

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Automatic English Vocabulary Question Generation for Efficient Measurement of Learner Proficiency
著者(和文)	ユニ スサンティ
Author(English)	Yuni Susanti
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第10995号, 授与年月日:2018年9月20日, 学位の種別:課程博士, 審査員:徳永 健伸,岡崎 直観,宮崎 純,村田 剛志,藤井 敦,宇佐美 慧
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第10995号, Conferred date:2018/9/20, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

(博士課程)
Doctoral Program

論文要旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	計算工学専攻 コース	系	申請学位 (専攻分野)： Academic Degree Requested	博士 Doctor of	(Engineering)
学生氏名： Student's Name	YUNI SUSANTI		指導教員 (主)： Academic Supervisor(main)	TOKUNAGA TAKENOBU	
			指導教員 (副)： Academic Supervisor(sub)		

要旨 (英文 800 語程度)

Thesis Summary (approx.800 English Words)

Conducting a language test is indispensable to evaluate the proficiency of the language learners. However, manual construction of questions is a difficult task that requires a high level of skill from experts. Hence, an automatic question generation system can be a breakthrough by assisting the experts in making questions; thus, it makes the question construction easier. This thesis, consisting of seven chapters, presents a study on automatic generation of multiple-choice English vocabulary questions for efficient measurement of language learner proficiency. It consists of four topics: (1) automatic question generation (AQG), (2) distractor improvement, (3) question difficulty control and (4) integration of AQG into the computerised adaptive test (CAT). We proposed a new method for each topic and conducted evaluations involving learners, teachers and experts.

In the first topic, we proposed a novel method for automatically generating English vocabulary questions, modelling the generated questions after the TOEFL vocabulary questions. In this type of question, determining the word sense of the target word in a reading passage is crucial to creating the question options (the correct answer and distractors). We could use word sense disambiguation (WSD) techniques to identify the word sense. However, the accuracy of the state-of-the-art WSD method remains about 70-80%, which is far from satisfactory to the question generation task. Thus, we proposed a method that avoids word sense disambiguation. Instead, we took an information retrieval approach where given a target word and one of its word sense, we search a passage that uses the target word with the given word sense. We conducted two kinds of evaluation for assessing the quality of the generated questions: 1) test taker-based evaluation and 2) expert-based evaluation. In the test taker-based evaluation, we administered the machine-generated questions together with human-made questions to the real students. Both evaluations showed that the machine-generated questions have a comparable ability to the human-made questions in measuring the student proficiency, and the English teachers were not able to distinguish more than half of the machine generated-questions from human-made questions.

Through the evaluation, we found that distractors are the primary source of low-quality machine-generated questions. Thus, the second topic focuses on improving the distractors. We proposed a new method to generate distractors that aggregates both semantic similarity and word collocation information. The method finds distractors which are close to the target word but far from the correct answer in their meaning, and also collocate with the adjacent words of the target word in the given context (the reading passage). The evaluation showed that the proposed method removes the problematic distractor candidates better than the baseline, and the generated distractors have comparable quality to the original human-made distractors. A further error analysis showed that we could use the problematic distractors generated by the proposed method for a real test despite their low score by the human expert.

Toward an efficient measurement of the language learner proficiency, we proposed to integrate the AQG with CAT, which presents items tailored to the test taker proficiency, e.g. the item difficulty suits to the test taker proficiency. Therefore, CAT needs a big collection of items with their item difficulty known in advance. The conventional CAT estimates item difficulty from the test taker's responses in a pretesting phase. However, this process is expensive and poses a risk of exposing the item before the real test. To cope with this problem, we proposed to control the difficulty of the generated question with the three predetermined factors: 1) target word difficulty (TWD), 2) similarity between the correct answer and distractors (SIM) and 3) distractor word difficulty level (DWD). The analysis of test taker-based evaluation revealed that we could control the item difficulty with the predetermined factors, and the SIM factor contributes the most to the item difficulty.

We conducted simulation-based experiments on the AQG and CAT integration using two types of item difficulty, i.e. the item difficulty estimated from the test taker's responses and that controlled during the question generation process. We evaluated the performance of the simulations by looking at the mean squared error (MSE) between the true proficiency of the test takers and the proficiency estimated by each simulation. The result showed that all proposed CAT simulations with the controlled item difficulty yielded smaller MSEs than the baseline (a linear test) simulation. Moreover, their MSEs were close to the MSE of the gold standard, i.e. the CAT simulation with the estimated item difficulty by test taker responses. This is an encouraging result showing a possibility of integrating the CAT and AQG with controlling item difficulty, which can eliminate the pretesting.

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).