Exploring Constructive Learning Activity in Online Programming Discussion Forums

I-Han Hsiao, Arizona State University, Sharon.Hsiao@asu.edu
Piyush Awasthi, Arizona State University, pawath1@asu.edu

Abstract: In this paper, we explored constructive learning activities in an online programming discussion forum. We built a logistic regression model based on the underlined cognitive processes in constructive learning activities. The results confirmed that the crowd perceived Accepted Answers were likely to contain more constructive words. More importantly, users had as many interactions with Answers as with Accepted Answers, disregarding the quantity of constructive words. They especially bookmarked more and up voted more in difficult Answers when the content had more constructive words. The findings supported that passive-proactive behavior exists and suggested that detecting constructive content can be a helpful classifier in discerning relevant information to the users, and in turn optimizing learning opportunities.

Keywords: learning activity, engagement activity, discourse analysis, constructive learning, forum

Introduction
The juncture of Intelligent Tutoring Systems/Artificial Intelligence in Education (ITS/AIED) & Learning Science/Computer Supported Collaborative Learning (LS/CSCL) literature has successfully demonstrated that students can learn from a wide range of dialogue-based instructional settings, such as dialogic-based tutor, asynchronous discussion forums, etc. (Aleven, McLaren, Roll, & Koedinger, 2006; Chi, Roy, & Hausmann, 2008; Muldner, Lam, & Chi, 2014; VanLehn et al., 2007). Recently, studies show an alternative instructional context by learning from observing others learn (Chi et al., 2008) and is considered as a promising learning paradigm (Muldner et al., 2014). It suggests lurkers can still learn by reading the postings-and-replies exchanges from others due to the constructive responses in the content and if the lurkers also engage in some sort of activities (Chi & Wylie, 2014). Such paradigm addresses a major limitation on development time in ITSs & liberated the domains from procedural skills to less structured fields. However, to what extend can we capitalize passive-proactive learning activity (i.e. reading others’ constructive dialogues voluntarily and later engage in some sort of learning activity)? In the context of programming learning, can we successfully model users’ learning activities in such large-scaled open corpus environment? In this paper, we focus on researching the passive-proactive behavior and exploring the associated constructive learning activities in an online programming discussion forum. We aim to use quantitative approach to evaluate and harness students’ learning activities in massive online discussion forums. To do so, we first apply ICAP (Interactive, Constructive, Active, Passive) learning activity framework, which defines “learning activities” as a broader and larger collection of instructional or learning tasks. It allows educational researchers to explain subtle engagement activities (invisible learning behaviors) (Chi, 2009; Chi & Wylie, 2014; Muldner et al., 2014). The framework examines comparable learning involvement, where Interactive modes of engagement achieve the greatest level of learning, then the Constructive mode, then the Active mode, and finally, at the lowest level of learning, the Passive mode. It will allow us to engineering algorithms to capture learning activity features in predicting learning outcome and estimating knowledge transformation. At current stage, we focus on constructive learning activities and neglect other activities. We consider constructive activities that include the following possible underlying cognitive processes, inferring, creating, integrating new with prior knowledge, elaborating, comparing, contrasting, analogizing, generalizing, including, reflecting on conditions, explaining why something works. Based on these cognitive processes, we build a constructive lexicon library. For instance, comparing & contrasting words, explanation, justification & elaboration words, etc.)

Preliminary findings
We sampled one year (year 2013) of forum posts in topic Java from StackOverflow site through StackExchange API (http://stackexchange.com, a question and answer website network for various fields). The data pool was selected from the top 10 frequent tagged questions due to most the posts in this section contained at least one accepted answer. It will allow us to build a baseline on the answer quality according to crowdsourced votes. There are total 16,739 posts, including 3,725 questions, 13,014 answers, with 3,718 accepted answers.
We consider two dimensions of features of constructive engagement activities 1) Social aspects features, including posting votes, poster reputation, poster status & 2) Content related features, including code snippets, content syntactic (length, average sentence per thread), and most importantly the constructive lexicons.

We performed logistic regression analysis. The full model was able to successfully predict user engagement at 0.05 level, $\text{adjusted-R}^2 = 0.256$. We tested the goodness of the models reserving 20% of the observations for testing with 10-fold cross validation (MAE$_{10FOLD}$=7.08) and selected a final model.

We found that there are significant more constructive words within Accepted Answer ($M=0.827$, $SE=1.334$) than Answers ($M=0.583$, $SE=1.005$), $p<0.01$ (Table 1). The result confirmed that the answers accepted by the crowd not only agreed as correct among the best available answers, but also contain higher constructive information. Accepted Answers also showed a positive correlation between user favorites and the amount of constructive words ($r=0.0781$, $p<0.01$), but we did not see such correlation between Questions/Answers and the amount of constructive words. This result is not surprising. It indicates the community tends to bookmark useful Accepted Answers, but not Questions nor Answers. However, we found the community provided as many votes to Answers and Accepted Answers, no matter how constructive the content was. This observation was very interesting and revealed that the community may not bookmark the Answers as frequent as they do to Accepted Answers, but it did show the effort to screen the Answers and provide votes to them.

We further divided the content into two categories, Easy & Difficult (based on the topics covered in CS1 or CS2 courses). Easy topics include Classes, Objects, Loops, ArrayLists etc.; difficult topics contain Inheritance, Recursion, Multithreading, User Interfaces etc. We found that easier content had slightly higher constructive words than difficult content, but it was not significant. It was understandable that simpler problems may be easier to provide examples and tougher problems may require more efforts to justify the answers. However, we found that among Answers, users bookmarked more and up voted more in difficult content when the content had also more constructive words. But we saw no such pattern in Accepted Answers or in Questions. This again showed important evidence that the users in the community spending efforts in locating relevant information to themselves, even the answers are not the accepted ones. These results support the passive-proactive learning behavior, which users did not just read the Accepted Answers, but also Answers, and further provided some sort of actions. The findings also suggest that detecting constructive content can be a helpful classifier in discerning relevant information to the users, and in turn providing learning opportunities.

Table 1: Constructive word counts by content types and difficulties

<table>
<thead>
<tr>
<th>Topic/Type</th>
<th>Question</th>
<th>Accepted Answer</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>0.956±1.253</td>
<td>0.959±1.385</td>
<td>0.646±1.035</td>
</tr>
<tr>
<td>Difficult</td>
<td>0.984±1.355</td>
<td>0.827±1.294</td>
<td>0.583±0.981</td>
</tr>
<tr>
<td>Average</td>
<td>0.971±1.309</td>
<td>0.827±1.334</td>
<td>0.583±1.005</td>
</tr>
</tbody>
</table>

References


