Liu, C.-C. et al. (Eds.) (2014). Proceedings of the 22nd International Conference on Computers in Education. Japan: Asia-Pacific Society for Computers in Education

Collecting Pairs of Word Senses and Their Context Sentences for Generating English Vocabulary Tests

Yuni SUSANTI[a*], Ryu IIDA[a] & Takenobu TOKUNAGA[a]
[a]*Department of Computer Science, Tokyo Institute of Technology, Japan*
*susanti.y.aa@m.titech.ac.jp

Abstract: This paper describes our ongoing work for automatically generating multiple-choice questions for English vocabulary tests. Vocabulary-type questions consist of four components: a target word, context sentences, a correct choice and distractors. To generate such type of questions, identifying the word sense of a target word in a context is crucial to create a correct choice in the context sentences. We propose a novel Backward method that utilizes the existing search engine API to obtain context sentences for a given target word sense and further applied word sense disambiguation to confirm the retrieved context sentences. Our preliminary experiment showed that the proposed method achieved 0.951 in accuracy in collecting word sense-context pairs.

Keywords: Question generation, multiple-choice vocabulary questions, language learning, word sense disambiguation

1. Introduction

As the demands of communication across diverse communities have been developing in the recent years, the use of English as the main international language has increased to interact with different societies both in business and academic settings. Owing to this, English proficiency tests such as TOEFL® and TOEIC® are important to measure the English skill of non-native English speakers. However, since the past questions of those tests are not freely distributed, test takers can only rely on a limited number of test samples and preparation books. It is beneficial for the test takers if similar questions of those tests can be automatically generated from vast amount of materials in the Internet.

To address that issue, this research works on automatic question generation for English proficiency test practice. As an example, we focused on the vocabulary-type questions because it is the majority type of questions in the TOEFL iBT® reading section and also appears in other English proficiency tests such as TOEIC®. TOEFL iBT® vocabulary-type questions consist of four components as shown in the right part of Figure 1: a target word, context sentences, a correct choice and distractors (wrong choices). To generate such type of questions, identifying the word sense of a target word in a given context is crucial especially to produce the choices (a correct choice and distractors) of the question. That is why, collecting pairs of a word used in a certain sense and its context (i.e. sentences in the passage that contain the target word, henceforth called *context sentences*) is indispensable. In this research, we propose a novel Backward (BW) method that utilizes the existing search engine API to obtain context sentences for a given target word from the Internet.

2. Related Work

Question generation in the language learning domain has been broadly studied, e.g. generation of fill-in-the-blank questions for completing a sentence, words collocation, synonym, antonym, etc. This paper describes our work on generating vocabulary-type questions which ask for "closest in meaning" of a target word. Similar work has been done by Brown et al. (2005), generating multiple-choice questions where their components were taken from WordNet (Fellbaum, 1998). Lin et al. (2007) also adopted WordNet filtered with web corpus searching for producing English adjective questions.

3. Method

To generate context sentences, we use the combination of the proposing BW method with Word Sense Disambiguation (WSD). The BW method chooses a sense of a target word randomly from the dictionary, and then uses a search engine to retrieve a snippet from the Internet that contains the target word with the chosen sense. WSD is then applied to the target word in the retrieved snippet to confirm that the predicted word sense is the same as the chosen sense. The confirmed snippet is used as context sentences for generating a question.

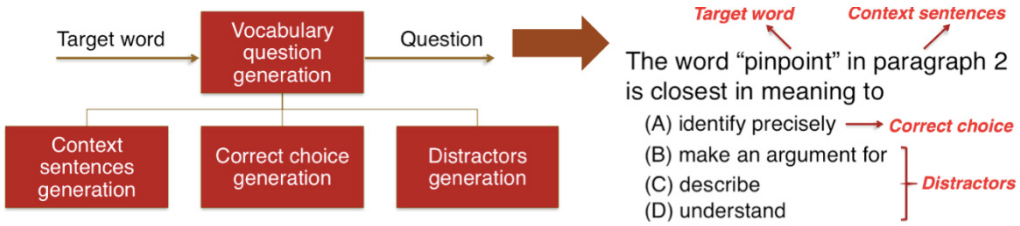

Figure 1. General workflow and components of TOEFL iBT® vocabulary-type questions.

3.1 Backward Method

In the BW method, given a target word, we start by choosing one of its word senses that has an example sentence in WordNet. Next, a query for the search engine is created from the example sentence by taking the target word and its adjacent words in both sides after removing the stop words such as *the, on, are,* etc. The third step submits the query to the search engine to retrieve snippets containing the target word with the chosen sense, and the last step selects one snippet which is the most suitable context sentences for the target word based on the following 3 scores: 1) word overlap between the example sentence and the snippet, 2) the number of adjacent query words to the target words in the snippet after removing the stop words, 3) the number of query words appearing in the snippet. The BW method is illustrated in Figure 2.

Figure 2. Backward Method.

3.2 Word Sense Disambiguation (WSD)

WSD is the task of identifying the meaning of words in context in a computational manner (Navigli, 2009). Vocabulary-type questions ask for "the closest in meaning" of a target word, thus to generate the correct choice we need to identify the meaning of the target word in a particular context. Therefore, WSD is very important in generating vocabulary-type questions, especially to generate a correct choice. The state of the art WSD methods as explained in McCarthy (2009) reaches around 0.37 in accuracy with knowledge-based approach, 0.88 with supervised and 0.82 with unsupervised approach. Further explanation on WSD can be found in survey papers by Navigli (2009) and McCarthy (2009). In this research we used Lesk Algorithm (Lesk, 1986) which chooses the sense that shares the highest number of words in its gloss in a dictionary and the current context. For example,

*Context sentence: I **inserted** the **key** and **locked** the door.*

*Sense 1: Metal device shaped in such a way that when it is **inserted** into the appropriate **lock** the **lock**'s mechanism can be rotated.*

Sense 2: Something crucial for explaining; "The key to development is economic integration."

In this case, sense 1 is the correct sense for the target word "key" because its definition has 3 words overlap (*insert, lock, lock*) with the context sentence, while sense 2 has 0 overlap.

4. Result and Discussions

We conducted a preliminary experiment on two target word sets: 98 target words from sample question and preparation books of TOEFL iBT®, and 98 target words from Senseval data which is data from WSD workshop. These two target word sets share no common word. The Bing Search API was used as the search engine. The result is presented in Table 1.

Table 1: Experimental result.

Methods	TOEFL iBT®	Senseval
WSD (Lesk Algorithm)	0.602	0.296
BW	0.885	0.745
BW + WSD	0.951	0.843

In the evaluation of WSD, WSD was applied to target words in the TOEFL questions or in the Senseval data to predict their word sense. In the BW case, provided a sense of the target word, the BW method tries to find the context sentences in which the target word is used with the given sense. The word senses used in this experiment were chosen randomly from the first 2 senses in WordNet. Manual judgment was done to check whether the retrieved context sentences include the target word in the given sense or not. We further applied WSD to the target words in the retrieved contexts by the BW method, and filtered out those cases that the given and predicted word senses did not match. This result is shown in the "BW+WSD" row.

The accuracy of the BW method reached 0.885 on TOEFL iBT® data. In addition, by combining with WSD the accuracy improved to 0.951. Although it is still a preliminary evaluation, the proposed BW method combined with WSD shows promising results in this research. In particular, this method is suitable for our domain (language learning) because WSD in this domain is in general easier compared with common WSD domains (used in Senseval data) as shown in Table 1.

5. Future Work

We proposed a novel method to collect pairs of word senses and their context sentences from the Internet as the early steps for generating multiple-choice English vocabulary tests. In forthcoming work we will generate correct choice by using the sense's synonym or definition and the distractors of the questions to complete all necessary components of multiple-choice vocabulary questions. Next we also plan to evaluate generated questions through real English test for English learning students.

References

Brown, J. C., Frishkoff, G.A., Eskenazi, M. (2005). Automatic Question Generation for Vocabulary Assessment. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing* (pp. 819-826).

Fellbaum, C. (1998). WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

Lesk, M. (1986). Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. *Proceedings of the ACM SIGDOC Conference* (pp. 24-26), Toronto.

Lin, Y. C., Sung, L. C., Chen and M. C. (2007). An Automatic Multiple-Choice Question Generation Scheme for English Adjective Understanding. *CCE 2007 Workshop Proc. of Modeling, Management and Generation of Problems / Questions in eLearning* (pp. 137-142).

McCarthy, D. (2009). Word Sense Disambiguation: An Overview. *Language and Linguistics Compass 3* (2), 537-558.

Navigli, R. (2009). Word Sense Disambiguation: A Survey. *ACM Computing Surveys* (41), 2, Article 10.