Question Generation: Papers from the 2011 AAAI Fall Symposium (FS-11-04)

Towards a Model of Question Generation for Promoting Creativity in Novice Writers

Julius Goth
North Carolina State University
jgoth@ncsu.edu

Eun Ha
North Carolina State University
eha@ncsu.edu

James Lester
North Carolina State University
lester@ncsu.edu

Abstract

Automated question generation has been explored for a broad range of tasks. However, an important task for which limited work on question generation has been undertaken is writing support. Writing support systems, particularly for novice writers who are acquiring the fundamentals of writing, can scaffold the complex processes that bear on writing. Novice writers face significant challenges in creative writing. Their stories often lack the expressive prose that characterizes texts produced by their expert writer counterparts. A story that is composed by a novice writer may also lack a compelling plot, may not effectively utilize a story's setting, characters, and props, and may describe events that play out in an unpredictable or confusing order. We propose an automatic question generation framework that is designed to stimulate the cognitive processes associated with creative writing. The framework utilizes semantic role labeling and discourse parsing applied to the initial drafts of the writer's passage to generate questions to promote creativity.

Introduction

Promoting creativity in novice writers is a challenging task. Novice writers face considerable obstacles crafting and refining their stories. To address this challenge, writing support systems could offer novice writers interrogative hints to provide them with the "aha" moment and to engage them creatively. In this vein, work has been conducted that shows students presented with trigger questions perform better than those who are not presented with such questions (Reynolds and Bonk, 1996). Building on foundational work on trigger questions, natural language processing techniques could be devised that have the ability to inform a question generation model based on the writer's initial drafts.

In this paper we first discuss key challenges posed by question generation for writing creativity. We then describe corpora collected to study narrative writing support, and introduce a framework for automatic question generation for novice writers. We then describe semantic and discourse analysis techniques that inform a question generation model to provide the creativity spark to motivate novice writers' to revise their writing and add depth and breadth to their stories.

Question Generation for Writing Creativity

Much of the recent research on question generation focuses on factoids contained within a source passage. To effectively promote creativity in writers, 'creativity-enhancing' questions should be posed. Certain types of questions may indeed trigger further revision in a passage. However, this may simply produce surface level elaboration and may not yield fundamentally more creative story-writing activities. For example, the novice writer might simply devise a slightly more detailed description of a character or setting, rather than enacting deeper revisions such as tweaking the plot to be captivating or flow more evenly.

To this end, it is useful to create a question taxonomy that helps to identify the types of questions that maximize creativity. Nielsen et al. (2008) define a question taxonomy drawn from several sources, e.g., (Bloom and Krathwohl, 1956), (Collins, 1985), (Graesser, Lang, and Horgan, 1988), for use in six subject areas across a variety of grade levels. For promoting creativity, these questions should be open-ended to provide students with a basis to reflect on alternative progression of their narrative.

Corpus Collection

We are exploring these issues in the context of the NARRATIVE THEATRE, a writing support system that is being designed for middle grade language arts education. The system targets the genre of fables. We have collected two corpora of student-written narratives using a multimedia interface to guide the students through the

Figure 1. NARRATIVE THEATRE Fable Composition Support Environment

planning and writing process. During the planning phase, users selected a moral, a setting, a cast of characters, and a set of objects for the story they will create. The NARRATIVE THEATRE (Figure 1) provides nine morals, four settings, ten characters, and twenty objects from which users may choose. Each setting has a visual descriptor associated with it, along with the ability for the student to enlarge the image by clicking on it.is annotated with salient features of the setting. Characters and objects are also visually represented by static graphics.

Once the choices have been made, users are presented with a planning area that allows them to view their previous selections and begin structuring their fables. At the top of the screen, there are icons illustrating the student's moral, setting, characters, and objects selection. For the setting, characters, and objects, the selection can be "zoomed in" by hovering over the appropriate selection. Students then craft a textual plan for the beginning (setting and characters are introduced), middle (introducing the conflict and problem), and end (conflict resolution) of their stories. After the planning information is entered, the user may begin writing. They then create the prose, which is entered as text on a notebook overlay image. The writing and revision phases are supported with a spelling error-correction facility. All student activities including interface selections and the text streams from planning and writing are logged and time-stamped. To avoid the potential of distracting the student from the writing task, the spell-correction facility was removed in the second version of the system.

The corpus collection activity spanned two days for each student. On the first day, the students were seated at a computer and asked to fill out a pre-survey questionnaire that included demographic and self-efficacy questions. They were given approximately twenty minutes to complete the survey. On the second day, the students were again assigned to a computer and presented with a login screen for the NARRATIVE THEATRE interface. Once they began, the students were presented with a short instructional video that described the features and operation of the interface. They were given fifteen minutes to complete the planning activity describe above (choosing a setting, characters, props, and deciding the beginning, middle, and end of their story). Once planning was completed, or time expired, the students were given another thirty-five minutes to write their fables. At the end of this block of time, the students were asked to compete a post-survey questionnaire, for which they were allotted

twenty minutes for completion. In total, the session lasted no more than eighty minutes.

Question Generation Architecture

The design of the question generation model that is being explored for the NARRATIVE THEATRE fable composition support user interface utilizes a natural processing pipeline (Figure 2). Its pipeline consists of a preprocessing grammatical disfluency corrector for missing punctuation and real-word spelling errors, a morphological processor, a semantic parser for extracting actions and supporting roles, and a discourse parser for analyzing the rhetorical structure of narratives.

Semantic Role Labeling

We will utilize semantic role labeling (Pradhan et al., 2004) to perform shallow semantic parsing on the initial drafts of students' stories. This semantic analysis task is essential in generating questions, such as "What if…", querying on the outcome of proposing alternate scenarios. The semantic role labeler will be supported by a coreference resolver, e.g., the OpenNLP framework (Baldridge, Morton, and Bierner, 2001), for pronoun resolution to resolve pronouns in students' stories that refer to characters and story-world objects.

Discourse Parsing

While a semantic role labeler provides a mapping from actions to the supporting roles of those actions within a story, to generate the most meaningful questions, it will also be important to identify the structure of rhetorical relationships among the narratives' events. To this end, we will use a discourse parser to organize sentences on salience and their relationships to one another.

Discourse parsing has been used effectively in automatic text summarization (Marcu, 1997). In our question generator, a discourse parser will extract nucleus elements from the global context and hypothesize questions based on this summarization. We propose to use Markov Logic Networks (Richardson and Domingos, 2006), a powerful tool used for relational learning, to inform our discourse parser. We will utilize our earlier work on temporal relation modeling for discourse analysis (Ha et al., 2010).

Discussion

Analyzing stories produced by novice writers for candidate questions poses significant NLP challenges because of various grammatical disfluencies. The stories in the fable corpus exhibit significant grammatical problems (Goth et al. 2010). To address this issue, our system has incorporated a grammatical correction component to address some of these errors, such as real-word spelling errors and punctuation elision. This component consists of a classification model trained used fable text that has been manually annotated with each error's corrected version.

Evaluation of an automatic question generation engine poses considerable challenges. First, one is faced with the

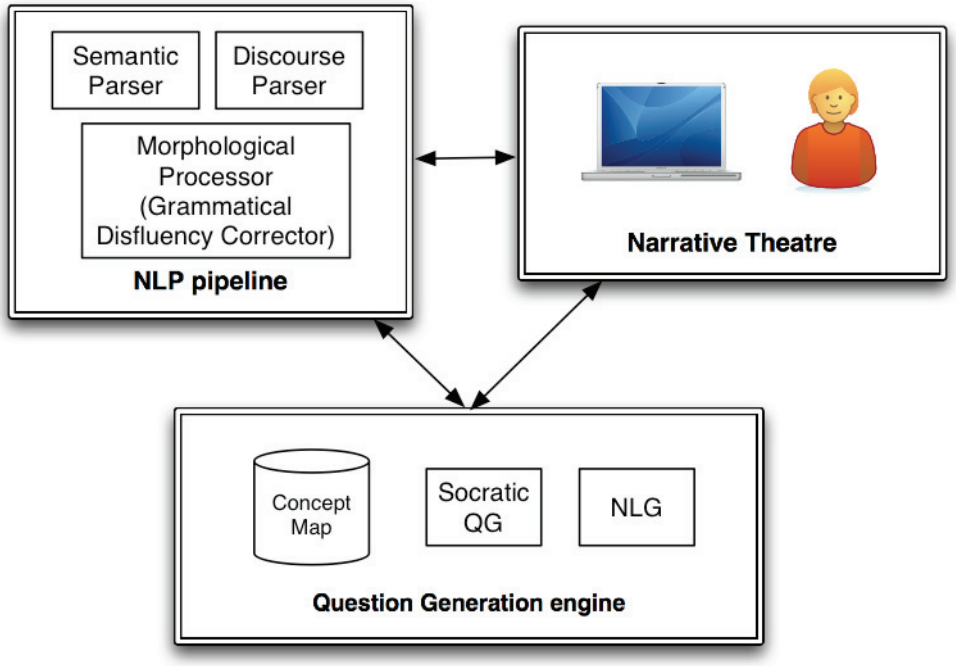

Figure 2. Question Generation Model

difficulty of identifying what constitutes an "interesting" question. While a consensus view of evaluation criteria have begun to materialize for question generation (Rus et al., 2010), it is not evident that the criteria apply well to the more open-ended Socratic question generation task. For selecting an n-best list of questions, statistical techniques such as overgenerate-and-select ranking system (Heilman and Smith, 2010) can be used .

To further evaluate the efficacy of a question generation model creating "What if…" questions, a guideline must be established to calculate the depth of students' revised text between pre-QG and post-QG passages. While high-level metrics can be utilized, such as examining lexical or syntactic differences, these metrics provide only a surface level analysis relating to the student's revision task. Deeper semantic- and discourse-related differences must be identified to establish what events in the story were altered.

Conclusion

Automatic question generation can play a central role in writing support systems for novice writers. We are designing a question generator that will utilize semantic role labeling and discourse parsing to inform the question generation process. Our goal is to design a robust question generation system that can extract the salient elements from novice writer's stories and generate meaningful Socratic questions to scaffold the cognitive processes of planning, writing, and revision.

We are currently designing the question generation framework, which will be instantiated in an implemented question generator that will be incorporated into the NARRATIVE THEATRE writing support system. Key to this process is the formal specification of a Composition-Centered Socratic writing question taxonomy. The question generator will be empirically studied with novice writers in a study in which writers will interact with the question generator throughout the writing process.

Acknowledgements

The authors wish to thank members of the North Carolina State University IntelliMedia Group for their assistance with NARRATIVE THEATRE. This research was supported by the National Science Foundation under Grant IIS-0757535. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Baldridge, J., Morton, T., and Bierner, G. 2001. The OpenNLP Maximum Entropy Package. http://maxent.sourceforge net.

Bloom, B. S., and Krathwohl, D. R. 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of college and university examiners. Handbook I: Cognitive Domain.* Longmans.

Collins, A. 1985. Teaching Reasoning Skills. *Thinking and Learning Skills*, Volume 2.

Goth, J., Baikadi, A., Ha, E., Rowe, J., Mott, B., and Lester, J. 2010. Exploring Individual Differences in Student Writing with a Narrative Composition Support Environment. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing: Writing Processes and Authoring Aids, 56-64*. Los Angeles, 56-64.

Graesser, A. C., Lang, K., and Horgan, D. 1988. A Taxonomy for Question Generation. *Questioning Exchange*.

Ha, E. Y., Baikadi, A., Licata, C. J., Mott, B. W., and Lester, J. C. 2010. Exploring the Effectiveness of Lexical Ontologies for Modeling Temporal Relations with Markov Logic Modeling Temporal Relations with Markov Logic. *Eleventh Annual SIGDIAL Meeting on Discourse and Dialogue*, 75-78. Tokyo.

Heilman, M. and Smith, N.A. 2010. Good Question! Statistical Ranking for Question Generation. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*, 609-617.

Marcu, D. 1997. The Rhetorical Parsing of Natural Language Texts. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, 96-103. Stroudsburg, PA.

Nielsen, R., Buckingham, J., Knoll, G., Marsh, B., and Palen, L. 2008. A Taxonomy of Questions for Question Generation. In *Proceedings of the Workshop on the Question Generation Shared Task and Evaluation Challenge*. Arlington, Virginia.

Pradhan, S., Ward, W., Hacioglu, K., Martin, J. H., and Jurafsky, D. 2004. Shallow Semantic Parsing using Support Vector Machines. In *Proceedings of the Human Language Technology Conference/North American chapter of the Association for Computational Linguistics annual meeting (HLT/NAACL-2004)*. Boston, MA

Reynolds, T. H., and Bonk, C. J. 1996. Computerized Prompting Partners and Keystroke Recording Devices: Two Macro Driven Writing Tools. *Educational Technology Research and Development*, *44*(3), 83-97.

Richardson, M. and Domingos, P. 2006. Markov Logic Networks. *Journal of Machine Learning*, 62(1-2): 107-136.

Rus, V., Wyse, B., Piwek, P., Lintean, M., Stoyanchev, S., and Moldovan, C. 2010. Overview of the First Question Generation Shared Task Evaluation Challenge. In *Proceedings of QG2010: The Third Workshop on Question Generation*, 45-57.