An Automatic Quiz Generation System for English Text

Li-Chun Sung, Yi-Chien Lin and Meng Chang Chen
Institute of Information Science, Academia Sinica, Taipei, Taiwan
E-mail: {mico63, jascha, mcc}@iis.sinica.edu.tw

Abstract

In this study, we design and prototype an automatic quiz generation system (auto-quiz for short) for a given English text to test learner comprehension of text content and English skills. The auto-quiz process parses an English text into a semantic network representation and enhances the semantic network iteratively with intrinsic knowledge, such as English grammar and writing styles, and extrinsic knowledge, such as the word relationship in WordNet and statistics from corpus or search engines. Then, the quiz generation process generates quiz from the text based on learner comprehension skills and according to the learner learning status and needs, such as English proficiency and frequent errors.

1. Introduction

Traditionally, quizzing is a primary manner to examine the learner's learning effect. For a well produced teaching material, a teacher can prepare manually a quiz for the purposes of learning effect evaluation, memory reinforcement, teaching strategy adjustment, etc.

In a ubiquitous English learning environment, learners can learn English while they browse through the Internet documents or read English documents of their professions. In such environment, it is not likely to have quiz prepared for each English text beforehand. Hence, an auto-quiz system is needed for a ubiquitous learning environment.

While researchers had proposed some strategies for automatic question generation, most of them focused on vocabulary assessment only. However, if a reader understands a reading, the learner will have the skill to comprehend the meaning of each vocabulary based on its context in the reading. Thus, we propose the system to evaluate learner’s comprehension skill.

2. The auto-quiz system

In our system, the relationship between a vocabulary and its context is represented by a semantic network (SemNet for short), which constitutes of players, actions, attributes, and relationships between them. To understand a given text, as well as learners, our system will attempt to generate such SemNets and integrate them into a coherent SemNet to promote further comprehension.

As shown in Fig. 1, the quiz of a given text is generated by the collaboration of the four parts. First, the top-left part in the figure is responsible for the generation of SemNet. In the preprocessor, sentences will be recomposed to avoid erroneously parsing caused by their specific structures, such as quotation. And then, each sentence will be translated into a SemNet by using Link Grammar Parser based on its syntactic structure. In the central-bottom part, the generated SemNets will be integrated and refined by SemNet refiners, such as anaphora resolution. In the sequel, in the top-right part, quiz generator, guided by user profile, will collect prerequisites for each question pattern inferred from the SemNet. If the sufficient prerequisites can be obtained for a question pattern, the question generator will produce the corresponding question. The generated questions will be organized to form a quiz.

Figure 1. The quiz generation process.
All the three subprocesses work by the assistance of the knowledge base which is extendable dynamically.
4 The question generation subsystem

In the auto-quiz system, there are several question generators that each generates one type of question. In this paper, we will introduce two question generators.

4.1 Question for sense comprehension of adjective

While a learner understands a text, the learner will be capable of inferring the meaning of a vocabulary from its context. This motivates us to develop the question generator to examine a learner for the sense of a vocabulary. The generator will extract attributes (adjectives) of players from the SemNet of a given text as questionnaire vocabularies and form multiple-choice cloze questions by scooping these vocabularies out of the corresponding sentences. An example is shown in Fig. 2. The options of a question include the right answer and the distracters. The right answer is substituted by the synonym or similar adjective of the applied sense of the questionnaire adjective. The whole substitute identification process is shown in Fig. 3. The candidates of the target substitute are acquired from WordNet. And then, most of them will be filtered out by two filtering steps. In the first step, we will check whether a candidate can be an attribute of the corresponding player in the knowledge base. In the second step, we will check whether the attributive relationship of the remaining candidate is raised frequently through similar texts.

4.2 Question for anaphor comprehension

A SemNet generated based on the structures of sentences of a text is often fragmental because each of the player nodes that correspond to anaphors forms an independent subnet. To understand the meaning of a text, a learner must integrate these subnets by connecting each anaphor with its antecedents. Thus, we design a question generator to examine whether a learner understand the association between an anaphor and its antecedent. The generator will identify the antecedent of an anaphor and form a multiple-choice cloze question by scooping the anaphor out of its sentence. The options comprise its antecedent and the distracters. An example is shown in Fig. 4.

The core of the process is the anaphora resolution scheme which is still a challenge issue today. In addition to simple reference restriction rules (e.g. the gender agreement, number agreement, etc.), we enhance the resolution by identifying the happens_before relationship between the joined events of the anaphor and the joined events of each antecedent candidate.

5. Conclusion

Currently, a prototype system has been implemented on the Java platform. However, automatic quiz generation is an open-ended task since the knowledge base required is tremendously large and cannot be completed with limited sources and time. Our future work is to solicit the helps of volunteers from the Internet for the construction of the knowledge base.