Proceedings of the Twenty-Third International Florida Artificial Intelligence Research Society Conference (FLAIRS 2010)

Summarization: Constructing an Ideal Summary and Evaluating a Student's Summary using LSA

Radhika Malladi[1], Irwin Levinstein[2], Chutima Boonthum[3], Joseph Magliano[4]

[1] Novo Solutions, USA
[2] Old Dominion University, Norfolk, VA, USA
[3] Hampton University, Hampton, VA, USA
[4] Northern Illinois University, DeKalb, IL, USA
{ radhimalladi@yahoo.com, ibl@cs.odu.edu, chutima.boonthum@hamptonu.edu, jmagliano@niu.edu}

Abstract

Summarization means getting the essence of the original document. The summary should be shorter than the original with all minor facts left out and only the main ideas extracted and paraphrased with new words. This work aims to construct an ideal summary from sentences chosen as targets for self explanations in prior experiments and to use this summary to assess the quality of students' summaries of the same text. This paper briefly describes (1) how ideal summary is systematically constructed based on the target sentences and (2) how LSA is used to measure the similarity between student's summary and ideal summary. Results indicate that target sentences are good indicators for an ideal summary. Comparison against human expert, the initial data shows that the system able to obtain 70 75% agreement in evaluating completeness of student summary and 0.65 0.74 correlation in evaluating quality of student summary.

Introduction

A summary is a condensed version of the original document. It is a concise statement of the most important information in a text. The main idea of the original should be expressed with new words and details or the minor facts should be omitted. The process of summarizing enables students to better grasp the original and the result shows the teacher that the student understood the original document. In addition, the knowledge gained allows the student to better analyze and critique the original and will provide an easy reference for later study. A student's summary may be incomplete because he or she may not have acquired a sufficient knowledge of the original document. In a typical experiment on which this work is based, students are shown a text one sentence at a time. Some sentences have been selected as target sentences and these they are asked to analyze or explain. At the end, they may be asked to summarize the text.

The goals of this work are (1) to find an ideal summary for a given text and (2) to evaluate the quality and completeness of student's summary. The student's summary is evaluated on its quality and the completeness of its coverage of the ideal summary. Ideal summary is a good summary to which the student's summary is compared. Evaluation of student's summary is done to in terms of its coverage (or completeness) and its quality. A good summary is ones that covers the ideal summary, *i.e.*, the main ideal of the text.

This paper describes the automated system that uses Latent Semantic Analysis (LSA; Landauer et al., 2007) to identify an ideal summary for a given text as well as to measure the quality of the student's summary. The data used in this paper was collected from the Reading Strategy Assessment Tool (R-SAT, Gilliam et al., 2007), an online automated assessment tool that identifies weaknesses in students' reading comprehension strategies.

Ideal Summary

A summary contains only the essence of the text. Instead of having human expert construct an ideal summary, our goal is to implement an automated process. This paper focuses on three forms of ideal summaries: (1) *Target Sentences* (TS): Target sentences are the human selected sentences from the text by using Causal analysis (Grishman and Kslezyk, 1990). These sentences have the highest comprehension-based connectedness to other sentences in the text; (2) *Target Sentences with no overlap* (TSN): Conceptual overlap between the target sentences may need to be removed to get an appropriate LSA cosine. If two target sentences have similar content, then there may be little change in the LSA score when one is removed and the result compared to the student summary. So, the target sentence set was modified by removing the overlap; and (3) *New Target Sentences*: New set of target sentences are considered to see if there is any improvement in LSA cosine. To select these new target sentences, all sentences in the text are compared with each other to get a LSA cosine. By looking at the LSA cosine, new set of target sentences are selected. The higher the LSA cosine, the more overlapping of this sentence-pair is. Overlap removal process is used to select the final set of new target sentences.

Student's Summary & Evaluation

Forty four (44) student's summaries were coded by a human expert (*human score*). The expert split the text into clauses and scored the students summaries based on the clause coverage. Each student's summary is given a score for each clause in the text; whether the student has covered a particular clause or not. This is a 3-point scale where 0, 0.5, and 1 means that this clause was not covered, partially covered, or fully covered, respectively, in the student's summary. Data are divided into two equal halves: training set and test set. They are randomly assigned with assumption that they should have equal number of good and bad summaries. Good summaries are those that have high similarity with ideal summary.

Each student's summary is evaluated in 2 ways: (1) *completeness of a summary*: during this evaluation, a target sentence that student missed or covered will be discovered; if LSA cosine is less than the threshold, then we said that the student missed that sentence; (2) *quality of a summary*: during this evaluation, a *regression analysis* is used to create a model for assessing quality of student's summary - indicating whether the summary is good or bad. Three models were chosen: Linear (L), Non-Linear/ Quadratic (Q), and Non-Linear/ Cubic (C).

The results in Table 1 show that all target sentences and all clauses of target sentences with any of three models (Linear, Quadratic, and Cubic) are performing very well based on the training set and test set. The LSA threshold doesn't pay any role in the summary's quality evaluation.

Conclusion

The results show that 70 to 75 percent agreement in evaluating completeness of student summary for target sentences and 0.65 to 0.74 of correlation in evaluating quality of student summary for target sentences. LSA results are good when considering small paragraphs or small sentences as it gives results for large texts.

LSA can be accompanied with word matching to get still better results. By increasing the number of words, LSA space can be improved to get a better LSA cosine. This project can be used for evaluating the student's summaries without the involvement of human expert.

Forms	**Mod**	**Pred**	**0.15**	**0.3**	**0.4**
TS					
Training	L	0.575	0.589	0.594	0.540
	Q	0.629	0.634	0.641	0.604
	C	0.658	0.658	0.685	0.625
Test set	L	0.674	0.674	0.670	0.664
	Q	0.723	0.717	0.717	0.711
	C	0.737	0.737	0.742	0.744
TSN					
Training	L	0.623	0.623	0.627	0.594
	Q	0.680	0.692	0.692	0.667
	C	0.707	0.707	0.715	0.683
Test set	L	0.426	0.426	0.439	0.427
	Q	0.411	0.467	0.467	0.453
	C	0.406	0.431	0.445	0.429
New					
Training	L	0.742	0.753	0.714	0.729
	Q	0.803	0.809	0.787	0.786
	C	0.833	0.837	0.797	0.814
Test set	L	0.598	0.600	0.612	0.537
	Q	0.506	0.516	0.540	0.493
	C	0.396	0.392	0.465	0.433

Table 1: Correlation for evaluating quality of students' summary
Mod: model of registration analysis (L: Linear, Q: Quadratic, C: Cubic)
Pred: prediction ; 0.15, 0.3, and 0.4: LSA threshold. If the value lowers than a threshold, then it's a missed; i.e. prediction value is 0.

References

Gilliam, S., Magliano, J.P., Millis, K.K., Levinstein, I.B., and Boonthum, C. 2007. Assessing the format of the presentation of text in developing a Reading Strategy Assessment Tool (R-SAT). *Behavior Research Methods Instruments and Computer*, 39 (20), 199-204.

Grishman, R. and Kslezyk, T. 1990. Causal and temporal text analysis: the role of the domain model. In *Proceedings of the 13th Conference on Computational Linguistics Volume 3* (Helsinki, Finland, August 20 - 25, 1990). H. Karlgren, Ed. International Conference On Computational Linguistics. Association for Computational Linguistics, Morristown, NJ, 126-131.

Landauer, T.K., McNamara, D.S., Dennis, S., and Kintsch, W. eds. 2007. *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum, Mahwah, NJ.