

Modelagem da Biodiversidade Utilizando Redes Neurais Artificiais

Fabrcio Augusto Rodrigues¹, Elisângela Silva da Cunha Rodrigues², Pedro Luiz Pizzigatti Corrêa¹, Ricardo Luis de Azevedo da Rocha², Antonio Mauro Saraiva¹

Departamento de Engenharia de Computação e Sistemas Digitais – Escola Politécnica da Universidade de São Paulo (EPUSP)

¹Laboratório de Automação Agrícola

²Laboratório de Linguagens e Técnicas Adaptativas

{fabricio.rodrigues, elisangela.rodrigues, pedro.correa, luis.rocha, antonio.saraiva}@poli.usp.br

Abstract. *Biodiversity conservation is one of the great contemporary challenges. Techniques and tools that contribute for the strategical planning of actions related to the environmental conservation are essential in the decision-making process. This paper presents an algorithm of biodiversity modeling based on Artificial Neural Networks, used to create models of biological species geographical distribution. Tests reached accuracy until 94%, showing that Artificial Neural Networks are suitable techniques to modeling process.*

Resumo. *A conservação da biodiversidade é um dos grandes desafios contemporâneos. Técnicas e ferramentas que contribuem para o planejamento estratégico de ações relacionadas à conservação ambiental são fundamentais no processo de tomada de decisão. Este artigo apresenta um algoritmo de modelagem da biodiversidade baseado em Redes Neurais Artificiais, utilizado para criar modelos de distribuição geográfica de espécies biológicas. Durante os testes, foram alcançadas taxas de acuidade de até 94%, demonstrando que as Redes Neurais Artificiais são técnicas adequadas ao processo de modelagem.*

1. Introdução

O desenvolvimento econômico, principalmente na última década, tem acelerado o processo de destruição de ecossistemas, aumentando a demanda por ferramentas de apoio à tomada de decisão no que se refere à conservação da biodiversidade. Uma nova área de pesquisa científica, denominada Informática para Biodiversidade [Canhos 2003], tem como desafio suprir esta demanda, fornecendo ferramentas computacionais aplicadas ao estudo da biodiversidade. Uma das tarefas, nas quais ferramentas computacionais têm sido aplicadas, é a modelagem de distribuição geográfica de espécies biológicas.

Uma vez que os recursos humanos e financeiros direcionados ao meio-ambiente geralmente são escassos, frente à demanda atual, ferramentas de modelagem podem contribuir para a conservação ambiental. Tais ferramentas podem ser aplicadas, por

exemplo, na identificação de áreas com alto risco de perda da sua biodiversidade, possibilitando assim que os recursos escassos sejam direcionados para essas regiões e suas espécies ameaçadas.

Os modelos de distribuição geográfica de espécies biológicas são criados por algoritmos que buscam, em um espaço de características, relações não aleatórias entre os dados de entrada. No processo de modelagem adotado neste trabalho, os modelos são criados a partir de dados georreferenciados de ocorrência da espécie e variáveis ambientais. Vários algoritmos têm sido utilizados para a criação dos modelos de distribuição, com destaque para algoritmos da área de Aprendizagem de Máquina, apresentados sucintamente na Seção 3.

Este artigo apresenta um algoritmo de modelagem baseado em Redes Neurais Artificiais (RNA) desenvolvido dentro do escopo do projeto *openModeller* [Muñoz et al. 2009], cujo objetivo principal é o desenvolvimento de um *framework* de código aberto para modelagem da biodiversidade. O algoritmo foi testado com dados de uma espécie vegetal típica do cerrado brasileiro. O objetivo dos testes foi avaliar de forma preliminar a aplicabilidade do algoritmo. A próxima etapa deste trabalho prevê a comparação do algoritmo desenvolvido com outros algoritmos de modelagem já utilizados.

Além do algoritmo de modelagem baseado em RNA, o *openModeller* oferece outros algoritmos de modelagem, ferramentas de tratamento e preparação de dados ambientais, opções de busca de dados de localização de espécies e interface gráfica para visualização de modelos de distribuição. Este conjunto de funcionalidades torna o *openModeller* um ambiente adequado a inclusão de novas técnicas de modelagem.

A Seção 2 apresenta resumidamente o processo geral de modelagem de distribuição geográfica de espécies biológicas. Alguns dos principais algoritmos utilizados no processo de modelagem são apresentados brevemente na Seção 3. Os principais aspectos do desenvolvimento do algoritmo de modelagem baseado em RNA são apresentados na Seção 4. A Seção 5 descreve os experimentos realizados e apresenta os resultados obtidos. A discussão final e propostas de trabalhos futuros são apresentados na Seção 6.

2. O Processo de Modelagem

Ferramentas de modelagem de distribuição geográfica de espécies biológicas podem contribuir de maneira importante para a conservação da biodiversidade, apoiando o processo de tomada de decisão. O objetivo deste tipo de modelagem é estimar a probabilidade da presença ou da ausência de uma determinada espécie em uma região geográfica. Isto é feito através da conversão de dados de ocorrência da espécie e dados ambientais em mapas de distribuição geográfica [Siqueira 2005].

Os dados de ocorrência são pontos georreferenciados (latitude e longitude) que representam a existência ou não da espécie em uma determinada região. São divididos em pontos de presença, que indicam a existência comprovada da espécie, e pontos de ausência, que representam a inexistência da espécie. Os dados de ocorrência normalmente resultam da observação e coleta em campo [Siqueira 2005]. Entretanto, pontos de ausência são muito difíceis de determinar e raramente estão disponíveis, pois a não observação de uma espécie em uma determinada região não é suficiente para determinar que aquela espécie não existe naquele local.

Os dados ambientais, também chamados de camadas ambientais, ou variáveis ambientais, representam as condições ambientais sob as quais a espécie pode existir naquela região. Esse conjunto de condições é normalmente chamado de nicho fundamental da espécie [Siqueira 2005]. Fatores como competição entre indivíduos e espécies, predação, parasitismo etc., não são considerados neste processo de modelagem. Temperatura, precipitação e altitude são alguns exemplos de variáveis ambientais frequentemente utilizadas no processo de modelagem.

O processo de modelagem de distribuição geográfica de espécies biológicas pode ser resumido nas seguintes etapas: (1) um conjunto de pontos de ocorrência (georreferenciados) é combinado com um conjunto de variáveis ambientais, dando origem a um conjunto de pontos de nicho, no qual cada ponto de nicho é formado pelos valores que as variáveis ambientais assumem em cada ponto de ocorrência; (2) um algoritmo de modelagem é utilizado para criar um modelo de nicho a partir do conjunto de pontos de nicho; (3) o modelo de nicho criado pelo algoritmo é projetado em uma determinada região geográfica, levando em consideração as mesmas variáveis ambientais utilizadas para a criação deste modelo de nicho.

O resultado do processo de modelagem é um mapa georreferenciado contendo, em cada ponto, a probabilidade de ocorrência da espécie, estimada pelo algoritmo de modelagem escolhido. Este mapa pode representar: cenários atuais, quando as variáveis ambientais utilizadas representam as condições climáticas atuais; cenários anteriores, quando as variáveis ambientais utilizadas representam condições climáticas do passado; e até mesmo cenários futuros, quando são utilizadas variáveis ambientais que simulam mudanças nas condições climáticas atuais [Hijmans et al. 2005]. Exemplos de mapas de distribuição podem ser vistos na Seção 5.

3. Algoritmos de Modelagem

A função principal dos algoritmos é executar o passo de inferência do processo de modelagem. O passo de inferência pode ser resumido na busca por associações não aleatórias entre os dados de ocorrência da espécie e os dados ambientais. O *openModeller* oferece diversos algoritmos de modelagem, alguns utilizam uma abordagem estatística enquanto outros são baseados em técnicas de Aprendizagem de Máquina.

Oferecer diferentes algoritmos numa mesma ferramenta de modelagem é um dos objetivos principais do projeto *openModeller*. Tal característica pode facilitar o processo de comparação entre os algoritmos disponíveis, além de permitir que os usuários possam selecionar o algoritmo mais adequado aos objetivos da modelagem. É importante destacar que, embora não seja meta deste artigo apresentar a comparação do algoritmo desenvolvido com outros algoritmos de modelagem, esse estudo comparativo já está sendo realizado.

Dentre os diversos algoritmos de modelagem oferecidos pelo *openModeller*, destacam-se o GARP (*Genetic Algorithm for Rule-set Production*) [Stockwell and Peters 1999], SVM (*Support Vector Machines*) [Vapinik 1995, Lorena et al. 2008] e Entropia Máxima [Rodrigues et al. 2010, Phillips et al. 2006]. Uma vez que os algoritmos citados apresentam bons resultados no processo de modelagem, é natural que a aplicabilidade de outras técnicas clássicas de Aprendizagem de Máquina seja investigada. Neste contexto, um algoritmo de modelagem baseado em Redes Neurais

Artificiais foi desenvolvido e integrado ao *openModeller*. Aspectos sobre o desenvolvimento desse algoritmo são abordados na próxima seção.

4. Modelagem com Redes Neurais Artificiais

O Processo de modelagem de distribuição geográfica de espécies biológicas pode ser visto como um problema de predição, uma vez que o objetivo principal é determinar a probabilidade da existência de uma determinada espécie biológica em cada ponto de um mapa georreferenciado. Como descrito anteriormente, este processo de predição é baseado em dados ambientais e dados de ocorrência da espécie.

A subseção 4.1 apresenta resumidamente alguns trabalhos que utilizaram RNA em alguma etapa do processo de modelagem. Detalhes sobre o algoritmo desenvolvido e integrado ao *openModeller* são apresentados na subseção 4.2.

4.1. Trabalhos Relacionados

A aplicação de Redes Neurais Artificiais no processo da modelagem de distribuição geográfica de espécies biológicas pode ser observada em alguns trabalhos, motivando novas pesquisas. Geralmente, tais trabalhos discutem os modelos de distribuição gerados do ponto de vista biológico, utilizando a técnica computacional apenas como ferramenta, ou seja, não discutem os aspectos de implementação e ajuste dos parâmetros do algoritmo.

O BIOMOD [Thuiller et al. 2009] é uma plataforma direcionada à modelagem de distribuição geográfica de espécies que inclui várias técnicas, dentre elas RNA.

Em [Ximenes et al. 2009], foi proposto o uso de Mapas Auto-Organizáveis (SOM - *self-organizing map*), uma abordagem de RNA, para a análise de dependência espacial de variáveis ambientais, utilizadas na modelagem de distribuição de espécies. Esta análise contribui para a pré-seleção de variáveis, reduzindo assim a quantidade de dados a serem manipulados e diminuindo o custo computacional da modelagem.

Em [Jiménez-Valverde et al. 2009], modelos de distribuição de espécies artrópodes foram criados utilizando RNA do tipo Perceptron de Múltiplas Camadas. Neste trabalho, os autores investigaram se os modelos de distribuição criados podem ser utilizados como substitutos da abundância de espécies artrópodes na Ilha Terceira do arquipélago dos Açores.

RNA também foram utilizadas no mapeamento da intensidade de luz no Nepal. Este mapeamento contribui com o monitoramento da propagação do arbusto *Chromolaena odorata* [Joshi et al. 2005]. Segundo os autores do trabalho, este arbusto é considerado uma das piores espécies invasivas do mundo.

Em [Mastrorillo et al. 1997], RNA foram utilizadas para estimar a presença/ausência de três espécies de peixe a partir de dez variáveis ambientais. Os resultados obtidos com esta abordagem foram comparados com os resultados obtidos através do uso de DFA (*discriminant factorial analysis*) para a mesma modelagem. As RNA apresentaram melhor capacidade de separação dos exemplos que o uso de DFA, para as três espécies.

4.2. Algoritmo baseado em Redes Neurais no *openModeller*

Para a implementação do algoritmo baseado em RNA e sua consequente

integração ao *openModeller*, foi adotado o modelo Perceptron de Múltiplas Camadas (*Multilayer Perceptron - MLP*), usando o algoritmo de treinamento de Retropropagação de Erro (*Backpropagation*) [Rumelhart et al. 1986]. A ideia foi utilizar um modelo clássico, bem testado e adequado à tarefa de modelagem. Além disso, as classes dos dados que formam os conjuntos de treinamento e teste são conhecidas e divididas em pontos de presença e pontos de ausência, tornando o treinamento supervisionado adequado, como é o caso do *Backpropagation*. Mesmo quando os dados são formados apenas por pontos de presença, uma estratégia de geração de pontos de pseudo-ausência é utilizada, permitindo o uso do treinamento supervisionado.

No algoritmo desenvolvido, a arquitetura da rede neural é organizada em três camadas, conforme o esquema mostrado na Figura 1. A camada de entrada recebe os valores das variáveis ambientais para cada exemplo de treinamento. Portanto, o número de neurônios dessa camada corresponde ao número de variáveis ambientais utilizadas na etapa de inferência e é definido automaticamente pelo algoritmo. A camada de saída é formada por um único neurônio, responsável por apresentar o valor de saída da rede (um valor entre 0 e 1), ou seja, a probabilidade da presença da espécie em cada ponto do mapa georreferenciado. Apenas uma camada escondida é utilizada na implementação atual do algoritmo e a quantidade de neurônios dessa camada é definida pelo usuário. Cada neurônio da rede está conectado unidirecionalmente a todos os neurônios da camada seguinte. No modelo MLP, o sinal de ativação dos neurônios da camada escondida é calculado através de uma função não-linear que recebe a soma ponderada dos valores de entrada. No algoritmo integrado ao *openModeller* foi adotada a função logística (sigmoide) como função de ativação.

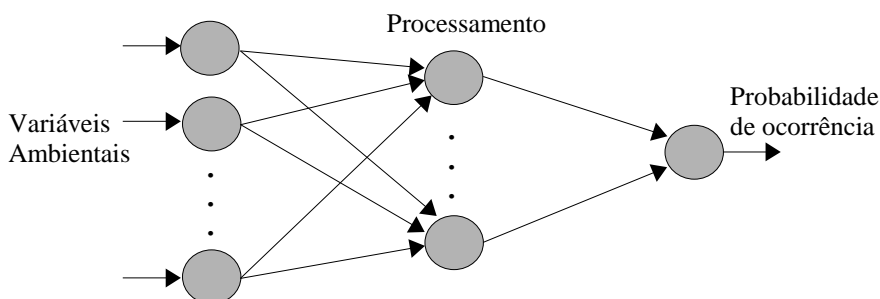


Figura 1. Esquema de uma rede neural para o processo de modelagem de distribuição geográfica de espécies biológicas

Durante a etapa de treinamento dois critérios de parada são utilizados: número limite de épocas de treinamento [Haykin 2001] e um valor que representa o erro mínimo esperado. Esta estratégia é adotada uma vez que o treinamento com o *Backpropagation* não garante que o erro da rede atingirá um valor mínimo estabelecido, podendo ficar preso em regiões de mínimos locais. O algoritmo de treinamento é executado até que um dos critérios de parada seja satisfeito. Dessa forma, mesmo que o erro mínimo esperado não tenha sido alcançado, quando o número de épocas previamente estabelecido no início do treinamento é atingido, o algoritmo de treinamento é interrompido.

Além do número de neurônios da camada escondida, do número de épocas e do erro mínimo esperado, outros parâmetros definidos pelo usuário são: (1) a taxa de aprendizagem; e (2) a constante do termo de *momentum*. É importante ressaltar que a inclusão do termo de *momentum* é uma modificação na regra original do algoritmo

Backpropagation. Detalhes sobre o ajuste desses parâmetros podem ser encontrados em [Haykin 2001].

O algoritmo de modelagem baseado em RNA foi implementado em linguagem C++, tornando mais direta a sua integração ao *openModeller*, uma vez que este *framework* foi totalmente escrito nesta linguagem. A implementação envolveu um trabalho de pesquisa e seleção de uma biblioteca de RNA, utilizada como base para o desenvolvimento do algoritmo. É importante destacar que a própria filosofia de desenvolvimento do *openModeller* tem como princípio a aplicação e o reuso de código aberto.

A biblioteca escolhida pode ser encontrada em [Chopra, 2008]. A simplicidade do código, permitindo uma melhor compreensão, e os métodos disponíveis foram pontos fundamentais na escolha da biblioteca. No entanto, várias alterações no código fonte da biblioteca foram realizadas para tornar possível a integração do algoritmo ao *openModeller* e tornar o algoritmo mais robusto e completo. Modificações na regra de atualização dos pesos da rede neural durante o treinamento foram realizadas, como por exemplo a inclusão do termo de *momentum*. Além disso, outra modificação foi a utilização do número de épocas de treinamento juntamente com o valor do erro mínimo como critérios de parada, conforme descrito acima. A biblioteca original considera apenas o número de épocas como critério de parada para a etapa de treinamento da rede neural.

A próxima seção descreve os experimentos realizados com o algoritmo de modelagem desenvolvido e apresenta os resultados alcançados.

5. Experimentos e Resultados

Os experimentos foram realizados em duas etapas com o objetivo de validar o algoritmo desenvolvido. A segunda etapa foi estabelecida após a verificação de possíveis ruídos nos dados. Esses ruídos foram detectados no primeiro conjunto de testes. As subseções seguintes descrevem os dados utilizados nos experimentos, a metodologia aplicada nos testes e os resultados obtidos.

5.1. Dados

Foram utilizados 100 registros de ocorrência, sendo 50 pontos de presença e 50 pontos de ausência da espécie *Stryphnodendron obovatum*, uma planta da família *Mimosaceae* presente no cerrado brasileiro. Os registros de ocorrência foram fornecidos por uma bióloga e coletados conforme descrito em [Lorena et al. 2008].

Foram utilizadas sete camadas ambientais recortadas para o estado de São Paulo, Brasil. Todas as camadas pertencem a mesma região com 30 segundos de arcos de resolução espacial (aproximadamente 1 km²) e estão disponíveis no *WorldClim – Global Climate Data* [Hijmans et al. 2005]. As camadas usadas foram temperatura média anual, escala diurna média, temperatura máxima do mês mais quente, temperatura mínima do mês mais frio, precipitação anual, precipitação do mês mais úmido e precipitação do mês mais seco.

5.2. Metodologia

As medidas estatísticas coletadas para análise dos modelos e validação do algoritmo de modelagem baseado em RNA foram: acuidade, porcentagem de erros de omissão,

porcentagem de erros de comissão, AUC (*Area Under the ROC Curve*) e quantidade de épocas do treinamento.

A acuidade corresponde à porcentagem de acertos do algoritmo em relação ao conjunto de dados e foi medida tanto durante o treinamento quanto com conjuntos de testes independentes. Os erros de omissão correspondem à porcentagem de pontos de presença que não foram corretamente classificados, enquanto os erros de comissão correspondem à porcentagem de pontos de ausência que não foram classificados corretamente. A AUC fornece uma medida de desempenho do modelo, representada por um valor entre 0 e 1. Quanto mais próxima de 1 melhor o modelo [Fielding and Bell 1997].

Os dados foram divididos em cinco partições, cada uma contendo 10 pontos de presença e 10 pontos de ausência. A partir dessas partições foram criados dez conjuntos de treinamento e dez conjuntos de teste. Em cada conjunto de treinamento foram agrupadas três partições, ou seja, 60% dos dados e em cada conjunto de teste foram agrupadas as duas partições restantes, correspondendo a 40% dos dados. Desta forma, cada modelo gerado a partir de uma partição de treinamento foi testado com a partição de teste, mutuamente exclusiva, correspondente.

A primeira etapa de testes mostrou a possível presença de ruído nos dados, ou seja, pontos de ocorrência que não são distinguidos pela rede neural. Após uma análise dos resultados desta primeira etapa de testes, quatro pontos de ausência foram identificados e considerados pontos com ruído. Esses pontos foram removidos do conjunto de dados, restando assim 50 pontos de presença e 46 pontos de ausência, a partir dos quais novos testes foram realizados.

Todos os experimentos foram realizados com os parâmetros de entrada calibrados com os mesmos valores. Testes preliminares mostraram que valores adequados para os parâmetros são: (1) 4 neurônios na camada escondida; (2) taxa de aprendizagem = 0,7; (3) constante de *momentum* = 0,4; (4) erro mínimo = 0,02. A quantidade de neurônios na camada escondida foi definida através da média aritmética entre o número de entradas (7 entradas) e o número de saídas (1 saída). Na primeira etapa de testes o número limite de épocas de treinamento foi igual dois milhões. Este valor foi definido em razão da demora na convergência por erro mínimo, causada pela possível presença de ruído nos dados. Com a remoção do ruído, o número limite de épocas de treinamento foi igual a 500.000, uma vez que a convergência se tornou muito mais rápida.

5.3. Resultados

As médias das medidas estatísticas coletadas durante o treinamento são apresentadas na Tabela 1 e as médias das medidas estatísticas coletadas a partir de conjuntos de teste independentes são mostradas na Tabela 2.

Analisando os dados coletados, é possível observar que a qualidade dos exemplos de treinamento interfere na taxa de aprendizagem do algoritmo. Mesmo que seja apresentado um conjunto de treinamento menor à rede, a qualidade dos modelos pode ser superior se os dados forem livres de ruídos. Esta sensibilidade aos dados de treinamento é uma característica inerente ao modelo neural adotado no algoritmo de modelagem.

O maior valor de acuidade registrado no treinamento foi 100% nas duas etapas.

Este valor indica que todos os dados de treinamento foram aprendidos pela rede neural. Na primeira etapa de experimentos, o maior valor de acuidade registrado quando apresentado um conjunto de teste independente foi 87,5%. Na segunda etapa de experimentos, o maior valor de acuidade registrado com um conjunto de teste independente foi 94,44%. As Figuras 2 e 3 apresentam os modelos que alcançaram os maiores valores de acuidade e AUC em cada etapa de experimentos respectivamente.

Tabela 1. Média das estatísticas de treinamento

Etapa	Acuidade (%)	Erro de Omissão (%)	Erro de Comissão (%)	AUC
1ª	98,50	0	3,00	0,98
2ª	99,11	0	1,92	0,99

Tabela 2. Média das estatísticas de teste

Etapa	Acuidade (%)	Erro de Omissão (%)	Erro de Comissão (%)	AUC
1ª	75,25	26,5	23	0,77
2ª	81,61	16	22	0,81



Figura 2. Modelo de distribuição que apresentou os maiores valores de acuidade (87,5%) e AUC (0,9) na primeira etapa de testes.

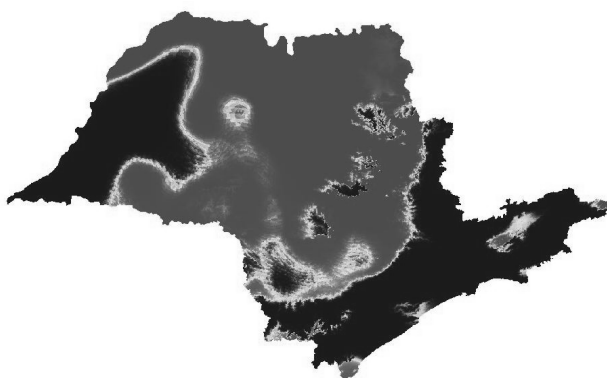


Figura 3. Modelo de distribuição que apresentou os maiores valores de acuidade (94,44%) e AUC (0,94) na segunda etapa de testes.

É importante destacar, nas Figuras 2 e 3, que as regiões representadas pela cor preta foram classificadas com baixa probabilidade de presença da espécie, podendo até apresentar probabilidade nula. Regiões representadas pela cor cinza apresentam alta probabilidade de presença da espécie. Por fim, regiões representadas por cores próximas do branco são regiões de fronteira entre altas e baixas probabilidades de presença da espécie, ou seja, probabilidades intermediárias.

6. Discussão e Trabalhos Futuros

O objetivo principal deste trabalho foi apresentar um algoritmo de modelagem de distribuição geográfica de espécies biológicas baseado em Redes Neurais Artificiais. O algoritmo apresentado está disponível no *openModeller*, um *framework* para modelagem que oferece diversos algoritmos de modelagem entre outras funcionalidades.

O algoritmo foi testado com dados de ocorrência da espécie *Stryphnodendron obovatum*, uma planta do cerrado brasileiro, e dados bioclimáticos. Os dados de ocorrência utilizados estavam divididos em pontos de presença e pontos de ausência, característica importante para a estratégia adotada no treinamento da rede neural.

Os modelos gerados foram testados com dados independentes, ou seja, dados que não faziam parte dos conjuntos de treinamento. Os resultados alcançados demonstraram que o algoritmo desenvolvido é adequado ao processo de modelagem, com taxas de acuidade de até 94%.

Quando pontos de ausência não estão disponíveis, o algoritmo utiliza um método de seleção de pontos de pseudo-ausência. Entretanto, tal estratégia pode gerar ruído nos dados, afetando as taxas de acuidade dos modelos criados. Além disso, ruído nos dados pode afetar o treinamento da rede, impossibilitando a convergência. Assim, uma proposta de trabalho futuro é o desenvolvimento de algoritmos de modelagem baseados em outros modelos neurais, capazes de aprender a partir de conjuntos de dados que contenham apenas pontos de presença.

A próxima etapa deste trabalho terá como objetivo a comparação do algoritmo desenvolvido com outras técnicas já difundidas na comunidade de modelagem, por exemplo, GARP, SVM e Entropia Máxima. É importante destacar que o desenvolvimento do algoritmo apresentado neste artigo é apenas uma etapa inicial de uma pesquisa mais abrangente que envolve, além de outras atividades, a comparação entre as técnicas de modelagem.

Referências

- Canhos, V. P. (2003) Informática para biodiversidade: padrões, protocolos e ferramentas. *Ciência e Cultura*. [online]. v. 55, n. 2, pp. 45-47. <http://cienciaecultura.bvs.br/pdf/cic/v55n2/15530.pdf>.
- Chopra, P. (2008) <http://www.paraschopra.com>. Downloaded in 2008.
- Fielding, A. H. and Bell, J. F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models, *Environmental Conservation* 24: 38 – 49.
- Haykin, S. (2001) *Redes Neurais – Princípios e Prática*. Trad. Paulo Martins Engel. – 2. ed. – Porto Alegre: Bookman.

- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965 – 1978.
- Lorena, A. C., Siqueira, M. F., Giovanni, R., Carvalho, A. C. P. L. F. and Prati, R. C. (2008) Potential Distribution Modelling Using Machine Learning. IEA/AIE '08: Proceedings of the 21st International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Wroclaw, Poland. Lecture Notes in Computer Science, Springer-Verlag Berlin Heidelberg, v. 5027, p. 255–264.
- Muñoz, M.E.S., Giovanni, R., Siqueira, M.F., Sutton, T., Brewer, P., Pereira, R.S., Canhos, D.A.L. & Canhos, V.P. (2009) "openModeller: a generic approach to species' potential distribution modelling". *GeoInformatica*. DOI: 10.1007/s10707-009-0090-7.
- Phillips, S. J.; Anderson, R. P.; Schapire, R. E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231 – 259.
- Rodrigues, E. S. C., Rodrigues, F. A., Rocha, R. L. A. and Corrêa, P. L. P. (2010) An Adaptive Maximum Entropy Approach for Modeling of Species Distribution. *Memórias do WTA 2010 – 4º Workshop de Tecnologia Adaptativa*. Escola Politécnica da Universidade de São Paulo, São Paulo. ISBN 978-85-86686-56-6.
- Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986) Learning representations by back-propagating errors. *Nature*, 323:533-536.
- Siqueira, M. F. (2005) Uso de modelagem de nicho fundamental na avaliação do padrão de distribuição geográfica de espécies vegetais. Tese de Doutorado. Escola de Engenharia de São Carlos da Universidade de São Paulo. São Carlos, São Paulo.
- Stockwell, D.; Peters, D. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, vol. 13, nº. 2, 143 – 158.
- Vapnik, V. (1995) *The Nature of Statistical Learning Theory*. Springer-Verlag.
- Thuiller, W., Lafourcade, B., Englerand, R. and Araújo, M. B. (2009). BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* 32: 369–373.
- Mastrorillo, S., Lek, S., Dauba, F. and Belaud, A. (1997). The use of artificial neural networks to predict the presence of small-bodied fish in a river. *Freshwater Biology* 38, 237–246.
- Ximenes, A. C., Amaral, S., Arcoverde, G. F. B., Monteiro, A. M. V. (2009). Redes neurais para a seleção de variáveis ambientais no processo de modelagem de distribuição de espécies na região Norte do Brasil. *Anais XIV Simpósio Brasileiro de Sensoriamento Remoto*, Natal, Brasil, 25-30 abril 2009, INPE, p. 5531–5538.
- Joshi, C., de Leeuw, J., Skidmore, A. K., van Andel, J., Lekhak, H. D. and van Duren, I. C. (2005). Remote Sensing and GIS for Mapping and Management of Invasive Shrub *Chromolaena odorata* in Nepal. *Proceedings of Eighth AGILE Conference on Geographic Information Science* Estoril, Portugal (2005), pp. 71–80.
- Jiménez-Valverde, A., Diniz, F., de Azevedo, E. B. and Borges, P. A. V. (2009) Species distribution models do not account for abundance: the case of arthropodes on Terceira Island. *Ann. Zool. Fennici* 46: 451-464.