Rodríguez-Cuadrado, J., Delgado-Gómez, D., Laria, J. C. & Rodríguez-Cuadrado, S. (2020). Merged Tree-CAT: A fast method for building precise computerized adaptive tests based on decision trees. *Expert Systems with Applications, 143*, 113066.

DOI: 10.1016/j.eswa.2019.113066

Merged Tree-CAT: A fast method for building precise Computerized Adaptive Tests based on Decision Trees

Javier Rodríguez-Cuadrado [1], David Delgado-Gómez[1], Juan C. Laria[2], and Sara Rodríguez-Cuadrado[2]

[1] *Universidad Carlos III de Madrid, Department of Statistics,Leganés, Spain*
[2] *Universidad Autónoma de Madrid, Department of Developmental and Educational Psychology, Cantoblanco, Spain*

April 14, 2021

Corresponding author.

Abstract

Over the last few years, there has been an increasing interest in the creation of Computerized Adaptive Tests (CATs) based on Decision Trees (DTs). Among the available methods, the Tree-CAT method has been able to demonstrate a mathematical equivalence between both techniques. However, this method has the inconvenience of requiring a high performance cluster while taking a few days to perform its computations. This article presents the Merged Tree-CAT method, which extends the Tree-CAT technique, to create CATs based on DTs in just a few seconds in a personal computer. In order to do so, the Merged Tree-CAT method controls the growth of the tree by merging those branches in which both the distribution and the estimation of the latent level are similar. The performed experiments show that the proposed method obtains estimations of the latent level which are comparable to the obtained by the state-of-the-art techniques, while drastically reducing the computational time.

Keywords

Computerized Adaptive Tests, Decision Trees, Linear Programming

1 Introduction

Computerized Adaptive Tests (CATs) have been a significant advance in the field of psychometrics by being able to estimate the examinee's latent level θ (e.g., reading comprehension, IQ, and so forth) with greater precision than the classical linear tests using a smaller number of items (Weiss, 2004). This is due to the fact that the examinee receives a personalized test. Moreover, this customization of the test, together with the possibility of incorporating various mechanisms that control for the exposure rate of each item, hinders their leakage among the participants.

Concisely, CATs estimate the latent level of the examinee each time he/she responds to an item. This is conducted by means of a probabilistic model that uses the responses given to the previously administered items, and the characteristics of those. This estimation is then used to select the next item to be administered, among those that satisfy the established exposure condition, by using an optimization criterion. As an example, some of the proposed criteria are Minimum Expected Posterior Variance (MEPV) (Van der Linden and Pashley, 2009), Maximum Fisher Information (MFI) (Lord, 2012; Weiss, 1982), Kullback-Leibler Information (KLI) (Chang and Ying, 1996) and Maximum Likelihood Weighted Information (MLWI) (Veerkamp and Berger, 1997). Unfortunately, as stated in the literature (Ueno and Songmuang, 2010), the high computational time required by some of these criteria when selecting the next item implies an excessive waiting time for participants. This makes it difficult to use CATs that apply these criteria in practical settings.

Over the past few years, several articles have proposed decision trees (DTs) as an alternative to CATs. The fact that DTs build the test before its administration solves the aforementioned computational issue. Among those contributions, Ueno and Songmuang (2010) proposed a DT that obtains more precise scores than the estimates of the CATs using a smaller number of items per individual. Michel et al. (2018) used this proposal to estimate the quality of life of patients with multiple sclerosis, obtaining a DT with a smaller bias in the selection of items. Also on these lines, Yan et al. (2004) proposed a regression tree to estimate the test subjects' scores. The novelty of this work lies in the proposal to merge those nodes that meet a similarity criterion. This criterion, based on a t-statistic, enables to have a sufficient number of observations in each node. However, the main problem of the former proposals is that they estimate test dependent scores rather than latent levels. This fact prevents the comparison of estimates between different tests, which is one of the fundamental characteristics of CATs.

Recently, Delgado-Gómez et al. (2019) mathematically showed an equivalence between CATs and DTs by proposing the Tree-CAT method. This method integrates the advantages of both techniques: A precise estimation of the actual latent level of the examinee plus the construction of the test before its administration. In their proposal, each branch has an associated density function characterizing the distribution of the latent variable of the examinees progressing through that branch. This density function is used to assign an item to each node by sequentially solving linear optimization problems. The main disadvantage of this method is that the construction of the tree requires a high performance cluster due to its growth, taking about a week to create a test with only 10 items per examinee.

In the current article, we propose the Merged Tree-CAT method, which builds upon the previous method so that a tree of any depth can be created on a personal computer in a few minutes. To that end, the growth of the tree is controlled by merging branches with similar distributions and estimates of the associated latent level. This idea was already conceptualized by Yan et al. (2004), although in the present article the objective and merging criterion are different, as we are working with distributions instead of samples.

The rest of the article is structured as follows. Section 2 describes the proposed Merged Tree-CAT method. Section 3 shows the results obtained in two experiments aimed at evaluating the proposed method. In those experiments, the performance of the Merged Tree-CAT method is compared to the original Tree-CAT method and with respect other three CAT techniques widely used in the literature. Finally, Section 4 concludes the article with a discussion on the benefits of the proposed method.

2 Merged Tree-CAT method

Before starting to describe the Merged Tree-CAT method we introduce the notation that will be used.

M:	Depth of the tree (number of tree levels).
m:	Tree level.
Z_m:	Number of nodes in level m.
N:	Number of items in the item bank.
i:	Item from the item bank.
K_i:	Number of responses to item i
k_i:	Response given by the examinee to item $i(k_i = 1, \ldots, K_i)$.
r_i:	Maximum exposure rate of item i.
c_i^m:	Capacity of item i after the creation of the level $m-1$. For the first level, $c_i^1 = r_i, i = 1, \ldots, N$.
E_i:	Fitness index of item i.
θ:	Latent level of the examinee.
$f(\theta)$:	Prior density function of the latent level.
$\hat{\theta}_n^{k_i}$:	Estimation of the latent level θ given the response k_i in the node n.
$f_n^{k_i}(\theta)$:	Posterior density function of the latent level θ given the response k_i in the node n.
$P_i(\theta, k_i)$:	Probability that the examinee with a latent level θ gives the response k_i to item i.
$P_i(k_i)$:	Probability that the examinee gives the response k_i to item i.
α_i:	Probability that the first item administered to the examinee is the item i.
K^*:	Maximum number of branches per level.
δ:	Minimum similarity between distributions of two branches to merge.
L_1:	Lower limit of an interval containing the latent level θ with probability p.
L_2:	Upper limit of an interval containing the latent level θ with probability p.
$D_n^{k_i}$:	Probability that an examinee is in the node n and gives the response k_i to item i.
$A_n^{k_i}$:	Set of items answered by those examinees that in the node n gave the response k_i to item i.

2.1 Tree structure

The Merged Tree-CAT method generates a tree with as many M levels as the maximum number of items to be administered to each examinee. Each m level $(m = 1, \ldots, M)$ is composed of Z_m nodes, with each node having the structure shown in Figure 1.

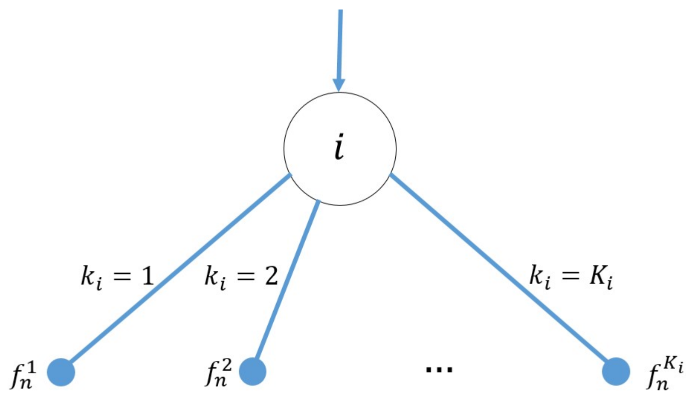


Figure 1: Node structure.

This figure shows how a node n of the tree is composed by an item i and a set of K_i branches corresponding to each of the possible k_i responses to the item ($k_i = 1, \ldots, K_i$). Each one of these branches has associated the posterior density function $f_n^{k_i}$ of the latent level from the set of participants that have reached this node and have chosen the response k_i. In addition, the node also implicitly contains the estimation of the latent level $\hat{\theta}_n^{k_i}$ of these participants, given by (Bock and Mislevy, 1982):

$$\hat{\theta}_n^{k_i} = \int_{-\infty}^{\infty} \theta f_n^{k_i}(\theta) d\theta. \tag{1}$$

The top line in Figure 1 represents the linkage of this node with another previous node in the tree. For the first level $m = 1$, this line joins the root of the tree which contains only the prior density function $f(\theta)$ of the population.

2.2 Building of the first level $(m = 1)$

The Merged Tree-CAT method initiates the building of the tree by determining the item with which each examinee begins the test. To do so, firstly, the E_i fitness index is calculated for each of the N items from the item bank. This index represents how far the item is from the optimum according to an established criterion. As an example from the criterion MEPV, which will be used later for experiments, E_i is given by (Delgado-Gómez et al., 2019):

$$E_i = \int_{-\infty}^{\infty} \left(\sum_{k_i=1}^{K_i} \left(\theta - \hat{\theta}^{k_i} \right)^2 P_i(\theta, k_i) \right) f(\theta) d\theta, \tag{2}$$

where $\hat{\theta}^{k_i}$ is the estimation of the latent level when the examinee gives the k-th response to item i.

Once the E_i fitness indexes, $i = 1, \ldots, N$ have been calculated for each item, the assignation of items is carried out by solving the following linear programming problem:

$$\min \sum_{i=1}^{N} \alpha_i E_i \tag{3}$$

s.t.

$$\sum_{i=1}^{N} \alpha_i = 1 \tag{4}$$

$$\alpha_i \leq r_i \qquad i = 1, \ldots, N \tag{5}$$

$$\alpha_i \geq 0 \qquad i = 1, \ldots, N \tag{6}$$

where α_i is the probability that the examinee will receive item i to start the test and r_i represents the maximum allowed exposure rate. Please note that when no exposure rate control is applied ($r_i = 1, i = 1, \ldots, N$), the Merged Tree-CAT method creates a single node containing the item with the highest fitness index. On the other hand, when $r_i < 1, i = 1, \ldots, N$, the method creates multiple nodes. Concisely, each of these nodes would contain one of the items for which $\alpha_{i^z} > 0$, where $\alpha_{i^z} = \min \{r_{i^z}, 1 - \sum_{s=1}^{z-1} \alpha_{i^s}\}$ and i^z is the z-th item with higher fitness index, $z = 1, \ldots, N$. The examinee would start the test randomly among these nodes according to these probabilities α_{i^z}. Therefore, in this case, it can be understood that the method simultaneously generates a forest of trees.

Once the items have been assigned to the nodes $n = 1, \ldots, Z_1$ of the first level, the posterior density functions $f_n^{k_i}$ of the latent level can be calculated, that according to Bayes' theorem, are given by:

$$f_n^{k_i}(\theta) = f(\theta|k_i) = \frac{P_i(\theta, k_i) f(\theta)}{\int_{-\infty}^{\infty} P_i(\theta, k_i) f(\theta) d\theta}. \tag{7}$$

Next, the Merged Tree-CAT method analyzes the possibility of merging branches between the different nodes of the first level according to a conditional criterion based on the similarity of estimates and distributions. This criterion seeks to limit the growth of the tree so that it is computationally tractable and the construction time of the tree is reduced.

For this purpose, the criterion determines whether the number of branches of the nodes at this level exceeds the maximum number K^* of branches per level. That is, if,

$$\sum_{n=1}^{Z_1} K_{i_n} > K^*, \tag{8}$$

where i_n is the item assigned to the node n. When this condition is met, two branches with associated estimations $\hat{\theta}_u^{k_s}$ and $\hat{\theta}_v^{k_t}$, are merged if

$$\left| \hat{\theta}_u^{k_s} - \hat{\theta}_v^{k_t} \right| < \frac{L_2 - L_1}{K^*}, \tag{9}$$

where L_1 and L_2 are the limits of the interval containing the prior latent level of the examinee with a probability p:

$$\int_{-\infty}^{L_1} f(\theta)d\theta = \int_{L_2}^{\infty} f(\theta)d\theta = \frac{1-p}{2}. \tag{10}$$

On the other hand, when condition (8) is not met, the criterion merges two branches if condition (9) and the following intersection condition (Cha, 2007) are met:

$$\int_{-\infty}^{\infty} \min\{f_u^{k_s}(\theta), f_v^{k_t}(\theta)\}d\theta > \delta. \tag{11}$$

That is, the similarity between the distributions given by the intersection must exceed a prefixed minimum similarity δ. In this way, the growth of the tree is constrained in a probabilistic manner given K^* and δ. On one hand, increasing the number of branches per level K^* results into a higher accuracy of estimates and a greater computational cost. On the other hand, a high value of δ allows that only those branches with very similar distributions are merged, resulting in a lower number of merged branches, and therefore, into higher accuracy of estimates and a greater computational cost. The sequential application of the criteria given by (9) and (11) is due to the fact that the former is less computationally expensive and reduces the number of evaluations in (11). Therefore, two branches are merged if the criteria given by (8) and (9) or (9) and (11) are met. The merged branch will come from two different nodes and the density function $f_{u,v}^{k_s,k_t}$ will be a mixture of the density functions of the two merged branches. Being $D_u^{k_s} = \alpha_s P_s(\theta, k_s)$ the probability that the examinee will take the branch associated with the response k_s to the item s assigned to the node u, the mixture density function is given by:

$$f_{u,v}^{k_s,k_t} = \frac{D_u^{k_s}}{D_u^{k_s} + D_v^{k_t}} f_u^{k_s}(\theta) + \frac{D_v^{k_t}}{D_u^{k_s} + D_v^{k_t}} f_v^{k_t}(\theta). \tag{12}$$

In this case, the probability $D_{u,v}^{k_s,k_t}$ of taking this merged branch is $D_{u,v}^{k_s,k_t} = D_u^{k_s} + D_v^{k_t}$. Moreover, being $A_u^{k_s}$ the set of items answered by the examinees who gave the response k_s in the node u, for the merged branch we have $A_{u,v}^{k_s,k_t} = A_u^{k_s} \cup A_v^{k_t}$.

After evaluating the merger of each possible pair of branches (including already merged branches) and carrying out the corresponding ones, the first level $m = 1$ is built. Finally, the capacity of each item i that will be employed in the construction of the level $m = 2$ is calculated. This item capacity, c_i^2, represents the updated exposure rate of item i: $c_i^2 = r_i - \alpha_i, i = 1, \ldots, N$

2.3 Building the m-th level

Assume that the test has been built up to the level $m - 1$. Ideally, the Merged Tree-CAT method will associate a single node at the end of each branch of the nodes belonging to level $m - 1$. However, due to the exposure control constraints, it may be the case that several nodes are associated at the end of the same branch.

To build level m, the Merged Tree-CAT method initially assumes that there is a single node n at the end of each branch, and firstly calculates the fitness index G_i^n of assigning the item i to that node:

$$G_i^n = \begin{cases} \int_{-\infty}^{\infty} \left(\sum_{k_i=1}^{K_i} \left(\theta - \hat{\theta}_n^{k_i} \right)^2 P_i(\theta, k_i) \right) f_u^{k_s}(\theta) d\theta, & \text{if } i \notin A_u^{k_s} \\ \infty, & \text{if } i \in A_u^{k_s} \end{cases}, \tag{13}$$

where $f_u^{k_s}$ and $A_u^{k_s}$ are the density function and the set of items previously used, which are linked to the branch ending in the node n. Please note that under the above definition it is impossible to administer the same item more than once to the same examinee.

Once that the fitness indexes G_i^n have been calculated for each item $i = 1, \ldots, N$ and node $n = 1, \ldots, Z_m$, the Merged Tree-CAT method calculates the probability α_i^n that an examinee arrives at node n and receives the item i, solving the following optimization problem:

$$\min \sum_{i=1}^{N} \sum_{n=1}^{Z_m} \alpha_i^n G_i^n \tag{14}$$

s.t.

$$\sum_{i=1}^{N} \alpha_i^n = D_u^{k_s} \qquad n = 1, \ldots, Z_m \tag{15}$$

$$\sum_{n=1}^{Z_m} \alpha_i^n \leq c_i^m \qquad i = 1, \ldots, N \tag{16}$$

$$\alpha_i^n \geq 0 \qquad i = 1, \ldots, N, \quad n = 1, \ldots, Z_m \tag{17}$$

where c_i^m is the updated capacity of the item i after the creation of the level $m - 1$.

For a node n, each item i that satisfies $\alpha_i^n > 0$ is assigned to that node. In the case of multiple items $i^1, \ldots, i^W$ such as $\alpha_{i^1}^n, \ldots, \alpha_{i^W}^n > 0$, several nodes are associated at the end of the corresponding branch. Concisely, each one of these nodes will have an item i^j associated that will be accessed with a probability $\alpha_{i^j}^n / \sum_{w=1}^{W} \alpha_{i^w}^n$.

As before, once the items have been assigned to the nodes $n = 1, \ldots, Z_m$ of level m, the posterior density functions $f_n^{k_i}$ of the latent level are calculated:

$$f_n^{k_i}(\theta) = f_u^{k_s}(\theta | k_i) = \frac{P_i(\theta, k_i) f_u^{k_s}(\theta)}{\int_{-\infty}^{\infty} P_i(\theta, k_i) f_u^{k_s}(\theta) d\theta}. \tag{18}$$

Following, the Merged Tree-CAT method evaluates the possible merger between each pair of branches of the nodes of the current level according to a combined criterion of similarity of distributions and estimates as seen in the previous section (equations (8)-(11)). Finally, the capacities c_i^{m+1}, the probabilities of accessing each branch and the sets of selected items are updated.

3 Experimental results

In this section, the results of three experiments comparing the performance of the proposed Merged Tree-CAT with respect to the original Tree-CAT method are presented[1]. The first two experiments recreate those conducted by Delgado-Gómez et al. (2019). In addition, these experiments include three other widely used techniques developed for building exposure-controlled CATs: The Restricted method which forbids the administration of items that have exceeded their exposure rate (Revuelta and Ponsoda, 1998), the Eligibility method which restricts the probability of administering an item to a given exposure rate (van der Linden, 2003) and the Randomesque method which randomly chooses the next item to administer among the X items with higher fitness index (Kingsbury and Zara, 1989). On the other hand, the last experiment shows a direct comparison between the Merged Tree-CAT and the Tree-CAT method.

The item selection criteria taken in the five methods is the MEPV (2). This criterion was selected because of its high computational cost and because it is equivalent to minimizing the mean squared error (Delgado-Gómez et al., 2019). In the Randomesque method, the size of the group of items with the highest fitness index was fixed at $X = 6$. Lastly, in the Merged Tree-CAT method, the following parameters have been taken for both experiments: $\theta \sim N(0,1), K^* = 200, p = 0.9, \delta = 0.98$.

3.1 Experiment 1: Simulated data

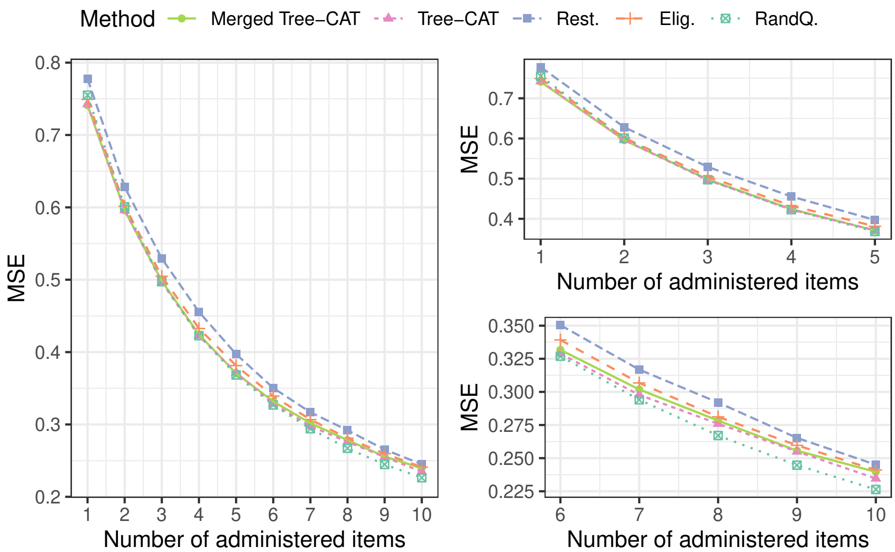
In this first experiment, an item bank composed of 100 items with three answers each was used. The items' parameters were generated according to the Samejima's graded response model (Samejima, 1969): The discrimination parameter was set to a log normal distribution with mean 0 and standard deviation 0.1225, and the difficulty parameters were generated according to a standard normal distribution (Magis and Raîche, 2011). Each item's exposition rate was set at $r_i = 0.3$ and the number of items to administer to each participant was set at $M = 10$.

To compare the abovementioned methods, the latent levels θ of a group of 500 examinees were generated, according to a standard normal distribution, where their responses to each item of the bank were simulated (Magis et al., 2012). Given that the Merged Tree-CAT, Tree-CAT, Eligibility and Randomesque methods have a random component in the administration of the test, this is repeated 25 times per examinee. Finally, this process is repeated 10 times to eliminate the dependence of the results with respect to the simulated

[1]The Merged Tree-CAT method is included in the cat.dt package created with the software **R** available in the CRAN repository or by contacting the corresponding author.

responses and parameters. The left panel in Figure 2 shows the mean squared error (MSE) of the estimates obtained by each method during the administration of the test. The panels on the right show this same image split into the first and last five items so that the MSE of each method is better appreciated.

Figure 2: Average MSEs for the Alternative Techniques for Simulated Data



It can be observed that at the end of the test, the proposed Merged Tree-CAT method obtains estimates that are as accurate as those of the Eligibility method and more precise than those of the Restricted method. However, those estimates are slightly less accurate than those of the Tree-CAT method, which does not limit tree growth. Finally, the Randomesque method shows the most accurate results. This is because this method does not satisfy the established exposure control; its overlap rate defined as the proportion of common items shared by two random examinees (Barrada et al., 2007) is 0.538, whereas for the Merged Tree-CAT, Tree-CAT, Restricted and Eligibility are 0.260, 0.283, 0.268 and 0.275 respectively.

Table 1 shows the average time used to construct the test by the various methods. The Merged Tree-CAT and Tree-CAT methods build the test before it is administered (0 s building time during administration) whereas the Restricted, Eligibility and Randomesque methods construct the test during its administration (0 s building time before administration).

This table shows that the Merged Tree-CAT method is the fastest of all the evaluated methods. Please note the difference with the Tree-CAT method: while the Tree-CAT method needed seven days of computation using 128 cores of a cluster with a Xeon 2630 processor and 32 GB of RAM, the Merged Tree-CAT method required only 40 seconds to create the test on a standard personal

Table 1: Test Building Times for Simulated Data

Method	Before administration	During administration
Merged Tree-CAT	$\approx$ 40 s	0 s
Tree-CAT	$\approx$ 7 days	0 s
Randomesque	0 s	$\approx$ 16.8 h (120 s$\times$500)
Eligibility	0 s	$\approx$ 23.6 h (170 s$\times$500)
Restricted	0 s	$\approx$ 16.8 h (120 s$\times$500)

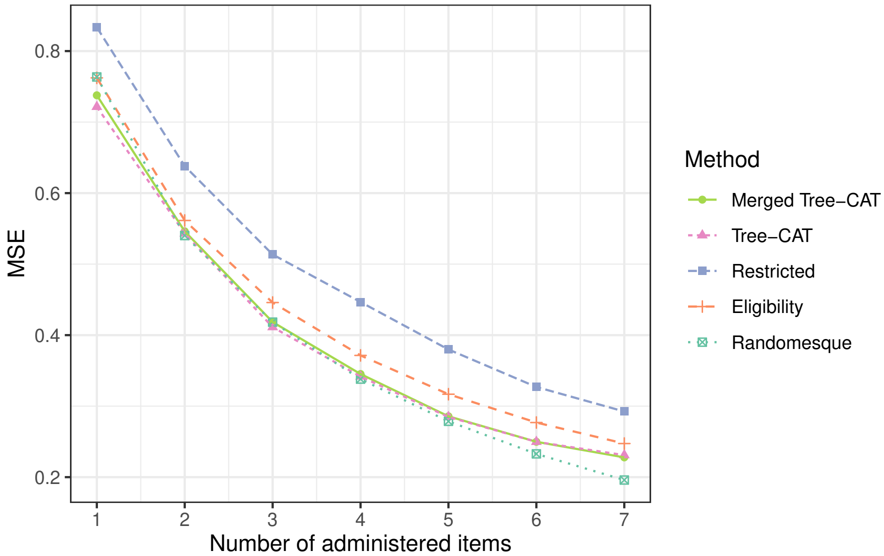
computer. With respect to the remaining methods, each examinee waits 12 seconds on average to receive the next item when using the Restricted and Randomesque techniques, and 17 seconds when using the Eligibility technique. In contrast, the Merged Tree-CAT method administers the next item immediately after an examinee answers the current item.

3.2 Experiment 2: Real data

In this second experiment, the bank used was the Emotional Adjustment Bank (Rubio et al., 2007), containing 28 items. In total, the answers were obtained from 792 participants, considering three categories for each item: "agree", "neutral" and "disagree". The answers provided by the participants were randomly divided into a training and a test set of equal size. The responses from the training set were used to calibrate the parameters of the items and construct the corresponding CATs with each method. On the other hand, the answers from the test set were used to obtain estimates of latent levels using the CATs created by the different methods. These estimates were compared with the estimates of the latent levels that were obtained after administering the 28 items to each participant. Similarly to the experiment with simulated data, the item selection criterion was the MEPV. Moreover, the items' exposition rate was set at $r_i = 0.3$ and the number of items to be administered was seven. Finally, the administration of the test to each participant for the Merged Tree-CAT, Tree-CAT, Eligibility and Randomesque methods was repeated 25 times, performing this process 10 times.

Figure 3 shows the MSE of the estimates obtained by each method during the administration of the test. It can be seen that the Merged Tree-CAT method obtains estimates as accurate as those from the Tree-CAT method, improving the Restricted and Eligibility methods. Only the Randomesque method obtains a smaller error, because as in the previous case, it exceeds the overlap rate: 0.58 versus 0.29, 0.28, 0.28 and 0.29 from Merged Tree-CAT, Tree-CAT, Restricted and Eligibility, respectively.

Figure 3: Average MSEs for the Alternative Techniques for Real Data



Regarding the average computation time, results are shown in Table 2. Computation times have been considerably reduced with respect to the previous experiment due to the fact that both the number of items in the item bank and the number of administered items are lower. Again, the proposed Merged Tree-CAT method is the fastest of all those analyzed.

Table 2: Test Building Times for Real Data

Method	Before administration	During administration
Merged Tree-CAT	$\approx$ 5 s	0 s
Tree-CAT	$\approx$ 36 min	0 s
Randomesque	0 s	$\approx$ 103 min (15.6 s$\times$396)
Eligibility	0 s	$\approx$ 117 min (17.7 s$\times$396)
Restricted	0 s	$\approx$ 103 min (15.6 s$\times$396)

3.3 Experiment 3: Sensitivity Analysis

In this last experiment, a sensitivity analysis was performed in order to compare the performance of the Merged Tree-CAT method with respect to the original Tree-CAT method. In order to do so, several simulation scenarios are set according to the number N of items of the item bank, the exposure rate r_i and the merger parameters K^* and δ. These last two parameters differentiate the Tree-CAT method from the Merged Tree-CAT method, since the first is a particular case of the latter for $K^* = \infty$ and $\delta = 1$. Concretely, the values $N \in \{100, 200, 500\}$, $r_i \in \{0.3, 0.6, 0.9\}$, $K^* \in \{25, 50, 100, 250, 500, 1000\}$ and $\delta \in \{0, 0.5, 0.98\}$ have been used.

For each scenario, the items' parameters and the latent levels θ of the examinees were simulated exactly as in Experiment 1 (see section 3.1). In this concrete case, the answers from 10000 examinees were generated in order to increase the accuracy of the results. In addition, the test administration was repeated 25 times per examinee, where this whole process was repeated 10 times. Finally, the number of items to administer to each participant was set at $M = 7$.

Table 3 shows the MSE of the final estimations obtained by both methods in each scenario. It can be observed that the Tree-CAT method obtains the smallest error in all scenarios. This is because the Merged Tree-CAT method limits the number of branches per level. However, it can also be noticed that the difference between both methods is minimal or even non-existent, depending primarily on the used K^* value: As the maximum number K^* of branches per level increases, the difference between both methods decreases. For instance, this difference is null when $K^* = 1000$ and there is enough variety in the item bank ($N = 500$). On table 3, it can also be appreciated that K^* is the decisive parameter to improve the performance of the Merged Tree-CAT method. Although the increase in the δ parameter tends to reduce the error, the improvement is small.

Table 3: Sensibility analysis results

$r_i = 0.3$				
Method		N. of items in bank		
		100	200	500
Tree CAT		0.297	0.275	0.259
Merged Tree CAT	$K^* = 25$, $\delta = 0$	0.308	0.285	0.269
	$K^* = 25$, $\delta = 0.5$	0.308	0.285	0.269
	$K^* = 25$, $\delta = 0.98$	0.308	0.284	0.268
	$K^* = 50$, $\delta = 0$	0.302	0.279	0.263
	$K^* = 50$, $\delta = 0.5$	0.303	0.279	0.264
	$K^* = 50$, $\delta = 0.98$	0.302	0.279	0.263
	$K^* = 100$, $\delta = 0$	0.301	0.278	0.261
	$K^* = 100$, $\delta = 0.5$	0.301	0.278	0.261
	$K^* = 100$, $\delta = 0.98$	0.300	0.278	0.261
	$K^* = 250$, $\delta = 0$	0.300	0.277	0.261
	$K^* = 250$, $\delta = 0.5$	0.300	0.277	0.261
	$K^* = 250$, $\delta = 0.98$	0.300	0.277	0.261
	$K^* = 500$, $\delta = 0$	0.300	0.276	0.260
	$K^* = 500$, $\delta = 0.5$	0.300	0.276	0.261
	$K^* = 500$, $\delta = 0.98$	0.299	0.276	0.260
	$K^* = 1000$, $\delta = 0$	0.299	0.276	0.260
	$K^* = 1000$, $\delta = 0.5$	0.299	0.276	0.260
	$K^* = 1000$, $\delta = 0.98$	0.299	0.277	0.259

$r_i = 0.6$				
Method		N. of items in bank		
		100	200	500
Tree CAT		0.285	0.262	0.250
Merged Tree CAT	$K^* = 25, \quad \delta = 0$	0.296	0.273	0.261
	$K^* = 25, \quad \delta = 0.5$	0.295	0.273	0.261
	$K^* = 25, \quad \delta = 0.98$	0.294	0.273	0.258
	$K^* = 50, \quad \delta = 0$	0.291	0.269	0.255
	$K^* = 50, \quad \delta = 0.5$	0.291	0.269	0.255
	$K^* = 50, \quad \delta = 0.98$	0.291	0.268	0.253
	$K^* = 100, \quad \delta = 0$	0.289	0.266	0.253
	$K^* = 100, \quad \delta = 0.5$	0.290	0.266	0.253
	$K^* = 100, \quad \delta = 0.98$	0.288	0.266	0.253
	$K^* = 250, \quad \delta = 0$	0.288	0.264	0.251
	$K^* = 250, \quad \delta = 0.5$	0.288	0.264	0.251
	$K^* = 250, \quad \delta = 0.98$	0.287	0.264	0.251
	$K^* = 500, \quad \delta = 0$	0.287	0.264	0.251
	$K^* = 500, \quad \delta = 0.5$	0.286	0.264	0.251
	$K^* = 500, \quad \delta = 0.98$	0.286	0.264	0.250
	$K^* = 1000, \quad \delta = 0$	0.286	0.263	0.251
	$K^* = 1000, \quad \delta = 0.5$	0.286	0.264	0.251
	$K^* = 1000, \quad \delta = 0.98$	0.286	0.263	0.250

$r_i = 0.9$				
Method		N. of items in bank		
		100	200	500
Tree CAT		0.278	0.258	0.247
Merged Tree CAT	$K^* = 25$, $\delta = 0$	0.294	0.271	0.257
	$K^* = 25$, $\delta = 0.5$	0.294	0.271	0.257
	$K^* = 25$, $\delta = 0.98$	0.291	0.270	0.257
	$K^* = 50$, $\delta = 0$	0.288	0.266	0.253
	$K^* = 50$, $\delta = 0.5$	0.288	0.266	0.253
	$K^* = 50$, $\delta = 0.98$	0.286	0.265	0.252
	$K^* = 100$, $\delta = 0$	0.287	0.264	0.250
	$K^* = 100$, $\delta = 0.5$	0.287	0.264	0.250
	$K^* = 100$, $\delta = 0.98$	0.285	0.264	0.250
	$K^* = 250$, $\delta = 0$	0.282	0.263	0.250
	$K^* = 250$, $\delta = 0.5$	0.282	0.263	0.250
	$K^* = 250$, $\delta = 0.98$	0.283	0.261	0.248
	$K^* = 500$, $\delta = 0$	0.281	0.261	0.248
	$K^* = 500$, $\delta = 0.5$	0.281	0.261	0.248
	$K^* = 500$, $\delta = 0.98$	0.281	0.259	0.248
	$K^* = 1000$, $\delta = 0$	0.281	0.261	0.248
	$K^* = 1000$, $\delta = 0.5$	0.281	0.261	0.248
	$K^* = 1000$, $\delta = 0.98$	0.280	0.259	0.247

4 Conclusions

In this article, we have presented the Merged Tree-CAT method for building CATs using a DT structure. Like the original Tree-CAT method, this method creates the test before it is administered, instantly supplying the items to the participants. The main difference between the Merged Tree-CAT and the Tree-CAT method is that the former constrains the growth of the tree by merging those branches whose estimates and distributions of the latent level are similar. The objective of these unions is to fast build tests that require little memory space without losing precision in the estimates.

The Merged Tree-CAT has two fundamental advantages over the Tree-CAT method. First of all, Merged Tree-CAT can build and administer the test on a standard computer. This is an important difference with respect to the Tree-CAT method since, although the built test can be administered on a personal computer, it requires a cluster for its construction. Secondly, Merged Tree-CAT performs the construction of the CAT in seconds, while Tree-CAT may take from several minutes up to months, depending on the length of the test and the size of the item bank. This calculation speed also surpasses that of other alternative CAT building techniques, with the advantage that the Merged Tree-CAT does not incur calculation time during test administration.

The speed of the Merged Tree-CAT method does not imply any significant loss of accuracy. The errors obtained by this method are practically the same as those obtained by the Tree-CAT method. In addition, with respect to the other alternative techniques, the Merged Tree-CAT method achieves maximum precision by controlling the overlap rate.

In conclusion, the Merged Tree-CAT method significantly improves the Tree-CAT method, enabling to quickly build CATs on any personal computer. Therefore, this proposed method is an ideal tool for the building of tests that must be administered simultaneously to a large number of participants.

Acknowledgments

Numerical experiments were conducted in Uranus, a supercomputer cluster located at Universidad Carlos III de Madrid and jointly funded by EU-FEDER funds and by the Spanish Government via the National Projects No. UNC313-4E-2361, No. ENE2009-12213- C03-03, No. ENE2012-33219, No. ENE2012-31753 and No. ENE2015-68265-P. This article was also funded by the Spanish National Project No. RTI2018-101857-B-I00.

References

Barrada, J. R., Olea, J., and Ponsoda, V. (2007). Methods for restricting maximum exposure rate in computerized adaptative testing. *Methodology*, 3(1):14–23.

Bock, R. D. and Mislevy, R. J. (1982). Adaptive eap estimation of ability in a microcomputer environment. *Applied psychological measurement*, 6(4):431–444.

Cha, S.-H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 1(4):300–307.

Chang, H.-H. and Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20(3):213–229.

Delgado-Gómez, D., Laria, J. C., and Ruiz-Hernández, D. (2019). Computerized adaptive test and decision trees: A unifying approach. *Expert Systems with Applications*, 117:358–366.

Kingsbury, G. G. and Zara, A. R. (1989). Procedures for selecting items for computerized adaptive tests. *Applied measurement in education*, 2(4):359–375.

Lord, F. M. (2012). *Applications of item response theory to practical testing problems*. Routledge.

Magis, D. and Raîche, G. (2011). catr: An r package for computerized adaptive testing. *Applied Psychological Measurement*, 35(7):576–577.

Magis, D., Raîche, G., et al. (2012). Random generation of response patterns under computerized adaptive testing with the r package catr. *Journal of Statistical Software*, 48(8):1–31.

Michel, P., Baumstarck, K., Loundou, A., Ghattas, B., Auquier, P., and Boyer, L. (2018). Computerized adaptive testing with decision regression trees: an alternative to item response theory for quality of life measurement in multiple sclerosis. *Patient preference and adherence*, 12:1043–1053.

Revuelta, J. and Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. *Journal of Educational Measurement*, 35(4):311–327.

Rubio, V. J., Aguado, D., Hontangas, P. M., and Hernández, J. M. (2007). Psychometric properties of an emotional adjustment measure: An application of the graded response model. *European Journal of Psychological Assessment*, 23(1):39–46.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement*.

Ueno, M. and Songmuang, P. (2010). Computerized adaptive testing based on decision tree. In *2010 10th IEEE International Conference on Advanced Learning Technologies*, pages 191–193. IEEE.

van der Linden, W. J. (2003). Some alternatives to sympson-hetter item-exposure control in computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, 28(3):249–265.

Van der Linden, W. J. and Pashley, P. J. (2009). Item selection and ability estimation in adaptive testing. In *Elements of adaptive testing*, pages 3–30. Springer.

Veerkamp, W. J. and Berger, M. P. (1997). Some new item selection criteria for adaptive testing. *Journal of Educational and Behavioral Statistics*, 22(2):203–226.

Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied psychological measurement*, 6(4):473–492.

Weiss, D. J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, 37(2):70–84.

Yan, D., Lewis, C., and Stocking, M. (2004). Adaptive testing with regression trees in the presence of multidimensionality. *Journal of Educational and Behavioral Statistics*, 29(3):293–316.