

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Utilização de Algoritmo Genético e Redes Neurais Artificiais para Classificação de Blastocistos Bovinos.

Felipe J. Passalia¹

Laboratório de Matemática Aplicada (MaAP), Departamento de Ciências Biológicas, Unesp, Assis, SP

Felipe D. Matos²

Institut de Biologie de l'École Normale Supérieure (IBENS), Paris, Île-de-France

Marcelo F. G. Nogueira³

Laboratório de Micromanipulação Embrionário (LaME), Departamento de Ciências Biológicas, Unesp, Assis, SP

José C. Rocha⁴

Laboratório de Matemática Aplicada (MaAP), Departamento de Ciências Biológicas, Unesp, Assis, SP

Resumo. A classificação de blastocistos bovinos apresenta grande importância comercial, já que está diretamente ligada às taxas de gestação. A aplicação da técnica de RNA, juntamente com AG, pode ser uma solução para evitar a classificação subjetiva. Foi construído um software baseado em AG para obter arquiteturas ótimas de RNA.

Palavras-chave. Rede Neural Artificial, Algoritmo Genético, Blastocisto Bovino.

1 Introdução

Devido ao rebanho brasileiro ser estimado em cerca de 200 milhões de cabeças de gado, segundo o Ministério da Agricultura, a bovinocultura representa dois segmentos economicamente lucrativos, a carne e o leite. Com esta atividade sendo realizada em todo o território brasileiro, além do valor bruto de produção, a bovinocultura representa um importante papel social e econômico no país [7].

Embriões bovinos são introduzidos em animais quando chegam ao estágio de blastocisto e diversos fatores podem influenciar a taxa de gestação destes embriões. Na espécie bovina é recomendado pela Sociedade Internacional de Transferência de Embriões

¹ felipepassalia@hotmail.com

² delestro@biologie.ens.fr

³ marcelo@assis.unesp.br

⁴ jcelso@assis.unesp.br

(IETS) um sistema de classificação de embriões, o qual apresenta três graus de qualidade, bom, regular ou pobre [3].

A análise dos blastocistos bovinos é baseada na visualização da morfologia e depende da acurácia e da experiência do embriologista. Não mensurando variáveis objetivas, a determinação da classificação do embrião torna-se subjetiva e com pouca repetibilidade [2], no entanto, a aplicação de Redes Neurais Artificiais (RNA) como forma de reconhecimento de padrões [4, 5] pode evitar esta subjetividade.

A RNA é uma técnica baseada no modo como o neurônio humano transmite e processa informações. A arquitetura de RNA pode ser organizada na forma de multicamadas de neurônios tendo uma camada de entrada, camadas ocultas e uma camada de saída, sendo a saída representada como a resposta da rede, neste caso a rede é denominada *multilayer perceptron* (MLP), onde as camadas de entrada e saída são formadas pelo conjunto de dados, entrada/saída, apresentados a RNA. Através de um processo de aprendizado, neste caso supervisionado, a rede neural é capaz de obter uma arquitetura ótima para os padrões de entrada/saída propostos, podendo assim prever modelos futuros [1, 5].

A técnica de Algoritmo Genético (AG) é baseada em um processo biológico evolutivo [10], no qual os indivíduos que formarão a população inicial de cromossomos representam as arquiteturas de uma RNA, dentro de padrões pré-estabelecidos. Durante o processo evolutivo, está população, no caso as diversas RNAs, são treinadas, através do algoritmo *backpropagation*, e testadas, sendo as mais aptas, à solução do problema em questão, selecionadas. Após a seleção das mais aptas os operadores de crossover e mutação são aplicados gerando descendente até que uma nova população de igual tamanho da inicial seja estabelecida. Este procedimento é repetido quantas vezes o usuário estabelecer.

Como forma de solucionar a subjetividade da classificação de blastocistos bovinos, foi desenvolvida por Matos [6] uma técnica de extração de informações que serviram de padrões de entrada/saída para a Rede Neural Artificial, além da aplicação de Algoritmo Genético para a obtenção da melhor arquitetura de RNA para a solução do problema de classificação de blastocistos bovinos. No entanto, apesar de uma boa precisão quando analisado de forma geral, o software desenvolvido não apresentou uma acurácia suficientemente boa para a classificação do grau 2 de blastocistos bovinos.

A utilização de algoritmo genético para obtenção de arquiteturas de RNA ótimas apresenta resultados rápidos e eficientes para diversos problemas [8, 9], no entanto, durante o processo do AG estabelecido por Matos, os indivíduos (as arquiteturas de RNA) convergiram para um padrão que apresentaram resultados muito bons (79,2% de acerto total para as três classificações), porém, este padrão pode não representar a melhor solução, uma vez que para a classificação do grau 2, o acerto foi de apenas 16,7%, representando um mínimo local e não um mínimo global.

Em busca do mínimo global diversas alterações no cromossomo, para a composição da população, foram feitas, como por exemplo, alteração no número de camadas intermediárias, alteração da função de transferência em cada camada e na camada de saída, além do aumento na quantidade de funções de treinamento e transferências usadas, sendo agora sete funções para cada parâmetro. Ainda foram feitas alterações em relação à determinação dos melhores indivíduos, priorizando o acerto na classificação 2, sem que haja diminuição das classificações 1 e 3, não considerando mais o menor erro total no teste.

2 Propósito

O presente trabalho visou à obtenção de uma acurácia do software de classificação de blastocistos bovinos para a identificação do grau 2 (regular), sem que haja uma diminuição significativa da acurácia já obtida por Matos [6] para a identificação dos graus 1 e 3, bom e pobre, respectivamente.

O software desenvolvido utiliza a técnica de Algoritmo Genético, para a obtenção da melhor arquitetura de RNA possível, a fim de que a classificação de embriões bovinos, obtido por esta técnica, possa ser viável comercialmente, solucionando o problema da classificação subjetiva de blastocistos bovinos.

3 Metodologia

O software foi desenvolvido em ambiente *MatLab*[®] e baseado em um estrutura de “cromossomo”, cada configuração de uma determinada RNA é salva em uma *struct* como um valor numérico. O cromossomo (*struct*) apresenta os seguintes parâmetros para a configuração de arquitetura de RNA: quantidade de neurônios na primeira camada oculta; quantidade de neurônios na segunda camada oculta; quantidade de neurônios na terceira camada oculta; função de transferência para a primeira camada oculta; função de transferência para a segunda camada oculta; função de transferência para a terceira camada oculta; função de transferência para a camada de saída; função de treinamento a ser utilizada e quantidade de camadas ocultas a ser utilizada.

A figura 1, a seguir, ilustra a estrutura do cromossomo com seus parâmetros.

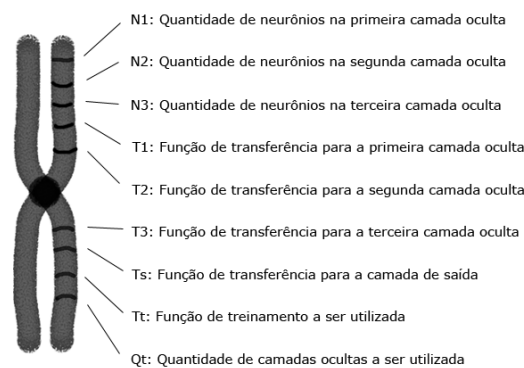


Figura 1: Estrutura do cromossomo utilizado pelo Algoritmo Genético.

Para iniciar o software, é preciso que o usuário determine alguns parâmetros no próprio ambiente *MatLab*[®], sendo eles: a quantidade de indivíduos que irá compor a população, um valor mínimo e máximo que representam o intervalo da quantidade de neurônios, em cada camada oculta, as variáveis de entrada e saída, a lista das funções de

transferência e de treinamento a serem utilizadas, um valor de mutação, a quantidade de gerações que o processo irá se repetir, além do valor que representa a quantidade de camadas ocultas das combinações de RNA geradas. Após a definição destes parâmetros, o algoritmo segue as seguintes etapas:

1. Constituição da população inicial: É gerada uma população inicial de indivíduos com valores e funções aleatórias, baseado em dados numéricos fornecidos pelo usuário, sendo que cada indivíduo representará uma arquitetura diferente de RNA. A quantidade de indivíduos a ser gerado é determinada pelo usuário, assim como as funções de transferência e treinamento, tais funções são definidas pelo usuário a partir de uma lista pré-definida em ambiente *Matlab*. Neste projeto a lista consta com as seguintes funções de transferência: *logsig*, *purelin*, *tansig*, *hardlim*, *tribas*, *radbas* e *satlin* e as seguintes funções de treinamento: *trainrp*, *trainscg*, *traincgb*, *traincgf*, *traincgp*, *traingdm*, *traingd*. [1]

Assim como as funções de transferência e treinamento, o número de neurônios para cada camada é sorteado aleatoriamente a partir de um intervalo pré-definido pelo usuário, sendo entre 10 e 100 neurônios neste caso, e a quantidade de camadas ocultas é escolhida aleatoriamente entre o intervalo de um e três para esta pesquisa [4], podendo ser ampliado pelo usuário caso necessário.

2. Teste, seleção e migração de população: Após a geração das RNA (os indivíduos), a população toda é treinada e testada utilizando os padrões de entrada/saída que é dividido entre treinamento (60% a 70% dos dados), validação e teste (15% a 20%). Após o teste, a população toda é analisada e caso nenhum indivíduo tenha um acerto maior que 30% para as classificações de grau 1, 2 e 3, é gerada uma população de migração. A partir destas duas populações, são selecionados aleatoriamente progenitores da população inicial e da população de migração e estes progenitores serão responsáveis pelo repovoamento de uma nova população. Cada característica é escolhida aleatoriamente entre os genitores para formar uma nova *struct* (crossover) e, neste processo, há a possibilidade de ocorrer uma mutação (taxa de 5%), possibilitando uma variabilidade genética.

3. Replicação e repovoamento: Na geração seguinte, os indivíduos que foram os mais aptos, para o nosso caso, com acerto maior do que 30% para as classificações de grau 1, 2 e 3, serão os responsáveis pela geração de uma nova população. Destes selecionados, são escolhidos aleatoriamente indivíduos, os genitores, que através dos operadores de crossover e mutação, irão compor as *structs* da nova população, assim como a replicação da etapa anterior. O processo de geração de novos indivíduos continua até que quantidade de indivíduos da nova população se iguale a população inicial

4. Obtenção da melhor arquitetura: O usuário determina a quantidade de gerações (iterações) que o processo terá, sendo que em cada iteração, os melhores indivíduos (acerto maior que 30% para as classificações de grau 1, 2 e 3) são selecionados e servirão de base para a recombinação das próximas gerações, além da melhor rede ser salva ao final de cada iteração, para posterior comparação. A determinação da melhor rede é baseada na rede com melhor acerto para a classificação grau 2 sem que os acertos para os graus 1 e 3 diminuam em relação à melhor rede selecionada anteriormente.

4 Resultados

Neste trabalho, o software foi usado para resolver um problema de classificação de blastocistos bovinos, apresentando três possíveis respostas, analisadas por uma matriz de confusão. O principal problema era o resultado para a classificação de grau 2 de blastocistos bovinos, e por isto, o software desenvolvido iniciava o processo de seleção da melhor RNA apenas quando havia acerto maior ou igual a 30% para todas as três classes. A partir dessa primeira RNA selecionada a próxima rede só seria selecionada se a classificação de grau 2 fosse maior que a anterior e desde que as classificações de grau 1 ou de grau 3 fosse no máximo 5% menor do que a RNA anterior. Desta forma o melhor resultado obtido para a classificação de blastocistos bovinos foi de 70,8% de acerto no teste para a classificação de grau 1, 41,7% para a classificação de grau 2 e 80% para a classificação do grau 3. Todos os resultados foram obtidos a partir de uma matriz de confusão, gerada a partir dos dados de teste de cada arquitetura de RNA. Na figura 2 a seguir estão as matrizes de confusão geradas ao final das etapas do algoritmo, sendo que para o resultado b, o algoritmo não realizou a etapa de migração.

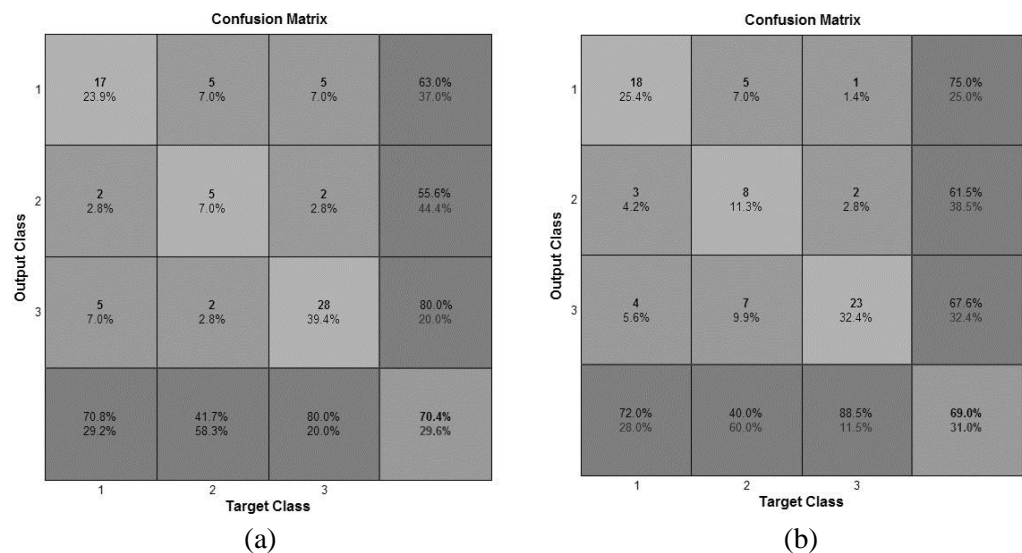


Figura 2: Matrizes de confusão para os resultados de teste da RNA.

5 Discussão

O trabalho iniciado por Matos [6] apresentou bons resultados para a classificação de blastocistos bovinos, no entanto, não obteve acurácia para a classificação de grau 2 de blastocistos (acerto de 16,7% para o grau 2). Como forma de melhorar os resultados obtidos por Matos, sem aumentar a quantidade de padrões de entrada/saída para a RNA, foram feitas modificações na metodologia utilizada para o Algoritmo Genético e para a obtenção da arquitetura de RNA, apresentando resultados significativamente melhores (41% de acerto

para o grau 2).

Uma das dificuldades que tivemos na construção do software foi manter os bons resultados já obtidos por Matos [6] para a classificação de blastocistos bovinos de graus 1 e 3. Isto só foi possível quando delimitamos um valor mínimo de acerto para todas as classificações dos blastocistos, no caso 30%, fazendo que a técnica de Algoritmo Genético buscasse efetivamente uma melhor RNA para os parâmetros previamente fornecidos.

As diferenças entre os resultados obtidos por Matos [6] e o desta pesquisa para a classificação de blastocistos bovinos pode ser observada através das matrizes de confusão representadas nas figuras 3, onde apresentamos a matriz de confusão obtida por Matos e a obtida nesta pesquisa, respectivamente.

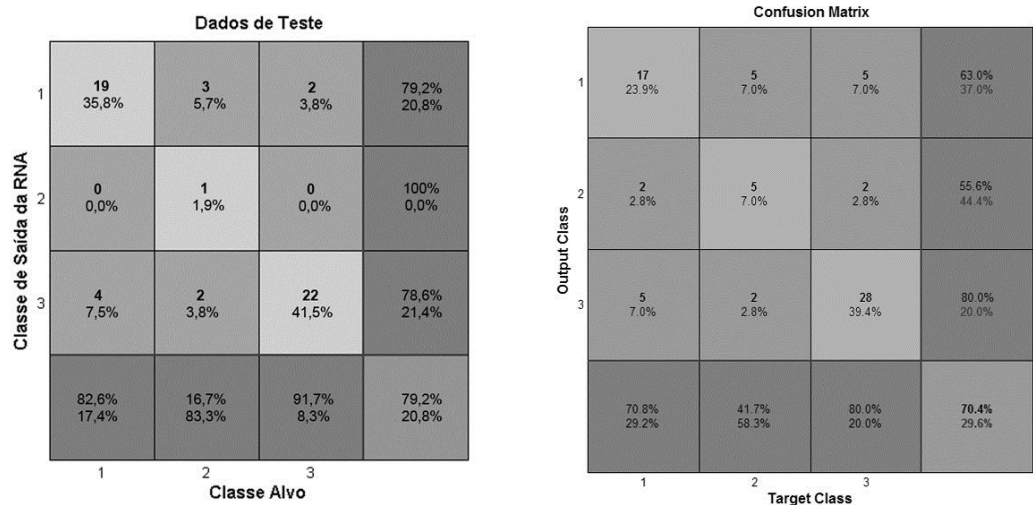


Figura 3: Comparação dos valores das matrizes de confusão.

6 Conclusão

Os resultados obtidos com a aplicação de técnica de Algoritmo Genético e RNA para a classificação de blastocistos bovinos, mostraram uma melhoria significativa em relação a classificação de grau 2, de 16,7% para 41,7%, objetivo deste estudo. Vale ressaltar aqui que mesmo embriologistas experientes tem enorme dificuldade na classificação precisa de grau dois para embriões bovinos.

Com os bons resultados obtidos a aplicação das técnicas de AG e RNA para a classificação de embriões bovinos mostram-se promissoras, proporcionando uma nova metodologia para esta classificação, uma vez que até os dias de hoje, elas são feitas de forma subjetiva, com pouca repetibilidade, baseada na visualização da morfologia do embrião e dependente da experiência do embriologista.

Referências

- [1] M. H. Beale, M. T. Hagan and H. B. Demuth, Neural Network Toolbox™ User's Guide, vol. 8.3, (2015).
- [2] B. Bényei, I. Komlósi, A. Pécsi, G. Pollott, C. H. Marcos, A. de Oliveira Campos, M. P. Lemes, The effect of internal and external factors on bovine embryo transfer results in a tropical environment, *Anim Reprod Sci*, v. 93, 268-79, (2006).
- [3] G. Bó, and R. Mapletoft, Evaluation and classification of bovine embryos. *Anim. Reprod*, v. 54, n. 935, 344–348, (2013).
- [4] P. L. M. Goethals, A. P. Dedecker, W. Gabriels, S. Lek and N. De Pauw, Applications of artificial neural networks predicting macroinvertebrates in freshwaters. *Aquatic Ecology*, v. 41, n. 3, 491–508, (2007), DOI: 10.1007/s10452-007-9093-3.
- [5] S. Haykin, *Redes Neurais: princípios e prática* 2ª ed. Porto Alegre, RS, Bookman. (2001).
- [6] F. D. Matos, Software baseado em rede neural artificial desenvolvida por meio de algoritmo genético para a classificação morfológica de blastocistos bovinos, *Dissertação de Mestrado em Biociências*, Unesp, (2014).
- [7] Ministério da Agricultura, Bovinos e Bubalinos, Disponível em: <<http://www.agricultura.gov.br/animal/especies/bovinos-e-bubalinos>>, Acesso em 28 dez. (2014).
- [8] J. C. Rocha, F. D. Matos and F. Frei, Utilização de redes neurais artificiais para a determinação do número de refeições diárias de um restaurante universitário, *Rev. Nutr.* vol. 24, n. 5, 735-742, (2011).
- [9] M. B. Takahashi, J. Leme, C. P. Caricati, A. Tonso, E. G. F. Núñez and J. C. Rocha, Artificial neural network associated to UV/Vis spectroscopy for monitoring bioreactions in biopharmaceutical processes, *Bioprocess and biosystems engineering*, (2015), DOI:10.1007/s00449-014-1346-7.
- [10] J. Tanomaru, *Motivação, fundamentos e aplicações de algoritmos genéticos. II Congresso Brasileiro de Redes Neurais*, Curitiba, (1995).