

Title: A Personalized Formative Assessment System for E-book Learning

Name: Albert Ming YANG

Abstract: Assessment is a common approach for evaluating students' comprehension and retention after learning knowledge. In contrast to summative assessments (such as quizzes or exams) that focus on one-time evaluation, formative assessments aim to improve student learning through repeated practice, receiving feedback, and adjusting behavior. However, several challenges may occur when employing formative assessment activities in real-world settings. First, generating additional questions for formative assessments can burden teachers due to time and resource constraints. Students may also demonstrate various behaviors during the formative assessments due to a lack of teacher surveillance. Unusual and nonstandard behaviors can affect the success of formative assessment. Furthermore, student engagement with the assessment may be reduced due to a lack of background knowledge, extremely difficult questions, or time management. Therefore, a system that helps teachers generate questions, identify students' patterns of formative assessment behavior, and motivate students to participate actively in practice is essential. With the development of artificial intelligence (AI) technology, studies have applied AI techniques for automatic question generation, behavioral analysis, and computerized adaptive testing (CAT) tasks to solve the above problems.

While existing studies have proposed advanced techniques for question generation and achieved promising results, most studies were conducted in laboratory settings or using standard datasets for performance competitions. The educational value of question generation remains to be further investigated as most approaches did not concern with the sentence selection process but mainly with converting a given sentence or paragraph into a question. In addition, although previous studies have identified several effective and nonstandard behaviors, most only considered basic features such as the frequency of attending practice, and the relationship between nonstandard behaviors and learning performance is lacking. Finally, most existing CAT methods are designed for accurate knowledge assessment rather than for learning, making it challenging to develop formative assessment systems. The current work aims to address these challenges by applying text summarization, question generation, CAT, and learning memory cycle models to propose a system that can automatically generate questions and recommend questions. Several experiments were conducted to examine the effectiveness of the proposed system in the e-book learning contexts. Meanwhile, clustering analysis was employed to analyze students' formative assessment behaviors of using the proposed system, and its influence on learning performance was explored.

The system involves four processes: sentence selection, question generation, assessment behavioral analysis, and question recommendation. The current thesis first compared three text summarization models for key sentence selection and examined whether the selected sentences can be used to score students' text marking skills, which is an essential reading skill. Then, a formative assessment system is proposed for students to practice their learned knowledge. We applied TextRank to select keywords for the selected sentences to generate fill-in-the-blank questions. Next, we employed hierarchical clustering to analyze students' formative assessment behaviors while using the system and identified behaviors that influenced learning performance. Finally, to create personalized formative assessments, the system integrates a CAT approach and

a learning memory cycle model to recommend questions based on students' retention of information learned from each item.

The thesis evaluated the effectiveness of each process in promoting e-book learning through four experiments. Participants in all experiments learned the materials on an e-book reading system, BookRoll. We measured whether students' reading skills, engagement, and comprehension could be enhanced by taking machine-generated formative assessments. Students' patterns of formative assessment behaviors were explored, and their relationship to learning performance was investigated. In addition, the study examined whether the adaptive feature of the assessment system can increase students' learning engagement and performance.

The results shows that students' reading skills, engagement, and comprehension improved significantly after repeatedly taking the generated formative assessments than those who restudied the materials. In addition, we found that the effectiveness of the assessment was more pronounced when students actively participated in the practice tests and did not exhibit nonstandard behaviors during the assessment process. The evaluation of the proposed adaptive assessment system shows that the recommendation based on CAT and the learning memory cycle model was more effective than the recommendation based on CAT alone in improving students' engagement in assessment and reading e-books and their learning performance.

The paper contributes to the related fields by proposing a personalized formative assessment system and investigating its impact on e-book learning. The proposed system can save teachers time in creating questions for formative assessments. For students, the system provides them with additional resources to review what they have learned. Identifying formative assessment behaviors allows teachers to provide intervention for students who demonstrate nonstandard behaviors. Personalized recommendations can motivate students to take assessments to improve knowledge retention. Therefore, the proposed system can significantly assist students' e-book learning. The comparisons between adaptive mechanisms for formative assessment and the exploratory analyses of students' behavioral patterns also provide insights for other researchers to design similar systems.