Synthesizing Quantitative Predictors for Interaction in an Asynchronous Online Course

Daniel Zingaro, Murat Oztok, Jim Hewitt – OISE – University of Toronto, 252 Bloor St.W. Toronto, Canada
Email: [daniel.zingaro,murat.oztok, jim.hewitt]@utoronto.ca

Abstract: The effectiveness and potential of asynchronous online courses hinge on sustained, purposeful collaboration. And while many factors affecting interaction have been uncovered by prior literature, there are few accounts of the relative importance of these factors when studied in the same online course. In this paper, we develop a literature-informed model of six predictors on the likelihood that a note receives a reply. We corroborate earlier findings (such as the impact of the date that the note was posted) but also obtain one contradictory result (that reading ease does not appear to be a significant predictor). We offer hypotheses for our findings and suggest future directions for this type of research.

Introduction

From a social constructivist perspective, online learning is seen to be effective when it permits the exchange of views and allows students to learn through social communication (Swan, 2005). Lengthy discussions do not inherently suggest deep processing or collaborative meaning-making, but it is unlikely that such processes could occur in the absence of sustained discourse. Many authors, therefore, have offered hypotheses and implemented supports toward the growth of online discussion. For example, the specific discussion tool used by students (Guzdial, 1997) and habitual reading and writing practices (Hewitt, 2003) dramatically influence thread structures and lengthy communication. In the current paper, we add to this body of knowledge through a study of note-level features related to interaction. Why do some notes get replies but others do not? We are interested in a binary (yes/no) classification model, the level of prediction accuracy we can reach with such a model, and the importance of individual predictors in determining whether a reply is predicted. When a note does not receive any replies, the thread of discussion of which that note is a part also necessarily terminates. We believe that understanding which notes are likely to receive replies is a step toward understanding which lines of discussion are likely to be elaborated upon by students and which are likely to falter.

Measured Variables

We searched the literature for measurable aspects of notes that might contribute to whether they would receive a reply. We arrived at six literature-informed hypotheses and used a logistic regression model to assess the significance of these possible predictors in an online education graduate course taking place at a large Canadian research university. Our six independent variables are as follows:

- *Posting date.* The course we analyzed was structured into weekly discussion modules, and such a structure is not amenable to discussions that extend past their designated week (Hewitt, 2003). We predicted a reply advantage for those notes written early in the week. The variable name in the model is **date**, suffixed with 2 (mid-week), 3 (weekend), or 4 (beginning of next week) to indicate its value compared to the baseline of 1.
- *Active participant.* Research suggests that, in terms of percentage of notes posted, no single student dominates the communication, and that most notes are produced by the collectively low-posting students (Guzdial, 1997). In spite of this, we wondered if high-posting students may have a more indirect effect on the discourse, in the sense that their messages might be more likely to receive replies. We therefore coded notes into two categories depending on whether they were written by a frequent poster (top half of the class in terms of post count, and the baseline value of the variable) or infrequent poster. The variable name is **inactive**, and compares inactive participants to the active-participant baseline.
- *Reading ease.* In a study of 37 graduate courses, Hewitt and Peters (2007) found that courses with high reading ease were more interactive (i.e. contained more notes written as replies) than courses with low reading ease. We have added Flesch reading ease to our model (variable name **ease**, mean 101.9, sd 14.4) in order to determine whether it is a robust predictor of replies, or whether it is measurable only as a broad, course-wide effect.
- *Word count.* Prior literature suggests that students are more likely to quickly skim longer notes as compared to shorter notes. As notes reach 500 words, for example, half of their readings occur at a rate indicative of skim-reading (Hewitt, Brett, & Peters, 2007). We hypothesized that word count (variable **wordcount** below) would be a negative predictor of reply-likelihood, based on a guess that students would be less likely to reply to a note that they quickly skimmed. The mean word count was 180.8, with sd 142.3.

- *Student or teacher.* In the course analyzed here, the instructor quantitatively had only moderate presence: 61 notes were written by the instructor, compared to 1095 by students. We wondered whether instructor-authored notes, even in minority, would shape dialogue relatively more forcefully than the bulk of the notes written by students. This teacher-student variable is referred to as **prof**; the baseline is that the note was written by a student.
- *Contains a question.* We suspected that notes containing at least one question would garner more replies than notes containing no questions. Question-asking is a common and important discussion-facilitation technique (Hew & Cheung, 2008). Lengthy threads often contain at least one note that asks students for clarifications or personal opinions on a matter, and it is hypothesized that notes with questions obligate students to reply (Hew & Cheung, 2008). Below, the variable **question** is coded as "no" (no question in the note, the baseline value) or "yes", containing 660 and 496 notes, respectively.

Results

The overall model was significant (chi-square = 123.00, **df** = 8, p = 0); Nagelkerke's pseudo R-squared = 0.128. The coefficients, standard errors of the coefficients, Wald Z-statistics, p-values, and odds change in reply probability for a one unit increase in each predictor are given in table 1. (**wordcount** was log-transformed in order to satisfy the logit-linearity assumption.)

Table 1: Logistic regression model

Predictor	**B**	**SE**	Wald **Z**	**p**	**Exp(B)**
intercept	-3.01	0.86	-3.48	0	0.05
date2	-0.40	0.19	-2.06	.039	0.67
date3	-1.07	0.20	-5.37	0	0.34
date4	-0.98	0.27	-3.70	.0002	0.37
wordcount	0.47	0.08	5.51	0	1.60
prof	-0.35	0.29	-1.22	0.22	0.70
ease	0.01	0.005	1.83	0.07	1.01
inactive	-0.02	0.15	-0.12	0.91	0.98
question	0.56	0.13	4.24	0	1.76

Word count and reading ease are the two results discordant with our hypotheses; the latter also contradicts prior literature. Perhaps the fact that longer notes are more likely to receive a reply is due to these notes containing more information and hence more opportunities for students to find something of interest. Even though such notes are more often scanned, it is possible that students search such notes for isolated aspects to which they can reply. On reading ease, we suggest that it is the social nature of the course at large, rather than reading ease arising from that sociality, that impacts resultant interaction.

Conclusion

Taken together, our findings largely concur with the literature on which our model was based. Long notes, notes written early in the week, and notes containing at least one question all lead to higher probabilities that a note receives a reply. Reading ease of notes did not have a significant effect on the likelihood that a note would receive a reply. Overall, our binary classification model predicted 64.4% of the notes correctly.

References

Guzdial, M. (1997). Information ecology of collaborations in educational settings: Influence of a tool. In *Proceedings of the 1997 conference on computer support for collaborative learning* (pp. 83–90). Toronto, Ontario, Canada: International Society of the Learning Sciences.

Hew, K. F., & Cheung, W. S. (2008). Attracting student participation in asynchronous online discussions: A case study of peer facilitation. *Computers & Education*, 51(3), 1111–1124.

Hewitt, J. (2003). How habitual online practices affect the development of asynchronous discussion threads. *Journal of Educational Computing Research*, 28(1), 31–45.

Hewitt, J., Brett, C., & Peters, V. L. (2007). Scan rate: A new metric for the analysis of reading behaviors in asynchronous computer conferencing environments. *American Journal of Distance Education*, 21(4), 215–231.

Hewitt, J., & Peters, V. (2007). The relationship between student interaction and message readability in asynchronous online discussions. *In Proceedings of the 2007 international conference on computer supported collaborative learning* (pp. 292–294). New Brunswick, New Jersey, USA: International Society of the Learning Sciences.

Swan, K. (2005). A constructivist model for thinking about learning online. In J. Bourne & J. Moore (Eds.), *Elements of quality online education: Engaging communities* (pp. 13–30). Needham, MA: Sloan-C.