Toward Automated Support for Teacher-Facilitated Formative Feedback on Student Writing

Jennifer Sabourin
SAS Institute
100 SAS Campus Dr.
Cary, NC 27513
1.919.531.3313
Jennifer.Sabourin@sas.com

Lucy Kosturko
SAS Institute
100 SAS Campus Dr.
Cary, NC 27513
1.919.531.3430
Lucy.Kosturko@sas.com

Kristin Hoffmann
NC State University
Raleigh, NC 27695
1.919.515.7061
klhoffma@ncsu.edu

Scott McQuiggan
SAS Institute
100 SAS Campus Dr.
Cary, NC 27513
1.919.531.1119
Scott.McQuiggan@sas.com

ABSTRACT

Formative, content-level feedback on student writing has been shown to have positive impacts on both writing and learning outcomes. However, many teachers struggle to provide this type of feedback to large classrooms of students. This paper takes an initial step towards supporting teacher-facilitated feedback through the use of automated and user-directed topic discovery. 114 student essays were collected from a local underperforming middle school as part of a pilot study for Write Local, a digital repository and workspace for authentic problem-based learning activities. Predictive models were built and evaluated to explore the impact of different topic discovery approaches as well as correction of student spelling errors on model accuracy. The resulting models provide promising direction for scaffolding teachers in providing formative feedback on content-level features of students' problem-based writing.

Keywords

Problem-based writing, formative feedback, teacher-facilitated feedback, automated writing assessment, topic discovery

1. INTRODUCTION

Problem-based writing tasks seek to elicit high-quality student writing by contextualizing the purpose of the task and providing an authentic audience [4]. These tasks also tend to extend across several days or learning periods offering more opportunities for formative assessment and feedback, which is expected to yield improved writing outcomes [1]. However, it is often difficult for teachers to focus on high-level features such as the focus, accuracy, and organization of student writing when working with a large classroom of students. Instead, teachers are more likely to focus on surface level features such as spelling, grammar, and mechanics. This is especially true in underperforming schools [2].

This work serves as an initial investigation into automated assessment of student writing in order to scaffold teachers in providing higher-level formative feedback. A pilot study was conducted as an initial step in the Write Local project. Write Local is intended to be a digital repository and workspace to facilitate both teachers and students in authentic problem-based writing activities. As part of a pilot study, 114 student writing samples were collected from students at an underperforming [3], local middle school as part of a multi-day problem-based learning activity. Student essays were manually coded for essay focus and accuracy. A variety of models for predicting these features were constructed and evaluated as an initial exploration for scaffolding teacher-facilitated feedback. In particular, this work sought to explore the role of automated and user-directed topic discovery in predicting content-level essay features. Additionally, we sought to investigate the importance of correction of student spelling mistakes prior to model construction. The results indicate that these initial models can serve as a starting point for supporting teachers in providing feedback on content-level features in problem-based writing and inform several directions for future work.

2. PILOT STUDY

This investigation uses data collected during a pilot study of Write Local. Write Local seeks to employ crowdsourcing to ensure teachers and students have immediate access to a large repository of writing prompts that cover the entire spectrum of text types and audiences—persuasive, informative/explanatory and narrative. Local businesses, and in particular, those employing STEM-related positions, can post various letters of need as well as any supplemental documentation such as images or vocabulary lists. Teachers can then select a call from the repository and assign the project to their students. Students will use the integrated workspace to plan, research, document, draft, revise, present, and submit their response in one central space.

The entire sixth grade from a local, underperforming [3] middle school (54% free/reduced lunch) participated in this study as part of their regular social studies class. Of the 168 participants, 86 were male and 82 were female with a mean age of 11.5. Of the 168 participants, 114 completed all components of the procedure. For the remaining analyses only data from these 114 students is used.

For the study, students were divided by class into one of two conditions: experimental and control. On the first day of the study, students in the experimental condition viewed a 3-minute introduction video that contained problem context: a frozen yogurt company plans to open a new location and asked students to write a letter with their researched opinions about 1) which 5 toppings should be available on the topping bar and 2) where the new shop should be located. Students used authentic data and a map of the area to make their decisions. Students in the control condition were given a similar task without real-world contextualization. Students in both conditions were given two full 50-minute class periods to plan and write their letters.

Three researchers then transcribed and coded the essays with sufficient inter-rater reliability (k = .89). Essays were given a composite score for essay focus and accuracy. Using the final composite scores, students were divided into 3 evenly distributed categories (High, Medium, and Low) for both focus and accuracy. These groupings are intended to be presented to teachers to inform formative feedback for their students.

3. TEXT ANALYSIS AND MODELING

The first step in building predictive models of student essay content classifications was to extract meaningful features from the student text. In total, the corpus for analysis included 114 student essays. The average length of the essays was 130.0 (SD = 91.4) words and 9.6 (SD = 7.6) sentences. The writing samples provided by the students were analyzed using SAS® Text Miner® and SAS® Enterprise Miner®.

For the purpose of this analysis we focused on the document topic analysis features of SAS Text Miner. The text topic procedure identifies terms that are strongly associated within the corpus. It also provides a strength of each topics' presence within the document. Topics can be automatically learned from the corpus or they can be provided or fine-tuned manually. Both approaches were used for this work. For automatic topic discovery, the limits were set at 25 multi-term topics. Manually-created topics were generated by highlighting terms in the text of the prompt and identifying whether each term applied to the problem context, the problem request, or the task instructions. In total 27 terms were identified; 8 context terms, 13 request terms, and 6 instruction terms. These terms were provided as user-created topics to the topic discovery procedure. In addition, up to 25 multi-term topics could be automatically generated; though because the engine tries to remove correlated topics, only 22 new topics were created. Of the 27 user-provided topics, only 17 occurred in the corpus of student data; 6 context terms, 9 request terms, and 2 instruction terms.

During essay transcription and coding, it was noted that there were a significant number of spelling errors present in the corpus. This may be due to the fact that essays were handwritten without the support of automated spell checking tools that many students are familiar with. In order to investigate the importance of correct spelling in modeling content-level features such as essay focus and accuracy, we chose to build models using different levels of spelling correction. Three different corpora of student essays were provided to the text topic discovery procedures: 1) the students' original texts, 2) an automatically spell-corrected version of the text, and 3) a manually spell-corrected version of the text.

For this exploration, we evaluated models across both topic discovery type (fully-automated and user-facilitated) and spelling correction type (manual, automated, and no correction). Additionally, we built separate models to predict both essay focus classification and essay accuracy classification. Finally, we used three modeling approaches for each corpus: logistic regression, decision tree, and neural network.

Each model was evaluated using 10-fold cross validation and predictive accuracies were compared against a baseline of most frequent class. This measure was 33.0% and 40.4% for essay focus and accuracy, respectively. The most common class for each evaluation type was Medium. With one exception, all models outperformed baseline with statistical significance at the 0.05 level (Table 1).

Overall, the models built using manual spelling correction and prompt-based topics outperformed other models in predicting essay focus and accuracy. This suggests that the prompt-based topics centered on the components of problem-based learning activities were beneficial in improving predictive accuracy. Unfortunately, this step requires manual annotation for each prompt. At present, this task, while manual, is not particularly labor intensive and can scale as we assess whether this benefit holds for future, unseen prompts. However, since the objective of Write Local is to scale with a large number of problem-based prompts, this may no longer be feasible. If we determine that this type of prompt annotation continues to be beneficial for predicting essay accuracy and focus we may investigate possible methods for automating or facilitating this task.

Table 1. Predictive accuracy for essay focus and accuracy using (a) discovered topics and (b) prompt-based topics

Discovered Topics

Model	Spelling Correction: Manual	Auto	None
Neural Net	F: 57.4 A: 55.0	F: 48.9 A: 51.9	F: 45.5 A: 52.6
Log. Reg.	F: 46.5 A: 56.1	F: 45.5 A: 46.4	F: 44.6 A: 47.3
Decision Tree	F: 51.3 A: 55.2	F: 46.4 A: 45.3	F: 47.3 A: 43.9
Average	F: 48.9 A: 55.7	F: 46.9 A: 47.9	F: 45.8 A: 47.9

Prompt-Based Topics

Model	Spelling Correction: Manual	Auto	None
Neural Net	F: 61.4 A: 56.1	F: 50.8 A: 57.1	F: 55.4 A: 50.0
Log. Reg.	F: 56.1 A: 68.4	F: 46.4 A: 62.5	F: 49.1 A: 52.7
Decision Tree	F: 53.5 A: 61.4	F: 50.0 A: 54.5	F: 46.5 A: 57.1
Average	F: 57.0 A: 62.0	F: 49.1 A: 58.0	F: 50.3 A: 53.3

Secondly, we note that the models using manual spelling correction tended to outperform models using automatic or no spelling correction, though this finding was less reliable. Since the "manual" spelling correction was done primarily using feedback from a word processor, it may be the case that had the essays been written digitally with spell check options available, many of the errors that were corrected would have been found by the student themselves. Future work will be necessary to determine if word processor spell check features are sufficient for this task.

4. REFERENCES

[1] Graham, S. et al. 2015. Formative Assessment and Writing: A Meta-Analysis. *The Elementary School Journal*. 115, 4 (2015), 523–547.

[2] Matsumura, L.C. et al. 2002. Teacher Feedback, Writing Assignment Quality, and Third-Grade Students' Revision in Lower-And Higher-Achieving Urban Schools. *The Elementary School Journal*. 103, 1 (2002), 3.

[3] North Carolina School Report Cards: *http://www.ncpublicschools.org/*. Accessed: 2016-03-01.

[4] Purcell-Gates, V. et al. 2007. Learning to read and write genre-specific text: Roles of authentic experience and explicit teaching. *Reading Research Quarterly*. 42, 1 (2007), 8–45.