Now We're Talkin':
Leveraging the Power of Natural Language Processing to Inform ITS Development

Laura K. Allen
Tempe, AZ, USA
Arizona State University
LauraKAllen@asu.edu

Erica L. Snow
Tempe, AZ, USA
Arizona State University
Erica.L.Snow@asu.edu

Danielle S. McNamara
Tempe, AZ, USA
Arizona State University
Danielle.McNamara@asu.edu

ABSTRACT

In the current study, we utilize natural language processing techniques to examine relations between the linguistic properties of students' self-explanations and their reading comprehension skills. Linguistic features of students' aggregated self-explanations were analyzed using the Linguistic Inquiry and Word Count (LIWC) software. Results indicated that linguistic properties of self-explanations were predictive of reading comprehension ability. The results suggest that natural language processing techniques can serve as stealth assessments of abilities within intelligent tutoring systems.

Keywords

Intelligent Tutoring Systems, natural language processing, stealth assessment, student modeling, reading comprehension

1. INTRODUCTION

In the field of intelligent tutoring systems (ITSs), there exists some debate to determine when it is most optimal to assess students' performance, skills, and affect during learning tasks. System developers aim to avoid repeatedly questioning and testing students, as it may disrupt their learning flow [1]. However, it is crucial to gather student information because such variables can affect the adaptability and sophistication of these systems. One way to collect student information (without directly testing students) is through the use of stealth assessments [1]. A stealth assessment is a measure of student information (e.g., engagement, affect, skills, etc.) that is embedded within a particular task and seemingly "invisible" to users [2].

Stealth assessments can serve to inform student models within adaptive environments and, accordingly, improve system feedback and instruction. By modeling the behavioral and cognitive states of students without explicit surveys or tests, ITSs can improve student models without disrupting the learning flow of the users. This information can then be used to guide the pedagogical content that is presented to each student [3].

1.2 iSTART

The Interactive Strategy Training for Active Reading and Thinking (iSTART) tutor is an ITS that was developed to teach reading comprehension strategies to high school and college students [4]. The primary focus of the system is on the strategy of self-explanation, which has been shown to benefit students on a number of higher-level tasks [5]. Within this ITS, there are introduction, demonstration, and practice modules that explain the purpose and demonstrate the use of these strategies.

2. STUDY

The goal of the current study is to examine the extent to which the linguistic and semantic properties of students' natural language input can be used as a stealth assessment of their reading comprehension skills. To accomplish these goals, we collected students' self-explanations from the iSTART system and aggregated the individual, sentence-level self-explanations across each text that was read. Students' aggregated self-explanations were then analyzed using the Linguistic Inquiry and Word Count (LIWC) software. We utilized this tool in the current study so that we could investigate relations between students' reading comprehension ability and the semantic properties of their natural language input.

Participants were 126 high-school students from a mid-south urban environment who participated in iSTART training. Students' reading comprehension skills were measured using the Gates-MacGinitie (4^{th} ed.) reading skill test (form S) level 10/12.

2.1 Text Analyses

The linguistic features of students' aggregated self-explanations were calculated using LIWC. LIWC is a text analysis tool that uses categorical word dictionaries to provide information about texts that corresponds to thematic and rhetorical language use [6].

To extract linguistic and semantic information from students' self-explanations, individual (sentence-level) self-explanations were combined for each text read during training. Thus, each student was left with one aggregated self-explanation file for each text that they read during their time in the iSTART system. This aggregation method is discussed in greater detail in previously published work [7].

LIWC indices were then calculated for each of the aggregated self-explanation files. For each student, this LIWC output was averaged across texts to create an average score on each of the linguistic measures. These scores provide a measure of students' aggregated self-explanations at multiple linguistic levels.

3. RESULTS

To examine the relations between the LIWC linguistic scores and students' reading comprehension performance, a correlation was calculated between students' reading comprehension scores and their LIWC scores. A stepwise regression model was then calculated to assess which properties were most predictive of students' comprehension skills. A training and test set approach was used for both regressions (67% for the training set and 33% for the test set) to validate the analyses.

There were 13 LIWC variables that significantly correlated with reading comprehension scores (see Table 1). We tested for multicollinearity among these variables; however, no variables were correlated with each other above $r = .90$.

Table 1. Correlations between reading comprehension scores and LIWC linguistic scores

LIWC variable/category	*r*	*p*
Word count	.350	<.001
Words per sentence	-.316	<.001
Number words	.266	<.001
Past words	.224	<.010
Certainty words	.222	<.050
Filler words	-.212	<.050
Second person pronouns	-.211	<.050
Quantitative words	.201	<.050
Third person pronouns	.197	>.050
Ingestion words	-.195	>.050
Home words	.194	>.050
Social words	-.182	>.050
Vision words	.177	>.050

A stepwise regression analysis was conducted on the 90 self-explanations files with the 13 LIWC variables as predictors of reading scores (see Table 2) and yielded a significant model, $F(4, 85) = 9.865$, $p < .001$, $r = .563$, $R^2 = .317$ with four predictors: word count [$\beta = .38$, $t(4, 85)=4.383$, $p < .001$], words per sentence [$\beta = -.29$, $t(4, 85)=-3.129$, $p = .002$], second person pronouns [$\beta = -.24$, $t(4, 85)=-2.58$, $p = .012$], and ingestion words [$\beta = -.22$, $t(4, 85)=-2.393$, $p = .019$]. The test set yielded $r = .490$, $R^2 = .240$.

Table 2. LIWC regression analysis prediction comprehension scores

Entry	Variable added	R^2	ΔR^2
Entry 1	Word count	.120	.120
Entry 2	Words per sentence	.228	.090
Entry 3	Second person pronouns	.271	.043
Entry 4	Ingestion words	.317	.046

4. DISCUSSION

We leveraged NLP to develop stealth assessments of students' reading comprehension skills. A subset of LIWC indices were related to reading comprehension scores – namely, high reading ability students were more likely to have longer self-explanations (with shorter individual sentences) with an emphasis on numbers, words related to the past, and words related to home and vision. With regards to writing style, these students self-explained more confidently (certainty words), using a greater number of third person pronouns and fewer second person pronouns. Follow-up regression analyses indicated that word count, words per sentence, second person pronouns (e.g., you), and ingestion words (e.g., dish, eat, taste) provided the most predictive power in this model, accounting for 32% of the variance. Importantly, most of these indices were basic indices, rather than semantic categories. Thus, while many semantic lexical categories were significantly related to students' comprehension scores, they provided less predictive power than basic indices. The ingestion words index was the only semantic LIWC variable that was retained in the final model. This is likely an effect of the specific content presented within the iSTART passages; perhaps better readers provided more specific, on-topic information in their self-explanations. This question will be investigated more thoroughly in future, qualitative analyses.

These results are important, as they suggest that students' abilities manifest in the way that they explain concepts in texts. Therefore, linguistic and semantic properties of self-explanations may provide crucial information about students' cognitive processes during text comprehension. Here, we only analyzed pretest reading ability. However, these methods could be applied to model a number of relevant student features, such as their affective states and prior knowledge. Overall, the results of this study (and similar studies) can be used to help researchers develop assessments and models that provide more nuanced information about students for the purpose of increasing personalized instruction and adaptability.

5. ACKNOWLEDGMENTS

This research was supported in part by: IES R305G020018-02, IES R305G040046, IES R305A080589, and NSF REC0241144, NSF IIS-0735682.

6. REFERENCES

[1] Shute, V. J. 2011. Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), Computer Games and Instruction. Information Age Publishers, Charlotte, NC, 503-524.

[2] Shute, V. J., Ventura, M., Bauer, M. I., and Zapata-Rivera, D. 2009. Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody & P. Vorderer (Eds.), Serious games: Mechanisms and effects. Routledge, Mahwah, NJ, 295-321.

[3] Vanlehn, K. 2006. The behavior of tutoring systems. International journal of artificial intelligence in education, 16 (2006), 227-265.

[4] McNamara, D. S., Levinstein, I. B., and Boonthum, C. 2004. iSTART: Interactive strategy trainer for active reading and thinking. Behavioral Research Methods, Instruments, & Computers, 36, (2004), 222-233.

[5] McNamara, D. S. 2004. SERT: Self-explanation reading training. Discourse Processes, 38, (2004), 1-30.

[6] Pennebaker, J. W., Booth, R. J., and Francis, M. E. 2007. Linguistic Inquiry and Word Count: LIWC [computer software]. Austin, TX.

[7] Varner, L. K., Jackson, G. T., Snow, E. L., and McNamara, D. S. 2013. Does size matter? Investigating user input at a larger bandwidth. In C. Boonthum-Denecke & G. M. Youngblood (Eds.), Proceedings of the 26th Annual Florida Artificial Intelligence Research Society Conference, AAAI, St. Petersburg, FL, 546-549.