

An Adaptive Maximum Entropy Approach for Modeling of Species Distribution (Janeiro 14, 2010)

E. S. C. Rodrigues, F. A. Rodrigues, R. L. A. Rocha, P. L. P. Corrêa

Abstract – One of the most used methods in modeling biological species geographical distribution is the Maximum Entropy Algorithm. This paper presents an Adaptive Maximum Entropy (AME) approach for modeling biological species, which aims to reduce the amount of features used in the construction of models. The approach presented here is an alternative to the classical algorithm. Instead of using the same set of features at each iteration, the AME approach tries to find a new set of features that converge faster. The preliminary experiments were well performed. However, statistical evaluations and comparisons with other algorithms are beyond the scope of this paper. These needed studies are proposed as future works.

Index Terms – Adaptive systems, Maximum Entropy methods, biological system modeling.

I. INTRODUCTION

The world is suffering the consequences of the lack of environmental conservation. Several environmental problems need attention and efficient strategies for its solution. The modeling of environmental systems can assist in the decision making process to solve these problems. There are several Artificial Intelligence (AI) techniques that are being used for modeling environmental systems. AI techniques have large potential to solve inference complex problems because they try to simulate some human characteristics, such as reasoning and learning.

Some examples of AI techniques used for modeling environmental systems are: (a) Case-Based Reasoning applied to forest fire management [1], air quality prediction of urban areas [2], reduction of environmental impact caused by chemical process [3] and cyclone forecasting [4]; (b) the Rule-Based Systems can be employed in pests diagnostic [5], [6], [7] and diseases [8]; (c) Artificial Neural Networks can be applied to water resources management [9], water quality prediction [10], [11], ozone concentrations prediction [12] and modeling of biological species geographical distribution [13], [14]; (d) Genetic Algorithms are a very known and used technique in modeling biological species geographical distribution [15], but it can also be applied in predictions of air quality [16], water quality models calibration [17] and water

management in irrigated agriculture [18]; (e) Cellular Automata can be applied to species migration modeling [19], landscape modeling [20] and modeling of seismic activity [21]; (f) Fuzzy systems can be employed in water management [22], soil erosion prediction [23] and forest fire risk estimation [24]; (g) Multi-Agents Systems has been applied in natural resource management [25], [26] and forest management [27]. Other techniques such as Swarm Intelligence, Reinforcement Learning and hybrid systems have also been used in environmental systems modeling. In [28] there is an overview about these AI techniques with additional references.

Since Brazil is the country with the richest flora and fauna on Earth, about one sixth of the total, the motivation and support to researches directed to natural resources management and sustainable development are very important. The modeling of biological species geographical distribution can assist decision-making processes, planning and accomplishing actions aiming at environmental conservation.

There are some free tools available for modeling biological species geographical distribution and several algorithms with different approaches were already implemented for this aim. OpenModeller is a framework for modeling biological species geographical distribution with several resources [29]. One of the algorithms available in openModeller is a Maximum Entropy (MaxEnt) algorithm. This is one of the most used algorithms by researchers of biological species modeling.

The MaxEnt algorithm implemented in openModeller chooses a feature from a set at each iteration and updates its parameters. This choice is based on a minimization function with all features. Thus, features influence the algorithm performance, that is, inserting or removing features can improve the parameters fitting and convergence can be reached faster. However, in the algorithm implemented in openModeller, the set of features used along the learning process is static. The approach proposed here inserts features in the set or removes features from it along the learning algorithm. Adaptive Devices were used here to represent this dynamic characteristic. Adaptive Devices can modify their own structure without external interferences. The only algorithm available in openModeller tool that has an adaptive version is GARP [30].

An Adaptive Maximum Entropy (AME) approach for modeling biological species geographical distribution is presented, which aims to reduce the amount of features used in the construction of models. The AME approach was based on the developed MaxEnt algorithm available in openModeller. The proposed approach uses an ambitious strategy because it

Thanks to CAPES for support the first two authors.

E. S. C. Rodrigues and F. A. Rodrigues are students of the Doctoral Program in Electrical Engineering of Polytechnic School of the University of São Paulo (e-mail: elisangela.rodrigues@poli.usp.br; fabricao.rodrigues@poli.usp.br).

R. L. A. Rocha and P. L. P. Corrêa are professors of Department of Computer and Digital Systems Engineering of Polytechnic School of the University of São Paulo (e-mail: luiz.rocha@poli.usp.br; pedro.correa@poli.usp.br).

tries to find out a set of features that seems to be the best at each iteration instead of using the same set of features all along the model learning.

Section II describes the modeling process of biological species geographical distribution as well as the openModeller tool. Section III presents the main idea of adaptive devices. Section IV describes the Maximum Entropy Principle and its application to modeling biological species. Section V presents the Adaptive Maximum Entropy approach. The methodology used in the experiments is presented in Section VI, as well as the results obtained. Final discussion and proposals for future works are presented in Section VII.

II. MODELING BIOLOGICAL SPECIES

The amount of available data about biological species is increasing and it is becoming largely disseminated in the World Wide Web. In [32] there are several websites addresses that supply data on species distribution. Modeling tools can process these data, for example, and the results can generate a lot of information to assist environmental conservation planning.

The set of ecological conditions necessary for a species to keep populations is known as species ecological niche [33]. This is the main concept related to modeling of biological species geographical distribution. The fundamental ecological niche is the set of all conditions that allow the survival of species for a long period of time. A niche-based model, produced by a modeling tool, is an approximation of the species fundamental ecological niche.

A modeling tool uses two kinds of data to produce a niche-based model: occurrence data and environmental data. The occurrence data are georeferenced points – latitude and longitude – recorded where the species were observed. These occurrence records are also called presence points. The species occurrences are determined by the environmental conditions in the region where it occurs. Sometimes, there are records of the species absence, indicating its inexistence in a given region. However, absence data are rarely available [34]. Environmental data are also known as environmental layers and they represent the species ecological niche [35]. All environmental layers should be in the same geographic area and they are georeferenced, too [34]. Some examples of environmental layers commonly used in modeling are temperature and precipitation.

The main purpose of a modeling tool is to find a probability function that represents the relation between the suitable environmental conditions for the species and the given environmental layers [34]. Some modeling tools are freely available on the web, such as DesktopGarp [36], MaxEnt [37] and openModeller [27], [38], in which this work is inserted. In these tools, several algorithms have been applied to modeling biological species, such as GARP (Genetic Algorithm for Rule-set Production) [39], [40], [41], [15], Maximum Entropy [34], [42], SVMs (Support Vector Machines) [43], [44], Neural Networks [14], [45] and others [46].

[1] *openModeller*

The openModeller is a framework developed to support all the modeling process. It offers several functionalities, such as search and preparation of data, pre-analysis modules, modeling algorithms and visualization of results [29]. OpenModeller is an open source tool, written in C++, which runs in different platforms and provides several modeling algorithms [38].

Each algorithm in the openModeller tool has specific input parameters that can be changed by users. Although a little knowledge about these parameters is desirable, all algorithms have default input parameters that were widely tested. The input data are the same for all algorithms: an occurrence data set and an environmental data set. It allows the user to make different experiments with a variety of algorithms but with the same data, making the result analysis and the comparison among them easier.

Figure 1 shows the modeling process used by openModeller. The niche points are environmental layer values at each georeferenced point where the species were recorded. Thus, the occurrence points are transformed into points in the environmental space. The algorithm receives the niche points as input data and, after processing, gives a probability function that maps the environmental suitability for the species to a domain in the environmental layers space. The probability function given as output by the algorithm represents a niche-based model. The model generated is projected in a geographical area, producing a georeferenced map with the probability function of the species. All areas in the map satisfying the environmental conditions of its fundamental niche represent the potential distribution of a species.

In a modeling with the openModeller tool, the user must specify the input occurrence data, the input environmental layers set and the algorithm(s) that will be used. The occurrence data must be in a file containing the record identifier, the species name, the coordinates (longitude and latitude) where the species was observed and the abundance, that assume value 1 when the record is a presence point and 0 when the record is an absence point. Each georeferenced point must be in a new line and a tab separates the fields. Each environmental layer in a GIS format (standardized codification of geographical information) must be stored in a different file.

Besides that, the user must choose the modeling algorithm and set its parameters or keep the default one. There are several help appliances and ways to use XML configuration files [38]. One of the most used algorithms in the environmental modeling community is the Maximum Entropy-based one. This was one of the motivations for the study of a new approach of a MaxEnt method. Another motivation for the development of this work was that both Adaptive Devices and MaxEnt methods tend to be largely applied to modeling and AI problems.

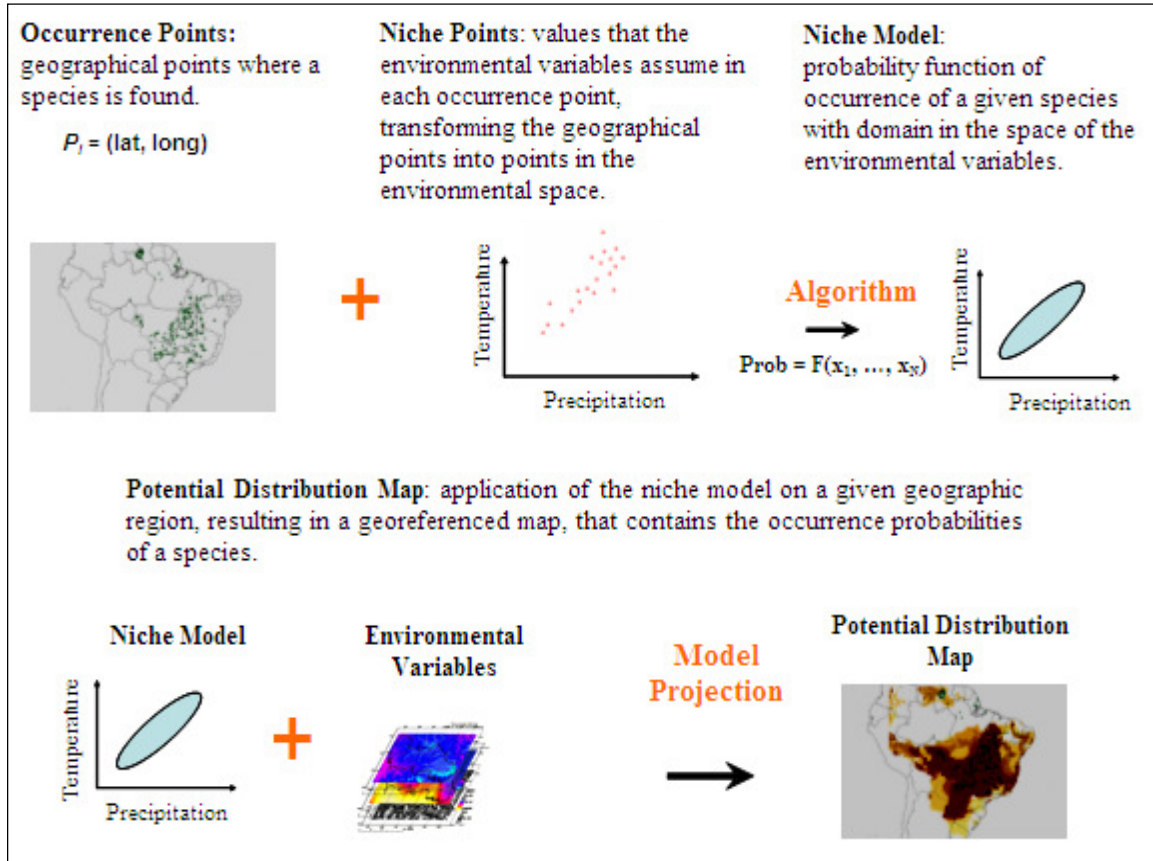


Fig. 1. Modeling Process used in openModeller (adapted from [35])

III. ADAPTIVE DEVICES

Adaptivity is the capacity that a system has to modify its own structure without external interferences [47]. Adaptive Devices are abstract descriptions of problems that have dynamic behavior. These descriptions are associated with non-adaptive subjacent devices that represent problems with static behavior [48].

Non-adaptive devices have their behavior defined by a static set of rules. A non-adaptive subjacent device is improved by the addition of a set of adaptive actions. These actions characterize the operations needed for making the system behavior adaptive [47].

Any system that has its behavior defined by a set of rules and that has dynamic behavior can use adaptive devices as its abstract description. There are several adaptive formalisms that can be used as descriptions, such as adaptive automata [49], [50], [51], grammars, state charts, Markov chain, decision tables and decision trees [47].

Adaptive actions allow alterations in the set of rules that defines the system. There are three elementary actions in adaptive devices: searching, erasing and inserting [52]. Searching actions try to find a rule according to some pattern. These actions do not modify the set of rules. Erasing actions remove from the set of rules all rules matching a given pattern. Inserting actions add rules with a given pattern in the set of rules. Adaptive actions can be executed before or after the application of the underlying non-adaptive rule that they are

associated.

In AME approach proposed here, the underlying device is the structure of the method. The adaptive actions will modify the number of environmental variables used along the learning process.

IV. MAXIMUM ENTROPY PRINCIPLE

The Maximum Entropy principle originates from statistical mechanics [53], [54] and has been successfully applied in several research areas, such as Natural Language Processing [55] and modeling of biological species geographical distribution [34], [42]. However, it is possible to produce better results and to find solutions using Adaptive Devices.

The main idea of the MaxEnt principle is: from a probability distribution set that satisfies some constraints, to find the one that has maximum entropy. This principle can be considered a constrained optimization problem, that is, the aim is to find a solution maximizing or minimizing a function.

Entropy (1) is a very important concept in Information Theory. It measures the amount of uncertainty associated to the possible states of an event. The amount of information of an event is inversely proportional to its probability of occurrence. The entropy is defined as [34]

$$H(p) = - \sum_{k=1}^N p_k \log p_k \quad (1)$$

where p is the probability distribution over the set of possible states of an event, N is the total number of possible states of the event and p_k is the occurrence probability of the k -th state.

A. Maximum Entropy to Modeling of Species Distribution

The main advantage of applying Maximum Entropy to modeling of biological species geographical distribution in comparison with other methods is that it needs just presence data, besides the environmental layers. Furthermore, it is possible to use both categorical and continuous layers [34].

In modeling biological species geographical distribution, suppose that the finite set of pixels representing the area of interest is X . The set of points x_1, \dots, x_m pertaining to X represents the presence points of a species. The aim is to find a probability distribution p^* that approximates p , the potential distribution of the species.

The environmental layers are treated as *features*, that is, a set of functions f_1, \dots, f_n , so that $f_j: X \rightarrow \mathbb{R}$. These features can also be functions derived from the environmental layers, such as the square of a continuous environmental layer or a product of two continuous environmental layers. Thus, each feature sets a real value $f_j(x)$ to each point in X . The constraints are that the expectations of each feature matches its empirical average, denoted by $\tilde{p}[f_j]$, where

$$\tilde{p}[f_j] = \frac{1}{m} \sum_{i=1}^m f_j(x_i). \quad (2)$$

The MaxEnt probability distribution can be proved to be equivalent to the Gibbs distribution, that is, an exponential distribution with a vector of feature weights that parameterizes it [56]. This probability distribution is defined as

$$q_\lambda(x_i) = \frac{\exp(\sum_{j=1}^n \lambda_j f_j(x_i))}{Z_\lambda} \quad (3)$$

where λ is a vector of the feature weights, with real values, and Z_λ is a normalizing constant that guarantees that the probability distribution sums to one over the area of interest. The MaxEnt probability distribution is also equivalent to minimize the log loss, that is the negative log likelihood [34], [42], [56]. The log loss is

$$-\lambda_j \cdot \tilde{p}[f_j] + \log(Z_\lambda) \quad (4)$$

Thus, both Gibbs distribution and log loss are used as objective functions.

B. Maximum Entropy in the openModeller

There are several algorithms to estimate the Maximum Entropy parameters, such as Generalized Iterative Scaling, Improved Iterative Scaling and limited memory variable metric [57]. The estimating algorithm available in the openModeller tool is similar to the sequential algorithm used in the MaxEnt program [34], [42]. It was proved to converge to Maximum Entropy probability distribution in [58].

This algorithm is called sequential because it chooses one feature at each iteration and adjusts its parameters. This procedure is executed until either the convergence is reached or the number of iterations is reached. This estimating method was implemented in openModeller because it has shown to be suitable for modeling biological species geographical distribution [34], [42]. However, there is a similar estimating algorithm that updates all feature weights at each iteration [58]. Figure 2 shows the high-level algorithm to estimate the Maximum Entropy parameters implemented in openModeller. The update consider

$$\delta = \log \frac{(\tilde{p}[f_j](1 - q_\lambda[f_j]))}{((1 - \tilde{p}[f_j])q_\lambda[f_j])} \quad (5)$$

and, in the second *for*, j is the feature's label that will be updated and j' is the current label.

The algorithm has four input parameters: number of iterations, number of background pixels and convergence. The default values of each parameter defined empirically are: number of iterations = 500, number of background pixels = 10,000 and convergence = 10^{-5} . Background pixels are georeferenced points that are used to delimit the area of study, but these points are not interpreted as pseudo-absences, as in other techniques [34]. Currently, only the linear feature is implemented, that is, the raw environmental layer values.

V. ADAPTIVE MAXIMUM ENTROPY APPROACH

The underlying device used here is the structure of the method presented in Figure 2. The adaptive actions do not change the program code but they change the method structure.

Input: feature functions f_1, \dots, f_n
 occurrence points x_1, \dots, x_2

Output: vector of feature weights, λ .

for $j = 1$ **to** n
 $\lambda_j = 1$;
for $k = 1, 2, \dots$
 1. Let $(j, \delta) = \arg \min_{(j, \delta)} F_j(\lambda_{k,j}, \delta)$, where
 $F_j(\lambda_{k,j}, \delta)$ is the log loss (4) and
 δ is the expression in (5);
 2. if $(j = j')$
 $\lambda_{k+1,j'} = \lambda_{k+1,j} + \delta$;
 else
 $\lambda_{k+1,j'} = \lambda_{k+1,j'}$

Fig. 2. High-level algorithm to estimate MaxEnt parameters (adapted from [58]).

Figure 3 presents the general view of the training procedure. This procedure can be applied to any maximum entropy algorithm because the changes occur in the number of features used at each iteration.

The algorithm begins with all features chosen by the user

and all feature weights are set to 1. Instead of just choosing the best feature to adjust its parameters, the adaptive approach searches for the best set of features based on the log loss of each possible set.

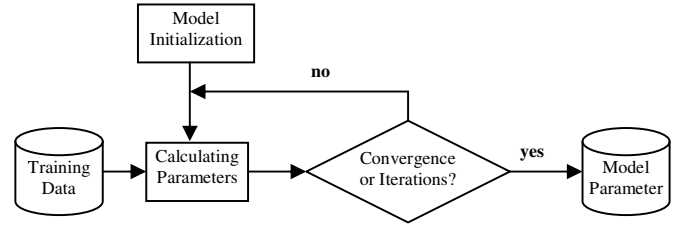


Fig. 3. General view of the training procedure.

Figure 4 shows how the procedure chooses the best set of features at each iteration. For all possible subset of features, F_j is calculated and the minimum value is chosen to update the respective parameter. At each moment of time, there is a set of feature weights that produces a probability distribution. The log loss of every subset of features is calculated and the subset with the smallest value is chosen to be the new set of features. This procedure is repeated until the convergence or the number of iterations is reached. Thus, the features can be inserted or removed from the set according to the log loss. This variable set of features characterizes the dynamic behavior of the task, which tries to find the probability distribution with maximum entropy as soon as possible.

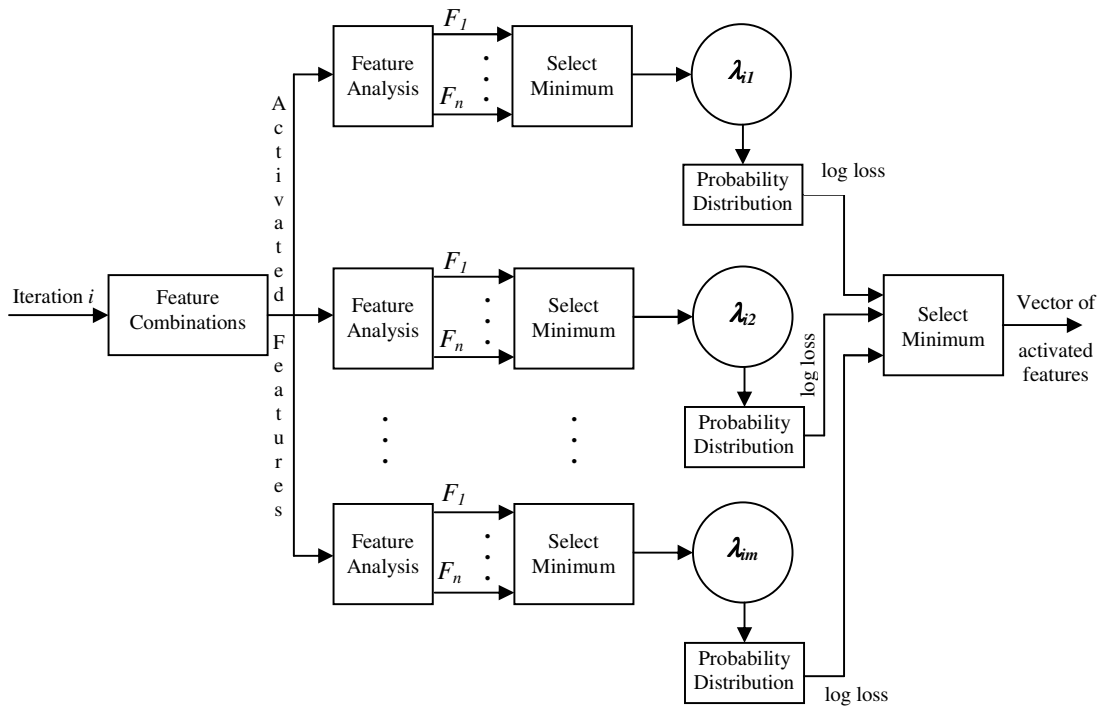


Fig. 4. Adaptive Maximum Entropy approach.

VI. EXPERIMENTAL METHODOLOGY AND RESULTS

A. Occurrence Data

Occurrence data of two species were used for tests here: *Byrsonima intermedia* and *Xylopia aromatica*. The first one is a shrub of the family *Malpighiaceae*. It is popular known as small-murici and is a native medicinal species of Brazilian Cerrado. The second one is a small tree of the family *Annonaceae*. It is popular known as malagueto and is found in Brazilian Cerrado. It is commonly used for firewood.

The occurrence data are derived from SinBiota – environmental information system for the program Biota/Fapesp [59]. The tests were carried out with 38 records of *Byrsonima intermedia* species and 33 records of *Xylopia aromatica* species. These species were chosen because they had been used in other experiments in openModeller.

B. Environmental Data

A biologist suggested the environmental layers. The models were generated for the Sao Paulo state, Brazil. Thus the environmental layers were from the same region with spatial resolution of 30 arc-second (approximately 1 km²) provided by WorldClim – Global Climate Data [60]. The layers used were annual mean temperature, mean diurnal range, maximum temperature of warmest month, minimum temperature of coldest month, annual precipitation, precipitation of wettest month and precipitation of driest month, totaling 7 layers.

C. Experiments

The aim of the experiments was to validate the AME approach. All experiments were carried out in a computer with Core 2 Duo Intel processor of 1.66 GHz and 2 GB of RAM. The Operational System used in this architecture was Ubuntu 7.04, a Linux distribution. In all experiments, the *time* command available in Linux was used. This command was used aiming to evaluate the impact of the AME approach in the openModeller performance.

The *time* command measures the execution time of the application, the time spent by the system functions during the application execution, the total time from the beginning until the end of the execution, the CPU percentage that the application got (application time + system time/total time), number of files read and written by the process and the number of page faults during the process execution. There are other output options that can be activated through the command line [61]. Here, only the total time of the application execution was considered.

Besides the execution time, the accuracy of all the models generated was considered to evaluate if the AME approach had a better performance than the classical algorithm. Both AME and the classical algorithms were run 5 times for each species and the considered values were the average of the recorded values.

D. Results

AME approach spent 9.47 seconds in average for *Xylopia aromatica* species and 9.57 seconds in average for *Byrsonima*

intermedia species, whereas the classical approach spent 14.25 seconds for the first species and 14.38 seconds for the second one.

The model average accuracy was 45.45% with the AME approach for *Xylopia aromatica* species and 56.31% for *Byrsonima intermedia* species. The classical approach generated models with 42.42% of accuracy in average for the first species and 33.68% for the second one.

The difference between the results of each run was insignificant. It is because the Maximum Entropy algorithm is deterministic. Thus, it is not necessary to run the algorithm more than once. However, the algorithms were run 5 times for each species because a different set of background pixels is generated at each run. This different set of background pixels can generate small differences in the algorithm's statistics.

Since the AME approach tests all possible combinations of features for choosing the best one, it is an algorithm computationally expensive. Its complexity is exponential because as the number of features increase, the number of combinations grows exponentially. However, AME approach searches for the best combination at each iteration, removing or inserting features. Therefore, AME approach reaches the convergence faster than classical approach. That is why it ran faster than the classical one.

A significant reduction in the execution total time of the AME approach can be observed, approximately 33.5% faster than the classical algorithm, in average. The accuracy was about 6.7% better with AME approach for *Xylopia aromatica* species than the classical algorithm. However, the accuracy for *Byrsonima intermedia* species with AME approach was 40.19% better than the classical one. These results indicate that the proposed strategy showed to be adequate for modeling species geographical distribution.

Figures 5 and 6 show a distribution model for the *Xylopia aromatica* and *Byrsonima intermedia* species, respectively, generated by the AME algorithm.

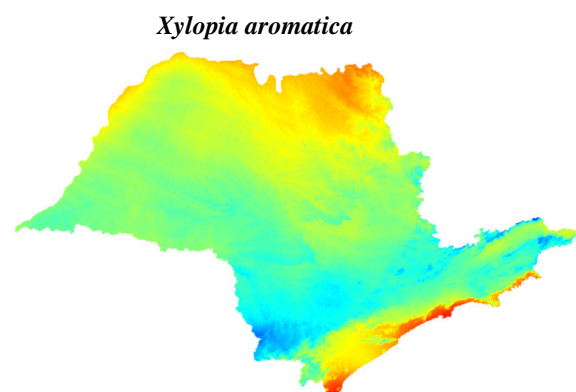


Fig. 5. Distribution model of *Xylopia aromatica*.

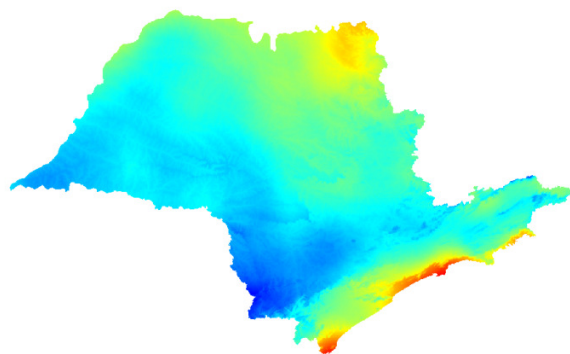


Fig. 6. Distribution model of *Byrsonima intermedia*.

The hot colors in the map represent more suitable environmental conditions for the species. Both distribution models are similar because the species are from the same biome. Thus, the environmental conditions for the species survival are the same. However, the biological analysis of the models generated is beyond the scope of this paper.

VII. FINAL DISCUSSION AND FUTURE WORKS

The aim was to present a new Adaptive Maximum Entropy approach for modeling biological species geographical distribution. Besides the AME approach integrated to the openModeller tool, some experiments were carried out to show that this approach works as well as the classical algorithm.

With the experiments carried out, it was possible to validate the use of an Adaptive Maximum Entropy approach for modeling biological species geographical distribution, since the results obtained were very significant. It was observed that, for this data set, the AME approach was about 33.5% faster than the classical algorithm, in average. Besides this, the accuracy had an increasing of 6.7% for *Xylopia aromatica* species and 40.19% for *Byrsonima intermedia* species.

These results obtained motivate the continuity of the researches in this area. Since the increasing in accuracy was very different for the two used species, one of the future works will be to test the proposed approach with a larger set of species.

Besides the improvement of the developed approach, another future work is the parallelization of this algorithm. The openModeller project has a cluster with 80 cores and there are just a few algorithms running there. The MPI library (Message Passing Interface) will be used for this implementation. Some algorithms available in the openModeller tool were already parallelized using this library. Thus, the same strategy will be used to guarantee the

Byrsonima intermedia

compatibility. In the parallelization of AME approach, several processes will work at the same time computing the log loss of a different set of features.

ACKNOWLEDGMENT

Thanks to biologist Marinez Ferreira de Siqueira from the

Botanical Garden Research Institute of Rio de Janeiro for suggesting the most suitable environmental layers for the studied species.

REFERENCES

- [1] P. Avesani, A. Perini and F. Ricci. Interactive case-based planning for forest fire management. *Applied Intelligence* 13, pp. 41–57, 2000.
- [2] E. Kalapanidas and N. Avouris. Short-term air quality prediction using a case-based classifier. *Environmental Modeling and Software*, Volume 16, Number 3, pp. 263–272, 2001.
- [3] J. M. P. King, R. Bañares-Alcántara and Z. A. Manan. Minimising environmental impact using CBR: An azeotropic distillation case study. *Environmental Modelling and Software*, Volume 14, Number 5, pp. 359–366, 1999.
- [4] J. S. Pedro, F. Burstein and A. Sharp. A case-based fuzzy multicriteria decision support model for tropical cyclone forecasting. *European Journal of Operational Research*, 160, pp. 308–324, 2005.
- [5] B. D. Mahaman, H. C. Passam, A. B. Sideridis and C. P. Yialouris. DIARES-IPM: a diagnostic advisory rule-based expert system for integrated pest management in Solanaceous crop systems. *Agricultural Systems*, Volume 76, Number 3, pp. 1119–1135, 2003.
- [6] E. El-Azhary, H. A. Hassan and A. Rafea. Pest control expert system for tomato (PCEST). *Knowledge and Information Systems*, Volume 2, Number 2, pp. 242–257, 2000.
- [7] B. D. Mahaman, P. Harizanis, I. Filis, E. Antonopoulou, C. P. Yialouris and A. B. Sideridis. A diagnostic expert system for honeybee pests. *Computers and Electronics in Agriculture*, Volume 36, Number 1, pp. 17–31, 2002.
- [8] F. Zetian, X. Feng, Z. Yun and Z. XiaoShuan. Pig-vet: a web-based expert system for pig disease diagnosis. *Expert Systems with Applications*, Volume 29, Issue 1, pp. 93–103, 2005.
- [9] L. S. Iliadis and F. Maris. An Artificial Neural Network model for mountainous water-resources management: The case of Cyprus mountainous watersheds. *Environmental Modelling & Software*, Volume 22, Issue 7, pp. 1066–1072, 2007.
- [10] Z. Qing and S. J. Stanley. Forecasting raw-water quality parameters for the North Saskatchewan River by neural network modeling. *Water Research*, Volume 31, Number 9, pp. 2340–2350, 1997.
- [11] D. Pozdnyakov, R. Shuchman, A. Korosov and C. Hatt. Operational algorithm for the retrieval of water quality in the Great Lakes. *Remote Sensing of Environment*, Volume 97, Issue 3, pp. 352–370, 2005.
- [12] S. I. V. Sousa, F. G. Martins, M. C. M. Alvim-Ferraz and M. C. Pereira. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*,

Volume 22, Number 1, pp. 97–103, 2007.

- [13] H. R. Maier, G. C. Dandy and M. D. Burch. Use of artificial neural networks for modeling cyanobacteria *Anabaena* spp. In the River Murray, South Australia. *Ecological Modeling* 105, pp. 257–272, 1998.
- [14] F. A. Rodrigues, A. O. Avilla, E. S. C. Rodrigues, P. L. P. Corrêa, A. M. Saraiva and R. L. A. da Rocha. Species distribution modeling with neural networks. *e-Biosphere* 2009, London-UK, 2009. Retrieved August 27, 2009, from <http://www.e-biosphere09.org/assets/files/e-Biosphere%20Abstracts%20Volume%20-%20FINAL.pdf>.
- [15] L. Persona, P. L. P. Corrêa and A. M. Saraiva. Environmental Niche Modeling in Biodiversity with Genetic Algorithms. In: 2nd International Information and 68 Telecommunication Technologies Symposium, Florianópolis. Proceedings of the IEEE - I2TS'2003, 2003. (In Portuguese).
- [16] E. Kalapanidas and N. Avouris. Feature selection for air quality forecasting: a genetic algorithm approach. *AI Communications*, Volume 16, Issue 4, pp. 235–251, 2003.
- [17] G. J. Pelletier, S. C. Chapra and H. Tao. QUAL2Kw – a framework for modeling water quality in streams and rivers using a genetic algorithm for calibration. *Environmental Modelling & Software*, Volume 21, pp. 419–425, 2006.
- [18] A. V. M. Ines, K. Honda, A. das Gupta, P. Droogers and R. S. Clemente. Combining remote sensing-simulation modeling and genetic algorithm optimization to explore water management options in irrigated agriculture. *Agricultural Water Management*, Volume 83, Number 3, pp. 221–232, 2006.
- [19] B. Schönfisch and M. Kinder. A fish migration model. *Lecture Notes in Computer Science*, Volume 2493, pp. 210–219, 2002.
- [20] S. El Yacoubi, A. El Jai, P. Jacewicz and J. G. Pausas. LUCAS: an original tool for landscape modeling. *Environmental Modelling and Software*, Volume 18, Number 5, pp. 429–437, 2003.
- [21] I. G. Georgoudas, G. C. Sirakoulis, E. M. Scordilis and I. Andreadis. A cellular automaton simulation tool for modeling seismicity in the region of Xanthi. *Environmental Modelling & Software*, Volume 22, Issue 10, pp. 1455–1464, 2007.
- [22] M. Schluter and N. Ruger. Application of a GIS-based simulation tool to illustrate implications of uncertainties for water management in the Amudarya river delta. *Environmental Modelling & Software*, Volume 22, Issue 2, pp. 158–166, 2007.
- [23] G. Metternicht and S. Gonzalez. FUERO: foundations of a fuzzy exploratory model for soil erosion hazard prediction. *Environmental Modelling & Software*, Volume 20, Issue 6, pp. 715–728, 2005.
- [24] L. S. Iliadis. A decision support system applying an integrated fuzzy model for long-term forest fire risk estimation. *Environmental Modelling & Software*, Volume 20, Issue 5, pp. 613–621, 2005.
- [25] F. Bousquet and C. Le Page. Multi-agent simulations and ecosystem management: a review. *Ecological Modelling* 176, pp. 313–332, 2004.
- [26] D. Borri, D. Camarda and A. De Liddo. Mobility in environmental planning: an integrated multi-agent approach. *Lecture Notes in Computer Science*, Volume 3675, pp. 119–129, 2005.
- [27] H. Purnomo, G. A. Mendoza, R. Prabhu and Y. Yasmi. Developing multi-stakeholder forest management scenarios: a multi-agent system simulation approach applied in Indonesia. *Forest Policy and Economics*, Volume 7, Issue 4, pp. 475–491, 2005.
- [28] S. H. Chen, A. J. Jakeman and J. P. Norton. Artificial Intelligence techniques: An introduction to their use for modelling environmental systems. *Mathematics and Computers in Simulation*, Volume 78, Issue 2–3, pp. 379–400, 2008.
- [29] M. E. S. Muñoz, R. Giovanni, M. F. Siqueira, T. Sutton, P. Brewer, R. S. Pereira, D. A. L. Canhos and V. P. Canhos. openModeller: a generic approach to species' potential distribution modeling. *GeoInformatica*. DOI: 10.1007/s10707-009-0090-7. 2009.
- [30] C. B. Pariente, J. J. Neto and F. S. Santana. Towards an Adaptive Implementation of Genetic Algorithms. I Taller Latinoamericano de Informática para la Biodiversidad (INBI) - CLEI 2007, San José, Costa Rica, 9-12 Octubre, 2007.
- [31] F. A. Rodrigues, E. S. C. Rodrigues, L. M. Sato, E. T. Midorikawa, P. L. P. Corrêa and A. M. Saraiva. Parallelization of the Jackknife Algorithm Applied to a Biodiversity Modeling System. Proceedings of the 7th International Information and Telecommunication Technologies Symposium – I2TS'2008, pp. 58–65, 2008.
- [32] C. H. Graham, S. Ferrier, F. Huettman, C. Moritz and A. T. Peterson. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, Volume 19, Number 9, pp. 497–503, 2004.
- [33] G. E. Hutchinson, Introduction to Ecology of Populations. Barcelona, Editorial Blume, 492p., 1981. (In Spanish)
- [34] S. J. Phillips, R. P. Anderson and R. E. Schapire. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231 – 259, 2006.
- [35] M. F. Siqueira, Use of Fundamental Niche Modeling in the Pattern Evaluation of Vegetal Species Geographic Distribution. PhD Thesis. Department of Ambient Engineering of University of Sao Carlos. Sao Carlos, Sao Paulo, 2005. (In Portuguese)
- [36] R. S. Pereira. DesktopGarp, 2002. Retrieved August 27, 2009, from <http://www.nhm.ku.edu/desktopgarp/index.html>
- [37] S. J. Phillips, M. Dudík and R. E. Scaphire. Maxent, 2009. Retrieved August 27, 2009, from <http://www.cs.princeton.edu/~schapire/maxent>
- [38] T. Sutton, R. Giovanni and M. F. Siqueira. Introducing openModeller – a fundamental niche modeling framework.

- OSGEO Journal, Vol.1., 2007. Retrieved August 27, 2009, from <http://www.osgeo.org/ojs/index.php/journal/article/viewFile/109/90>.
- [39] D. Stockwell and D. Peters. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, vol. 13, no. 2, 143–158, 1999.
- [40] D. R. B. Stockwell and I. R. Noble. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation*, 32, 249–254, 1991.
- [41] D. R. B. Stockwell. Machine learning and the problem of prediction and explanation in ecological modelling. Doctoral Thesis, Australian National University, Australia, 1992.
- [42] S. J. Phillips, M. Dudík and R. E. Schapire. A Maximum Entropy Approach to Species Distribution Modeling. *Proceedings of the 21st International Conference on Machine Learning*, Banff, Canada, 2004.
- [43] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.
- [44] A. C. Lorena, M. F. Siqueira, R. Giovanni, A. C. P. L. F. Carvalho and R. C. Prati. Potential Distribution Modelling Using Machine Learning Classifiers. In: *The Twenty First International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems*, 2008, Wroclaw. *Lecture Notes in Artificial Intelligence - Proceedings The Twenty First International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems*. v. 5027. p. 255–264, 2008.
- [45] S. Haykin. *Neural Networks and Learning Machines*. Prentice Hall, 3rd edition, 2008.
- [46] J. Elith, C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. Overton, A. T. Peterson, S. J. Phillips, K. S. Richardson, R. P. Scachetti, R. E. Schapire, J. Soberón, S. Williams, M. S. Wisz and N. E. Zimmermann. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, pp. 129–151, 2006.
- [47] J. J. Neto. A Survey of the Adaptivity Evolution and Adaptive Technology. *IEEE Magazine – Latin America*, Vol. 5, Num. 7, ISSN: 1548–0992, pp. 496 – 505, 2007. (In Portuguese)
- [48] J. J. Neto. A Glossary about Adaptivity. *Memories of WTA'2009: third Workshop of Adaptive Technology*. ISBN 978-85-86686-51-1. São Paulo: s.n., 2009. (In Portuguese)
- [49] J. J. Neto. Contributions to Methodology of Compilers Construction. Thesis of Free Professor, Engineering School of USP, São Paulo, 1993. (In Portuguese)
- [50] R. L. A. Rocha and J. J. Neto. Adaptive Automata, limits and complexity in comparison with Turing Machine. In: *Proceedings of the Second Congress of Logic Applied to Technology – LAPTEC'2000*, São Paulo, pp. 33 – 48, 2001. (In Portuguese)
- [51] J. J. Neto. Solving Problems Efficiently with Adaptive Automata. *CIAA 2000 – Fifth International Conference on Implementation and Application of Automata*, London, Ontario, Canadá, July 2000.
- [52] J. J. Neto. Adaptive Rule-Driven Devices – General Formulation and Case Study”, *Lecture Notes in Computer Science*, Watson, B. W. and Wood, D. (Eds.): *Implementation and Application of Automata 6th International Conference*, CIAA 2001, Vol. 2494, Pretoria, South África, Springer Verlag, pp. 234 – 250, July 23 – 25, 2002.
- [53] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, vol. 106, pp. 620–630, 1957a.
- [54] E. T. Jaynes. Information theory and statistical mechanics II. *Physical Review*, vol. 108, pp. 171–190, 1957b.
- [55] A. L. Berger, V. J. D. Pietra and S. A. D. Pietra. A maximum entropy approach to natural language processing. *Computational Linguistic*, Volume 22, Issue1, pp. 39–71, 1996.
- [56] S. J. Phillips and M. Dudík. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161–175, 2008.
- [57] R. Malouf. A comparison of algorithms for maximum entropy parameter estimation. In *Proceedings of the Sixth Conference on Natural Language Learning (CoNLL-2002)*, pp. 49–55, 2002.
- [58] M. Dudík, S. J. Phillips and R. E. Schapire. Performance Guarantees for Regularized Maximum Entropy Density Estimation. In *Proceedings of the 17th Annual Conference on Computational Learning Theory*, ACM Press, New York, pp. 655–662, 2004.
- [59] SinBiota – Environmental Information System for Biota. [on-line] available in: <http://sinbiota.cria.org.br/>
- [60] R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones and A. Jarvis. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965 – 1978.
- [61] MANPAGE OF TIME. Searched in 12/11/2007. [on-line] available in <http://www.ncsa.uiuc.edu/~gbauer/time.1.html>.

BIOGRAPHIES

Elisângela S. C. Rodrigues obtained her degree in Computer Science from the Federal University of Pelotas in 2000 and her master's degree in Computer Science from the Federal University of Campina Grande in 2002. She initiated her teaching activities in 2002, at the Faculty of Technology and Sciences of Feira de Santana/BA, where she acted until 2005. In the same year, she moved to College ÁREA1 in Salvador/BA, where acted until 2007. In 2006, she participated as researcher (supported by Petrobrás) in the Automated Well Management System Project (SGPA), at the Federal University of Bahia. Currently, she is a regular

doctorate candidate at the Engineering School of University of São Paulo.

Fabício A. Rodrigues obtained his degree in Information Systems from the Potiguar University in 2000 and his master's degree in Computer Science from the Federal University of Campina Grande in 2002. He initiated his teaching activities in 2002, at the Faculty of Technology and Sciences in Feira de Santana/BA, where he acted until 2005. In the same year, he moved to College ÁREA1 in Salvador/BA, where taught until 2007, besides acting as the coordinator of the Computation Nucleus. Currently, he is a regular doctorate candidate at the Engineering School of University of São Paulo.

Ricardo L. A. da Rocha was born in Rio de Janeiro / RJ in 29/05/1960. He obtained his degree in Electronics Engineering, focus on Digital and Computer Systems, from the Pontifícia Universidade Católica do Rio de Janeiro - PUC-RJ, in 1982. He was granted his MSc (1995) and PhD (2000) in Computing Engineering from the Engineering School of the University of São Paulo - EPUSP. He is currently a professor at the Engineering School, and his research interests are Adaptive Techniques and Technologies, Computational Models, Intelligent Systems, Kolmogorov Complexity and Foundations of the Theory of Computation. Dr. Rocha has been a member of the Association for Computing Machinery – ACM since 2001, and also of the Brazilian Computing Society - SBC since 2000.

Pedro L. P. Corrêa obtained his degree in Computer Science in 1987 - ICMSC / USP / Brazil, his master's degree in Computer Science in 1992 - ICMSC / USP / Brazil, his Ph.D. in Electrical Engineering in 2002 - EPUSP / Brazil. He worked as consultant in Development Systems with companies in the financial areas, through the Interchange Services (company of EDI - Electronic Data Interchange - Brazil) and with the coordination of teams and systems design for the electronic government Secretary of the State of São Paulo, through the UNDP (UN Development Program) from 1998 to 2002. He is currently a professor at the Engineering School - University of São Paulo, at the Department of Computer Engineering. Areas of interest: Distributed Database, Information Systems, Biodiversity Informatics, Service Oriented Architecture and Electronic Government.