construct explanatory answers facilitates learning. Educational Psychologist, 27(1):91–109