Research Article

An Adaptive Test Sheet Generation Mechanism Using Genetic Algorithm

Huan-Yu Lin,[1] Jun-Ming Su,[2] and Shian-Shyong Tseng[3]

[1] *Department of Computer Science, National Chiao Tung University, 1001 University Road,*
Hsinchu 300, Taiwan
[2] *Department of Information and Learning Technology, National University of Tainan, 33, Section 2,*
Shu-Lin Street, Tainan 700, Taiwan
[3] *Department of Information Science and Applications, Asia University, 500 Lioufeng Road, Wufeng,*
Taichung 41354, Taiwan

Correspondence should be addressed to Jun-Ming Su, junming.su@gmail.com

For test-sheet composition systems, it is important to adaptively compose test sheets with diverse conceptual scopes, discrimination and difficulty degrees to meet various assessment requirements during real learning situations. Computation time and item exposure rate also influence performance and item bank security. Therefore, this study proposes an Adaptive Test Sheet Generation (ATSG) mechanism, where a Candidate Item Selection Strategy adaptively determines candidate test items and conceptual granularities according to desired conceptual scopes, and an Aggregate Objective Function applies Genetic Algorithm (GA) to figure out the approximate solution of mixed integer programming problem for the test-sheet composition. Experimental results show that the ATSG mechanism can efficiently, precisely generate test sheets to meet the various assessment requirements than existing ones. Furthermore, according to experimental finding, Fractal Time Series approach can be applied to analyze the self-similarity characteristics of GA's fitness scores for improving the quality of the test-sheet composition in the near future.

1. Introduction

With the rapid developments of information and assessment technology, the computerized testing is generally used to assess, predict, and diagnose learners' learning statuses because it is able to effectively analyze examinees' abilities and learning barriers. The test quality offered by a computerized testing system depends on not only the quality of test items but also the satisfied test sheets to meet the various requirements of assessment parameters, such as the

difficulty degree, the discrimination degree, the associated concepts, and the expected testing time. Thus, how to efficiently assist teachers in composing and generating an appropriate test sheet to meet the diverse assessment requirements has become an important research issue.

Hwang [1] applied the dynamic programming technique to solve this issue, but the solution is inefficient for a large-item bank because of the exponential growth of time and space complexity. Su and Wang [2] developed an assistance system to provide teachers with statistic information for assisting teachers in manually composing the desired test sheets, but manually selecting appropriate test items in a large item bank is still inefficient and difficult to ensure the qualities of test sheets. Therefore, the pressing problem of automatic test item allocation is emerging and it can be regarded as a combinatorial optimization problem, which is proven an NP-hard problem [3]. Therefore, Hwang et al. [4] formulated this problem as a mixed integer programming model and proposed approximate solutions by using the Genetic Algorithm (GA) approach [5]. The experimental results show that their proposed approach can efficiently automatically compose a good enough test sheet for a large-scale test.

However, the aforementioned studies mainly aim to automatically generate a test sheet with a highest discrimination degree and to meet the constraints in terms of expected testing time and concept relevance. These mechanisms are suitable for the large-scale test only, but their natures are difficult to satisfy various purposes of assessments during the real learning situation. In order to efficiently understand the students' learning problems, it is important to compose the test sheets with diverse conceptual scopes ($\mathbb{C}$), discrimination ($\mathbb{D}$) and difficulty ($\mathbb{P}$) degrees, such as displacement and summative assessments (with normal distribution $\mathbb{C}$ and $\mathbb{P}$), and formative and diagnostic assessments (with various or specific $\mathbb{C}$ and $\mathbb{P}$) [6–9]. Moreover, the computation time of the test-sheet composition process and the Item Exposure Rate are our concerns as well. A long computation time will decrease the performance of test-sheet composition system and a high-item exposure rate will decrease the qualities of test items and Item Bank Security [10, 11]. Accordingly, to consider not only the various assessment requirements but also the computation time and item exposure rate, this study defines a new problem of automatic test item allocation, called an Adaptive Test Sheet Generation problem. To solve it, this research proposes Adaptive Test Sheet Generation (ATSG) mechanism, consisting of a Candidate Item Selection Strategy (CISS) and an Aggregate Objective Function (AOF). CISS process can adaptively determine candidate test items set and the conceptual granularities according to the desired concept scope, and AOF applies GA algorithms to solve the mixed integer programming problem. The evaluation results show that the proposed approach can generate test sheets to meet the various assessment requirements.

2. Related Work

The original issue of the test sheet generation problem is identified for the large-scale tests, where these test items covering all required concepts and having the highest degree of discrimination are selected from a test item bank. Hwang [1] proposed an algorithm based on dynamic programming technique to find optimal test sheets, but the exponential time complexity causes the efficiency issue for a large number of candidate test items. Therefore, the researchers formulated this problem as a mixed integer programming model and applied a genetic algorithm [4] to figure out the approximate solution. In this paper, assume that a

set of test items, which are related to m concepts, should be selected from n items in the item bank. Each test item Q_i is defined as

$$Q_i = (t_i, d_i, r_{ij}),$$ (2.1)

where Q_i is a test item in the item bank (IB) and has a set of parameters including the expected time t_i needed for answering, the degree of discrimination d_i, and the degree of association r_{ij} between Q_i and a concept C_j.

The assessment requirement of a Test Sheet (TS) includes the lower bound l and upper bound u of the totally expected answering time, and the lower bound h_j of the total relevance of each concept C_j. To formulate the problem, a decision variable x_i is defined as a *Kronecker delta*, that is,

$$x_i = \begin{cases} 1, & \text{if } Q_i \in \text{TS}, \\ 0, & \text{if } Q_i \notin \text{TS}. \end{cases}$$ (2.2)

The goal of this problem is to maximize $Z = (\sum_{i=1}^{n} d_i x_i)/(\sum_{i=1}^{n} x_i)$.

Subject to the concept range $\sum_{i=1}^{n} r_{ij} x_i \geq h_j$ for $j = 1$ to n and the testing time limitation $l \leq \sum_{i=1}^{n} t_i x_i \leq u$.

A Genetic Algorithm (GA) approach [5] is used to solve this problem, where a chromosome is represented as an n-bit binary string $[x_1, x_2, \ldots, x_n]$ and the fitness rank is the summation of selected items' discrimination degrees subtracted by the penalty scores. The penalty scores are the degrees about the violation of expected time and concept ranges constraints. The genetic algorithm iteratively generates new generation of chromosomes by the Crossover and Mutation processes, as Random Functions, and finds the best chromosomes according to their fitness ranks. In the Crossover, chromosomes of the next iteration are generated by combining halves of two chromosomes, which are randomly selected from the chromosomes in the current iteration. A chromosome can be more probably selected because it has a higher fitness rank. Mutation is the other operation of changing a chromosome, where the change of an arbitrary bit is randomly raised to a chromosome. This kind of evolutionary algorithm can iteratively approach to the optimal solution and use some random operations, such as the operations of Crossover and Mutation, to prevent falling into the local optimal solutions. According to the evaluation, the test sheet generation approach based on a GA can really provide good solutions among more than ten thousand test items in an acceptable response time. Furthermore, the greedy algorithm approach [12], the tabu search algorithm [13], and the discrete particle swarm optimization algorithm [14] were subsequently applied to enhance the computation efficiency of test sheet generation based on the aforementioned problem formulation.

Besides, the test sheet composition problem was extended to a parallel test sheets composition problem, where multiple test sheets are generated at one time. These sheets must have similar concept relevance, discrimination, and difficulty degrees but contain no common test items. The problem was solved by extending the existing tabu search algorithm [15] and the particle swarm optimization algorithm [16].
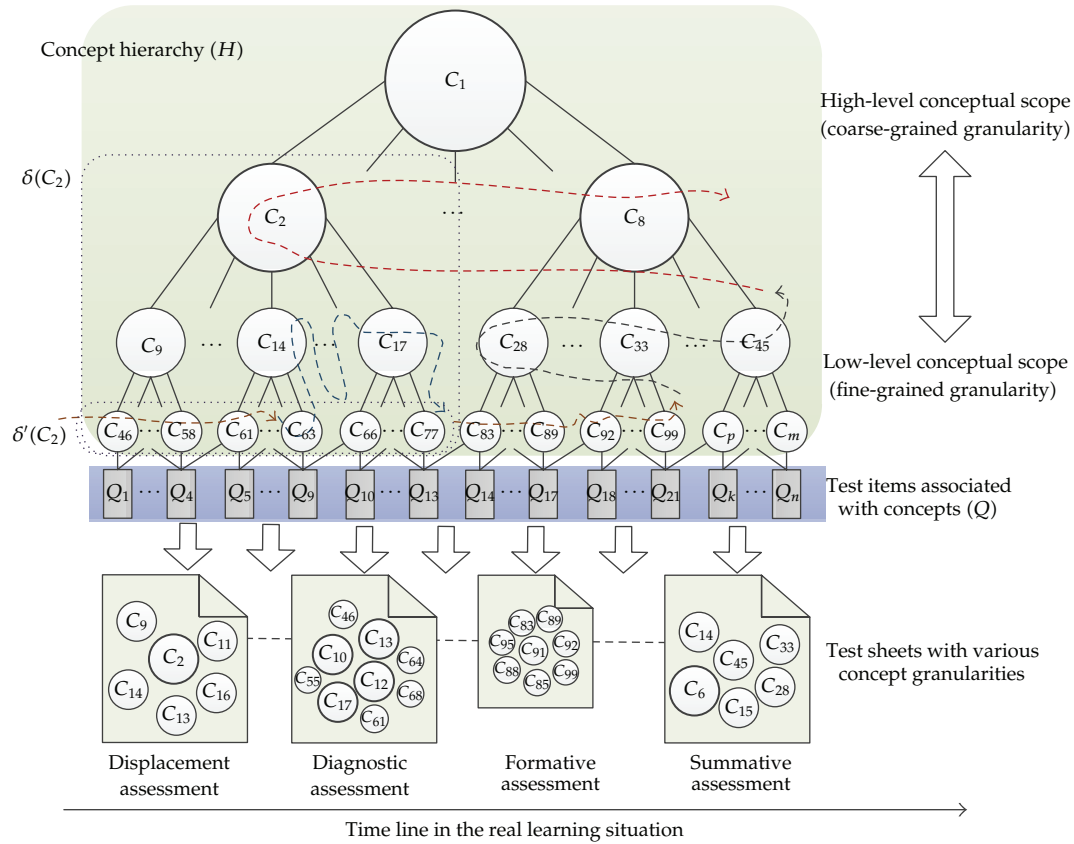
Figure 1: Test sheet types to meet various assessment requirements.

3. Adaptive Test Sheet Generation Problem

In order to efficiently understand the students' learning problems, the parameters of a test sheet including conceptual scopes ($\mathbb{C}$), discrimination ($\mathbb{D}$), and difficulty ($\mathbb{P}$) degrees should be adaptively composed according to the various assessment purposes, such as displacement and summative assessments (with normal distribution $\mathbb{C}$ and $\mathbb{P}$), and formative and diagnostic assessments (with various or specific $\mathbb{C}$ and $\mathbb{P}$). As illustrated in Figure 1, for the formative assessment, like a small-scale test, a test sheet with the specific and detailed concepts, that is, low-level conceptual scope/fine-grained granularity, is required to evaluate the students' specific conceptual capabilities during the learning; for the diagnostic assessment, like a specific-scale test, a test sheet with the diverse conceptual scopes and granularities is used to diagnose the students' learning problems; for the displacement and summative assessments, like a large-scale test, a test sheet with the high-level conceptual granularities is required to evaluate the students' learning performance before and after the learning, respectively.

However, as seen in Figure 2, the existing approaches did not take the adaptive requirements, that is, $\mathbb{C}$, $\mathbb{P}$, and $\mathbb{D}$ into account, and only focus on the highest $\mathbb{D}$. Consequently, their composed test sheets may contain the *miss-* and *error-included concept* nodes and cannot meet the adaptive requirements. Moreover, they also need to spend much more computation
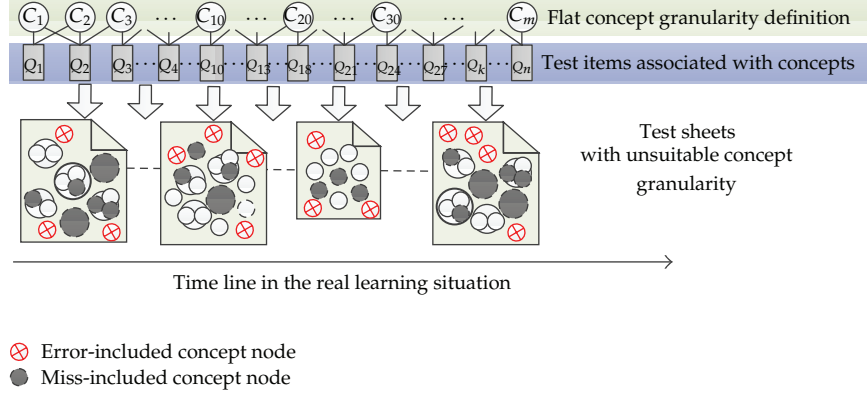
Figure 2: Issues for existing test sheet generation mechanisms.

time to select candidate test items in the item bank because they have no item selection strategy to filter the irrelevant ones in advance. Besides, item exposure rate, which denotes the number of a test item used in the test sheets, also needs to consider for enhancing the Item Bank Security.

Therefore, three issues are required to be solved for satisfying the adaptive requirements of a test sheet:

(i) how to generate a test sheet to precisely meet the adaptive requirements in terms of conceptual granularities, discrimination, difficulty, and expected test time parameters;

(ii) how to speed up the test sheet generation process for reducing the computation time;

(iii) how to consider the item exposure rate issue to enhance the Item Bank Security.

An Adaptive Test Sheet Generation Problem Is Defined as Follows

Assume that a set of test items should be selected from n items in the item bank $Q = \{Q_1, Q_2, \ldots, Q_n\}$. All items should be related to the concepts in a concept hierarchy H, a tree of concepts as shown in Figure 1. The tree H contains m concepts as the tree nodes C, namely, $C = \{C_1, C_2, \ldots, C_m\}$. δ is a descendent function, where $\delta(C_i) \subset C$ is a set of descendent nodes of C_i, and $\delta'(C_i) \subset C$ is a descendent leaf function, where C_j belongs to $\delta'(C_i)$ if and only if C_j is a leaf concept of H and the descendent of C_i.

$$C_j \in \delta'(C_i) \quad \text{iff} \ \left(C_j \in \delta(C_i)\right) \wedge \left(\delta(C_j) = \{\emptyset\}\right). \tag{3.1}$$

Based on the Q_i definition in Section 2, the item exposure times e_i and the degree of difficulty p_i are taken in account in this study. Thus, each test item Q_i is defined as follows.

$$Q_i = \left(p_i, t_i, d_i, r_{ij}, e_i\right), \quad \text{where } 0 \le d_i, p_i, r_i \le 1, \quad t_i, \ e_i \in \mathbb{N} \text{ (Natural Number)}. \tag{3.2}$$

An example is provided in Figure 3, where the concept hierarchy H is a tree of concept C_j and the test item set Q is a set of test items Q_i. A weight r_{ij} denotes relevance degree between
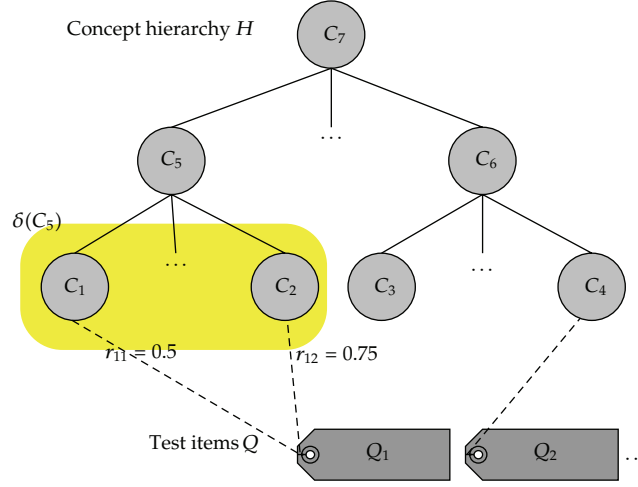
Figure 3: Concept hierarchy H and its related test items Q.

the concept C_j test item Q_i, for example, the relevance of C_2 and Q_1 is $r_{12} = 0.75$. The $\delta(C_j)$ denotes the subtree of the concept C_j, for example, C_1 and C_2 belong to the $\delta(C_5)$.

Therefore, in this study, a test sheet (TS) can be defined as follows:

$$TS = (Qs, t', p', C', r'),\tag{3.3}$$

where TS includes the expected test time t' of the test sheet, target difficulty degree p', target concepts $C' \subset C$, and the lower bound of average concept relevance r'. Based on the definitions of existing studies mentioned in Section 2, a decision variable $X = [x_1, x_2, \ldots, x_n]$ is defined where x_i is 1 if the test item Q_i is selected to the test sheet; 0, otherwise.

The goal of the adaptive test sheet generation problem is to generate a test sheet to

(i) approach all the target parameters p' and t',

(ii) have the highest average discrimination degree,

(iii) have the balanced concept relevance weight sum of each required conceptual granularity and its descents among the required concept range C' and the average relevance to be higher than r',

(iv) have the lowest average item exposure rate.

This is a multiobjective optimization problem, and the objective functions are defined as follows.

The objective function of the discrimination degree is inversed to the average discrimination degree of the test sheet:

$$D(X) = 1 - \left(\frac{\sum_{i=1}^{n} d_i x_i}{\sum_{i=1}^{n} x_i} \right).\tag{3.4}$$

The objective function of the expected test time is the distance between the sum of expected test time and the target expected time:

$$T(X) = \left| \left(\sum_{i=1}^{n} t_i x_i \right) - t' \right|.$$

(3.5)

The objective function of the difficulty degree is the distance between the average difficulty degree and the target difficulty degree:

$$P(X) = \left| \frac{\sum_{i=1}^{n} p_i x_i}{\sum_{i=1}^{n} x_i} - p' \right|.$$

(3.6)

Let $\bar{r}(X)$ be the average sum of relevance degree of each concept in the test sheet:

$$\bar{r}(X) = \frac{\sum_{j=1}^{m} \sum_{i=1}^{n} r_{ij} x_i}{|C'|}.$$

(3.7)

Let the generalized concept relevance $\mathfrak{R}_{ij}$ denote the maximum concept relevance of a test item toward the concept C_j or its descendent concepts:

$$\mathfrak{R}_{ij} = \text{Max}(r_{ik}) \mid C_k \in \delta(C_j).$$

(3.8)

The objective function of concept relevance is the distance between the sum of generalized concept relevance degrees and the average sum $\bar{r}(X)$. This objective function shows the imbalance degree of the concept relevance:

$$R(X) = \sum_{j=1}^{m} \left| \sum_{i=1}^{n} \mathfrak{R}_{ij} x_i - \bar{r}(X) \right|.$$

(3.9)

The objective function of the item exposure rate is the average exposure times:

$$E(X) = \frac{\sum_{i=1}^{n} e_i x_i}{\sum_{i=1}^{n} x_i}.$$

(3.10)

The multiobjective optimization problem is to find a test sheet X to minimize all the values of objective functions and subject to the lower bound of average concept relevance r', as shown in the following:

$$\min_{X} \quad [D(X), T(X), P(X), R(X), E(X)]^T$$

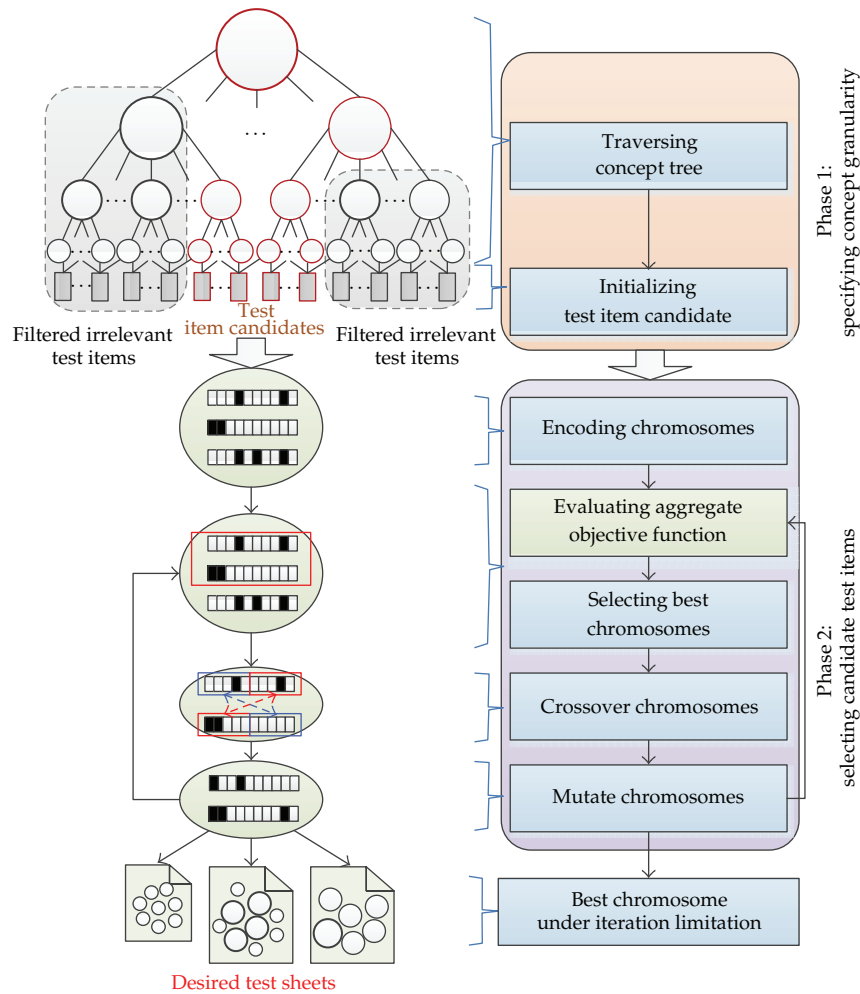Subject to $\quad \bar{r}(X) \geq r'.$

(3.11)

Figure 4: The flowchart of the CISS process.

4. Methodology

To solve the Adaptive Test Sheet Generation Problem, an Adaptive Test Sheet Generation (ATSG) mechanism has been proposed. ASTG mechanism consists of a Candidate Item Selection Strategy (CISS) to adaptively determine candidate test items set and the conceptual granularities according to the desired concept scope, and an Aggregate Objective Function (AOF) to apply Genetic Algorithm (GA) to figure out the approximate solution of mixed integer programming problem for the test-sheet composition. CISS process is illustrated in Figure 4.

4.1. Candidate Item Selection Strategy (CISS)

CISS process includes two phases: (1) specifying Concept Granularity and (2) selecting Candidate Test Item Set.

Phase 1: Specifying Concept Granularity

Concepts associated with a test sheet might be in various granularities for specific educational situations, so the conceptual granularities should be determined before generating a test sheet. Because the required concepts $C_i \in C'$ might be in various granularities, the most specific required concepts should be selected as the target concept set to precisely express the requirements. Let $\overline{C}'$ denote the target concept set, where no concepts in the set are the other concepts' ancestors, and the goal of the first phase is determining the concepts in $\overline{C}'$:

Phase 2: Selecting Candidate Test Item Set

Let θ be the candidate test item set, where the inner test items should be related to the target concept set. In Phase 2, test items whose related concepts are out of $\overline{C}'$ are filtered:

$$C_i \in \overline{C}' \quad \text{iff } (C_i \in C') \wedge (\neg \exists j, C_i \in \delta(C_j)). \tag{4.1}$$

Besides, the generalized concept relevance degrees $\Re$ of all test items toward all concepts in $\overline{C}'$ are calculated.

$$Q_i \in \theta \quad \text{iff } \exists j, \left(\left(C_j \in \overline{C}' \right) \wedge (\Re_{ij} > 0) \right) \wedge \neg \exists k, \left(\left(C_k \notin \overline{C}' \right) \wedge (\Re_{ik} > 0) \right). \tag{4.2}$$

After this phase, the search space can be reduced from Q to θ.

An example of CISS process is provided in Figure 5, where assume the required concepts set $C' = \{C_4, C_5, C_9, C_{10}\}$. In Phase 1, C_4, C_9, and C_{10} are selected into $\overline{C}'$ for expressing the most specific required concepts. In Phase 2, only the test items which are associated with the subtrees of concepts in $\overline{C}'$ can be selected to the candidate item set θ, so Q_3 and Q_4 are filtered before solving the optimization problem.

4.2. Aggregate Objective Function

An aggregate objective function $F(X)$ is defined to solve the multiobjective optimization problem:

$$F(X) = S_D + (1 - P_t) + (1 - P_p) + (1 - P_r) + (1 - P_e) + (1 - P_r'). \tag{4.3}$$

The aggregate objective function includes the discrimination score S_D and the penalty scores of the expected time P_t, the difficulty degree P_p, the concept relevance P_r, the concept relevance lower bound P_r', and the exposure times P_e. All score and penalty score are normalized to the range from 0 to 1.

The discrimination score S_D is inversed to the objective function $D(X)$:

$$S_D = 1 - D(X). \tag{4.4}$$
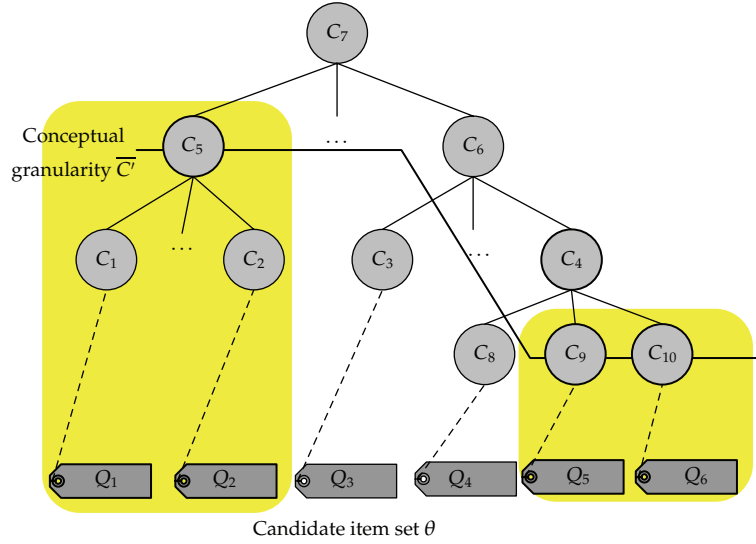
Figure 5: An example of the candidate item selection strategy (CISS) process.

The penalty score of the expected time is the percentage of the distance between the sum of expected test time and the target expected time over the target expected time. If the penalty score is greater than 1, 1 is assigned the penalty score:

$$P_t = \min\left(\frac{T(X)}{t'}, 1\right). \tag{4.5}$$

The penalty score of the difficulty degree is the value generated by the objective function of the difficulty degree:

$$P_p = P(X). \tag{4.6}$$

The penalty score of the concept relevance balance degree is the average distance between the sum of relevance degrees and the average sum of a concept:

$$P_r = \frac{R(X)}{|C'|}. \tag{4.7}$$

The penalty score of the concept relevance lower bound is greater than 0 if the average concept relevance is lower than the concept relevance lower bound and the value the percentage of the distance over the concept relevance lower bound. If the penalty score is greater than 1, the penalty score will be set as 1:

$$P_r' = \min\left(\frac{|\mathrm{Max}(r'(X) - \bar{r}, 0)|}{r'}, 1\right). \tag{4.8}$$

The penalty score of the exposure times is the percentage of the average of exposure times over the exposure times parameter e', which denotes the maximum exposure times to be considered. If the average of the exposure times is greater than e', the penalty score will be set as 1:

$$P_e = \min\left(\frac{E(X)}{e'}, 1\right). \tag{4.9}$$

Thus, a single aggregate objective function $F(X)$ can be defined to integrate all the score and penalty scores to a single objective score as (5.1).

The genetic algorithm (GA) can be applied to solve the Adaptive Test Sheet Generation Problem by maximizing the aggregate objective function $F(X)$. The overall process of the GA algorithm is shown in Figure 4. The CISS process can adaptively determine the desired concept scopes and granularities, and the out-of-scope test items, that is, *error-included concept nodes* in Figure 2, can be adaptively filtered to reduce the problem space of the test sheet generation. The candidate test items can be encoded into chromosomes, which is an N-bit binary string $[x_1, x_2, \ldots, x_N]$, where N is the amount of candidate test items and $x_i = 1$ denotes the test item i selected into the test sheet. In the beginning, a set of chromosomes, each whose bit value is randomly set, are generated as the initial selection states. Then, each chromosome is evaluated by the aggregate objective function $F(X)$. The higher score the chromosome gets, the more probability the chromosome can be reserved to generate the next generation. In the Crossover step, the chromosomes with higher score of $F(X)$ are selected to generate new chromosomes. Two chromosomes are both broken into two segments in the randomly selected segment lengths and the new chromosomes are generated by exchanging a segment with each other. Further, in the Mutation step, a random bit of a random chromosome in the new generation is inversed in order to prevent falling into the local optimal solutions. Then, return to the Crossover step to further generate next generation until the iteration limitation is achieved. Finally, the chromosome having the highest score of $F(X)$ among the whole process is the approximate solution.

5. Experiment and Evaluation

In order to evaluate the effectiveness of the proposed methodology in support of various purposes of assessments during the real learning situation, three experiments have been conducted. Firstly, various sizes of item banks are used to evaluate the efficiency and fitness scores of the proposed ATSG mechanism. Secondly, various levels of target concepts C' are used to evaluate the performance and the satisfaction degree of concepts in ATSG mechanism. Thirdly, exposure times of selected test items are measured during the 50 times of use. The exposure times of test items are accumulated and the experiment can evaluate whether ATSG mechanism can prevent the generation of the test sheets with high exposure times. In the three experiments, a system of the control group has also been developed based on Hwang's methodology [4], where the objective function shown in (5.1) was modified to meet the experimental requirements:

$$F(X) = S_D + (1 - P_t) + (1 - P_p) + (1 - P_r) + (1 - P'_r). \tag{5.1}$$
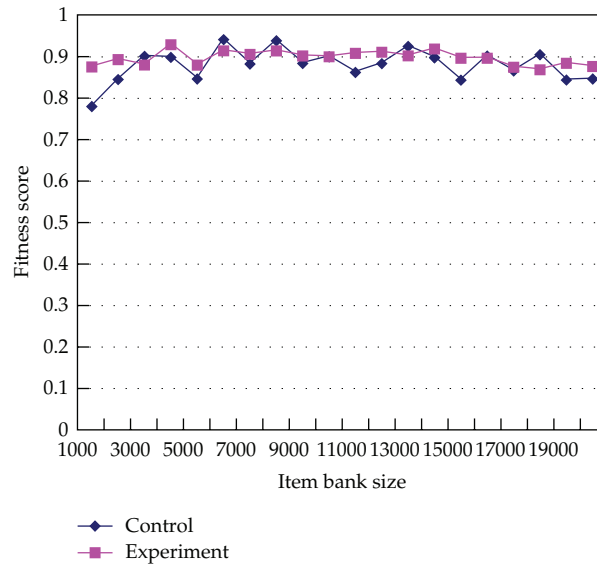
Figure 6: Fitness scores in various sizes of item banks.

Some differences in the system of control group are listed as follows:

(1) It does not run the CISS; all test items are considered in the GA algorithm.

(2) It does not consider the exposure times of test items.

(3) It does not calculate the generalized concept relevance, so the required concepts for control group are expended to all their descendent concepts.

The parameters of the GA algorithms used by the experimental and control systems were determined to balance the effectiveness and efficiency. In the three experiments, the GA algorithms were limited to 1,000 iterations and the mutation rate was 0.1. The population size was 30 and all initial bits of chromosomes were assigned to 0 because the amount of all test items was much larger than the amount of the selected test items.

5.1. Various Size of the Item Bank

The item banks having 1,000 to 20,000 test items are used to evaluate the systems' efficiency and effectiveness. In each item bank, 10 test sheets with randomly chosen parameters are generated by the control and experimental systems. The effectiveness is measured by the fitness score of the aggregate objective function $F(X)$. The result of effectiveness is shown in Figure 6, where the experimental system has more stable and generally higher fitness scores than those of the control system.

The experimental result of efficiency is shown in Figure 7, where the response time of the GA algorithm becomes higher if the size of item bank grows gradually. The reason is that if there are more candidate test items, much longer chromosomes will be used and the computing time dealing with all bits in chromosomes becomes much longer as well. Among the two systems, experimental system, which applies CISS process to dramatically reduce the size of candidate test items, can have much more efficient response time.
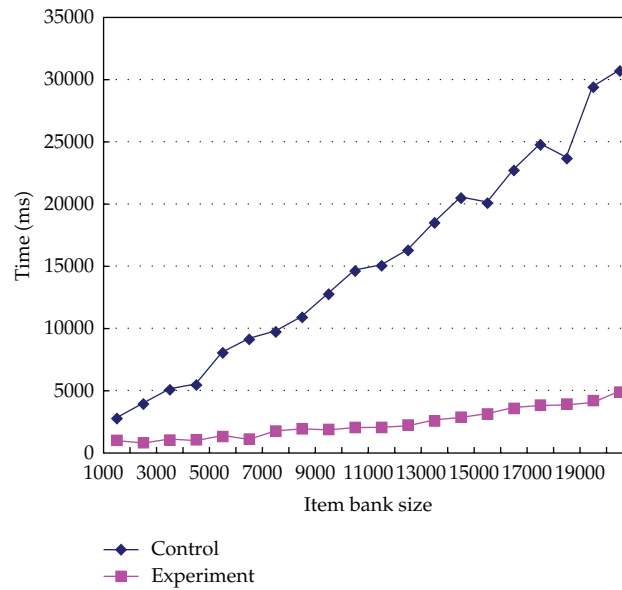
Figure 7: Response time in various sizes of item banks.

5.2. Various Levels of Target Concepts

This experiment demonstrates the systems' effectiveness of generating a test sheet for specific level of target concepts. Target concepts in the most coarse-grained level, level 1, to the most fine-grained level, level 6, are randomly chosen for the two systems. As shown in Figure 8, the concept relevance scores of the control system are much lower than those of the experimental system, especially when the concept level is fine grained. The reason is that without filtering out-of-scope test items, the GA algorithm of the control system is difficult to precisely choose the test items with accurate concepts. Figure 9 also shows that the test sheet generated by the control system contains many out-of-scope test items, which will seriously affect the test quality.

The result of response time in Figure 10 also reveals that the control system needs more computation times to generate a test sheet because many out-of-scope test items are also computed.

5.3. Exposure Times Measurement of Test Items

In the last experiment, 50 test sheets with similar target concept ranges are generated from the item bank containing 2,000 test items and the used test items are recorded to calculate the exposure times of each test item. Results of the average exposure times of test items are shown in Figure 11, where the control system and the experimental system have no noticeable difference. According to the analysis of each test sheet, although the experimental system can prevent the test items with high exposure times, the average exposure times are still accumulated due to the small range of target concepts. However, the out-of-scope test items are usually used in the test sheet generated by the control system, so the exposure times of a
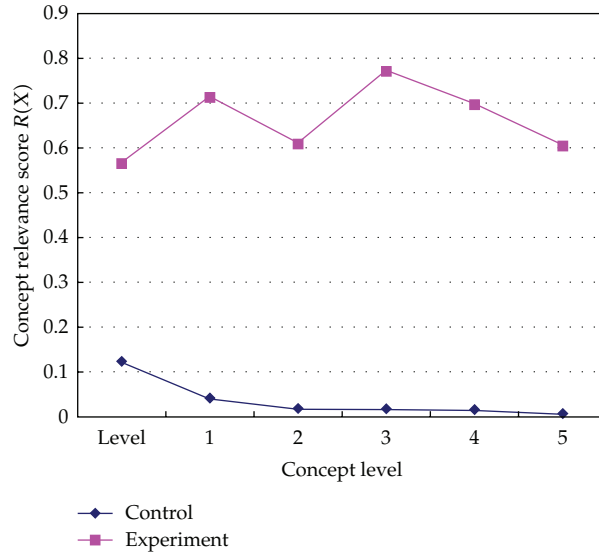
Figure 8: Concept relevance scores for various level of target concepts.
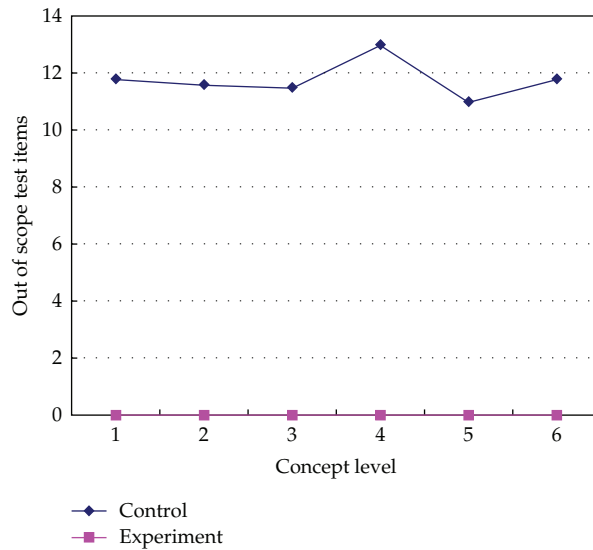

Figure 9: Amount of out-of-scope test items for various levels of target concepts.

single test item are accumulated slowly. That makes the exposure times of the experimental system are not better than those of the control system.

6. Discussion

The proposed ATSG mechanism is able to solve Adaptive Test Sheet Generation Problem in terms of the following aspects.
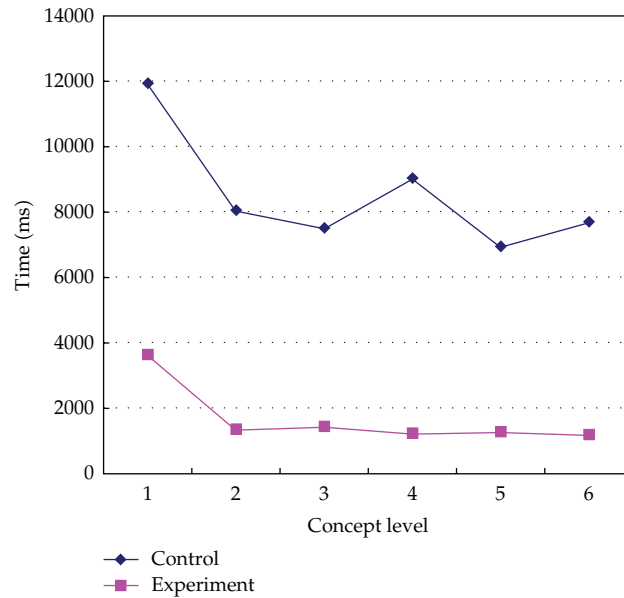
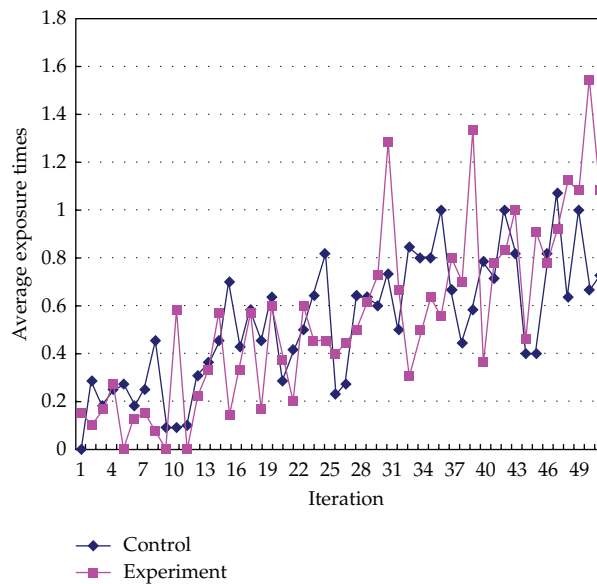**Figure 10:** Response time for various level of target concepts.

Figure 11: Average exposure times during 50 times of usage.

6.1. *The Control of the Concept Granularity of the Test Sheets and the Prevention of the Irrelevant Problem Space*

To simplify the discussion of this problem, assume that the concept tree is an L-level balanced tree, and the amount of branches in each level is B. Let an adaptive requirement of the test sheet contain n target concepts in level X. By applying the CISS mechanism, the problem

space of the test sheet generation problem can be reduced to n/B^{X-1} of the original problem space.

Proof. Assume that m items are related to a concept. The amount of candidate test items in the previous research is mB^{L-1}. By using the candidate item selection strategy, the amount of the candidate test items $\overline{C}'$ is mnB^{L-X}. Thus, the percentage of the new problem space over the previous problem space is $mnB^{L-X}/mB^{L-1} = n/B^{X-1}$.   □

6.2. The Generation of a Test Sheet to Precisely Fit the Target Concept Range, Difficulty, and Expected Test Time

In the new objective functions, the distances toward the target thresholds are used instead of the lower bound and upper bound in the previous studies. Thus, the difficulty and expected test item can be precisely fitted. Moreover, the candidate item selection strategy and the penalty score of the concept relevance balance degree can ensure that the test sheet contains balanced target concepts. As shown in Section 5.2, the concept relevance scores of the test sheets generated by the experimental system are also much higher than those of the control system.

6.3. The Consideration of the Item Exposure Rate

The penalty score of the exposure times P_e can prevent the high-exposure-rate items selected to the test sheet.

6.4. The Extensibility of the ATSG Mechanism

Most approaches mentioned in the related work section applied more efficient evolutionary algorithms, for example, the greedy algorithm approach [12], the tabu search algorithm [13], and the discrete particle swarm optimization algorithm [14] to enhance the computation efficiency of test sheet generation. However, these approaches did not yet take the conceptual granularity, exposure rates, and test item filtering into account. Therefore, these enhanced evolutionary approaches can thus be expected to replace the Hwang's methodology [4] for improving the efficiency of the Selecting Candidate Test Item Set phase (Figure 4) in the CISS process of ATSG mechanism.

6.5. The Future Work of the ATSG Mechanism

According to our observation and finding of experimental results, the degree of fitness score changes with the item bank sizes and the computation time (see Figure 6). Because the fitness scores directly affect the quality of the generated test sheet, a new important issue will be how to analyze the characteristics and predict the trends of fitness scores over times and item bank sizes for improving the quality of test sheet composition. However, this kind of time series problem may not be modeled by the conventional distribution model because the quality of the GA selection strategy seems to have the characteristics of self-similarity. Therefore, according to the study of Li [17], Fractal Time Series, which has the features of Long-Range Dependence (LRD) and obeys the Power Law, are a suitable mathematical approach to model

and analyze the features and phenomenon of self-similar series [18], for example, the data series in the cyber-physical networking systems [19], the time series of sea level [20] and molecular motion on the cell membrane [21], the DNA series [22], and the fractal lattice geometry using Iterated Function System (IFS) on simplexes [23]. Accordingly, in the near future, we are going to try to apply the fractal time series approach to analyze and model the series of fitness score for figuring out the characteristics of self-similarity.

7. Conclusion

In this paper, an Adaptive Test Sheet Generation (ATSG) mechanism is proposed, where the Candidate Item Selection Strategy (CISS) is come up to reduce the problem space of test sheet composition and an Aggregate Objective Function (AOF) based on the Genetic Algorithm (GA) is modeled to figure out the approximate solution. In this approach, the adaptive conceptual scope and granularity and item exposure rates have been considered to meet the various purposes of assessments during the real learning situation. Experimental results show that ATSG mechanism is able to more efficiently, precisely, adaptively generate the various test sheets than the existing approaches in terms of various conceptual scopes, computation time, and item exposure rates. Furthermore, in the near future, the fractal time series approach can be expected to be applied to analyze and model the series of GA's fitness score for figuring out the characteristics of self-similarity and improving the quality of test sheet composition according to the experimental finding.

Acknowledgment

References

[1] G. J. Hwang, "A test-sheet generating algorithm for multiple assessment requirements," *IEEE Transactions on Education*, vol. 46, no. 3, pp. 329–337, 2003.

[2] C. Y. Su and T. I. Wang, "Construction and analysis of educational assessments using knowledge maps with weight appraisal of concepts," *Computers & Education*, vol. 55, pp. 1300–1311, 2010.

[3] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, The MIT Press, Cambridge, Mass, USA, 3rd edition, 2009.

[4] G. J. Hwang, B. M. T. Lin, H. H. Tseng, and T. L. Lin, "On the development of a computer-assisted testing system with genetic test sheet-generating approach," *IEEE Transactions on Systems, Man, and Cybernetics C*, vol. 35, no. 4, pp. 590–594, 2005.

[5] W. Banzhaf, P. Nordin, R. E. Keller, and F. D. Francone, *Genetic Programming: An Introduction*, Morgan Kaufmann, San Francisco, Calif, USA, 1997.

[6] D. Fisher and N. Frey, *Checking for Understanding: Formative Assessment Techniques for Your Classroom*, Association for Supervision & Curriculum Deve, Alexandria, Va, USA, 2007.

[7] J. Leighton and M. Gierl, *Cognitive Diagnostic Assessment for Education: Theory and Applications*, Cambridge University Press, Cambridge, UK, 2007.

[8] K. Burke, *Balanced Assessment: From Formative to Summative*, Solution Tree, Bloomington, Ind, USA, 2010.

[9] C. Tovani, *So What Do They Really Know?: Assessment That Informs Teaching and Learning*, Stenhouse, Portland, Me, USA, 2011.

[10] S. W. Chang and B. Y. Twu, "A comparative study of item exposure control methods in computerized adaptive testing," Tech. Rep. 99-3, ACT Research Report Series, 1998.

[11] S. Y. Chen, R. D. Ankenmann, and J. A. Spray, "Exploring the relationship between item exposure rate and test overlap rate in computerized adaptive testing," Tech. Rep. 99-5, ACT Research Report Series, 1999.

[12] G. J. Hwang, B. M. T. Lin, and T. L. Lin, "An effective approach for test-sheet composition with large-scale item banks," *Computers & Education*, vol. 46, pp. 122–139, 2006.

[13] G. J. Hwang, P. Y. Yin, and S. H. Yeh, "A tabu search approach to generating test sheets for multiple assessment criteria," *IEEE Transactions on Education*, vol. 49, no. 1, pp. 88–97, 2006.

[14] K. H. Tsai, T. I. Wang, T. C. Hsieh, T. K. Chiu, and M. C. Lee, "Dynamic computerized testlet-based test generation system by discrete PSO with partial course ontology," *Expert Systems with Applications*, vol. 37, pp. 774–786, 2010.

[15] G. J. Hwang, H. C. Chu, P. Y. Yin, and J. Y. Lin, "An innovative parallel test sheet composition approach to meet multiple assessment criteria for national tests," *Computers $#38; Education*, vol. 51, pp. 1058–1072, 2008.

[16] T. F. Ho, P. Y. Yin, G. J. Hwang, S. J. Shyu, and Y. N. Yean, "Multi-objective parallel test-sheet composition using enhanced particle swarm optimization," *Educational Technology & Society*, vol. 12, no. 4, pp. 193–206, 2009.

[17] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.

[18] H. Sheng, Y. Q. Chen, and T. S. Qiu, *Fractional Processes and Fractional-Order Signal Processing*, Springer, New York, NY, USA, 2011.

[19] M. Li and W. Zhao, "Visiting power laws in cyber-physical networking systems," *Mathematical Problems in Engineering*, vol. 2012, Article ID 302786, 13 pages, 2012.

[20] M. Li, C. Cattani, and S. Y. Chen, "Viewing sea level by a one-dimensional random function with long memory," *Mathematical Problems in Engineering*, vol. 2011, Article ID 654284, 13 pages, 2011.

[21] H. Sheng, Y. Q. Chen, and T. S. Qiu, "Heavy-tailed distribution and local long memory in time series of molecular motion on the cell membrane," *Fluctuation and Noise Letters*, vol. 10, no. 1, pp. 93–119, 2011.

[22] C. Cattani, G. Pierro, and G. Altieri, "Entropy and multifractality for the myeloma multiple TET 2 gene," *Mathematical Problems in Engineering*, vol. 2012, Article ID 193761, 14 pages, 2012.

[23] C. Cattani, E. Laserra, and I. Bochicchio, "Simplicial approach to fractal structures," *Mathematical Problems in Engineering*, vol. 2012, Article ID 958101, 21 pages, 2012.