Chapter from the book *Self Organizing Maps - Applications and Novel Algorithm Design*
Downloaded from: http://www.intechopen.com/books/self-organizing-maps-applications-and-novel-algorithm-design

Interested in publishing with IntechOpen?
Contact us at book.department@intechopen.com
Computational Approaches as a Tool to Study Developmental Biology in New World Primates

Maria Bernardete Cordeiro de Sousa, Allan Medeiros, Dijenaide Chaves de Castro, Adriano de Castro Leão and Adrião Duarte Dória Neto

Universidade Federal do Rio Grande do Norte, Brazil

1. Introduction

Kohonen neural network or SOM (Self-Organizing Maps) and k-means are effective computational techniques for extracting information from a large dataset, in some cases, more suitable than traditional statistical methods (such as regression) for detecting multivariate data patterns. We used these two computational approaches to study the developmental biology of a large dataset of weight versus age in common marmosets, seeking to provide a more precise description of the onset and end of behaviors during ontogenesis. Both techniques perform clustering, a technique that falls into a group of undirected data mining tools. The goal of undirected data mining is to discover structure in the data as a whole, separating or discovering classes or groups that are not easily detected from direct data inspection. Both k-means algorithms and self-organizing maps are clustering methods that classify patterns without the need of previous information on data distribution, as occurs in supervised methods. When used in feature datasets (for instance, age and weight), these techniques allow a data mining process that identifies stages and substages of the data distribution being used. In particular, the SOM algorithm is able to map features from a high-dimensional space into a one- or two- (in fact any dimension less than the dimension of the data) dimensional space with preservation of topology. Thus, these tools are better for modeling and identifying stages in a sample pool and they enable us to visualize data distribution in a two-dimensional space, even when the number of features in the data is high (or the topology is complex). For a same database we found that both approaches converged to a same result, enabling the process of classification, thus achieving the goal of this study. Biomedical research using the common marmoset (Callithrix jacchus) continues to grow, through the investigation of different topics such as infectious diseases, neuroscience and development, toxicology and drug development, reproductive biology and behavior. Using this classification we found that the infantile stage of the current classification could be divided into three stages and that feeding behavior is statistically different among them. We also applied a statistical method, Principal Component Analysis (PCA) to evidence the sex difference. Using data collected from 9 animals (4 males and 5 females) during development evidenced a sex difference, where male behavior features are more dispersing than those of females, indicating that behavior is more changeable in males. These findings indicate that the use of clustering techniques provides a more precise description of changes during ontogenesis in common marmosets.
in addition to being useful in refining the follow-up of the ontogenetic development of common marmosets in terms of behavioral and physiological variables.

2. Computational tools and the study of animal behavior

There is currently great interest in computational intelligence and statistical methods applied to the study of animal behavior. These groups of techniques are very useful and powerful in analyzing natural behavior, simulating physiological models and measuring the complex sequential or isolated neural circuits of motor activities, vocalization or sensory processing among others.

According to Enquist & Ghirlanda (2005), the use of artificial neural networks (ANN) began as an effort to understand the nervous system and behavior based on the fundamental arrangement of the cytoarchitectural components of a system formed by a complex network of neurons. Moreover, this knowledge is multidisciplinary, encompassing mathematics, theories of intelligent machines and philosophy. Although artificial neural networks have provided evidence that this model is a powerful computational tool, applicable to a wide range of phenomena, Harris-Warrick & Marder, (1991), seeking to develop a model for an invertebrate brain, observed that high computational processing is necessary, even for these simple animals. However, the overwhelmingly promising findings compensate for these effects, and today many laboratories, mostly engineering, are working in this area. Enquist & Ghirlanda, (2005) also pointed out that students of animal behavior should have embraced the study of neural network models, but unfortunately, this did not occur, since traditional ethological and animal psychological thinking does not include a discipline related to this matter. Another important question involved with the use of integrative neural networks in neuroscience is the need to make a critical comparison of available theoretical approaches (Reeke and Sporns, 1993), considering their many applications. The contribution of neurophysiology to the development of artificial neural networks is very important and many situations, such as the demonstration of visual receptive fields by Hubbell & Wiesel (1962), the invertebrate models of learning developed by Kandel & Pittenger (1999) and the recent study by Tin et al., (2008) where the authors developed a Bayesian linear regression algorithm that confirms recent findings on the organization of the motor cortex, providing an incremental real-time version for real-time interfaces between brains and machines.

The use of statistical techniques to analyze animal behavior has many applications in a wide range of biological activities, from molecular mechanisms to complex animal behavior. Very recent studies using black lemurs (*Eulemur macaco*: Pozzi et al., 2010) showed that ANN’s describe all seven vocal categories and are more suitable than other statistical analysis methods, extracting from both temporal and spectral characteristics. Matthews (2009) studied the socially learned traditions white-fronted capuchin monkeys in the field (*Cebus albifrons*) by combining variations in multiple foraging techniques into pairwise behavioral similarity matrices. He found that cluster analysis makes it possible to predict specific features from the social learning hypothesis.

Animals living under natural conditions have to deal with social and environmental changes and unpredictable situations that trigger behavioral adjustments. A number of ethologists and comparative psychologist have studied animal behavior during adult life, when animals exhibit a more stable pattern of responses and the quantification of these variables is more predictable. During ontogenesis, i.e. during the development of individuals.
during their lifetime, a detailed approach is necessary to provide a more accurate picture of developing features, the onset and end of transient behaviors such as weaning and being carried, as well as the start and stabilization of most adult behaviors. However, even when the behavioral repertoire is fully developed, changes will occur owing to new challenges demanding requirements, becoming the expression of behavior that is flexible enough to allow survival in hostile environmental conditions. Thus, in both cases, in adult life and, mainly during ontogeny, individual behaviors also change with experience and when animals are faced with unpredictable situations.

Common marmosets (*Callithrix jaccus* - Linnaeus, 1758) are small-bodied non-human primates, endemic to Brazil with an adult weight of around 320 g. They are considered omnivores and in the last 10-15 years have extended their habitats beyond secondary forests to exotic species plantations, peridomiciliary areas and public parks. They live in free-ranging groups composed of between 5-17 different-aged individuals, including adult males and females (Rylands, 2006; Araujo, 1996). They exhibit a flexible mating system, depending on ecological constraints, showing that common marmosets are preferentially monogamous but on some occasions a polygynous system may arise (Arruda, 2005; Sousa et al., 2005).

Both mature and immature males and females emigrate from the natal group to build new social units, but this behavior is more frequent in females (Sousa et al., 2009). Subordinates are expelled by breeding (dominant) females or leave the group voluntarily showing or not functional ovulatory cycles (Albuquerque et al., 2001; Sousa et al., 2009; Yamamoto et al., 2010). Studies on social dynamics within free-ranging common marmoset groups have shown that breeding males and females form a nuclear structure (Koening and Rothe, 1991; Araujo, 1996) surrounded by satellite non-breeders that are inspected in terms of reproductive condition. Breeding pairs try to maintain reproductive exclusivity using reproductive inhibition by means of aggressive behavioral mechanisms that limit the mate’s copulation opportunities with other individuals (Araujo & Yamamoto, 1993; Yamamoto et al., 2010). Aggression is frequent when the home ranges of neighboring free-ranging groups overlap. Males and females become more aware, increasing scent marking, vocalizations and vigilant behaviors near territorial borders. On these occasions, competition is intense and sexual contact between males and females and extra-group individuals is frequently recorded (Lazaro-Pere et al., 2000; Lazaro-Perea, 2001). Dimorphic behavior in captive common marmoset males and females has been recorded in relation to infant carrying behavior, where males carried more during the first weeks of development, and in the physiological characteristics of females that may become pregnant again in the first weeks of the postpartum period (Kendrick & Dixon, 1986). Other studies using captive marmosets described additional sexual dimorphism in physiological variables, with females showing three times higher cortisol levels (Johnson et al., 1996; Raminelli et al., 2001) and higher aggressive behavior toward same sex individuals (Galvão-Coelho et al., 2008) than males.

The common marmoset is a very useful model in stress-induced situations. Galvão-Coelho et al., (2008) recently demonstrated that common marmosets exhibit behavioral changes during challenging situations. In adult subjects, basal cortisol profiles varied in a population (n= 48; 24 males), allowing their use as a model for studying mental disorders associated with high and low responsiveness of the hypothalamic-pituitary-adrenal (HPA) axis. Since the social behavior of common marmosets has been extensively studied in both situations (free-ranging and captive groups), computational tools seem to be a promising...
path to the discovery of new components or features of behavior that may provide fresh insights into the sociobiology of this species. Clustering tools such as k-means and principal component analysis techniques are alternatives to using artificial neural networks. In this chapter we present the results obtained regarding the effectiveness of k-means applied to the analysis of weight to characterize the infantile stage, corresponding to a period of rapid ontogenetic development in mammals. We also show the usefulness of PCA in identifying sex differences in the pooled data of individual and social behaviors expressed by common marmosets during development.

3. Is weight a good measure to characterize the early developmental stages of common marmosets?

Leão et al. (2009) have demonstrated that classifying the lifetime of common marmosets (*Callithrix jacchus*) using SOM and k-means techniques in experimental protocols where the variables “weight” and “age” are critical could generate new possibilities for refining data analysis during developmental stages previously characterized by behavioral recordings. The authors employed clustering techniques, where data mining is used to discover structure in the data as a whole, and substages in the infantile and juvenile stages described by Yamamoto, 1993. Both k-means algorithm (Duda et al., 2000) and self-organizing maps (Kohonen, 1982) are clustering methods that classify patterns without the need of previous information on data distribution, in contrast supervised methods. These techniques allow a data mining process that identifies stages and substages of data distribution. These tools are better for modeling and identifying stages in a sample pool and they enable us to visualize data distribution in a two- dimensional space, even when the data form high- dimensional spaces, i.e. when the input vector is high (Leão et al., 2009). Using these techniques, no target variable is required and, therefore, no distinction is made between independent and dependent variables. Thus, they are useful tools for identifying and classifying patterns immersed in a mass of data.

Figure 1 illustrates the distribution curve pattern for weight based in the Leão et al., (2009) dataset (n= 9,200 weight entries) and the 4 developmental stages used by Yamamoto (2003) (upper) and the curve pattern generated by using k-means to the data (below). The animals used by Leão et al., (2009) were living in outdoor cages at the Núcleo de Primatologia of the Federal University of Rio Grande do Norte. A database with recorded weights across the ontogenetic development of captive *Callithrix jacchus* born between 1985 and 2003 was used. The reference data used the age in days of the animals, as proposed by Yamamoto (1993), where infant class = 0-150 days; juvenile = 151 to 300 days; subadult = 301-600 and adult > 600 days. We also present in Figure 1 the application of k-means to the other classifications described in Ingram (1977), and Abbott et al. (2003) and Stevenson & Rylands (1988) using behavioral and physiological variables. Cluster techniques show that developmental stages cannot be divided based exclusively on weight information. Comparing the results with the behavioral classification that was used by the different authors, we see that clustering based only on weight generally tends to spread age classification further than conventional behavioral classification. For instance, an infantile stage based on a range of weights could be confounded with an older animal, suggesting that weight is not directly (linearly) related to age stage classification. Even though the animal’s behavior stabilizes in adulthood, ceasing to gain weight between 15 and 20 months, there is still enough time (in the lifespan of the animal) to allow the division of stages according to weight, i.e. weight no longer characterizes behavioral development.
Ingram and Stevenson and Rylands describe substages within the infantile stage. Using the number of infantile substages according to Ingram (n=3) and Stevenson & Rylands (n=7), characterized by behavioral and physiological data, k-means reveals that 3 different substages might be considered within the first 4 months, since statistical analysis of the 3 centroids shows significance among them. Furthermore, 2 substages were also found within the juvenile stage. Thus, according to Leão et al., (2009) using the 4 age stages proposed by Yamamoto (1993) resulted in 5 substages as follows: Infantile stage: 3 substages - infantile I (0 to 1 month) with mean weight of 40g (SD ± 11.79g); infantile II (>1 to 3 months) with mean weight of 98g (SD ± 26g) and infantile III (>3 to 4 months) with mean weight of 165g (SD ± 20g) and Juvenile stage: 2 substages – juvenile I (>5 to 7 months), with mean weight of 197.68g (SD ± 36g) and juvenile II (>7 to 10 months), with mean weight of 255.43g (SD ± 34g). These findings suggest that increased weight in early development follows behavioral progress in common marmosets. To explore this possibility we started to behaviorally characterize the infantile stage in common marmosets fitted to these 3 new infantile period substages.

Fig. 1. Comparison plot of age classification with k-means using the stages proposed by Yamamoto, 1993 (a); Ingram, 1977 (b); Abbott et al., 2003 (c), and Stevenson & Rylands, 1988 (d). Colors indicate the infantile to adult stages. Age limit values are shown in Table 2.
Table 1. Age limits (in months) to classify stages according to different classifications available in the literature and k-means clustering.

**Feeding behavior during the 3 infantile substages of captive common marmosets**

We monitored 9 common marmosets from birth to the young adult stage (16 months) to characterize the behavioral correlates of the classification proposed by Leão et al., (2009). We followed 9 captive common marmosets (5 females and 4 males) from the first to the 16th week of life. Data sampling occurred twice a week for 30 min, in the morning and afternoon, totaling 4 sessions per week in the infantile stage (0-4 months). According to this classification, in which the infant stage was divided into 3 substages, weaning takes place at the end of the infantile II stage (twelve weeks) and the first bouts of solid food ingestion for marmosets were recorded in the infantile I substage (Table 2). We also found significant
statistical differences between food episode frequency across the three stages (ANOVA: infantile I x infantile II p< 0.001; infantile II x infantile III, p = 0.022; infantile II x infantile III, p< 0.001) (see Figure 2), suggesting that the infantile stage must be investigated as, at least, a tripartite phase, since behavioral changes occur rapidly and the onset of most individual and social behaviors occurs at this stage. These results suggest that this classification provides more detailed information about the surge of behavioral development, providing new information for the management and care of these small primates in early stages of development.

Table 2. Developmental stages in common marmosets. Different colours represent the three successive substages (I, II and III) of the infancy phase.

4. Data analysis of the ontogenetic development of common marmosets using principal component analysis

The main goal in data analysis is to represent the information contained in the data. In most cases, the data consist of a table in which the rows are measurements of some process (the information source) and the columns are the attributes that are being measured (information container).
To perform this rotation, one must find the proper rotation angles (in several dimensions). PCA finds a new set of coordinates that correspond to the eigenvectors of the covariance matrix of the data, which can be computed as

$$\Sigma = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T$$

Where $x_i$ is the vector representing the features in the $i$’th column ($i$’th measure) and $\mu$ is the mean of the entire table, computed as

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

In both cases, $N$ is the number of points (or measures). The superscript $T$ stands for matrix transpose. The entire table can be represented as matrix $X$ as follows:

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix}$$

PCA can now be performed by projecting (or simply multiplying) the whole dataset by the transformation $U$ given by the eigenvectors of $\Sigma$ (each eigenvector of $\Sigma$ is a column of $U$).

$$Y = XU$$

Each column in the new table $Y$ corresponds to the new feature. Since the transformation matrix $U$ is formed by the eigenvectors of $\Sigma$, each column corresponds to an eigenvalue. Matrix $\Sigma$ is symmetric (by formation), implying that the eigenvalues are all real and positive values. It turns out that the size of the eigenvalue corresponds to the variance of the respective column in the new dataset. Therefore, the larger the eigenvalue the larger the amount of information contained in the new column. The columns where the eigenvalues are small can now be discarded.

**Sex signature using behavioral development data of common marmosets using PCA applied to the dataset**

To perform PCA analysis the data were collected from 9 common marmosets ($n=5$ females) from birth to young adulthood (16 months) as previously described. All the frequencies of social and solitary behaviors measured, classified as feeding (solid ingestion of food), affiliation (allogrooming), agonistic (scent-marking and piloerection) and individual (autogrooming) behaviors were typed into a template and Principal Component Analysis (PCA) was applied. One of the goals of applying PCA to this dataset is to perform feature extraction. We performed PCA in the dataset without the “sex” feature (removing the sex column) using only behavioral variables. Thus, only the first two columns of the
transformed data (two largest eigenvalues) were used. We then plotted the new features for each individual. Figure 2 illustrates that behavior features of common marmoset males (plots on the left) after PCA are more dispersing than females (plots on the right), suggesting that the variance in behavior in the original data is greater in males. Thus, analysis showed that computational tools exhibit behavioral differences between sexes during development.

For instance, males are more involved with territorial defense than females and adult males are the main caregivers, providing both transportation and sharing food with the young animals (Yamamoto et al., 2010). Females show more competitive behavior (agonistic and physiological inhibition) to other same-sex conspecifics (Abbott, 1984; Abbott et al., 1993; Araujo & Yamamoto, 1993; Alencar et al., 2006). Basal cortisol is higher in adult females than in males (Johnson et al., 1996; Raminelli et al., 2001) and a positive correlation is recorded between cortisol and both agonistic and anxiety behaviors when nonrelated females are living in same-sex dyads (Galvão-Coelho et al, 2008). With respect to the reproductive strategies of both sexes, it is considered that females compete and males cooperate to achieve reproductive success. As suggested by Yamamoto et al., (2010) the success of the common marmosets’ social system seems to depend on an equilibrium between both strategies.
Fig. 2. Plots for the first two features extracted with PCA from behavioral data of common marmosets (4 males and 5 females). First column shows plots of male subjects and the second column shows the same features for females. The gray circles are the features for the entire dataset and the red dots are the features for each animal (each one in a different plot).

Although the use of PCA on behavioral data of common marmosets during development did not identify which type of behavior differs between males and females using sociobiological and ecological approaches, it is known that common marmoset males and
females diverge in their strategies, as previously demonstrated. Further studies might reveal other behavioral differences between males and females in post hoc analysis.

5. Future directions

Given that the behavioral repertoire in developing common marmosets was recorded across the juvenile, subadult and adult phases, it would be very interesting to use computational tools to describe the subtle changes occurring in juvenile I and juvenile II substages. For instance, preliminary data pointed out that solitary and both types of social play - between the twins and between the focal animal and the older siblings - are very intense in the juvenile II substage. The association between behavioral measurement and hormone cortisol dosage and sexual steroids, mainly during the juvenile stage, will provide more detailed information on the endocrine modulation of the behaviors, since the puberty mechanisms are triggered in primates at this time.

Another very interesting point to be analyzed using cluster techniques might be the behavioral data obtained by continuous recording during the entire light phase and circadian rhythmicity could be demonstrated for the behaviors we are already measured and by adding other behavioral categories such as moving and stationary behaviors. Moreover, characterizing the growing stages of common marmosets using motor performance and sensory skill tests at different ages will allow developmental researchers to be aware of opportunities for studying critical periods for the acquisition and maturation of motor and sensory competencies.

6. References


Kohonen Self Organizing Maps (SOM) has found application in practical all fields, especially those which tend to handle high dimensional data. SOM can be used for the clustering of genes in the medical field, the study of multi-media and web based contents and in the transportation industry, just to name a few. Apart from the aforementioned areas this book also covers the study of complex data found in meteorological and remotely sensed images acquired using satellite sensing. Data management and envelopment analysis has also been covered. The application of SOM in mechanical and manufacturing engineering forms another important area of this book. The final section of this book, addresses the design and application of novel variants of SOM algorithms.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: