Assessing Student Understanding of the Text: Comparing Model-Based and Text-Based Approaches to Summary Evaluation

Min Kyu Kim, Kathryn McCarthy and Ali Heidari

July 12, 2020

Assessing Student Understanding of the Text: Comparing Model-Based and Text-Based Approaches to Summary Evaluation

Min Kyu Kim, Kathryn S. McCarthy, Ali Heidari

Department of Learning Sciences, Georgia State University

Author Note

The authors declare that there no conflicts of interest with respect to this preprint.

Correspondence should be addressed to Min Kyu Kim, Department of Learning Sciences
Georgia State University, 238 College of Education and Human Development
30 Pryor St SW, Atlanta, GA 30303, United States. Email: mkim120@gsu.edu

Abstract

In this study, we proposed two approaches to summary analysis (*model-based* and *text-based*) along three dimensions: surface, structure, and semantic. We investigated the power of the two approaches to assess changes in students' summaries. Results demonstrated the theoretical overlap of model-based and the text-based approaches and the potential for a more nuanced account of how students understand text.

Keyword: Reading comprehension, summary writing, automated summary evaluation

Assessing Student Understanding of the Text: Comparing Model-Based and Text-Based Approaches to Summary Evaluation

Summary writing is a common learning strategy for helping students to comprehend the materials they need for meaningful class discussion. However, it is challenging for teachers to evaluate the quality of individual students' summary and to support students to build sufficient understanding of the readings (Dunlosky et al., 2013).

To tackle this problem, technologies have been used to automatically assess student summaries through a process of analyzing a student text by a natural language processing (NLP) algorithm, generating a student model (i.e., parameterizing), which is then compared to an expert model (e.g., computing similarity values), and thereby providing formative feedback.

In this initial investigation, we propose two different analytic approaches (i.e., model-based and text-based) that technologies have taken to analyze student text data (e.g., summary). We then examine the extent to which the indices derived from these two different approaches detect critical changes in the quality of students' summaries.

Classification Framework

A review of state-of-the-art technologies reveals two different approaches: *model-based* and *text-based*. The model-based approach focuses on eliciting an underlying mental model of the summary in the form of a concept map. Model-based evaluators include technologies such as SMART (Kim et al., 2019), GISK (Kim, 2018), and AKOVIA (Ifenthaler, 2014). For example, SMART generates parameters along three different dimensions: surface, structure, and semantic (Kim et al., 2016) and these model-based parameters change as students improve their summaries (Kim et al., 2019).

In contrast, text-based approaches evaluate the summary on hundreds of indices in multiple dimensions (Crossley et al., 2016; Kyle et al., 2018). Technologies associated with the text-based approach include the Tool for the Automatic Analysis of Lexical Sophistication (TAALES, Kyle et al., 2018) and the Tool for the Automatic Analysis of Cohesion (TAACO, Crossley et al., 2016). These tools are not specific to summary evaluation, but rather provide a variety of indices at the word, sentence, and document level that can be leveraged to evaluate linguistic features of students' written responses.

This study builds on our previous studies (Kim et al., 2016; Kim et al., 2019) that validated a model-based framework used for the SMART technology to examine text-based approaches. We first selected features from TAACO and TAALES that were consistent with the three dimensions analyzed in SMART. We then analyzed initial and revised summaries using both model-based SMART indices and the text-based linguistic features derived from TAACO and TAALES.

Method

Our corpus included summaries of seven different texts. These summaries (initial drafts and multiple revisions) were written by 38 students during a graduate online course. On SMART, the students wrote summaries and revised their summary based on the SMART feedback as many times as desired. In this study, we analyzed the initial and the final version of the summaries from 47 cases in which students wrote at least two revisions. The model-based indices were generated by the SMART system. The summaries were also submitted to TAACO and TAALES to extract text-based indices related to the 3S indices (see Table 1). To test changes in the text-based indices, we used one-way repeated-measures Multivariate Analysis of

Variance (MANOVA). We then examined the correlations between the text-based and model-based indices at each of the three dimensions.

Table 1

Comparison of Model-Based and Text-Based Approaches to Summary Analysis

Model-Based Approach			Text-Based Approach		
Dimension	Similarity Index	Definition	Dimension	Similarity Index	Definition
Surface	Number of concepts*	Compare the number of concepts (nodes) in two models	Surface Code	Content TTR*	Compare the Type-Token Ratio (TTR) value of content words in two models
	Number of relations*	Compare the number of links (edges) in two models		Lemma TTR*	Compare the Type-Token Ratio (TTR) of lemma in two models
	Density of graphs*	Compare the density of the two models		-	-
Structure	Average Degree*	Compare the average number of degrees in two models	Text-based/ Situational Model	Overlap N-1S*	Compare the overlap index between nouns and synonym sets across an adjacent sentence in two models.
	Mean Distance*	Compare the mean distances in two models		Overlap N-2S*	Compare the overlap index between nouns and synonym sets across the next two sentences in two models
	Diameter*	Compare the largest geodesics in two models		Overlap CN-2S*	Compare the overlap index between content words and synonym sets across the next two sentences in two models
	-	-		Overlap Lemma-2S*	Compare the overlap lemmas across the next two sentences in two models
	-	-		Overlap ARG.-2S*	Compare the argument index across the next two sentences in two models.
Semantic	Concept Matching	Compare semantically identical concepts, including contextual and principle variables		Source Sim-LSA	Similarity between the words in a source text and a target text, building on latent semantic analysis (LSA).
	Propositional Matching	Compare fully identical propositions (edges) between two concept maps		Source Sim-W2V	Similarity between the words in a source text and a target text, building on Word2vector.
	Balanced Semantic Matching	Compare the balances calculated by dividing Propositional Matching by Concept Matching		-	-
	Recall-C	The proportion of key concepts that appear in a student summary		-	-
	Recall-P	The proportion of key relations that appear in a student summary		-	-

Note. * Parameters are compared to compute similarity values.

Results

Parameter Comparisons

Parameter comparisons allowed us to examine the nature of a students' summary in terms of different dimensions of structure and language. The text-based analytic technologies generate parameters in the surface and structure dimensions. Consistent with the findings of the model-based parameters (see Kim et al., 2019), MANOVAs indicated significant changes in the text-based parameters from initial to final summary across the two dimensions (see Table 2).

For the surface level dimension, univariate tests revealed that all TTR indices significantly decreased, but only Content TTR negatively related to Number of relations, indicating that students tended to use more important concepts (a decrease in lexical diversity; reduced TTR values) that connected with an optimal number of content words.

Table 2

Text-Based Parameters and Correlations with Model-Based Parameters

Index	Initial	Final	*F*	*p*	η^2	Correlation w/ Model-based Parameters[a]		
	M (*SD*)	*M* (*SD*)						
Surface						*N* Concepts	*N* Relations	Density
Content TTR	0.68 (0.10)	0.64 (0.07)	5.54	.023*	0.11	-0.21	-0.48**	-0.05
Lemma TTR	0.55 (0.10)	0.51 (0.05)	6.18	.017*	0.12	0.01	-0.28	-0.24
Structure						Avg. Degree	Mean Distance	Diameter
Overlap N-1S	0.14 (0.08)	0.15 (0.07)	1.43	0.24	0.03	0.55**	-0.38*	-0.43**
Overlap N-2S	0.62 (0.26)	0.72 (0.19)	7.96	0.01*	0.15	0.27	-0.38*	-0.32*
Overlap CN-2S	0.71 (0.26)	0.82 (0.17)	6.05	0.02*	0.12	0.22	-0.38**	-0.34*
Overlap lemma-2S	0.87 (0.24)	0.95 (0.09)	4.72	0.03*	0.09	0.79**	0.63**	0.52**
Overlap ARG-2S	0.74 (0.18)	0.63 (0.25)	8.21	0.01*	0.16	0.23	-0.34*	-0.33*

Note. * $p < .05$, ** $p < .01$.

All structure-related parameter showed a significant change except Overlap N-1S, indicating that those descriptive parameters might be sufficient to track changes within an individuals' summaries. For final versions, all the parameters negatively related to Mean Distance and Diameter, meaning that students who used more overlapped nouns, lemmas, content words, and arguments across the next two sentences built a concise and cohesive mental model (a shorter mean distance and diameter).

Similarity Comparisons

Similarity measures describe the degree to which a students' summary is similar to either the original text or to an expert (e.g., benchmark) summary. We computed similarity values and then examined critical changes in the similarity measures from initial to final summary (see Table 3). MANOVAs revealed that all the similarity measures across all three dimensions increased, meaning that so far as those indices, students wrote a revised summary closer to an expert summary.

Further analyses revealed interesting patterns of relations between text-based and model-based similarity measures. Similarity measures in the surface and semantic dimensions were correlated with model-based similarity measures. For example, the two semantic similarity measures showed a strong correlation with the four model-based semantic measures.

Overlap lemma-2S was the only structure similarity index significantly correlated with all the model-based structure similarity indices. In contrast to Overlap lemma-2S, other text-based similarity measures might describe different aspects of the changes in mental model structure.

Table 3

Text-Based Similarities and Correlations with Model-Based Similarities

Index	Initial M (SD)	Final M (SD)	F	p	η^2	Correlation w/ Model-based Similarity			
Surface						N Concepts	N Relations	Density	-
Content TTR	0.85 (0.10)	0.88 (0.08)	5.68	0.02*	0.11	0.12	0.30*	0.05	-
Lemma TTR	0.86 (0.11)	0.90 (0.08)	7.14	0.01*	0.14	0.13	0.42**	-0.06	-
Structure						Avg. Degree	Mean Distance	Diameter	-
Overlap N-1S	0.61 (0.24)	0.71 (0.21)	5.44	0.02*	0.11	0.52**	0.28	0.27	-
Overlap N-2S	0.71 (0.27)	0.79 (0.16)	3.99	0.05	0.08	0.21	0.11	0.09	-
Overlap CN- 2S	0.75 (0.27)	0.85 (0.15)	5.95	0.01*	0.12	0.12	0.29	0.09	-
Overlap lemma-2S	0.85 (0.25)	0.93 (0.08)	4.44	0.04*	0.08	0.83**	0.70**	0.78**	-
Overlap ARG-2S	0.73 (0.26)	0.81 (0.15)	4.60	0.03*	0.10	0.23	0.15	0.17	-
Semantic						Concept Match	Prop. Match	Recall-C	Recall-P
Source Sim-LSA	0.77 (0.14)	0.85 (0.09)	14.26	0.00**	0.23	0.74**	0.54**	0.70**	0.66**
Source Sim-Word2Vec	0.88 (0.08)	0.92 (0.05)	9.34	0.00**	0.18	0.71**	0.52**	0.67**	0.63**

Note. * $p < .05$, ** $p < .01$.

Conclusion

The findings of the study support the theoretical overlap of model-based and text-based approaches to analysis of student summaries. Strong correlations between the text-based and model-based indices indicate good validity, suggesting the potential of the three dimensions (surface, structure, and semantic) to classify text-based indices. Future studies should further investigate the relationships of the indices from the two approaches and assessment of reading comprehension in multiple dimensions to drive more specific feedback to learners.

References

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods, 48*(4), 1227-1237.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371-395.

Ifenthaler, D. (2014). AKOVIA: Automated Knowledge Visualization and Assessment. *Technology, Knowledge and Learning, 19*(1-2), 241-248.

Kim, K. (2018). An automatic measure of cross-language text structures. *Technology, Knowledge and Learning*, *23*(2), 301–314.

Kim, M., Zouaq, A., & Kim, S. (2016). Automatic detection of expert models: The exploration of expert modeling methods applicable to technology-based assessment and instruction. *Computers & Education, 101*, 55-69.

Kim, M., Gaul, C., Kim, S., & Madathany, R. (2019). Advance in detecting key concepts as an expert model: Using Student Mental Model Analyzer for Research and Teaching (SMART). *Technology, Knowledge and Learning*, 1-24. https://doi.org/10.1007/s10758-019-09418-5

Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior Research Methods, 50*(3), 1030-1046.