Hybrid Question Generation Approach for Critical Review Writing Support

Ming Liu[a*], Rafael A. Calvo[a] and Vasile Rus[b]
[a]School of Electrical and Information Engineering, University of Sydney, Australia
[b]Department of Computer Science, University of Memphis, USA
[*]Ming.Liu@sydney.edu.au

Abstract: Research towards automated feedback can build on the work in other areas. In this paper we explore question generation techniques. Most research in question generation has focused on generating content specific questions that help students comprehend a set of documents that they must read. However, this approach is not so useful in writing activities, as students would generally understand the document that they themselves wrote. The aim of our project is to build a system which automatically generates feedback questions for academic writing support, particularly for critical review support. This paper presents our question generation system which relies on both syntax-based and template-based approaches, and uses Wikipedia as background knowledge.

Keywords: Question Generation, Academic Writing Support, Natural Language Processing

Introduction

Critical review requires students to read relevant sources in enough detail so that they can present a fair and reasonable evaluation of the selected sources. The ability to critically review relevant literature is a key academic skill. Afolabi [1] identified the most common problems that students have when doing literature review which includes not being sufficiently critical, lacking synthesis and not discriminating between relevant and irrelevant materials.

Questions can be a very useful form of feedback to develop critical reviewing. An example of a generic feedback question is *Have you critically analyzed the literature you use? Instead of just listing and summarizing items, have you assessed them, discussing strengths and weaknesses?* However, such question is too general and not likely to be effective for helping a student write on a specific topic.

More specific feedback questions are needed to develop students' critical review skills and help them with their writing. Table 1 shows an example of two system-generated questions from an engineering research student's literature review document. These questions were triggered by the citation sentence, copied from a literature review document, shown at the top of the table. Q1 asks the student-writer to critically evaluate the citation sentence by asking for supporting evidence and other people's views. Q2 asks the writer to critically compare the technique Principal Component Analysis (PCA) with another similar technique called factor analysis.

Table 1: Example of system-generated feedback questions

Citation 1: Cannon challenge this view mentioning that physiological changes were not sufficient to discriminate emotions.
Q1: Why did Cannon challenge this view mentioning that physiological changes were not sufficient to discriminate emotions? Does any other scholar agree or disagree with Cannon?

key phrase: Principal Component Analysis (PCA)
Q2: *What do you think of the differences between PCA and factor analysis in relation to your project?*

1. The Automatic Question Generation Framework

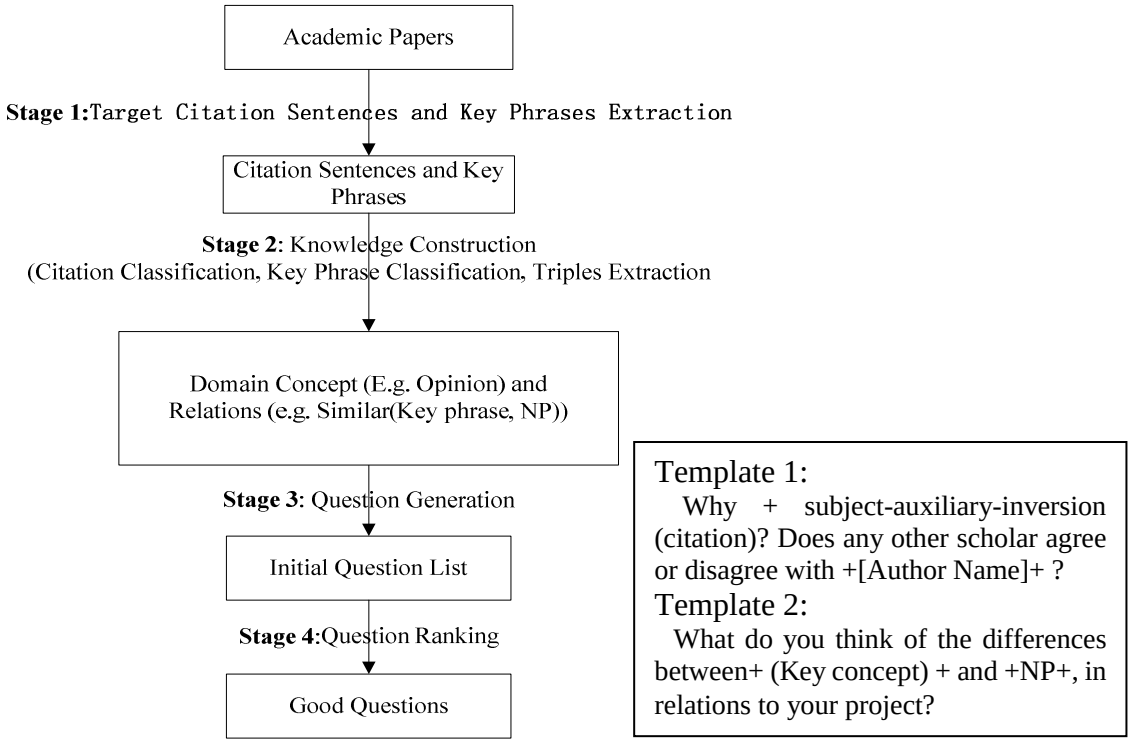

Figure 1: Four Stages Question Generation Processing

Our question generation approach is based on two important topic independent features of an article: citations and key phrases. The question generation process has been divided into four main stages as shown in Figure 1. In stage 1, the citations and key phrases are identified using regular expressions and the Lingo algorithm [2], respectively. In stage 2, according to the function of a citation, each citation is classified in one major citation category in the scientific register. We have defined the following 6 citation categories: Opinion, Result, Study Aim, System, Method and Application. Meanwhile, each key phrase is linked to a Wikipedia article and classified in a scientific instrument category based on the first section of the Wikipedia article. We defined 4 useful scientific instrument categories: Research Field, Technology, System and Term. If the key phrase is classified as one of the instruments, information extraction techniques are used to extract triples from the linked Wikipedia page. Each triple contains a key concept, another noun phrase (NP) or sentence (S), and their relation. We have defined 7 common relations including Has-limitation, Has-Strength, Include-Technology, Apply-to, Similar-to, Different-to and Kind-of. In stage 3, questions are generated based on these concepts and their relations. As we mentioned before, the goal is to generate specific questions using the syntax-based approach to transform the declarative citation sentence into questions. We also need deeper questions and use the template-based approach to fill the extracted citation or triples into pedagogical question templates. In stage 4, like Heilman and Smith's approach [3], a statistical question ranker is used to rank the quality of the generated questions in terms of readability.

For example, we assume that the citation and key phrase shown in Table 1 are extracted from stage 1. In stage 2, the citation 1 is classified as an Opinion concept. The Principle Component Analysis (Key Phrase) is linked to the Wikipedia article called PCA and

classified as a Technology Concept. From the Wikipedia article, we extracted a triple which contains a key phrase, a noun phrase (NP)/sentence (S) and its relation, which is denoted RelationName(Key Phrase, NP/S). In stage 3, Q1 is generated using the template 1 shown in Figure 1. In this case, we need to perform Subject-Auxiliary Inversion commonly used in the syntax-based approach to generate "Why did Cannon challenge this view mentioning that physiological changes were not sufficient to discriminate emotions?" We then use the template-based approach to fill the extracted author name in the question template ("Does any other scholar agree or disagree with Cannon?").

2. Discussion and Future Work

This paper presents a hybrid approach for generating specific deep questions which are used for critical review writing support. Citations and key phrases are the target elements for the question generation because they are important features of an article. Asking critical questions about citations can be effective for critical literature review writing because citations directly relate to the literature and such questions have high specificity. In this case, the questions generation process does not necessarily require any background knowledge. However, questions generated from key concepts (key phrases) used background knowledge extracted from Wikipedia. Wikipedia can be seen as huge lexical semantic resource that includes knowledge about domain specific terms. Therefore, it would be useful to ask questions about understanding these key concepts based on the knowledge extracted from Wikipedia.

Previous works [4-6] only separately proposed and evaluated the question generation based on citations or key phrases. Our future work will evaluate the new question generation framework and investigate the differences between questions generated from both.

Acknowledgements

This project was partially supported by a University of Sydney TIES grant, an Australian Research Council Discovery Project grant (DP0986873) and Google Research Award for measuring the impact of feedback on the writing process.

References

[1] Afolabi, M. (1992) "The review of related literature in research," *International journal of information and library research*, vol. 4, pp. 59-66.

[2] Osinski, S., Stefanowski, J., and Weiss, D. (2004),"Lingo: Search results clustering algorithm based on singular value decomposition," in *The International Conference on Intelligent Information Systems*, Zakopane, Poland, pp. 359--368.

[3] Heilman, M. and Smith, N. A.(2010) "Good Question! Statistical Ranking for Question Generation," presented at the Annual Conference of North American Chapter of the Association for Computational Linguistics - Human Language Technologies, Los Angeles.

[4] Liu, M., Calvo, R. A., Aditomo, A., and Pizzato, L. A. (to appear 2012) "Using Wikipedia and Conceptual Graph Structures to Generate Questions for Academic Writing Support," *IEEE Transactions on Learning Technologies*,

[5] Liu, M., Calvo, R. A., and Rus, V. (2012) "G-Asks: An Intelligent Automatic Question Generation System for Academic Writing Support," *Dialogue and Discourse: Special Issue on Question Generation*, vol. 3, pp. 101-124.

[6] Liu, M., Calvo, R. A., and Rus, V.(2010) "Automatic Question Generation for Literature Review Writing Support," presented at the 10th International Conference on Intelligent Tutorial Systems, Carnegie Mellon University, USA.