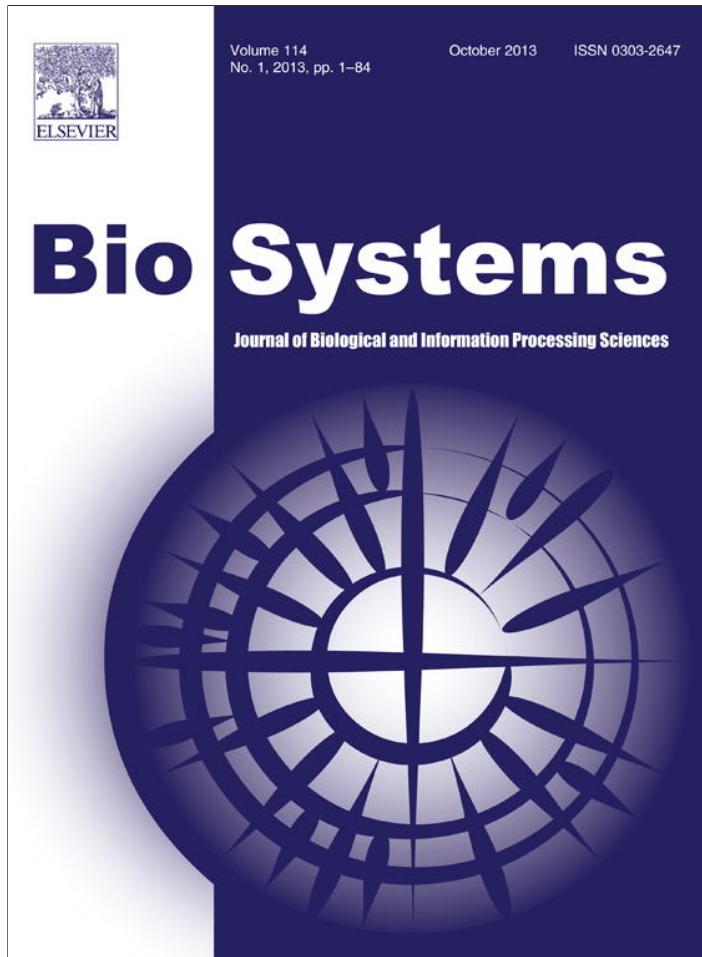


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

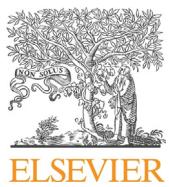


This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>



Contents lists available at ScienceDirect



Simulation of behavioral profiles in the plus-maze: A Classification and Regression Tree approach

Mauricio Molina-Delgado ^{a,*}, Michael Padilla-Mora ^b, Jaime Fonaguera ^b^a Psychological Research Institute, University of Costa Rica, Costa Rica^b Center of Neuroscience Research, University of Costa Rica, Costa Rica

ARTICLE INFO

Article history:

Received 19 September 2012

Received in revised form 10 April 2013

Accepted 3 July 2013

Keywords:

Simulation

Plus-maze

Anxiety

Decision trees

CART

Animal behavior

ABSTRACT

This article introduces a simulation model of rat behavior in the elevated plus-maze, designed through a Decision trees approach using Classification and Regression algorithms. Starting from the analysis of the behavior performed by a sample of 18 Sprague-Dawley male rats, probabilistic rules describing behavioral patterns of the animals were extracted, and were used as the basis of the model computations. The model adequacy was tested by contrasting a simulated sample against an independent sample of real animals. Statistical tests showed that the simulated sample exhibits similar behaviors to those displayed by the real animals, both in terms of the number of entries to open and close arms as well as in terms of the time spent by the animals in those arms. However, the performance of the model in parameters related to the behavioral patterns was partially satisfactory. Given that previous attempts in the literature have neither include this kind of patterns nor the time as a crucial model parameter, the present model offers a suitable alternative for the computational simulation of this paradigm. Compared with antecedent models, the present simulation produced similar or better results in all the considered parameters. Beyond the goal of establish an appropriate simulational model, extracted rules also reveal important regularities associated to the rat behavior previously ignored by other models, i.e. that specific rat behaviors in the elevated plus-maze are time dependent. These and other important considerations to improve the model performance are discussed.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

At the core of the behavioral neurosciences, the modeling of processes related to variations in the animal behavior, highlights the Elevated Plus-Maze as an important target for recent computational research projects (Salum et al., 2000; Giddings, 2002; Tejada et al., 2010). The Elevated Plus-Maze (EPM) is one of the most used and well validated paradigms for the analysis of rodent anxious behavior (Buccafusco, 2009; Lister, 1987; Pellow et al., 1985) and given its structure, the EPM also offers the possibility to develop simulational models of some behavioral parameters that could improve the understanding of the behavioral response in rats (Salum et al., 2000) and their underlying processes.

The standard EPM consists of two open arms, two closed arms and a central area where the animal can choose to enter at any of those four arms. The entire maze is elevated from the ground; and the test procedure usually involves the analysis of the free moving animal during five minutes, starting with the animal in the central area position. The frequency of entries and time spent on

every arm are usually registered. Commonly, high frequencies of entries and longer time lapses spent in open arms are associated with low anxiety states (Brenes-Sáenz et al., 2006); while other behavioral measures like grooming and rearing are recorded for a better characterization of the behavioral response (Holmes et al., 2000).

So far, computational modeling approaches to rat behavior in the EPM have shown partially satisfactory results (Miranda et al., 2009; Salum et al., 2000). In a seminal reference, Salum et al. (2000) based their proposal on the approach/avoidance theory of Montgomery (1955), and introduced the use of a neural network in which nodes corresponds to every possible position of the animal in the maze. Following Montgomery's statements, every node in the network was associated with a set of w_{ij} values which represents the tendency (w) of change from a position i to a position j . In this system, the node of the network that represents the actual position of the animal in a given state gets a value of 1, while all the other nodes maintain a value of 0. All the weights of the network were computed by a few algorithms directed to estimate those tendencies (Salum et al., 2000). Also, a random adjustment was added to introduce the effect of variation in the exploratory motivation exhibited by some animals. This later adjustment was not based on data, but on assumptions introduced by the authors, as were

* Corresponding author.

E-mail address: orescu@yahoo.com (M. Molina-Delgado).

the case with other parameters that were also re-adjusted for the network to reach an acceptable performance.

Later, a second computational proposal tried to improve the performance of this seminal model using an empirical based initiative (Giddings, 2002) not derived from theory driven algorithms, but from the analysis of the probabilities observed in the performance of a group of real rats. Giddings (2002) analyze the evidenced probabilities of seven different situations that frequently occur in the maze such as: a rat moving toward the entry of an open arm, toward the end of an open arm, toward the entry of a closed arm or toward the end of a closed arm, among others. Then, the empirically registered probabilities associated to those situations, direct the changes in the movements of a virtual rat through a simulation process (Giddings, 2002). Nevertheless, besides the important claim in favor of a more empirically based approach, the segregation of rat behavior in seven different situations was established based on the author ad hoc judgment and not as result of any empirical data analysis. In addition, the effect of time was not considered; while some relevant model parameters were corrected using a trial and error strategy until it reaches an acceptable result.

A third approach was proposed by Tejada et al. (2010) using Markov chains. Conceiving the antecedent models as evidence about the plausibility of modeling the rat behavior in the EPM as a probabilistic problem of event transitions, Tejada et al. (2010) proposed a Markov chain model. The states in this model correspond to places in the maze, while transitions represent movements to adjacent locations. Using this interesting approach, the authors verify that the proposed model reproduced the generic features of the exploration transition patterns of real rats in an EPM. But the model does not include the influence of time on the patterns of transition among different locations nor the display of animal behavior (i.e. grooming and rearing) as actual parameters.

The present article tries to take these former models a step further, and describes a simulational model of the rat behavior in the EPM based on a Classification and Regression Trees (CART) approach (Hastie et al., 2009). Starting from a database with locational and behavioral records of real animals in the EPM, this approach uses CART analyses to extract a set of rules that characterizes the different conditional probabilities of animal movements and behavioral transitions at different time intervals through the test length. Then, those sets of rules are used to attribute the movements and behaviors of a virtual sample through a simulational architecture. Hence, the approach tries to improve the gaps in the empirical foundations of antecedent models, and offers a new simulation model that for the first time: (a) introduce the time as a relevant parameter to predict the pattern of the animal performance in the maze, and (b) bring in the behaviors commonly exhibited by the rats in this maze as other kind of potentially valuable parameters. For the validation process, the performance of a simulated sample is contrasted against the performance of a new sample of real animals, not used for the establishment of the model.

2. Methods

2.1. Subjects and housing conditions

Recorded videos of 40 Sprague-Dawley male rats (28 days old) on the EPM were used for this study. The Animals were obtained from LEBI Laboratories (University of Costa Rica) and were housed in the colony room (room temperature at $22^{\circ}\text{C} \pm 2.8^{\circ}\text{C}$, 68–91% of relative humidity, 10 air cycles per hour and 12:12 h light-dark schedule) during 1 week before the behavioral measurement. At the moment of the behavioral measurement, the animals showed a mean weight of $64.08\text{ g} \pm 3.07\text{ g}$ (mean \pm S.E.M.). All the procedures

were approved by the Institutional Committee for Animal Care and Use of the University of Costa Rica (Session 3-AE-450).

2.2. Behavioral testing

The EPM apparatus was made of wood and consisted of four arms of equal dimensions ($50\text{ cm} \times 10\text{ cm}$) connected by a central area ($10\text{ cm} \times 10\text{ cm}$) and elevated at 50 cm from the floor. Two arms enclosed by walls (40 cm high), were perpendicular to two opposed open arms. To avoid falls, the open arms were surrounded by a Formica rim (0.5 cm high). Testing room was dimly illuminated with two 25 W red bulbs located 150 cm above the maze. At the beginning of the test, each rat was placed in the central area facing to a predefined closed arm. Each animal experienced the EPM for 5 min. All testing sessions occur between 8 a.m. and 11 a.m. and were digitalized as individual videos for posterior analysis. The maze was cleaned with 70% alcohol between rat sessions to reduce odor cues.

2.3. Data codification

Following the procedure of Salum et al. (2000), the EPM area was divided in 13 zones for a better characterization of the transitions among different locations. Behavioral and positional data registered in each video were codified using two different modalities: (a) locational transitions of the animal through different EPM areas were codified using an automatic video-tracking system (Stoeling, Any-maze 4.63 using the location of the 88% of the animal body as the criteria to determine the current area), while (b) behavioral parameters like grooming times/frequencies, rearing time/frequencies, and stretch-attempt postures (as described in Brenes et al., 2009) were codified by trained human observers. Inter-rater agreement was assessed using 30% of all the video recorded data. Agreement reliability was >0.85 for each behavioral category. Given the relevance of both the number of zones used as surface area divisions in the simulation, and the percentage of animal body used as the criteria for the current area determination, these parameters were explored as part of the preliminary analysis (see Section 3). Also, following Giddings (2002) the maze was divided in regions A and B, where A identifies the region that includes the initially predefined closed arm that the animal is facing at the beginning of the test and the open arm to the right; while B identified the two other arms. This was implemented to empirically identify possible bias in arm selection favoring the initially predefined closed arm.

Both kinds of data (behavioral and locational) were coded for every second that the animal spent in the maze, and were included as relevant information for the analysis of the behavior of each rat. Later, half of the animal's videos randomly selected were used to implement the simulation model, and the remaining half were used for the validation procedure.

2.4. The simulation model

The implementation of the simulation model begins with the analyses of a database that includes the behavior and position of 18 rats (2 animals were removed after being identified as outliers, using the criteria of Mahalanobis distances higher than $\chi^2(17)=33.40$) at every second in the EPM (300 s for each animal). Using this data, computations were made to establish other relevant variables like: *animal movement* (set to 1 if the animal moved to an adjacent area from one second to the next, or 0 otherwise), *current zone location* (set to 0, 1 or 2, if the animal location belongs to the central area, closed or open arms, respectively), *distance from center* (set to 0 when the animal was located in the central area, or to 1, 2 or 3, if the animal was located at the entrance, middle or

end of any arm respectively) and *region selection* (set to 1 if the animal enter the region that includes the initially predefined closed faced arm, or 0 otherwise). Subsequently, the complete database was analyzed using a CART approach (Hastie et al., 2009) performed with the SPSS 17.0 Decision Tree Module, and the Classification and Regression Tree algorithm as the growing method. From these analyses, four different CARTs were obtained. The simulational model combines the use of those different CARTs (see Section 3) to calculate every new state. The first tree determines the occurrence probability of animal movement to a different area (dependent variable), based on the conditional probabilities associated to: time lapse, current arm location, current behavior and distance from center (as independent variables). The second tree determines the probability associated to the possible direction of the movement (toward or away the center). While the third tree determines the occurrence probability of different behaviors (i.e. grooming, rearing or stretch-attempt postures); in both cases the same set of independent variables used in the first tree were included again. Finally, the fourth tree determines the probability of entrance to an open or close arm for those cases in which animal were located in central position. Here, the independent variables were the time lapse, the current behavior and the number of previous entrances of the animal in the open arms. All decision trees with the exception of the fourth, shared the following parameters: minimum number of cases for parent and child node was 100 and 50 respectively, while the maximum tree depth was set to 5. Given that the fourth tree only includes those cases where the animal was in central position, the minimum number of cases was set to 40 and 12 for the parent and the child nodes respectively. The impurity measure used was the Gini's Index with a minimum change in improvement of 0.0001. A 10 fold cross-validation was applied.

2.5. Decision trees with CART method for behavioral analysis

As mentioned, the simulation model was implemented using four different decision trees that used the Classification and Regression Tree (CART) algorithm as the growing method (Hastie et al., 2009). In the CART algorithm, there can be a y that refers to a

nominal variable that assumes values in $\{a_1, a_2, a_3, \dots, a_m\}$, and X to a vector (x_1, x_2, \dots, x_p) where each x_i represents an independent nominal or metric variable. Hence, the algorithm allows the possibility to consider classification problems in terms of finding partitions of the p -dimensional space conformed by the x_i 's in n non overlapped multidimensional regions; where those regions were as homogenous as possible in relation to the dependent variable y . Thus, the CART algorithm recursively divides the p -dimensional space selecting a first variable x_j from the vector X and cutting the range of values of the variable in a point x_{j0} , producing two intervals: $x_j < x_{j0}$ and $x_j > x_{j0}$. The improvements in the predicted probabilities associated with these intervals are temporary stored in memory, while a new variable x_k from the vector X is selected, split in a point x_{k0} , and compared to those previously stored in term of their associated improvements. For any variable, the process continues until the new intervals improvement values do not exceed a given threshold. Those variables and intervals associated with major improvements in the prediction probabilities are selected as relevant nodes for each CART.

The application of CART algorithms to the biological problem of animal behavioral prediction, allows the consideration of the EPM dynamics in terms like the following: let's consider a case with only 3 relevant variables for the prediction of a rat's next movement in the EPM,

$$y \in \{\text{stay, move}\}$$

$$x_1 \in \{\text{open, close}\}$$

$$x_2 \in \{1, 2, \dots, 300\}$$

where y represents the dependent variable (the rat will move to another position or will stay at least one more second in the same position); x_1 the current arm (open or close) and x_2 the time (in s).

The CART analysis for a theoretical sample of 20 animals is shown as an example in Fig. 1. The Node 0 represents the starting point, where half of the units of analysis shows movement (set as 1) and the other do not (set as 0). Thus, the probability associated to a

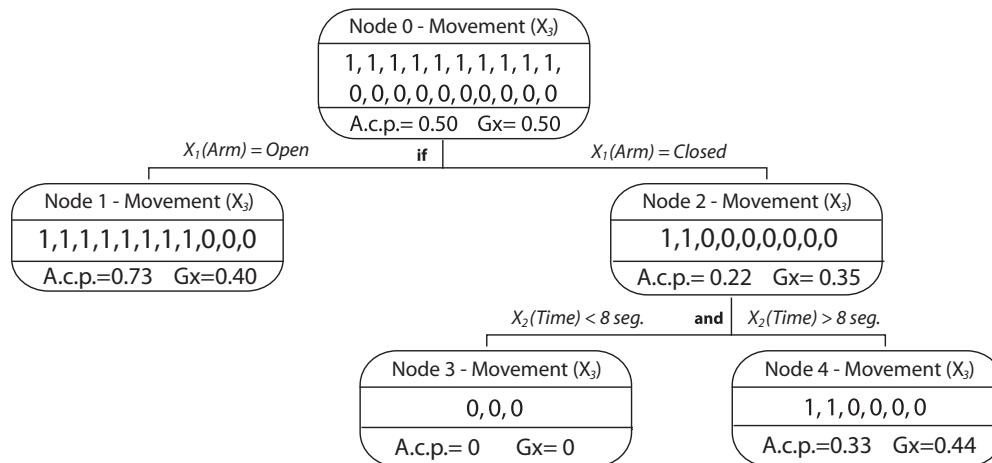


Fig. 1. Extended example of a Decision Tree Structure for the prediction of Animal Movement using two independent variables ($n = 20$). In the example, the initial value of Gini's index for the 20 cases was: $1 - (0.50^2 + 0.50^2) = 0.50$. The partition that minimizes Gini's Index is in the variable x_1 . In the Nodes 1 and 2 the index was 0.40 and 0.35 respectively. Thus, the gain of including this partition would be: $0.50 - (0.55 \times 0.40 + 0.45 \times 0.35) = 0.50 - 0.38 = 0.12$ (the gain is defined as the difference between the initial model and the model with the new partition, taking into account the sum of the final nodes' Gini's index weighted by the number of elements in each node). The recursive application of this procedure promotes the division of the right branch according to the time in seconds. That new partition showed a Gini's index of 0 (Node 3) and 0.44 (Node 4) and represents a gain of: $0.38 - (0.55 \times 0.40 + 0.15 \times 0 + 0.30 \times 0.44) = 0.38 - 0.35 = 0.03$. At this moment the process stops because of the previously established restrictions (Node 3 has 3 items and Node 4, 6 items). Similarly, the tree will not grow in the right branch because the minimum gain .01 has been obtained. The 3 final nodes could be expressed as rules in the following terms: (2.3) IF($x_1 = \text{open}$) THEN $move, p = 0.73$; (2.4) IF($x_1 = \text{close}$) AND ($x_2 \geq 8$) THEN $move, p = 0.33$; (2.5) IF($x_1 = \text{close}$) AND ($x_2 < 8$) THEN $move, p = 0.00$. According to this, the Rule (2.3) means that when the animal is in the open arm it will move with a probability expressed by the value $p = 0.73$. Similarly, Rule (2.4) expresses that if the animal is in the closed arm and the time is equal or superior to 8 seconds, it will move with a probability of 0.44. Finally, Rule (2.5) means that an animal in the closed arm before 8 seconds must stay in the same position until the next second. Notes: A.c.p.: associated conditional probability (of movement), Gx: Gini's Index.

the prediction of movement from Node 0 is 0.50. Later, the Node 1 groups those cases from Node 0 where the animals were in the open arms and, adding this consideration, the associated probability of movement for this Node is 0.73. Then, following the same procedure, other nodes like Node 3 or Node 4 also evidence different associated probabilities (0 and 0.33 respectively) by considering those cases where the animals were in the closed arms and the time was under or above 8 s. Importantly, the exemplar CART in Fig. 1 contemplates the following consideration: every parent node should have 9 or more elements; every final node should have at least 3 elements and, the minimum gain tolerated (defined as the difference in the given information index between a parent node and the weighted average of its sons) for any node was 0.01 (for a detailed description see Fig. 1).

The CART procedure also include an information index criteria for the quantification of the convenience associated to the establishment of a new partition (or branch), in order to produce a new homogeneous region in y . Particularly, Gini's index (see Eq. (2.1)) is the preferred criteria in CART because of its mathematical properties (Hastie et al., 2009). Where p_k corresponds to the proportion of the N elements in a given region with k representing the m possible categories assumed by y_i (see Eq. (2.2)); where $I(y_i = k)$ is equal to k and $I(y_i \neq k) = 0$ if y_i is not equal to k .

$$G = 1 - \sum_{k=1}^m (p_k)^2 \quad (2.1)$$

$$p_k = \left(\frac{1}{N} \right) \sum_{i=1}^N I(y_i = k) \quad (2.2)$$

2.6. Validation method

The performance of the simulation model was tested via comparison with the behavioral parameters of a different subsample of real animals ($n = 19$ after the removal of one outlier using the criteria of Mahalanobis distances higher than $\chi^2(17) = 33.40$, not included in the establishment of the model). The behavioral performance of this second subsample of real animals was contrasted to an artificial subsample of equal n size, generated through the simulation model. The overall performance of both subsamples in relevant EPM locational and behavioral parameters was analyzed using a Multivariate Analysis of Variance (MANOVA) with $p < 0.05$, the standard statistical procedures reported in the literature.

3. Results

3.1. Preliminary analysis

As mentioned previously, given their relevance for the simulational model, both the percentage of animal body used as the criteria for the current area determination, as well as the number of zones used as surface area divisions in the simulation were analyzed.

The first analysis contrasted the differences associated to the use of 88% of animal body as criteria for current area determination (as used in the automatic video-tracking) against the use of the 'four paws into the area' criteria (a common parameter generally used by trained human observers). The same set of 18 videos of animals in the plus-maze was registered by video-tracking and human observers using the above-mentioned criteria respectively. A paired sample analysis of the resulting times spent in open and closed arms according to those different criteria were contrasted. The results did not show statistically significant

differences between both criteria in the time spent in open arms, $t(17) = -0.54$, $p = 0.60$; nor in the times spent in closed arms, $t(17) = -0.69$, $p = 0.50$.

Also, a posterior analysis was performed using a different number of area divisions to analyze possible differences resulting from the adoption of this parameter. Thus, the results obtained using the 13 zone division was contrasted (zone by zone) against the results obtained by the implementation of a 25 zone division (the central area was maintained as unitary, but each of the remaining 12 zones were disaggregated as 2 subdivisions (i.e. 1-NE/1-AE, 2-NE/2-AE, etc.; NE and AE for subdivision Near the arm Entrances or Away from the Entrance)). Thus, significantly different times spent between two subdivisions derived from the same area were considered as evidenced favoring a gain of information by subdivision. Nevertheless, a repeated measure MANOVA using the times spent in every subdivision as dependent variable and three intra-subject variables (subdivision, arm, distance from center and region) does not show a significant principal effect associated to the subdivisions (Wilks's lambda = 0.98, $F(1,17) = 0.30$, $p = 0.58$, partial $\eta^2 = 0.02$). Interestingly, the analysis of the MANOVA interactions exhibited significant results (subdivision \times region \times distance from center \times arm: Wilks's lambda = 0.24, $F(2,16) = 13.59$, $p = 0.02$, partial $\eta^2 = 0.44$). When reviewed in detail, these interactions point toward differences between 4 (in 12) pairs of maze subdivisions. Nevertheless a posterior visual inspection of videos showed that these differences were not produced by the locational transitions of the animal, but were derived from the vertical/horizontal changes in the axis of the animal associated to rearing and stretch attempt posture behaviors. Hence, the 13 areas division was conserved as an adequate representation of the maze surface in the following.

The *region selection* parameter also was explored empirically. After the analysis of the probability associated to the selection of the A region (that includes the initially predefined faced arm) for those cases where the animal is located at the central area, this value was set to 0.669 for the simulation.

3.2. CART trees

Fig. 2 shows the structure of the first and deepest tree derived from the analysis of the real subsample. This tree was generated for the prediction of *animal movement* using time lapse, *distance from center*, current behavior and current location at the beginning of each second as independent variables. The tree evidenced a total of 85% correct predictions (risk = 0.54, SE = 0.018) for those cases where the rats actually move, with a slightly minor percentage of correct predictions for the total of cases (80%).

As can be seen, *distance from center* demonstrated the most important contributions followed by time lapse; whilst current behavior and location showed lower relevance (See Fig. 2).

Similarly, Fig. 3 shows the structure of the decision tree for the prediction of *animal direction of movement* for those cases where the rat is not located at the central position (otherwise the direction is fixed away from center), using the same independent variables introduced in the previous tree. The main result of this tree shows that the condition of being in the outer or middle part of an open arm is associated with a higher probability of movement toward the central area, when compared to the same probability when the animal is in the outer or middle part of a closed arm. The percentage of correct predictions was 71% (risk = 0.29, SE = 0.02).

Fig. 4 exhibits the structure of the decision tree for the prediction of *animal behavior* using the same independent variables introduced in the previous tree. Here the main effect detected was for the current behavior, plus some minor contributions

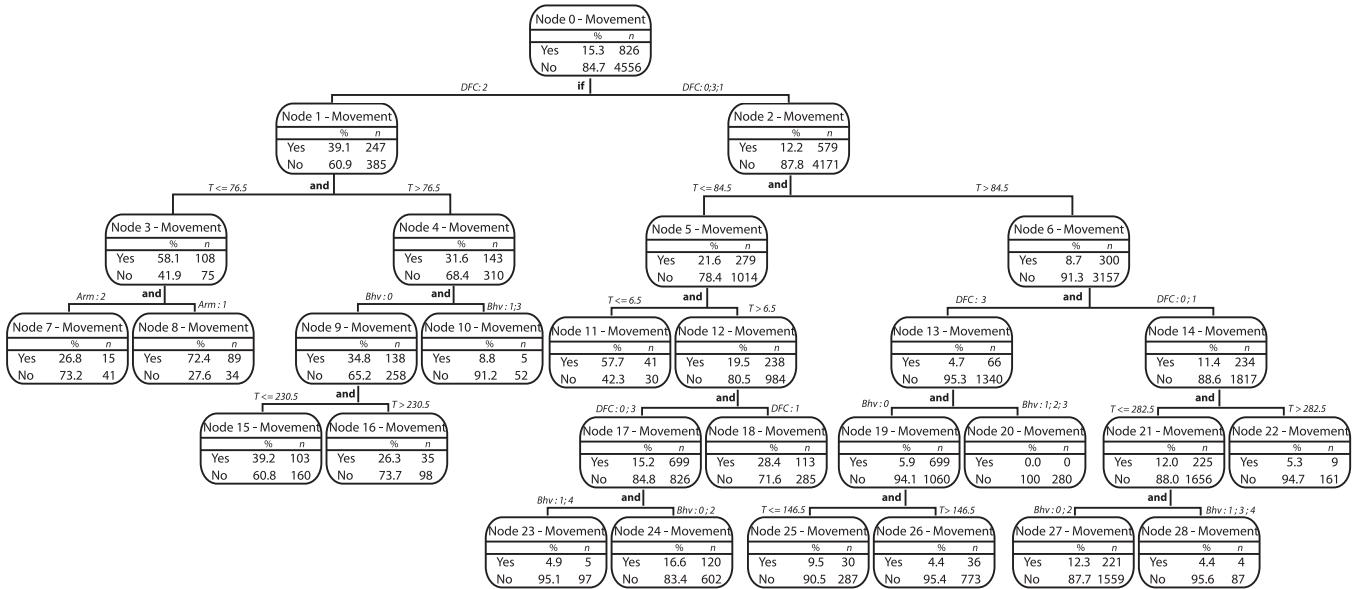


Fig. 2. Decision Tree Structure for the prediction of *animal movement* using time lapse, distance from center, current behavior and current location at the beginning of each second as independent variables ($n = 18$). Notes: Node 0 shows the general unconditional probability of animal movement for the entire sample through the total length of time. Subsequent Nodes shows the conditional probabilities for specific cases were: T = time in seconds, Bhv = Current Behavior (0 = none, 1 = rearing, 2 = stretched attend posture, 3 = grooming, 4 = head-deepening), DFC = Distance from Center (0 = at the central area, 1 = at the entrance of the arm, 2 = at the middle of the arm, 3 = at the end of the arm), Arm (1 = closed, 2 = open).

associated to time lapse. The percentage of correct predictions was 94% (risk = 0.06, SE = 0.003).

Finally, Fig. 5 shows the decision tree for the prediction of the rat's arm choice (open or close) from central position. The main effect was for the time lapse, followed by the number of previous entrances in open arm. The current behavior was excluded by the model. The rate of correct predictions was 93% (risk = 0.17, SE = 0.03) for those cases in which the animal moved to the close arm, and 70% when it moved to the open arm.

The probabilities associated with the leaves of these trees, were computed and implemented in Python 2.6 code as a set of production rules (similar to those presented in Fig. 1) for the implementation of the simulation program (see Appendix 1). Additionally, a single restriction was imposed to the program. In specific, head-deepening in the middle of closed arms behavior (actually implausible in real conditions) was restricted.

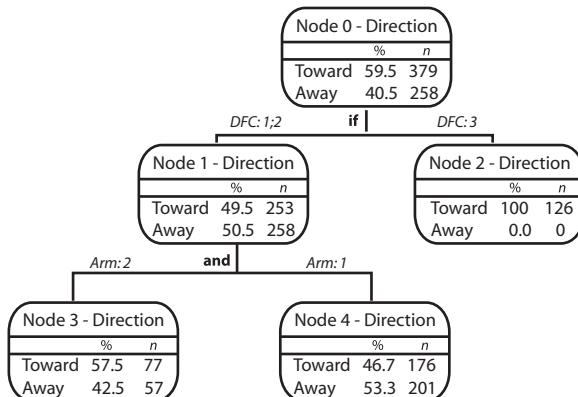


Fig. 3. Decision Tree Structure for the prediction of animal direction using time lapse, current arm location, previous behavior and distance from center as independent variables ($n = 18$). Notes: DFC = Distance from Center (1 = at the entrance of the arm, 2 = at the middle of the arm, 3 = at the end of the arm), Arm (1 = closed, 2 = open).

3.3. Validation results

Using the simulation architecture based on the rules extracted from the CARTs a simulated sample of 19 rats was produced. This sample was contrasted with an independent sample of 19 real rats that were not used as part of the implementation of the program. Then a series of MANOVAS tests were conducted between the real and the simulated sample, using times and frequency of visits to arms and behaviors exhibited as dependent variables. As expected, the first MANOVA using entries frequency in open arms, closed arms and central area as dependent variables between both samples did not show statistically significant differences ($p < 0.05$, Wilks' lambda 0.88, $p = 0.11$, $\eta^2 = 0.12$ (see Fig. 6a). Similarly, there were not statistically significant differences in another MANOVA using the time spent in the same areas as dependent variables (see Fig. 6b), Wilks' lambda 0.99, $p = 0.89$, $\eta^2 = 0.01$. Nevertheless, a third MANOVA exhibited partially satisfactory results. The overall effect of the sample on the time spent in different behavior was significant (see Fig. 6c), Wilks' lambda 0.75, $p = 0.02$, $\eta^2 = 0.25$, as well as the effects on stretch-attempt postures, $F(1, 36) = 9.39$, $p < 0.01$, $\eta^2 = 0.21$. In contrast, the specific effects on grooming, $F(1, 36) = 1.23$, $p = 0.27$, $\eta^2 = 0.03$ and rearing, $F(1, 36) = 0.26$, $p = 0.62$, $\eta^2 = 0.01$, were not significant. The overall effect of the sample on the frequencies of different behaviors was significant on the last MANOVA (see Fig. 6d), Wilks' lambda 0.47, $p < 0.01$, $\eta^2 = 0.53$. Unexpectedly, there were significant differences on grooming, $F(1, 36) = 16.45$, $p < 0.01$, $\eta^2 = 0.21$; rearing, $F(1, 36) = 19.99$, $p < 0.01$, $\eta^2 = 0.36$, and stretch-attempt posture, $F(1, 36) = 10.26$, $p < 0.01$, $\eta^2 = 0.22$.

4. Discussion

The results suggest that the Classification and Regression Trees approach for the simulation of the Elevated Plus Maze, offers a suitable alternative for the computational modeling of this classic paradigm in behavioral neuroscience. The simulational model was able to produce similar results to those revealed by the real sample in central variables for the cognitive-behavioral

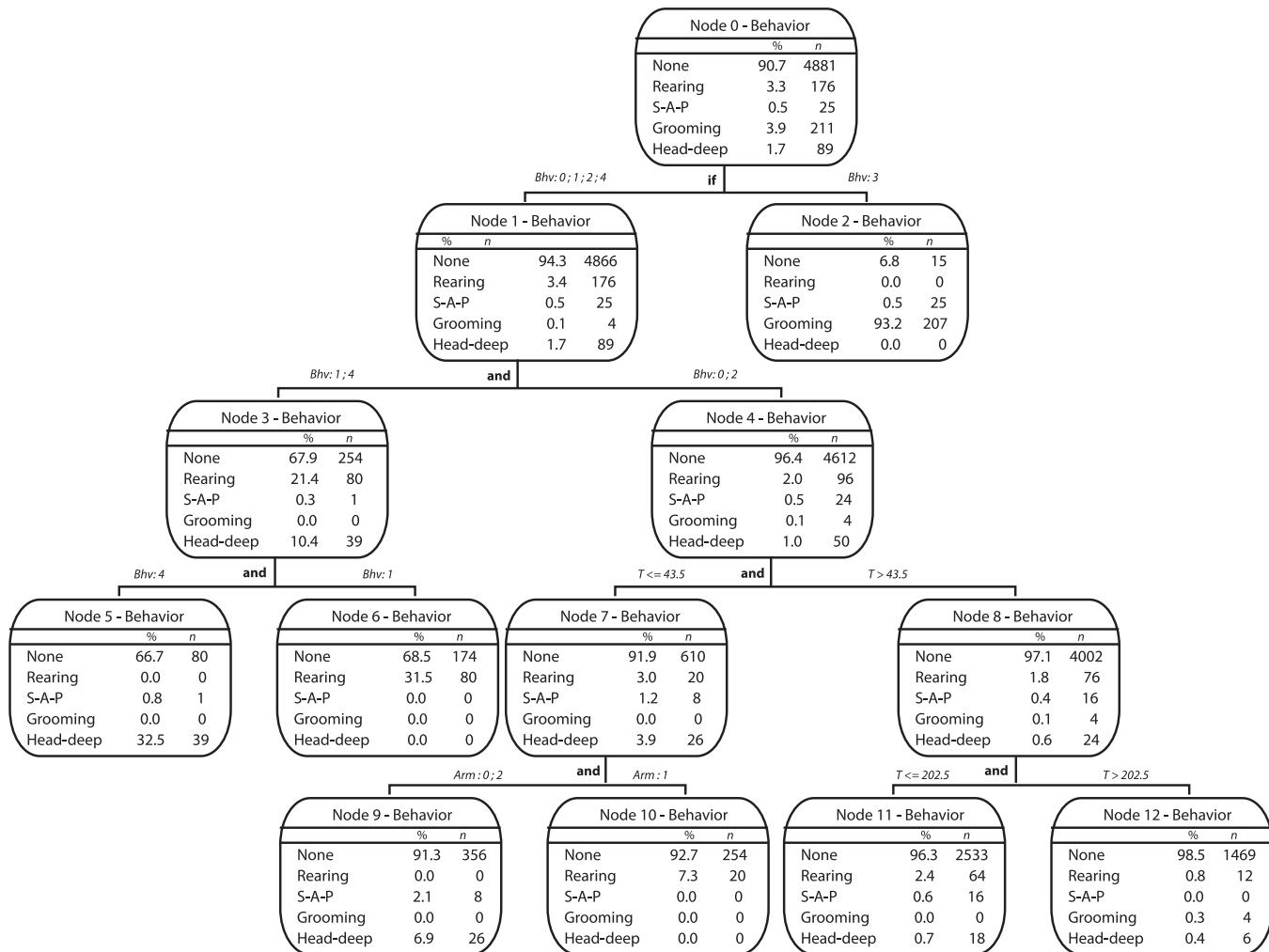


Fig. 4. Decision Tree Structure for the prediction of animal behavior using time lapse, current arm location, previous behavior and distance from center as independent variables ($n=18$). Notes: T = time in seconds, Bhv = Current Behavior (0 = none, 1 = rearing, 2 = stretched attend posture, 3 = grooming, 4 = head-deeping), Arm (1 = closed, 2 = open). n = number of cases, % = percent of total probability.

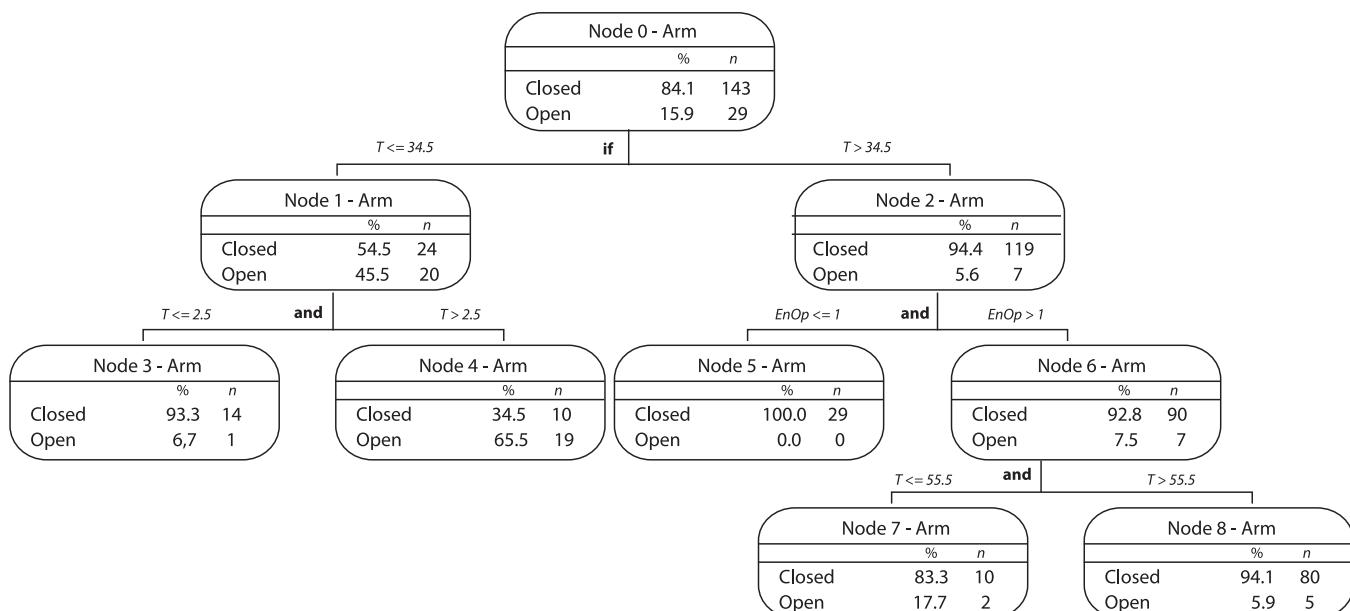


Fig. 5. Decision Tree Structure for the prediction of Entry to Arm using lapse, current arm location, and number of previous entries in Open Arm ($n=18$). Notes: T = time in seconds, $EnOp$ = Previous Entries in Open Arm. n = number of cases, % = percent of total probability.

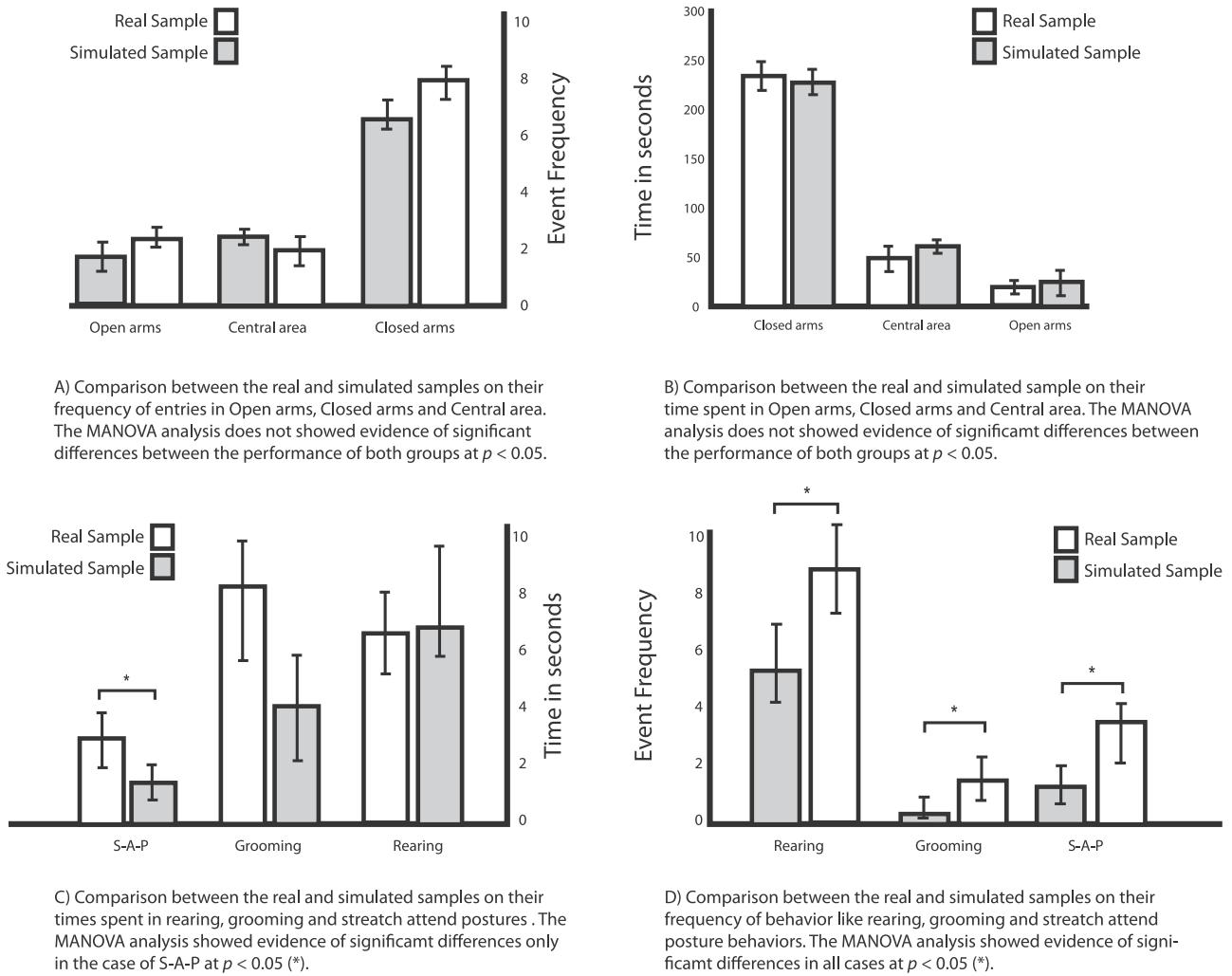


Fig. 6. MANOVA results for the analysis of statistical differences between the real and simulate ($n = 19$ for each group) using frequencies of entries and times spent on different areas of the maze, as well as frequencies and time spent at different behaviors as dependent variables.

characterization of the behavioral profiles commonly displayed on this maze (Brenes et al., 2009; Holmes et al., 2000). Essential parameters like time spent in open and closed arms, demonstrated similar values between the real and virtual sample after the validation process. Furthermore, the model also showed satisfactory results in terms of the generated number of entries to open and closed arms, another subset of relevant parameters (Hogg, 1996).

The CART approach also allows the identification of time as a significant parameter for the prediction of the animal performance, not previously considered by other models (Salum et al., 2000, 2003; Tejada et al., 2010). Actually, several specific conditions associated to different nodes in most CARTs were time dependent, a result that remarks the need to take into account the eventual contribution of this parameter for an improvement in the accuracy of the simulational models. In general, the extraction of these and other kinds of parameter regularities from the raw data and through the CART method, facilitate a reduction in the establishment of ad hoc parameters at the basis of the formal models.

Nevertheless, even when the architecture also achieved the expected results in terms of the generation of behavioral parameters like time spent in grooming and rearing, this was not the case for the time spent in stretch-attempt posture nor for the total event frequencies of these variables, where the model showed statistically significant differences against the real sample. These unexpected results can be explained given different tenets. It may

be the case that these behavioral parameters demand a different strategy of analysis, since it was evident that some behaviors (i.e. grooming) appeared in clusters of consecutive seconds at different time intervals for different animals, impeding their accurate characterization through a CART procedure (Keedwell and Narayanan, 2005; Timofeev, 2004). In this case, future proposals should consider mixed strategies of analysis, where different algorithms can be applied to the prediction of different parameters. Here, an eventual introduction of a Cox regression model for the specific prediction of behavioral frequencies, may improve the proportion of the model's correct predictions, by an analysis of their specific time lengths (Kattan, 2006).

Besides, another consideration points to the number of cases included in the model establishment. Because of the low frequency exhibited by certain behaviors, it may be expected that a bigger sample size can lead to a more accurate representation of the conditional probabilities derived from the CART's. In this sense and following the results, an increment in the accuracy associated with the probabilities of grooming and stretch-attempt postures initiation, could improve the model performance against the real sample. Another important consideration remarks the acquisition of the behavioral measures, which in contrast to the positional measures were not automatized but registered by trained observers (see Section 2.3). Even when reliability between observers reach satisfactory coefficients (Watkins and Pacheco, 2001), this points at

the necessity to consider new initiatives to contrast different sets of automatized parameters and their possible effects on the simulation results. Despite the fact that the present article also analyzed possible significant effects associated with parameters like the percentage of animal body used as criteria for current area determination and the number of zone division, this should be extensively explored in the future. Here, neither of those parameters evidenced a significant contribution when compared to a trained observer criteria ('four pawns into the area') or to a major number of division (25 zones), respectively. Nevertheless, these are just a few relevant values among plenty others that may be assessed in searching for model optimizations.

Interestingly, the CARTs also offer an innovative way to visualize and examine the behavior of the real animals in the EPM. The results highlight a subset of time intervals and conditions that may demand more attention from researchers to specify the nature of the anxiety-like performance of the animals in this paradigm. As shown in Nodes 11 and 12 from Fig. 2, as well as in Nodes 3 and 4 from Fig. 5; the very first seconds on the EPM are crucial. From their initial position at the central area, the probability for an animal to move to another location change from 0.57 (if the transition is made during the first 6.5 s, Node 11 in Fig. 2) to 0.19 (if this transition is made posterior to that time interval, Node 12 in Fig. 2). Among those transitions executed before 6.5 s, there is another important consideration for the cases where the transition is made during the first 2.5 s. In this latter case (Node 3 in Fig. 5) the probability to move toward a closed arm is 0.93 in contrast with a probability of 0.34, for those transitions executed from the central area after 2.5 s (Node 4 in Fig. 5).

This kind of evidence can help to better characterize an initial avoidance response, or reactivity, as a main factor in the modulation of the anxiety-like behavior of the animal through their maze experience. A factor putatively elicited by the novel and aversive character of the maze structure, or derived from the drive to escape from the initial human handling. Simulational models where these initial parameters were altered produce remarkable changes in the percentages of time spent in open and closed arms (data not shown), claiming the need for a more detailed analysis of these initial performances.

In addition, the fact that *distance from center* emerges as a relevant dimension for the prediction of movement, denotes the particular character of the middle of the arms as passage zones with high probability of movement associated. Most important, this also contributes to quantify the differences between the middle of open and closed arms (Nodes 7 and 8, in Fig. 2). Based on expert criteria and intuition, recent literature has started to take into account the quantification of the explorations beyond the middle of the open arm as an alternative indicator to better characterize the anxious response of the animal in the EPM (Rico et al., 2009). Here, the difference in the movement probability associated to both arms, offers support to the use of this kind of indicator and claims for the exploration of the temporal variations related to that parameter, given that certain specific probabilities associated to it were only present during the initial 76 s.

Finally, another informative feature revealed by the CARTs is exposed through Nodes 2, 5 and 6 from Fig. 4. There, conditional probabilities of maintaining a grooming, head-deepening or rearing behavior from one second to the next are shown. Interestingly, the eventual consideration and inclusion of these values and their changes through time as parameters or dynamical attractors, could improve the performance of future models, opening new modeling perspectives based on dynamic systems theories (Montebelli et al., 2008; Poucet and Save, 2005).

Other alternatives to improve the model general performance should be also contemplated. Some relevant options have been applied as part of previous models with good results. For example,

the introduction of locational and behavioral data from 3 or 5 antecedent seconds in the prediction of the next step could be an alternative to optimize the prediction of both locational and behavioral variables. This alternative could also improve the analysis of other kind of parameters like the inertia in animal transition, used with satisfactory results by antecedent research (Giddings, 2002; Tejada et al., 2009). Markov networks, as well, allow the introduction of specific occurrence probabilities associated to specific sequence chains. Therefore, the implementation of a Markov transition matrix to support the prediction accuracy of the architecture may be another good alternative that has been probed before (Tejada et al., 2010). Still, these and other possible alternatives should be carefully assessed, examining the given improvement in the prediction accuracy against the computational demands and complexity introduced in exchange (Keedwell and Narayanan, 2005). In another direction, the application of different methods of CART pruning should also be tested, given that none of them were used here in an effort to prevent the unnecessary introduction of as much ad hoc parameters as possible.

Another alternative worth to mention is the neural network approach. A neural network could allow the simultaneous introduction of some of the previous considerations, and has been used in two earlier reports with partial satisfactory results (Salum et al., 2003; Miranda et al., 2009). Although neither of them included behavioral parameters, a neural network may consider both locational and behavioral data, as well as specific nodes for n previous states of positions and behaviors in order to categorize a given next step.

In conclusion, the present article offers a CART based architecture that simulates the behavior of rats in an Elevated Plus-Maze. The simulational model was able to generate and achieve good results in the production of relevant parameters like time spent in open and closed arms, as well as in terms of the number of entries to those areas. Plus, the model generates partially satisfactory results in terms of the prediction of different behaviors performed through the test length. Given that these parameters constitute the basis of the assessment commonly used in the behavioral neurosciences, the present results in addition to some of the antecedent research commented here, offers a good perspective toward the future applications and contributions of these approaches. Among these eventual applications, the present article showed that these models can facilitate new ways to visualize and analyze the animal behavior in detail, offering the possibility to identify alternative and informative indicators and/or endophenotypes that should be considered and contrasted both by *in vivo* and *in silico* procedures. Finally, as it has been proposed elsewhere (Knight, 2011; Scalesse and Issenberg, 2005; Thomas, 2009), the optimization of these models can be particularly important for the explorations of alternatives to reduce the number of animals used in experimental testing and teaching. However, in parallel to those efforts, improvements and systematic changes are needed for these models to really achieve a totally satisfactory performance. Some of those improvements have been commented here, but others should take advantage from the emerging new approaches deriving from the interdisciplinary convergence of neuroscience, behavioral science, computer science and bioinformatics.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.biosystems.2013.07.002>.

References

- Brenes-Sáenz, J., Rodríguez-Villagra, O., Fornaguera-Triás, J., 2006. Factor analysis of forced swimming test, sucrose preference test and open field test on enriched, social and isolated reared rats. *Behavioural Brain Research* 169, 57–65.
- Brenes, J., Padilla, M., Fornaguera, J., 2009. A detailed analysis of open-field habituation and behavioral and neurochemical antidepressant-like effects in post-weaning enriched rats. *Behavioural Brain Research* 197, 125–137.
- Buccafusco, J., 2009. Methods of Behavioral Analysis in Neuroscience. CRC Press, Florida.
- Giddings, J., 2002. Modeling the Behavior of Rats in an Elevated Plus-Maze. Acadia University, Wolfville, Nova Scotia, Canada, Master's Thesis (accessed 1.05.2012).
- Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York.
- Hogg, S., 1996. A review of the validity and variability of the elevated plus-maze as an animal model of anxiety. *Pharmacology Biochemistry and Behavior* 54, 21–30.
- Holmes, A., Parmigiani, S., Ferrari, P., Palanza, P., Rodgers, R., 2000. Behavioral profile of wild mice in the elevated plus-maze test for anxiety. *Physiology & Behavior* 71, 509–516.
- Kattan, M., 2006. Comparison of cox regression with other methods for determining prediction models nomograms. *Journal of Urology* 160, S6–S10.
- Keedwell, E., Narayanan, A., 2005. Intelligent Bioinformatics: The Application of Artificial Intelligence Techniques to Bioinformatics Problems. Wiley & Sons, Sussex.
- Knight, A., 2011. The Costs and Benefits of Animal Experiments. Palgrave Macmillan, New York.
- Lister, R., 1987. The use of a plus-maze to measure anxiety in the mouse. *Psychopharmacology* 92, 180–185.
- Miranda, D., Conde, C., Celis, C., Corzo, P., 2009. Modelado del Comportamiento de Ratas en Laberinto en Cruz Elevado Basado en Redes Neuronales Artificiales. *Revista Colombiana de Física* 41, 406–408.
- Montebelli, A., Herrera, C., Ziemke, T., 2008. On cognition as dynamical coupling: an analysis of behavioral attractor dynamics. *Adaptive Behavior* 16, 182–195.
- Montgomery, K., 1955. The relation between fear induced by novel stimulation and exploratory behavior. *Journal of Comparative and Physiological Psychology* 48, 254–260.
- Pellow, S., Chopin, P., File, S., Briley, M., 1985. Validation of open: closed arm entries in an elevated plus-maze as a measure of anxiety in the rat. *Journal of Neuroscience Methods* 14, 149–167.
- Poucet, B., Save, E., 2005. Attractors in memory. *Science* 308, 779–780.
- Rico, J., Castilho, V., Morato, S., Nobre, M., 2009. Previous spatial learning is required for the development of place preference in rats following chronic administration of and withdrawal from morphine. *Psychology & Neuroscience* 2, 59–65.
- Scalesse, R., Issenberg, B., 2005. Effective use of simulations for the teaching and acquisition of veterinary professional and clinical skills. *Journal of Veterinary Medical Education* 32, 461–467.
- Salum, C., Roque, A., Morato, S., 2000. Anxiety-like behavior in rats: a computational model. *Neural Networks* 13, 21–29.
- Salum, C., Roque, A., Morato, S., 2003. Conflict as a determinant of rat behavior in three types of elevated plus-maze. *Behavioural Processes* 63, 87–93.
- Thomas, S., 2009. Physiologically-based simulation modeling for the reduction of animal use in the discovery of novel pharmaceuticals. *Alternatives to Laboratory Animals* 37, 497–511.
- Tejada, J., Bosco, G., Morato, S., Roque, A., 2010. Characterization of the rat exploratory behavior in the elevated plus-maze with Markov chains. *Journal of Neuroscience Methods* 193, 288–295.
- Tejada, J., Bosco, G., Morato, S., Roque, A., 2009. Characterization of rat behavior in the elevated plus-maze using a directed graph. *Journal of Neuroscience Methods* 184, 251–255.
- Timofeev, R., 2004. Classification and Regressions Trees: Theory and Applications. Humboldt University, Berlin, Master's Thesis (accessed 1.06.2012).
- Watkins, M., Pacheco, M., 2001. Interobserver agreement in behavioral research: importance and calculation. *Journal of Behavioral Education* 10, 205–212.