Development of Automatic Decision System in the Next Item Selection for Computerized Adaptive Testing

Ratchakrit Tanapattanadol*, Seree Chadcham*, and Piyathip Pradujprom*

Abstract

The research was invented in order to develop auto-decision system in the next item selection for computerized adaptive testing using Ant colony system with Triangle decision tree (TDT). The process is proposed by three main steps. The first step is dividing group of items from item bank and involved rules design before Ant colony system process. The second step is process design of Ant colony system for make decision to select the next item properly with ability level of the examinees. The third step is the format operation of Computerized adaptive testing. The efficiency testing of aforementioned steps measures up the examinee ability estimate and True ability, which operate under Monte Carlo Simulation via item bank simulation in accordance with Item Response Theory (IRT). IRT use Three-Parameter Logistic Model (3PL) to simulate the results from item and True ability of the examinees afterwards calculate Root Mean Square Error (RMSE) and Average Bias compare with the next item selection procedure using Maximum Information Criterion (MIC). The results showed that the developed system is more efficient than the next item selection procedure using MIC, the developed system has RMSE = 0.112 and Average Bias = -0.036.

Keywords: The Next Item Selection, Computerized Adaptive testing, Ant Colony System.

1. Introduction

The item selection regard as the heart of the matter in Computerized adaptive testing [1]. The aim of item selection is the most accurate in examinee ability estimate and security of the item. If choosing the inefficient item selection, will cause to get the unsuitable item with ability level of the examinees, high deviation of ability of the examinee so it is used more the number of items and test time [2]. Besides it causes too much repetition in item selection and the item in item bank is unselected in testing. This occurrence is naturally obstacle to examinee ability estimate to match the reality so that the efficiency of the item selection procedure causes in direct variation with the efficiency of Computerized adaptive testing. Because the aim of aforementioned testing processes using computer in suitable item selection with each examinee ability using Item Response Theory (IRT) to basic calculation of various item information based on the principle that the more correct answer, the more difficult next item in the other hand the more incorrect, the easier next item.

According to mentioned above, the importance of the next item selection procedure of Computerized Adaptive Testing, so the system of Computerized Adaptive Testing has the standard next item selection procedure that is Maximum Information Criterion (MIC). MIC is the next item selection bringing guessing test factor to consider with ability of the examinees (θ) at that time. Then select the maximum item information that is close to ability of the examinees to the next item presented by Bimbaum [3]

The next item selection problem of Computerized Adaptive Testing is the problem in searching the group of items that is corresponding to ability or suitability of the examinees for using the exam. So Ant colony system which presented by Dorigo, et al. [4] is suitable to solving the next item selection problem of Computerized Adaptive Testing.

From the importance of the problem and the studying of

* *Department of Research and Statistics in Cognitive Science, College of Research Methodology and Cognitive Science, Burapha University.*

involving research, the researcher has idea to develop auto decision system in the next item selection for the new way of Computerized Adaptive Testing.

Hence this research proposes development of auto decision system in the next item selection for Computerized Adaptive Testing using Ant Colony System together with Triangle Decision Tree (TDT) structure causing more efficient in Computerized Adaptive Testing.

2. Theoretical Background and Related Researches

2.1 Computerized Adaptive Testing

Computerized Adaptive Testing includes knowledge in two parts, the education measurement theory and computer technology. The procedures of testing select the item from item bank suit to ability of the examinees. This procedure based on test answer preceding, cause each examinee get the different item and unequal item. The adaptive testing with ability of the examinees is the efficient testing because of save time, reducing the number of items by not reducing quality, but trending to increase the measurement quality both reliability and validity.

Computerized Adaptive Testing includes five steps, (1) create item bank, (2) first item selection, (3) next item selection which have to consider selecting the item in accordance with determined conditions by applying the previous answer, (4) calculate possible ability levels which is estimation after the examinees answered and this estimation that be approved namely Maximum Likelihood Estimation, Bayesian Estimation and Bayesian Updating Estimation, and (5) termination criterion, computerized adaptive testing will terminate when the past answer was in line with termination criterion [2].

2.2 Ant Colony Optimization

Ant Colony Optimization (ACO) received the basic idea from searching food source of ant colony in nature. All of ant colony cooperate to search food source and try to search the route from the nest to the nearest food source.

Searching the route, ants use the chemical that called Pheromone to find round-trip nest by spitting Pheromone to the past way for being data to others and the follower ants spit Pheromone over. The more Pheromone is spit on the route, the more ants pass and the route is shorter that the less Pheromone. Until finally the most chosen route is the shortest that makes Pheromone is the important data to search food source and go to the nest. Hence Ant colony optimization procedure is used in solving Combinatorial Optimization.

The efficiency improvement process of Ant colony optimization has 5 means include Ant system, Elitist system, Rank-base ant system, Max-min ant system and Ant colony system by process explanation of Traveling Salesman Problem (TSP) [5].

2.3 Related Research

The related research divide to 2 groups, the first group is Deterministic or Stochastic. The research in this group find relation of various variables that related with measurement and evaluation of item in equation form. Then the equation is used as decision of the next item selection by consideration from the result of the equation such as the next item selection using Kullback-Leibler Information (KL) [6], the next item selection using Efficiency Balanced Information (EBI) [7], the next item selection using Variable-length Item Selection [8] and the next item selection using Hurwicz Criterion and the item exposure control (HC-Ex) [9].

The second group is the next item selection in Artificial intelligence item classification. This means uses artificial intelligence technique in item classification to the most suitable various forms. For example, Decision tree, used for the way in the next item selection. The artificial intelligence mean that relate with development of the next item selection using Computerized Adaptive Testing such as Bayesian networks [10], Genetic Algorithms [1] and Genetic Fuzzy Expert System [11].

3. Research Methodology

Development of The automatic decision system in the next item selection for computerized adaptive testing using

Ant colony system that presented in this research, this process selected the item from item bank to order automatic form on Triangle Decision Tree (TDT) standard form. 1 chart TDT include 21 nodes and 6 level of TDT [1], then selected the most proper TDT for using computerized adaptive testing.

Hence the development procedure of The automatic decision system could select the item properly with ability of the examinees. The researcher designed development steps to 3 steps as follows.

3.1 Data Preparation and Involved Process Design

3.1.1 Group of Items Division from Item Bank

Group of items division from item bank was divided by item difficulty to 5 levels as follows (1) very easy, (2) easy, (3) medium, (4) hard and (5) very hard [12]. This group of items division, the researcher brought data division technique with K-Means Clustering which the used item was simulated parameter of item along item response theory [13]. The items were simulated from random number by WinGen3 program for simulating parameter of 500 items, each item included Discrimination Power Parameter: a is in range of 0.50 to 2.50 Difficulty Parameter: b is in range of -2.50 to 2.50 and Guess Parameter: c is not over 0.30.

Ant colony system selected the item from classified various difficulty in the item bank to arrange automatic form on TDT was shown in figure 1.

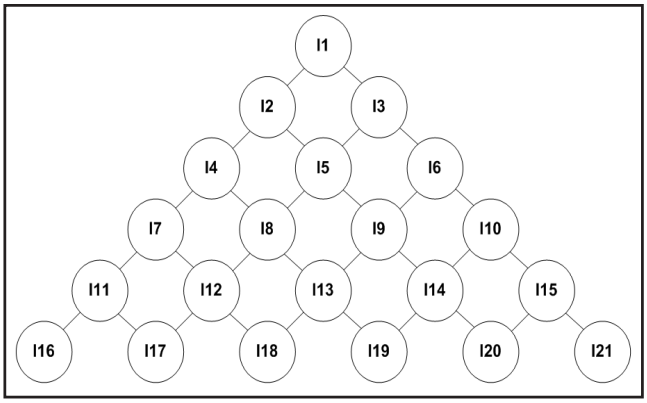

Figure 1. *MItem set arrangement on TDT stricture.*

From Figure 1 All items arranged on TDT. Each chart of standard form equal 21 items, the examinee can do the test limit at 6 items or require examinees to do the test by the maximum of items. So the number of TDT was arranged items by the following equation 1.

$$ST = \frac{L}{N} \quad (1)$$

Where ST is number of sets TDT, N is number of maximum items of each TDT and L is number of maximum items that used in all test.

3.1.2 Item Arrangement Rule on TDT

The item arrangement in the mentioned form had to arranged under 2 main rules, (1) In case of node were different level, the lower left node had to be less item difficulty parameter (b) in the other hand, the lower right node had to be more item difficulty parameter. (2) In case of node were same level, the right node is more difficulty than the left node. So ant colony system had to selected the item for arrangement on TDT according to the rule by following the equation 2-3.

$$\{b_{x+1,\,y}\} < \{b_{x,\,y}\} < \{b_{x+1,\,y+1}\} \quad (2)$$

$$\{b_{x,\,1}\} < \{b_{x,\,2}\} < \ldots < \{b_{x,\,y}\} \quad (3)$$

Where b is the item difficulty, x is the level of TDT and y is the order of arranged item in various levels on TDT from left to right.

3.2 Ant Colony System (ACS) Process Design

Finding answer process design of ACS in order to gaining the group of item on TDT that was proper with ability of the examinees. This process could propose detail of each step design as follows.

3.2.1 Suitability Measurement of Item Arrangement on Each Chart TDT of Each Ant

Suitability measurement of TDT was considered from 2 factors together.

1) Mean Square Error (MSE)

Suitability measurement process of TDT during ACS process by Local testing, was proceeded by simulating test results 6 items per TDT, each ant has already arranged item, the simulated test results used total 15 true ability of the examinees on various levels include high, medium and low ability on each 5 factors in test results simulation. After that

calculated MSE of true ability and estimated ability as the following equation 4 for using suitability decision criterion of each TDT.

$$MSE = \frac{1}{N}\sum_{m=1}^{n}\left(\hat{\theta}_m - \theta_m\right)^2 \quad (4)$$

Where n is number of all examinees, $\hat{\theta}_m$ is estimated ability of the m examinee, θ_m is true ability of the m examinee.

So MSE from mentioned equation was used in Pheromone calculation.

If any chart TDT arrange suitably items, will be the least MSE.

2) Total Distance of Item Parameter

Measured up distance of each item parameter on TDT by using equation 5.

$$f_d = \sqrt{w_a\left(a_x - a_y\right)^2 + w_b\left(b_x - b_y\right)^2 + w_c\left(c_x - c_y\right)^2} \quad (5)$$

Where f_d is the distance parameter of any couple items in level x, y on TDT, a b c is Discrimination Power Parameter, Difficulty parameter and Guess parameter respectively, w_a, w_b, w_c is the weighted of a, b, c which determined equal 0.25, 0.5 and 0.25 respectively and x, y is position of any item pairs on TDT structure that calculated the distance of parameter, x is the upper item and y is the lower item.

So f_d from mentioned equation was used in calculation of Heuristic Function.

Then calculating sum total of the distance of item parameter on TDT chart by the following equation 6.

$$F = \sum_{d=1}^{n} f_d \quad (6)$$

Where F is sum of distance parameter of couple item in level x, y and n is connection number of total couple item on TDT.

If any TDT chart arrange suitably, will be F at least because requiring to each item that arranged follow the rule was not too different on each parameter.

Hence suitability measurement of item arrangement on each chart TDT of each ant was calculated by the following equation 7.

$$Fitness = MSE + F_{normalize} \quad (7)$$

$F_{normalize}$ can be calculated by the following equation 8.

$$F_{normalize} = \frac{F - F_{min}}{F_{max} - F_{min}} \quad (8)$$

Where F_{min} is the minimum that is possible of f_d F_{max} is the maximum that is possible of f_d.

3.2.2 Route Design to Each Ant for Using The Route to Choose Item

The researcher design to 2 dimensions Array data structure form and n*n size. Where n is number of all item in each group which the group is divided difficulty level by K-Means Clustering, the X axis item means the next item and the Y axis item means the previous item.

3.2.3 The Next Item Selection of The Ant

The next item selection using ACS, the process select item by IRT theory which determined difficulty parameter (b) follow the rule (equation 2-3). While each ant is representatives of group items on TDT (Figure 1).

The next item selection (item number j) from the previous item (item number i) for placing determined various position on TDT was proceed by the following equation 9.

$$j = \begin{cases} \arg\max_{c_{il} \in N(s^p)} \left\{\tau_l^{\alpha} \cdot \eta_l^{\beta}\right\} & \text{if } q \le q_0, \\ J & \text{otherwise} \end{cases} \quad (9)$$

Where $N(s^p)$ is the set of item number i on to j which each ant still had not selected, c_{il} is item that each ant was selectable, l is item that each ant still had not selected, τ_{il} is Pheromone of item that ant was selectable, η_{il} is the distance of each item parameter on TDT which ant was selectable, α, β is coefficient of Pheromone and distance of item parameter respectively, determined to equal 1.

q_0 is constant ($0 \le q_0 \le 1$), q is random that range on [0, 1] (each route randomise new q every time).

If $q \le q_0$, the k ant will choose the j item that have the most Pheromone.

But if $q > q_0$, will choose the J item which J is the item order that obtained from random using probability of

selected each item was represented by p_{ij} which p_{ij} of each item can be calculated by the following equation 10.

$$p_{ij} = \frac{\tau_{ij}^{\alpha}.\eta_{ij}^{\beta}}{\sum_{c_{ij \in N(S^p)}} \tau_{il}^{\alpha}.\eta_{il}^{\beta}} \quad (10)$$

Where p_{ij} is probability of selected each item, $\tau_{ij}^{\alpha}.\eta_{ij}^{\beta}$ is Attractiveness of previous item (i) to next item (j). $\sum_{c_{ij \in N(S^p)}} \tau_{ij}^{\alpha}.\eta_{ij}^{\beta}$ is total of Attractiveness of previous item to every next item (l) that follow the rule in 3.1.2.

3.2.4 Local Pheromone Update

This process update all the route that each ant selected the next item selection (j) while at the present the ant is at the previous item (i) which amount of

Pheromone that connected between i and j (τ_{ij}) was updated by the following equation 11.

$$\tau_{ij} = (1-\varphi).\tau_{ij} + \varphi.\tau_0 \quad (11)$$

Where τ_0 is initial Pheromone, φ is coefficient of gradually disappeared Pheromone in local which $\varphi \in (0, 1]$

3.2.5 Global Pheromone Update

When each cycle work completed, the local route Pheromone of the ant that have the least Fitness ($Fitness_{best}$) was considered to adjust for Global Pheromone update by the following equation 12.

$$\tau_{ij} = \begin{cases} (1-\varphi).\tau_{ij} + p.\Delta\tau_{ij} & if(i,j)belong\ to\ best\ tour \\ \tau_{ij} & otherwise \end{cases} \quad (12)$$

Where $p \in (0, 1]$ is coefficient of gradually disappeared global Pheromone and $\Delta\tau_{ij} = \frac{1}{Fitness_{best}}$.

3.3 Process Format of Computerized Adaptive Testing

The procedure of computerized adaptive testing by the next item selection with Ant colony system, the researcher proposed algorithm as follows.

3.3.1 The item that past K-Means clustering brought to testing.

3.3.2 Group of items (medium level) were brought to ACS process for selecting items to arrange follow position on designed TDT.

3.3.3 The examinees did the next item follow ACS format.

3.3.4 Simulating Test Results

Because unreal examinees were in simulation, hence programming simulation had to simulate each test results that proceeded by random one number which had Uniform Distribution U (0, 1), then brought this number compared with probability that the examinees (which had ability, θ) could answer the i item correctly ($P_i(\theta)$) followed Three-Parameter Logistic Model (3PL) by the following equation 13.

The test results were determined from the conditions as follows [14]. In case of $P_i(\theta) \geq$ random number, determined test results was 1 (Correct) and in the other hand $P_i(\theta)$ < random number, determined test results was 0 (Incorrect).

$$P_i(\theta) = c_i + \frac{1-c_i}{1+e^{-Da_i(\theta-b_i)}} \quad (13)$$

Where $P_i(\theta)$ is probability that the examinees (which had ability) could answer the i item correctly, θ is ability level of the examinees, b_i is difficulty parameter of the i item, a_i is discrimination power parameter of the i item, c_i is guess parameter of the i item and D is constant, equal 1.7

3.3.5 Considering the results of each item together with estimating ability of the examinees using Bayesian updating [14].

3.3.6 Verifying Termination Criterion

Determination of testing was terminated when the examinees did the test 15 items altogether.

3.3.7 Verifying the item which the examinees did that was the last level on TDT or not? If this is not, back to process step 3 or If this is on TDT, follow step 7.

3.3.8 Determined to select the next group of item for ACS process

The principle of this process can explain using Figure 2 as the example.

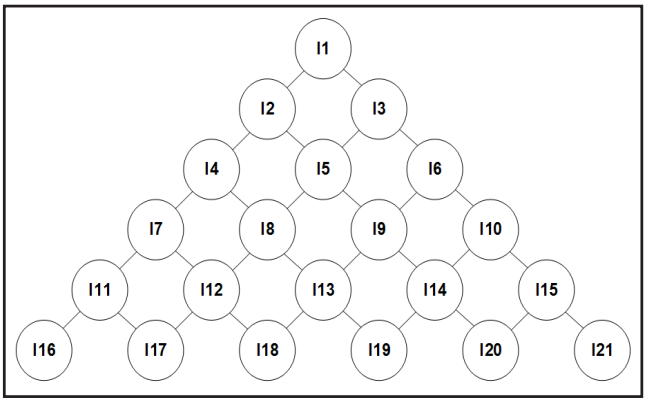

Figure 2. *The format of selected group items by ACS.*

From Figure 2, the conditions in selection of group items could be explained by dividing 3 cases as follows.

If the examinees do the test in the last level on TDT (The I16 or I17 item), the next group items are reduced difficulty to one level.

If the examinees do the test in the last level on TDT (The I18 or I19 item), the next group items are the same level.

If the examinees do the test in the last level on TDT (The I20 or I21 item), the next group items are increased difficulty to one level. When the processes have already done, repeat from step 3 again.

4. Results

Automatic decision development in the next item selection for computerized adaptive testing using ant colony system together with Triangle Decision Tree that designed above-mentioned, the researcher divided results to 3 parts as follows.

4.1 Parameter Simulation Results of Item and True Ability of The Examinees

4.1.1 Item bank was simulated from number random that have uniform distribution by WinGen3 for simulating item parameter (500 items). Each item includes 3 parameters (3PLM) such as (1) Discrimination Power Parameter: a (2) Difficulty Parameter: b and (3) Guess Parameter: c which can show simulation results by the following Table 1.

Table 1. *Basic statistic of item parameter in simulated item bank.*

Item Parameter	n	Min	Max	Mean	SD
a	500	0.502	2.497	1.479	0.575
b	500	-2.477	2.499	-0.101	1.455
c	500	0.001	0.300	0.149	0.086

4.1.2 The true ability of the examinees was simulated by determining random data from random number that have normal distribution N (0,1) 10,000 numbers, divided in 1000 parameters and all 10 sets. The results showed by the following Table 2.

Table 2. *Basic statistic of true ability of the examinees.*

θ	n	Min	Max	Mean	SD
Set 1	1,000	-3.224	4.255	.007	1.004
Set 2	1,000	-3.200	3.308	.035	1.004
Set 3	1,000	-3.301	3.463	.031	1.023
Set 4	1,000	-3.386	3.065	.010	1.003
Set 5	1,000	-2.784	3.080	.004	0.938
Set 6	1,000	-2.841	3.247	.034	1.000
Set 7	1,000	-3.598	3.418	-.002	1.004
Set 8	1,000	-3.249	3.098	.013	1.016
Set 9	1,000	-3.222	3.509	.004	1.019
Set 10	1,000	-3.530	3.617	.037	1.035

4.2 Item Classification Results by Difficulty

Aforementioned item classification (500 items) was divided 5 levels by difficulty the following Table 3.

Table 3. *Basic statistic of item classification from simulated item bank.*

Difficulty	Number of item	Min	Max	Mean	SD
Very easy	100	-2.477	-1.560	-2.064	0.280
Easy	123	-1.537	-0.522	-1.024	0.320
Medium	90	-0.493	0.453	0.009	0.284
Hard	87	0.516	1.470	0.991	0.261
Very hard	100	1.477	2.499	1.950	0.287

From Table 3 showed basic statistic of item classification form simulated item bank by RapidMiner Studio 8. The result found that the same class item was the same difficulty and the different class was also the different difficulty which this results conformed to K-Means Clustering.

4.3 Automatic Decision Development Results in The Next Item Selection Using Ant Colony System

In this proposed system, the researcher develop program using Java and run various conditions that determined on NetBeans IDE. This experiment used true ability of the examinees that simulated 10 different sets (Table 2), each set had 1,000 parameters of true ability including all used data was 10,000 parameters. The experiment was repeated 10 times (first to tenth times). In each time, used true ability of the examinees set by set (1,000 parameters) and every time of experiment used simulated item bank (same set and 500 items) that was classified (Table 3), which each result obtained estimated ability of the examinees and 1,000 parameters of tested items.

Besides this experiment adjusted various parameters involving with ACS such as ant population, cycle time in searching the answer in order to find the most suitable model for the next item selection problems then compared with Maximum Information Criterion (MIC)[3].

The results can conclude average of RMSE and Average bias when run the program 10 times of every ACS model by following Table 4.

Table 4. *Results of average when run the program 10 times of RMSE and Average bias from involved parameter adjustment.*

Ant Population	Iteration	Average of RMSE	Average of Average Bias
10	50	0.316	0.081
	100	0.218	-0.017
	200	0.166	-0.015
20	50	0.213	0.065
	100	0.128	-0.020
	200	0.132	-0.026
30	50	0.203	0.055
	100	0.118	-0.040
	200	0.112	-0.036

From Table 4 the results showed that ACS model that used 30 ants and 200 iterations was the least RMSE (0.112), Average bias closed to 0 most (-0.036) which had the estimated parameter ability of the examinees was lower than the true ability of the examinees slightly. Hence this ACS model is the most suitable.

4.4 Efficiency Comparison of The Next Item Selection

Efficiency comparison between the developed next item selection and MIC (the standard next item selection). The results showed average of RMSE and Average bias when run the program 10 times of every item selection mean, the comparison result was showed by the following Table 5.

Table 5. *Efficiency comparison result of the next item selection means.*

The item selection means	Average of RMSE	Average of Average Bias
MIC	0.894	0.063
ACS	0.112	-0.036

From table 5, the relationship between RMSE and Average Bias can explain. RMSE is accuracy of examinee ability estimate. Therefore, the small RMSE is the least different from true ability. While Average Bias is precision of examinee ability estimate, this is showed in average difference between ability estimate and true ability. Moreover, Average Bias can estimate the examinee ability direction that is higher or lower than true ability. If Average Bias close to zero, the efficiency of the next item selection is accurate.

Therefore, ACS is more efficient in ability of the examinees estimate than MIC because ACS is lower error.

5. Conclusion

Automatic decision development in the next item selection using Ant colony system together with TDT is the mean that designed system to think, analyze and decide for searching the most suitable answer automatically under determined conditions and scope of the possible answer extensively. Hence the most important thing is suitability measurement design of item selection of each ant in ACS process. The researcher experimented to use MSE of true ability and ability parameter getting from simulated test result estimation together with sum total distance of item's parameter on TDT. If any ant choose item to arrange on TDT then the mentioned

suitability parameter is the lowest, so this item (on TDT) will be the most suitable.

From all of the results, the automatic decision in the proposed next item selection is more overall efficient than the compared mean. This result revealed that the efficiency of this process design can select extensively all difficulty before termination. All items are arranged in position of TDT suitably. Besides the item selection is neither tufted on one difficulty nor one direction. Therefore, the examinees obtain the suitable item with ability and the efficiency of computerized adaptive testing can estimate ability of the examinees almost the same with true ability.

6. References

[1] M. Phankokkruad and K. Woraratpanya. "Item classification algorithm for computer adaptive testing based on web services." *International Journal of Engineering Research and Applications*, Vol. 2, No. 1, pp. 835-847, 2012.

[2] M. Reckase. *Multidimensional item response theory*. New York: Springer, 2009.

[3] A. Birnbaum. "Some latent trait models and their use in inferring an examinee's ability." In F.M. Lord and M.R. Novick (Eds.), *Statistical theories of mental test scores*, pp. 397-479, Reading (Mass.) Addison-Wesley, 1968.

[4] M. Dorigo and L. M. Gambardella. "Ant colony system: A cooperative learning approach to the traveling salesman problem." *Evolutionary Computation, IEEE Transactions*, Vol. 1, No. 1, pp. 53-66, 1997.

[5] M. Dorigo and T. Stützle. *Ant colony optimization*. London : The MIT Press., 2004.

[6] H. H. Chang and Z. Ying. "A-stratified multistage computerized adaptive testing." *Applied Psychological Measurement*, Vol. 23, No. 3, pp. 211-222, 2008.

[7] K.T. Han. "An efficiency balanced information criterion for item selection in computerized adaptive testing." *Journal of Educational Measurement*, Vol. 49, No. 3, pp. 225-246, 2012.

[8] A. Koedsri, N. Lawthong, and S. Ngudgratoke. "Efficiency of item selection method in variable-length computerized adaptive testing for the testlet response model: Constraint-weighted a-stratification method." *Procedia-Social and Behavioral Sciences*, Vol. 116, pp. 1890-1895, 2014.

[9] P. Premthongsuk, S. Chadcham, and P. Pradujprom. "Development of the next item selection procedure using hurwicz criterion and item exposure control for computerized adaptive testing, Veridian e-Journal." Vol. 4, No. 6, pp. 32-50. (in Thai), 2017.

[10] T. Y. Hsieh and B. C. Kuo. "Information-based Item Selection with blocking strategy based on a Bayesian network." *System*, Vol. 9, No. 7, pp. 1160-1169, 2010.

[11] E. Verdú, M. J. Verdú, L. M. Regueras, J. P. de Castro, and R. García. "A genetic fuzzy expert system for automatic question classification in a competitive learning environment." *Expert Systems with Applications*, Vol. 39, No. 8, pp. 7471-7478, 2012.

[12] A. Karahoca, D. Karahoca, and F. lnce. "ANFIS supported question classification in computer adaptive testing (CAT)." *In Soft Computing, Computing with Words and Perceptions in System Analysis*, Famagusta, 2009.

[13] R. J. de Ayala. *The theory and practice of item response theory*. London: The Guilford Press, 2009.

[14] N. A. Thompson and D. J. Weiss. "A Framework for the Development of Computerized Adaptive Tests." *Practical Assessment, Research and Evaluation*, Vol. 16, No. 1, pp. 1-9, 2011.