KSC-PaL: A Peer Learning Agent

Cynthia Kersey[1], Barbara DiEugenio[2], Pamela Jordan[3], Sandra Katz[4]

[1] Lewis University
[2] University of Illinois at Chicago
[3] LRDC and Biomedical Informatics, University of Pittsburgh
[4] LRDC, University of Pittsburgh

Abstract. We have developed an artificial agent based on a computational model of peer learning we developed. That model shows that shifts in initiative are conducive to learning. The peer learning agent can collaborate with a human student via dialog and actions within a graphical workspace. This paper describes the architecture and implementation of the agent and the user study we conducted to evaluate the agent. Results show that the agent is able to encourage shifts in initiative in order to promote learning and that students learn using the agent.

Key words: Peer Agent, Knowledge Co-construction, Initiative

1 Introduction

Research shows that collaboration promotes learning, potentially for all of the participants [2, 7, 12]. Similarly, studies in peer tutoring demonstrate that there are cognitive gains for both the tutor and the tutee [1, 5, 11]. However, the study of peer learning from a computational perspective is still in the early stages. Although some researchers have attempted to develop simulated peers [3, 14], there is very little research on what constitutes effective peer interaction to guide the development of effective peer learning agents.

In our previous work we derived a model of peer interactions that was suitable for incorporation in an agent. This model operationalizes *Knowledge Co-construction* [8] via the notion of initiative shifts in dialogue. We have incorporated this model in an innovative peer learning agent, KSC-PaL, that is designed to collaborate with a student to solve problems in the domain of computer science data structures.

This paper presents the details of the implementation and evaluation of KSC-PaL. We start by summarizing the computational model of peer learning that is incorporated into the agent, followed by a description of the system design and architecture. We conclude with the results of the user study we conducted to evaluate the agent.

2 Computational Model

We have performed an extensive corpus analysis [10] in order to derive a computational model of Knowledge Co-construction (KCC). This construct explains the effectiveness of peer learning by postulating that learning is enhanced when students work together to construct knowledge. An earlier study by Hausmann et al. [8] had extended the analysis of KCC by incorporating relations, such as elaborate and criticize, within KCC episodes. However, our analysis found that these relations were not only difficult to identify but did not correlate with learning in our corpus. Hence, we looked for simpler but principled correlates of KCC. We found those in the linguistically motivated notion of *initiative shifts* in dialogue. Our analysis found a strong relationship between initiative shifts and KCC episodes. A paired t-test showed that there were significantly more initiative shifts in the annotated KCC episodes compared with the rest of the dialogue ($t(57) = 3.32$, $p = 0.0016$). The moderate effect difference between the two groups (effect size = 0.49) shows that there is a meaningful increase in the number of initiative shifts in KCC episodes compared with problem solving activity outside of the KCC episodes. Additionally, we found moderate correlations of learning with both KCC ($R^2 = 0.14$, $p = 0.02$) and with initiative shifts ($R^2 = 0.20$, $p = 0.00$).

Since the corpus analysis showed a correlation between initiative and KCC and between initiative and learning, the next step was to identify ways for KSC-PaL to encourage such shifts in initiative. We explored two different methods to do so. One method is based on the observation that student uncertainty (hedging) may lead to a shift in initiative. The other is based on related literature [4, 15] which shows that certain conversational cues, other than hedging, lead to shifts in initiative. Our analysis showed that the following cues were most likely to lead to initiative shift and to increase knowledge score (which was computed using the student model described in section 3.3): hedging, using prompts, making mistakes intended to incite student criticism and requesting feedback.

3 KSC-PaL

Based on this analysis, we developed a peer learning agent, KSC-PaL. The core of KSC-PaL is the TuTalk system [9]. TuTalk is a dialogue management system that supports natural language dialogues for educational applications and allows for both tutorial and conversational dialogues. In developing the agent we extended TuTalk by adding a graphical user interface, replacing TuTalk's student model and augmenting TuTalk's planner to implement the model discussed above.

The interface manages communication between TuTalk and the student. Students communicate with the agent using typed natural language and graphical actions within a graphical user interface. The student input is processed by the interface and its related modules into an appropriate format and passed to TuTalk. Since TuTalk's interpretation module is not able to appropriately handle all student utterances and we wanted to avoid interpretation issues impacting our results, a human interpreter assists in this process. Additionally TuTalk requests assistance from the Student Model/Dialogue Planner (SMDP) to manage the dialogue in order to appropriately shift initiative and encourage learning. These modules are described below in more detail.

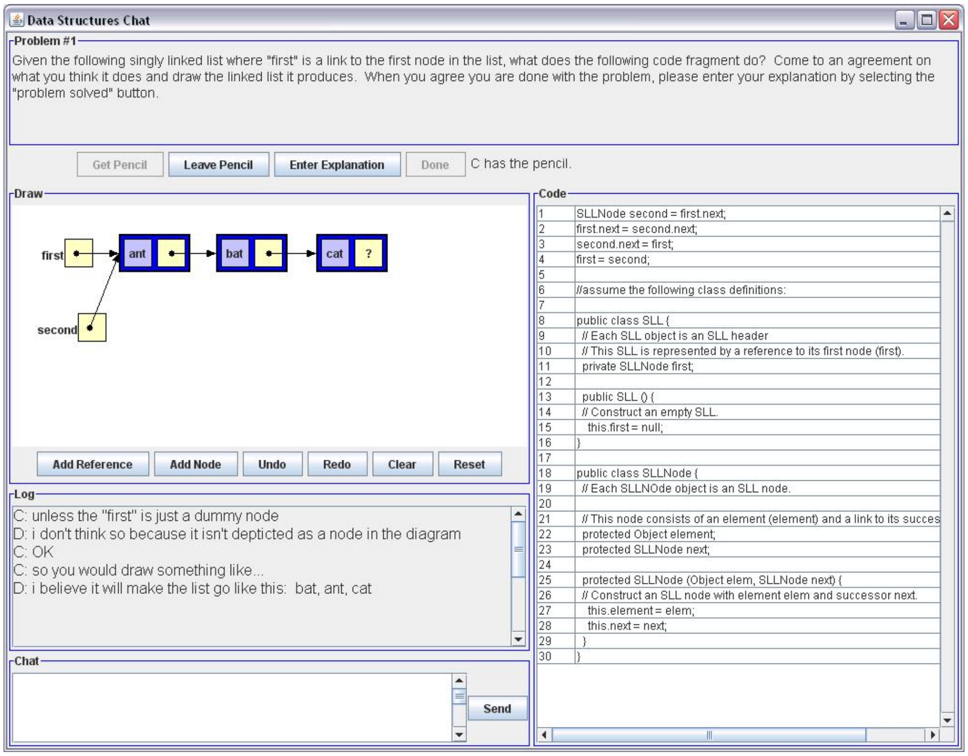

Fig. 1. KSC-PaL user interface

3.1 Interface

The user interface consists of four distinct areas (see figure 1) :

1. Problem display: Displays the problem description.
2. Code display: Displays the code from the problem statement.
3. Chat Area: Allows for typed user input and an interleaved dialogue history of the student and the agent.
4. Drawing area: Here users can diagram data structures to aid in the explanation of parts of the problem being solved. The drawing area has objects representing nodes and links that can be used to build lists.

The interface includes a preprocessor module which takes as input a student's utterances and actions and modifies them so that they can be recognized by TuTalk. This preprocessor consists of a spell corrector and a graphical actions interpreter that interprets the student's drawing and coding actions and passes them to TuTalk as natural language utterances.

3.2 Human Interpreter

Given the limitations of current technology for natural language understanding, a human interpreter was incorporated to assist in the disambiguation of student utterances. The interpreter receives a student utterance along with a list of possible matching concepts from TuTalk. The interpreter then selects the most likely matching concepts from TuTalk, thus assisting in natural language interpretation. If the student utterance doesn't match any of these concepts, a second list of concepts, containing student initiative utterances, are presented to the interpreter. If none of these match then all known concepts are presented to the interpreter for matching. Note that the interpreter has a limited, predetermined set of choices, corresponding to the concepts that TuTalk is aware of. In this way, his/her intervention is circumscribed.

The interpreter also plays a role in the interpretation of graphical actions. The natural language interpretation of the drawing or coding action is first sent to the interpreter. He/she then verifies that the interpretation is valid or selects an alternate interpretation from a list of known graphical actions and sends it on to TuTalk for processing.

Additionally, since interpreting student input of a solution would require extensive natural language processing, the interpreter also matches the solutions entered by the student to a limited range of possible solutions, such as *correct*, *incomplete* or *incorrect*.

3.3 Student Model/Dialogue Planner (SMDP)

KSC-PaL's planner selects scripts and responses to student initiative to manage initiative shifts. TuTalk uses *scenarios* for guiding the dialogues. These scenarios contain both the *recipe* (script) and *concepts*, which are linguistic concepts used to realize the dialogue. Scripts are hierarchical in nature and consist of a sequence of goals for addressing a topic. Goals usually involve multiple steps where each step consists of an initiation followed by one or more responses. Generally, the initiation is an agent utterance and responses are possible ways in which a student can respond. However, when using mixed-initiative, the initiation could represent a student utterance while the responses are potential agent replies to the student's utterance. Additionally, TuTalk allows for alternative recipes to achieve a goal.

In drafting the scripts for KSC-PaL, we authored goals that would encourage shifts in initiative as well as goals that would not encourage initiative shifts. Similarly in drafting responses to student initiative, we drafted both initiative-shifting responses as well as responses that would not likely shift initiative. The agent encourages initiative shifts by using prompts, hedging, requesting feedback from the student and encouraging student criticism by intentionally making errors in problem solving. TuTalk's planner does not manage these options to the level required by the agent, so a planning module was added to make choices on goal implementation and agent response with the objective of managing shifts in initiative.

This planner was combined with the student model to create the Student Model/Domain Planner (SMDP). The SMDP consists of a server that manages communication with TuTalk, a student model, an initiative module that tracks initiative shifts and a planner that makes decisions based on the current state of initiative and student knowledge.

Student Model The agent requires a student model to track the current state of problem solving as well as estimate the student's knowledge of concepts involved in solving the problem in order to guide its behavior. Since TuTalk's student model does not provide these capabilities, a student model which incorporates problem solution graphs [6] was added to the agent. Solution graphs are Bayesian networks where each node represents either an action required to solve the problem or a concept required as part of problem solving. A user's utterances and actions are then matched to these nodes. This provides the agent with information related to the student's knowledge as well as the current topic under discussion.

Initiative Tracker On receiving a student or agent utterance or action from the SMDP server, the initiative tracker codes the turn with either student initiative or agent initiative. The tracker contains a classifier for natural language utterances and a separate classifier for drawing and coding actions. Natural language utterances are parsed using the Stanford Maximum Entropy Tagger [13] to provide the appropriate features for use by the initiative classifier. When classifying a drawing or coding action, the initiative tracker retrieves the student knowledge score for use by the classifier. Once the turn is classified, it is determined whether a shift in initiative has occurred by comparing the current classification with the classification of the previous turn.

When requested by the planner, the initiative tracker returns the average level of initiative shifts. This is computed by dividing the number of initiative shifts by the total number of turns.

Planner Module Requests for goal implementation and requests for agent response are managed by the planner module. Two factors determine whether a goal implementation or response that encourages an initiative shift will be selected: (1) the current level of initiative shifts and (2) the change in the student's knowledge score. Initiative shifts are tracked using the initiative tracker module described above and knowledge levels are maintained in the student model. Goals or responses are selected to encourage initiative shifts when the average level of initiative shifts is less than 0.2117 (mean initiative shifts in KCC episodes as calculated from corpus data) and the student's knowledge level has not increased since the last time a request for goal implementation or response was requested.

If the planner has determined that an initiative shift should be encouraged, it selects among alternatives based on the holder of initiative in the previous utterance/action and a label associated with each of the potential goal implementations or responses. For example, to encourage an initiative shift when the initiative holder for the previous utterance/action was the student and the agent has the choice of goal implementations labeled *correct*, *partial-correct* and *incorrect*, the agent will select the goal implementation labeled *correct* because it is likely to result in a shift of initiative

4 Evaluation

We developed two versions of KSC-PaL to test the effectiveness of the model of KCC described above. In the *experimental* version of the agent (PaL), goal versions and responses to student utterances are selected by the planner to maintain a high level of shifts in initiative. In the *control* version (PaL-C), the planner is not consulted for goal versions or responses. Additionally, the script was modified to remove those utterances that were identified as likely to shift initiative: incorrect statements, hedges, prompts and requests for feedback.

4.1 User Study

We collected interactions of 25 students, where 13 interacted with PaL and 12 interacted with PaL-C. At the beginning of the session, each student was given a five question pre-test to evaluate his or her knowledge prior to interacting with the agent. Prior to problem solving, the students were given a short tutorial on using the interface. They then solved two linked list problems with the agent. At the conclusion of problem solving, students were given a post-test, identical to the pre-test. Additionally, they were asked to fill out a questionnaire to assess their satisfaction with the agent.

4.2 Effect on Learning

In order to investigate whether students learned using KSC-PaL, we first performed a paired t-test of pre-test and post-test scores. This analysis showed that overall students learn using KSC-PaL (see table 1). T-test analysis also shows that there is a significant difference between pre-test and post-test in the experimental condition and a trend toward a significant difference in the control condition. However, there is no significant difference between the gains in the two groups.

4.3 Initiative Shifts and Learning

In both conditions, the agent tracks the initiative holder in each utterance using the classifiers described above. A manual annotation of these utterances showed that the level of accuracy of the classifiers was as expected. 747 of 937 of utterances and drawing actions (80.15%) were correctly classified. However, in order to evaluate the effectiveness of initiative shifts, the following analysis uses the utterances manually annotated for initiative.

Table 1. Learning using KSC-PaL

Condition	N	Pre-test M	Post-test M	gain	t	p
KSC-PaL (all students)	25	0.61	0.68	0.07	2.90	0.01
PaL	13	0.60	0.66	0.06	2.55	0.02
PaL-C	12	0.62	0.69	0.07	2.03	0.06
PaL plus upper quartile PaL-C subjects	18	0.61	0.68	0.07	3.29	0.00
PaL-C less upper quartile PaL-C subjects	7	0.66	0.66	0.00	-0.96	ns

To examine the impact of initiative shifts on learning, we used two measures of shifts: (1) the number of shifts and (2) normalized initiative shifts, calculated by dividing the number of initiative shifts by the total number of utterances and drawing actions for the session.

Given that the control condition does not encourage initiative shifts but neither does it prevent them, we combined the experimental subjects with those control subjects whose interactions showed high levels of initiative shifts, i.e. where the amount of initiative shifts falls in the upper quartile of the number of initiative shifts for the combined group. As shown in table 1, when compared with the remaining control condition subjects there is a difference in learning between the two groups. A t-test performed on the gains between the two groups showed the difference is significant ($t = 2.35$, $p = 0.03$) . The effect size (d) is 0.18 which is considered a moderate difference.

Additionally, using multiple linear regression, the measures described above were used as predictors of post-test score after regressing out the impact of pre-test score. Table 2 shows that while the correlations are significant or trending toward significance, the impact is relatively small. If this same analysis is applied to those subjects with a pre-test score below the mean, there is a larger impact of initiative shifts on post-test score. Analysis of high pre-test subjects showed no significant correlation of post-test score with initiative shifts or normalized initiative shifts.

Table 2. Impact of Initiative Shifts on Learning

Predictor of Post-test	β	R^2	p
Initiative shifts	0.24	0.03	0.06
Normalized initiative shifts	0.28	0.01	0.02
Low pre-test subjects (n=14)			
Initiative shifts	0.45	0.07	0.09
Normalized initiative shifts	0.49	0.16	0.04

4.4 Agent's Ability to Shift Initiative

In the experimental condition, KSC-PaL attempts to shift initiative in order to maintain a certain level of initiative shifts. In retrospect, this threshold appears to be set too low, since KSC-PaL rarely selected responses or goal implementations that would encourage a shift in initiative. Only 34 of the 200 requests for response or goal implementation (17%) resulted in a selection to shift initiative.

Therefore, in order to examine the effectiveness of encouraging initiative shifts, we used an alternative method. As mentioned above, the script for the experimental condition included agent utterances that encourage initiative shifts, including instances where no request for agent response or goal selection would be made. These types of utterances were generally excluded from the script for the control condition. Thus, the students in the experimental condition were more likely to encounter those utterances that encourage initiative shifts. To examine the impact of this difference, the dialogues in the user study were semi-automatically annotated with the following encouragers of initiative shifts:

- hedge
- request for feedback
- incorrect statements
- prompts

This was accomplished by collecting all of the agent responses and identifying those responses that fall into one of the categories listed above. Since the agent has a limited set of responses, the transcripts were queried for matching utterances and automatically coded with the appropriate labels.

First we investigated whether the number of shifts encouraging utterances had an impact on learning by using multiple regression to predict post-test score using pre-test score + initiative shift utterances. This was not statistically significant.

We then ran a t-test to see if the number of utterances tagged as shift encouragers differed between control sessions and experimental sessions. An unpaired t-test showed that they were significantly different ($t = 3.28$, $p = 0.0036$). We then used linear regression to see if there was a relationship between the number of these shift inducing utterances and number of initiative shifts that occurred. This was also significant ($\beta = 0.40$, $R^2 = 0.16$, $p = 0.04$). This result suggests that these shift encouragers do have an impact on the number of initiative shifts that occur during a problem solving session.

4.5 Student Satisfaction

At the conclusion of problem solving, students were asked to complete a short survey related to their satisfaction using KSC-PaL. The survey consisted of statements to which the students were asked to rate their level of agreement with. Responses were on a 5 point Likert scale, with 1 representing strongly disagree and 5 representing strongly agree. The statements on the survey are shown in Table 3.

Table 3. Student Survey - Average Responses

Statement	Control Condition		Experimental Condition	
	M	sd	M	sd
The agent helped me learn about linked lists	3.54	0.97	3.08	1.38
Working with the agent is like working with a classmate	3.23	1.23	3.38	1.26
I would use the agent on a regular basis, for other topics (like trees)	3.92	0.86	3.31	1.11
The agent understands what I am saying	3.77	1.16	3.23	1.09
The agent responds appropriately to what I am saying	3.54	1.33	3.46	1.26
I found what the agent said repetitive	3.00	0.91	3.08	1.32
I felt like I had control over solving the problems, and the agent wasn't trying to take charge too often.	1.49	3.31	3.69	1.38

There were no significant differences between the responses to these questions for those in the control condition versus those in the experimental condition suggesting that attempting to shift initiative does not have a negative impact on student satisfaction with the agent.

5 Conclusion and Future Work

We implemented a peer learning agent, KSC-PaL, based on the results of an extensive corpus analysis that showed that KCC episodes could be identified from shifts in initiative. KSC-PaL is an innovative peer learning agent in that it attempts to shift initiative between itself and the student. Therefore, unlike other peer learning agents, it shifts roles from more experienced peer to less-experienced peer within a single problem-solving episode. Our evaluation of KSC-PaL found that students learned using the agent. Although there was no significant difference between the conditions, we found those students whose interactions with the agent had higher normalized initiative shifts, regardless of condition, learned more. We also found that this effect was more pronounced for students who began with a lower level of initial knowledge regarding linked lists. Additionally, in the experimental condition, KSC-PaL was successful in encouraging shifts in initiative using the identified shift encouraging cues and these attempts to shift initiative did not have a negative impact on student satisfaction with the agent.

Since in the current implementation of KSC-PaL, the agent choses to shift initiative based on a fixed level of average initiative shifts, future work will explore varying the threshold for initiative shifts. There may be some ideal level of initiative shifts that encourages learning without decreasing student satisfaction with the agent. Additionally, we plan to incorporate more sophisticated natural language understanding technology into KSC-PaL. With improved NLU, the human interpreter could be removed from the system. This would allow the system to be deployed in classrooms or potentially on the Internet.

References

1. M. W. Birtz, J. Dixon, and T. F. McLaughlin. The effects of peer tutoring on mathematics performance: A recent review. *B. C. Journal of Special Education*, 13(1):17–33, 1989.
2. A. L. Brown and A. S. Palincsar. *Guided, cooperative learning and individual knowledge acquisition*, pages 307–226. Lawrence Erlbaum Associates, Hillsdale, NJ, 1989.
3. Tak-Wai Chan and AB Baskin. Studying with the prince. In *Proceedings of the ITS-88 Conference*, pages 194–200, 1988.
4. Jennifer Chu-Carroll and Michael K. Brown. An evidential model for tracking initiative in collaborative dialogue interactions. *User Modeling and User-Adapted Interaction*, 8(3–4):215–253, September 1998.
5. P.A. Cohen, J.A. Kulik, and C.C. Kulik. Education outcomes of tutoring: A meta-analysis of findings. *American Education Research Journal*, 19(2):237–248, 1982.
6. Cristina Conati, Abigail Gertner, and Kurt VanLehn. Using Bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction*, 12(4):371–417, 2002.
7. E. Fisher. Distinctive features of pupil-pupil classroom talk and their relationship to learning: How discursive exploration might be encouraged. *Language and Education*, 7:239–257, 1993.
8. Robert G.M. Hausmann, Michelene T.H. Chi, and Marguerite Roy. Learning from collaborative problem solving: An analysis of three hypothesized mechanisms. In K.D Forbus, D. Gentner, and T. Regier, editors, *26th Annual Conference of the Cognitive Science Society*, pages 547–552, Mahwah, NJ, 2004.
9. Pamela W Jordan, Brian Hall, Michael A. Ringenberg, Yui Cue, and Carolyn Penstein Rosé. Tools for authoring a dialogue agent that participates in learning studies. In *Artificial Intelligence in Education, AIED 2007*, pages 43–50, 2007.
10. Cynthia Kersey, Barbara Di Eugenio, Pamela Jordan, and Sandra Katz. Ksc-pal: a peer learning agent that encourages students to take the initiative. In *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 55–63. Association for Computational Linguistics, 2009.
11. M. D. Rekrut. Teaching to learn: Cross-age tutoring to enhance strategy instruction. *American Education Research Association*, 1992.
12. Tan Bee Tin. Does talking with peers help learning? the role of expertise and talk in convergent group discussion tasks. *Journal of English for Academic Purposes*, 2(1):53–66, 2003.
13. Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. In *NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pages 173–180, Morristown, NJ, USA, 2003. Association for Computational Linguistics.
14. Aurora Vizcaíno. A simulated student can improve collaborative learning. *International Journal of Artificial Intelligence in Education*, 15(1):3–40, 2005.
15. Marilyn Walker and Steve Whittaker. Mixed initiative in dialogue: an investigation into discourse segmentation. In *Proceedings of the 28th annual meeting on Association for Computational Linguistics*, pages 70–78, Morristown, NJ, USA, 1990. Association for Computational Linguistics.