Using a Lexical Analysis of Students' Self-Explanation to Predict Course Performance

Nicholas M. Rhodes
Dept. of Computer Science
University of California Riverside
Riverside, CA 92521, USA
nrhod001@ucr.edu

Matthew A. Ung
Dept. of Mechanical Engineering
University of California Riverside
Riverside, CA 92521, USA
mung001@ucr.edu

Alexander E. Zundel
Dept. of Mechanical Engineering
University of California Riverside
Riverside, CA 92521, USA
azund001@ucr.edu

James Herold
Dept. of Computer Science
jhero001@ucr.edu

Thomas F. Stahovich
Dept. of Mechanical Engineering
stahov@engr.ucr.edu

ABSTRACT

Numerous studies have shown that self-explanation can lead to improved learning outcomes. Here we examine how the words which students use in their self-explanations correlate with their performance in the course as well as with the effort they expend on their homework assignments. We compute two types of numerical features to characterize students' work: vocabulary-based features and effort-based features. The vocabulary-based features capture the frequency with which individual words and n-grams appear within students' self-explanation. The effort-based features estimate the effort expended on each assignment as the amount of time spent writing a homework solution or self-explanation response.

We use the most predictive vocabulary-based and effort-based features to train a linear regression model to predict students' overall course grade. This model explains up to 19.4% of the variance in students' performance. Furthermore, the underlying parameters of this model provide valuable insights into the ways students explain their own work, and the cognitive processes students employ when asked to self-explain. Additionally, we use the vocabulary-based features to train linear regression models to predict each of the effort-based features. In doing so we demonstrate that the vocabulary employed by a student to self-explain his or her solution to an assignment correlates with the amount of effort that student expends on that particular assignment. Both of these findings serve as a basis for a novel automated assessment technique for evaluating student performance.

1. INTRODUCTION

Self-explanation is the process by which a student explains his or her solution process, summarizing his or her understanding. Prior work has demonstrated that self-explanation can improve a student's metacognitive skills, leading to improved learning gains. These studies have typically focused on summative assessments of students' learning, demonstrating, for example, that students who were asked to provide self-explanation of their homework solutions performed better on exams than students who did not provide self-explanation. In this paper, we present a novel technique which provides a formative analysis of self-explanation, identifying behaviors which correlate with good performance. In particular we employ machine learning techniques to identify successful patterns latent in students' self-explanations.

This analysis is enabled by our unique dataset of students' handwritten coursework. We conducted a study in which students in an undergraduate Mechanical Engineering Statics course generated handwritten self-explanations of the major steps they followed when solving each of their homework problems. The students completed the homework and self-explanations using Livescribe$^{\mathrm{TM}}$Smartpens. These devices produce a digital record of students' handwritten work in the form of time-stamped pen strokes, enabling us to see not only the final ink on the page, but also the order in which it was written.

We compute numerical features from this digital record which characterize the vocabulary used and the effort (time) expended, both in solving problems and writing self-explanation. Using these features we have computed a statistical model which predicts students' grades on various homework assignments. This model accounts for up to 19.4% of the variance in the students' performance. Furthermore, the underlying parameters of this model provide valuable insights into the ways students explain their own work, and the cognitive processes students employ when asked to self-explain.

Additionally, we use the vocabulary-based features to train linear regression models to predict each of the effort-based features. In doing so we demonstrate that the vocabulary employed by a student to self-explain his or her solution correlates with the amount of effort that student expends on that particular assignment.

2. RELATION TO PRIOR WORK

Chi et al. [4] have argued that "the metacognitive component of training is important in that it allows students to understand and take control of their learning process." Metacognition is the awareness of one's own learning process, and it serves as a major foundation for research performed on self-explanation. We use self explanation as a tool to improve students' metacognition.

Numerous studies have demonstrated the positive impact self-explanation has on student performance. Bielaczyc et al. [2] studied the impact of different self-explanation strategies on a student's ability to learn LISP programming. The experiment revealed a significant difference between the learning gains from the pre- to posttest performance of students who did and did not generate self-explanation. In this study students self-explained after viewing study materials but before solving problems. This differs from our study in which students generate self-explanation throughout their solution process.

Chi et al. [4] made comparisons between two groups of students: "poor" and "good" performing students. These students were asked to generate self-explanation after studying worked-out example problems. The results of this study demonstrated that students who perform poorly are typically unable to generate sufficient self-explanation of the worked-out example problems. This study indicates that a correlation may exist between the quality of students' self-explanation and their performance.

Hall and Vance [8] investigated the impact of self-explanation on student performance as well as self-efficacy in a Statistics course. This study showed that students who generated collaborative self-explanation performed significantly better at solving problems than students who did not self-explain. What these studies have in common is their use of summative performance assessments to show the positive impact of self-explanation on learning gains. To our knowledge, little prior work has focused on formative assessments which identify behaviors in students' self-explanations.

Prevost et al. [11] examined typed self explanations from an online system. Prevost et al. compared multiple choice responses versus constructed (free form) responses and found that constructed responses provided better insight into student thinking than multiple choice responses. Although mentioned, the authors did not examine the sequencing between individual words. Our paper focuses on analyzing sequences of words to predict student performance without manually scoring student self-explanations. While past research has typically examined data extracted from close-structured responses(e.g., multiple choice or check boxes), our paper examines free-form, handwritten responses in order to predict course performance. Our analysis is similar to that of Forbes-Riley et al. [5] in which the authors modeled students' spoken interactions with a tutoring system.

3. EXPERIMENTAL DESIGN

In the winter quarter of 2012, we conducted a study in which students enrolled in an undergraduate Mechanical Engineering Statics course were given Livescribe^{TM}Smartpens. These devices serve the same purpose as traditional pens, allowing students to handwrite their homework on paper. Additionally, these devices record a digital copy of the handwritten work as time-stamped pen strokes.

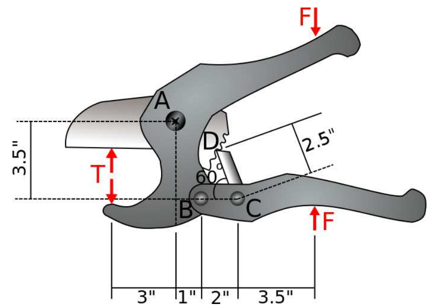

Figure 1: "The device shown is used for cutting PVC pipe. If a force, $F = 15$ lb, is applied to each handle as shown, determine the cutting force T. Also, determine the magnitude and the direction of the force that the pivot at A applies to the blade."

Thirty of the students in the course were asked to provide self-explanation on five of the homework assignments. A typical homework problem is shown in Figure 1. These problems required students to solve for unknown forces that result when external forces are applied to a system in equilibrium. Students were provided with three to five prompts eliciting explanations for each of their major solution steps. An example of a typical self-explanation prompt is, "Why did you select the system that you used for your free-body diagram?" Students handwrote their responses to these questions and submitted them along with their solutions.

4. DATA PROCESSING

We manually transcribed each handwritten self-explanation, producing 111 text documents. Each document contains all self-explanation written by a single student for a single homework assignment. During this manual transcription we made slight modifications to the students' explanations to make them suitable for later processing. First, we corrected any spelling mistakes, but did not correct grammatical errors. Second, we replaced each verb with its unconjugated form. For example, we replaced "pushed" (past tense) with "push" (infinitive). Our later analysis counts the number of occurrences of words based on exact spelling. These changes ensure that spelling variations do not prevent words from being correctly identified.

We also developed a thesaurus to replace synonymous words with a single, canonical word. Students use a variety of words to refer to a given concept or object. For example, when students described a free-body diagram, they often used the terms "system" and "body" interchangeably. To ensure that semantically identical words were identified as such, we manually developed a thesaurus that maps a canonical concept to each of the words that may be used to express that concept. For example, we created a "free-body diagram element" concept category that comprises every word that students used to refer to any component (body) in a free-body diagram, such as "jaw" or "handle". In this example, whenever the word "jaw" was found in a transcript, it was replaced with the token "FBD-Element". We developed a total of ten conceptual categories with the help of a Statics domain expert. There were approximately 1640 unique words used by students across all documents before correcting spelling or verb tense. After applying our thesaurus-based replacement, there were 750 unique words.

5. VOCABULARY, EFFORT, AND PERFORMANCE FEATURES

In this section we describe the numerical features which characterize: the vocabulary employed by a student in his or her self-explanation; the effort expended by a student on his or her solution and self-explanation; and each student's performance in the course.

We use the "term frequency - inverse document frequency" score (TF-IDF)[9] to characterize the importance of each word in a transcribed self-explanation document. The TF-IDF scores of all words encountered in all documents are used as features.

To characterize the sequence of words in students' self-explanation, we analyze the frequencies of the bigrams and trigrams which appear in each self-explanation document. In our analysis, we split word choices on periods and thus consider bigrams and trigrams within sentence boundaries. We use the N-Gram Statistics Package (NSP) [3, 1] to both identify and calculate the frequency of n-grams present in each document.

NSP provides a number of different methods for measuring the frequency of a given n-gram. We used the total mutual information (TMI) to score each gram. TMI scores an n-gram by computing the ratio of the log of the joint probability of all words in that n-gram over the marginal probability distributions of each word in an n-gram.

Additionally, we compute two features that characterize the effort spent on an assignment and the corresponding self-explanation. We used the average time spent drawing free-body diagrams, writing equations, and answering self explanation questions. This produced three separate effort based features.

6. FEATURE SUBSET SELECTION

Given that there are $\sim$ 750 unique words across all explanations, there would be over 19,000 TF-IDF, bigram, and trigram features computed for each of the 111 documents. This is too large a feature set and would lead to an over-fit model with inflated accuracy. To address this issue we use two feature subset selection algorithms to reduce the size of our feature set.

First, we apply the computationally inexpensive RELIEF [10] algorithm to prune our feature set to the top 500 features. The RELIEF algorithm scores each feature by its similarity to the nearest instance of the same class and to the nearest instance of each other class. Next, we apply the computationally expensive, but more rigorous Correlation Feature Selection (CFS) [7] algorithm to further reduce the feature subset. We use the RELIEF and CFS implementation available in the WEKA [6] machine learning software suite.

7. PREDICTING STUDENT PERFORMANCE AND EFFORT

We trained four separate linear regression models to predict students' course performance, equation effort, free-body diagram effort, and self-explanation effort respectively. Both of the aforementioned subset feature selection algorithms are executed separately for each model, resulting in four distinct feature subsets which range from 13 to 25 features. The effort-based features and vocabulary-based features are used as input to the feature subset selection for the performance-based model, and only the vocabulary-based features are used for the three effort-based models. Each of these four models was trained using the linear regression implementation available in WEKA [6].

Table 1 shows the coefficient weights for each of the features in the first regression model. The magnitude of each weight indicates the predictive power of that feature in determining either students' performance or effort. Similarly, the sign of the weight indicates whether or not that feature correlates positively or negatively with performance or effort. The performance, equation effort, free-body diagram, and self-explanation models are able to explain 19.4%, 17.8%, 20.0%, and 45.7% of the variance in their respective dataset. The final three models are computed, but not shown in this paper.

Table 1: Underlying parameters of the linear regression model used to predict students' overall course grade. Each row corresponds to a single feature which is the mutual information value of a single n-gram or the TF-IDF scores of single words. The attribute column presents the n-gram or word used to compute the feature and the weight column presents the weight of that feature in the linear regression model.

Weight	Attribute
-1174.8889	a<>force<>on
-395.8207	the<>twoforcemember
-358.6974	twoforcemember<>force
-300.7872	a<>force<>was
-292.9685	on<>an
-153.5296	action<>and<>i
-135.6537	the<>direction<>steeper
-135.6536	direction<>steeper<>angle
-135.6535	steeper
75.4336	solving<>solving<>of
87.9657	point<>i
92.4137	asked<>for
99.8558	interaction<>with
104.9789	knew
115.2228	body<>but
115.4557	interaction<>would<>be
125.317	and<>the<>boom
125.3176	we<>look<>at
125.3178	we<>are<>act
125.3181	act<>a<>force
125.3182	are<>act<>a
125.3186	act<>a
125.319	because<>we<>are
152.9983	to<>the
459.9756	body<>is<>a

8. DISCUSSION

The accuracy of our model for predicting student performance is encouraging. More interesting though, is the fact that the model and its parameters indicate the self-explanation behaviors that correlate with strong or weak performance. By manually investigating these behaviors, we are able to identify the metacognitive skills students demonstrate regarding their problem-solving processes.

Take, for example, the feature with greatest weight, the TMI score of the trigram "body<>is<>a". By manually inspecting the self-explanation responses that contain this trigram, we found that the trigram "body<>is<>a" is typically used by a student to identify a special type of body in a system. For example, one student's self-explanation response reads "The lever is a two-force system." In this example, the word "lever" belongs to the "body" thesaurus category. This provides strong evidence of a students' ability to both recognize and apply concepts learned in class to given homework problems. By identifying the "two-force system" the student is able to apply a particular technique from class which only applies to two-force systems.

Similarly, consider the difference between the trigram "a<>force<>was" and the bigram "point<>i". In examining the self-explanation responses, we found that responses which contained the trigram "a<>force<>was" were used passively in sentences, whereas the bigram "point<>i" was used actively in sentences. This provides evidence of the importance of active voice in self-explanations positively correlating with student performance while passive-voice sentences correlate negatively with performance.

Some attributes tended to reinforce our intuition regarding students' performance. The word "knew" indicates conceptual understanding and a student's confidence in their problem-solving. When we examined these self-explanation responses, the word "knew" expressed premeditation and certainty. For example, one such self-explanation transcript read, "I knew that by taking a moment about point A that I would cancel out forces at F."

Obvious grammatical errors tended to lead to poor performance. In our manual investigation, we found that the trigram "action<>and<>i" was primarily used in run-on sentences. Consideration of alternative solution paths correlated positively with performance. The bigram "body<>but" was typically used by a student to indicate that there was another way to solve a particular problem.

9. CONCLUSION

In this work, we have demonstrated a novel technique for analyzing students' handwritten self-explanations of their homework solutions. This technique is enabled by our unique dataset of student work. We conducted a study in which thirty students in an undergraduate Mechanical Engineering Statics course provided handwritten self-explanations of the major steps they followed when solving each of their homework problems. The students completed the homework and self-explanations using Livescribe$^{\text{TM}}$Smartpens. These devices produce a digital record of students' handwritten work in the form of time-stamped pen strokes, enabling us to see not only the final ink on the page, but also the order in which it was written.

We compute numerical features from this digital record which characterize the vocabulary used and effort expended in constructing handwritten self-explanations. We applied a heuristic subset selection algorithm to identify the optimal subset of features for predicting homework performance. Using this subset, we computed a linear regression model that predicts students' grades on homework assignments. This model accounts for 19.4% of the variance in the students' performance. While this is a strong correlation, what is more valuable are the insights that can be drawn from the underlying parameters of this model. The coefficient weights of the model may be used to guide manual analysis of the students' self-explanation responses, revealing patterns that provide insights into the types of self-explanation behaviors that are indicative of understanding or lack thereof.

10. REFERENCES

[1] S. Banerjee and T. Pedersen. The design, implementation, and use of the ngram statistics package. *Computational Linguistics and Intelligent Text Processing*, pages 370–381, 2003.

[2] K. Bielaczyc, P. L. Pirolli, and A. L. Brown. Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and instruction*, 13(2):221–252, 1995.

[3] W. B. Cavnar and J. M. Trenkle. N-gram-based text categorization. In *In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, pages 161–175, 1994.

[4] M. T. Chi, M. Bassok, M. W. Lewis, P. Reimann, and R. Glaser. Self-explanations: How students study and use examples in learning to solve problems. *Cognitive science*, 13(2):145–182, 1989.

[5] K. Forbes-Riley, D. Litman, A. Purandare, M. Rotaru, and J. Tetreault. Comparing linguistic features for modeling learning in computer tutoring. In *Proceedings of the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, pages 270–277, Amsterdam, The Netherlands, The Netherlands, 2007. IOS Press.

[6] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.

[7] M. A. Hall. *Correlation-based Feature Subset Selection for Machine Learning*. PhD thesis, University of Waikato, Hamilton, New Zealand, 1998.

[8] S. Hall and E. A. Vance. Improving self-efficacy in statistics: Role of self-explanation and feedback. *J. Stat. Educ*, 18(3), 2010.

[9] K. S. Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28:11–21, 1972.

[10] K. Kira and L. A. Rendell. A practical approach to feature selection. In D. H. Sleeman and P. Edwards, editors, *Ninth International Workshop on Machine Learning*, pages 249–256. Morgan Kaufmann, 1992.

[11] L. Prevost, K. Haudek, M. Urban-Lurain, and J. Merrill. Deciphering student ideas on thermodynamics using computerized lexical analysis of student writing. 2012.