

Previsão de Casos de Dengue no Município de Guarulhos com Redes Neurais Artificiais Multicamadas e Recorrentes

MunIQUE MittELMANN¹, DANIEL GOMES SOARES².

Resumo

A dengue é uma doença transmitida pela picada de mosquitos do gênero *Aedes*, sobretudo pela espécie *Aedes aegypti*. Ao possibilitar a tomada de ações preventivas para um controle efetivo da doença, a previsão da incidência da dengue adquire um papel social indispensável. Dentre as técnicas que podem ser utilizadas na elaboração de modelos de previsão, incluem-se as Redes Neurais Artificiais (RNAs). O propósito deste trabalho é desenvolver um modelo de previsão de dengue com RNAs para o município de Guarulhos – São Paulo. Foram modeladas e treinadas redes com diferentes parâmetros iniciais. Os resultados mostram que é possível realizar a previsão de dengue na área de estudo por meio de RNAs com erro e antecedência aceitáveis.

Palavras-chave: Previsão de Dengue; Redes Neurais Artificiais; Inteligência Artificial.

Abstract

Dengue is a disease transmitted by female *Aedes* mosquitoes, mainly by the specie *Aedes aegypti*. Being possible to take preventive actions for effective control of the disease, a prediction of the incidence of dengue acquires an indispensable social role. Among the techniques that can be used in the elaboration of prediction models, such as Artificial Neural Networks (ANNs). In this work, is developed a predictive model of dengue cases with ANNs in Guarulhos – São Paulo. The neural networks were modeled and trained with different initial bases. The results showed that ANNs were very effective in predicting the number of dengue confirmed-cases.

Keywords: Dengue Forecasting; Artificial Neural Networks; Artificial Intelligence.

¹Universidade Federal de Santa Catarina, Florianópolis - SC, Brasil, munyque.mittelmann@posgrad.ufsc.br

²Instituto Federal Catarinense, Rio do Sul - SC, Brasil, daniel.soares@ifc.edu.br

1 Introdução

Apesar da dengue ser um problema de saúde pública mundial, países tropicais, como o Brasil, são os mais atingidos pela doença em virtude das suas características ambientais, climáticas e sociais (RIBEIRO et al., 2006). De acordo com Gharbi et al. (GHARBI et al., 2011), um alerta precoce de surtos de dengue pode aumentar a eficácia de campanhas de controle do vetor e contribuir no direcionamento de ações preventivas. Deste modo, intervenções precoces tem a capacidade de atrasar a epidemia e assim reduzir seu impacto no sistema de saúde. Phung et al. (PHUNG et al., 2015) complementam que modelos de previsão podem ser uma ferramenta válida na preparação e controle de uma epidemia. Neste cenário, as instituições de saúde poderiam organizar a disponibilidade de leitos e a mobilização de recursos humanos e materiais. Estes fatores possibilitariam uma resposta mais adequada dos órgãos de saúde pública e, conseqüentemente, a minimização da taxa de mortalidade da doença (GHARBI et al., 2011).

Ao possibilitar a tomada de ações preventivas para um controle efetivo da doença, a previsão da incidência da dengue adquire um papel social indispensável. Uma das estratégias de previsão de séries temporais que vêm sendo aplicadas atualmente a este problema inclui a utilização de modelos univariados (DOM et al., 2013; EARNEST et al., 2012; PROMPROU; JAROENSUTASINEE; JAROENSUTASINEE, 2006). Modelos univariados baseiam-se apenas em uma única variável de série temporal para realizar a predição. Apesar da utilidade desta metodologia em situações nas quais há indisponibilidade de dados, o comportamento de uma série temporal pode depender de múltiplas variáveis. Nesta perspectiva, estudos evidenciam a relação de variáveis climáticas com a incidência de dengue (RIBEIRO et al., 2006; LIMA; FIRMINO; FILHO, 2008). Deste modo, a inclusão de informações climáticas em modelos de previsão de casos de dengue pode aumentar sua eficácia.

Redes Neurais Artificiais (RNAs) são uma alternativa aos métodos tradicionais na solução de problemas de previsão de séries temporais. Esta técnica permite múltiplas variáveis nas camadas de entrada e saída. Deste modo, é possível utilizar variáveis climáticas distintas na modelagem da previsão de casos de dengue aumentando, dessa forma, a eficácia do modelo predictor. Além disso, RNAs possuem a capacidade de mapear relacionamentos incertos entre variáveis (ADNAN et al., 2014). Ou seja, para construir um modelo de previsão com RNAs, não é necessário compreender detalhadamente as complexas relações entre casos de dengue e as demais variáveis do ambiente.

O uso de RNAs na previsão de séries temporais vem crescendo nos últimos anos. Modelos de previsão que utilizam esta técnica têm apresentado resultados com baixo erro em diversos tipos de aplicação, inclusive previsão de dengue, conforme pode ser observado no estudo de (ABURAS; CETINER; SARI, 2010).

Em um trabalho anterior (MITTELMANN; SOARES, 2017), foi descrito um modelo de previsão de casos de dengue com RNAs para o município de Itajaí, Santa Catarina. Neste trabalho, busca-se modelar a abordagem para uma área de estudo com maior incidência da doença e, conseqüentemente, maior necessidade de desempenhar uma previsão acurada. Devido às características de alta urbanização e a alta incidência de dengue, optou-se pela cidade de Guarulhos (São Paulo). Outra diferença do estudo em (MITTELMANN; SOARES, 2017) é que os dados incluem apenas casos autóctones de dengue, ou seja, os casos da doença que surgiram fora do município são desconsiderados. Este trabalho compara a utilização de diferentes arquiteturas de Redes Neurais, especificamente, redes Perceptron Multicamadas (MLP) e Redes Neurais Auto-Regressivas com Entradas Exógenas (NARX) no estudo de caso da cidade de Guarulhos.

Este trabalho está organizado da seguinte maneira: a Seção 2 apresenta conceitos de RNAs. A Seção 3 descreve a metodologia

deste estudo. A Seção 5 descreve as RNAs propostas. A seção 6 descreve os resultados obtidos. Por fim, a Seção 7 discute as considerações finais.

2 Redes Neurais Artificiais

Uma RNA pode ser definida como um processador paralelamente distribuído constituído de unidades de processamento simples, os neurônios artificiais, os quais possuem a capacidade de armazenar conhecimento experimental e disponibilizá-lo para uso (HAYKIN, 2001). Segundo Loesch e Sari (LOESCH; SARI, 1996), os neurônios artificiais, que são emulações simplificadas de neurônios biológicos, recebem informações de sensores ou de outros neurônios artificiais, em seguida produzem operações simples sobre estes dados e então passam o resultado para outros neurônios artificiais.

2.1 Redes Multilayer Perceptron

Redes Multilayer Perceptron (MLP) pertencem à arquitetura neural conhecida como *feedforward* de camadas múltiplas, sendo caracterizadas pela presença de pelo menos uma camada intermediária (escondida) de neurônios, situada entre a camada de entrada e a camada neural de saída (SILVA; SPATTI; FLAUZINO, 2010).

Os sinais são apresentados à rede em sua camada de entrada. As camadas intermediárias extraem a maioria das informações referentes ao seu comportamento e as codificam através dos pesos sinápticos e limiares de seus neurônios (SILVA; SPATTI; FLAUZINO, 2010). O processamento realizado por cada neurônio de uma determinada entrada é definido pela combinação dos processamentos realizados pelos neurônios da camada anterior que estão conectados a ele (BRAGA; CARVALHO; LUDERMIR, 2011). Os neurônios da camada de saída da rede recebem os estímulos vindos dos neurônios da última camada intermediária, produzindo um padrão de resposta que será a saída disponibilizada pela rede.

O processo de treinamento de redes MLP normalmente é realizado com o algoritmo *backpropagation*, que utiliza pares de entrada e saída para, por meio de um mecanismo de correção de erros, ajustar os pesos e limiares da rede (BRAGA; CARVALHO; LUDERMIR, 2011).

2.2 Redes NARX

Redes auto-regressivas não-lineares com entradas exógenas (*Nonlinear Auto-regressive with eXogenous inputs* - NARX) pertencem à arquitetura neural recorrente ou realimentada. Estas redes permitem a recuperação de respostas passadas a partir da realimentação de sinais produzidos em instantes anteriores (SILVA; SPATTI; FLAUZINO, 2010). Em uma NARX, a predição de valores futuros associados ao comportamento do processo, a partir do instante t , será também baseada em função dos valores anteriores que foram produzidos por suas saídas. Uma rede NARX executa, implicitamente, o mapeamento entre entradas e saídas de processos que sejam tanto não-lineares como também variantes no tempo. O treinamento de uma rede NARX é efetuado de maneira similar ao de uma MLP, ou seja, promovendo o ajuste dos pesos para minimizar o erro entre o valor esperado e a resposta estimada pela rede (SILVA; SPATTI; FLAUZINO, 2010).

3 Metodologia

3.1 Área de Estudo

A cidade de Guarulhos pertence à Grande São Paulo, região economicamente mais importante do Brasil. O Instituto Brasileiro de Geografia e Estatística (IBGE) estima que a população total de Guarulhos no ano de 2015 foi de 1.324.781 habitantes (IBGE, 2016). A densidade demográfica do município, no ano de 2010, foi de 3.832,51 habitantes/km². É uma cidade completamente urbanizada, tendo grau de urbanização estimado em 100%.

O município de Guarulhos é acometido pela dengue desde a década de 90, período no qual,

provavelmente, houve a instalação definitiva e do *Aedes aegypti* no Estado de São Paulo. Um estudo da dengue neste município entre os anos de 2002 a 2005 observou que, nos meses de fevereiro a maio, o número de casos da doença supera a média anual, enquanto que entre junho a janeiro a quantidade de casos permanece abaixo da média (MANSO, 2006).

3.2 Dados

As amostras de dados de Guarulhos – São Paulo (SP) são constituídas de cinco atributos de entrada e um de saída. Os dados meteorológicos disponíveis na região foram obtidos através do Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP), do INMET (INMET, 2016). Os dados disponibilizados são da estação de medição automática de Guarulhos.

A base de dados meteorológicos consiste em dados mensais de janeiro de 2009 a novembro de 2014 referentes a precipitação total (mm), temperatura máxima média (°C), temperatura compensada média (°C), temperatura mínima média (°C) e umidade relativa média(%).

A quantidade mensal de casos confirmados de dengue em Guarulhos foi obtida através do portal do Centro de Vigilância Epidemiológica “Prof. Alexandre Vranjac” (CVE) e corresponde ao período de janeiro de 2009 a novembro de 2014 (CVE, 2016). Para este estudo, foram selecionados apenas os dados quantitativos referentes aos casos autóctones de dengue na cidade, com o objetivo de alcançar maior correlação com as variáveis climáticas da região.

3.3 Normalização dos Dados

Com o objetivo de escalonar as amostras de dados para a faixa de valores das funções de ativação das camadas escondidas da RNA, os dados foram normalizados para os intervalos [0;1] e [-1;1], por meio das seguintes equações, respectivamente:

$$y_1 = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$$y_2 = \frac{2 \times (x - x_{min})}{x_{max} - x_{min}} - 1$$

Nas quais, y_1 é o valor normalizado entre [0;1], y_2 é o valor normalizado entre [-1;1], x é o valor original, x_{max} é o maior valor da amostra de dados e x_{min} é o menor valor.

3.4 Validação Cruzada por Amostragem Aleatória

Após a normalização dos dados, foi utilizada a técnica de validação cruzada por amostragem aleatória. Deste modo, 70% do conjunto total de dados disponíveis foi aleatoriamente escolhido para o subconjunto de treinamento. Por sua vez, os dados restantes (30%) integraram o subconjunto de teste e validação.

4 Métricas para análise da qualidade da previsão

Os índices para avaliação da qualidade da previsão adotados neste trabalho foram o RMSE (Raiz do Erro Quadrático Médio) e o MAPE (Erro Médio Percentual Absoluto).

O RMSE pode ser calculado por:

$$RMSE = \sqrt{\frac{\sum_{k=1}^n (t_k - o_k)^2}{n}}$$

Por sua vez, o MAPE é definido pela seguinte equação:

$$MAPE = \frac{1}{n} \times \sum_{k=1}^n \frac{|t_k - o_k|}{(t_k + o_k)/2} \times 100\%$$

Em ambas as equações, t_k representa a saída esperada na amostra k (valor alvo ou observado), o_k é a saída da rede (valor previsto) para a amostra k e n é o número total de amostras.

5 RNAs propostas

As RNAs construídas neste trabalho possuem seis entradas, uma camada intermediária com n neurônios e um neurônio na camada de saída. A Figura 1 ilustra o modelo de rede MLP empregado para a previsão de casos de dengue em Guarulhos.

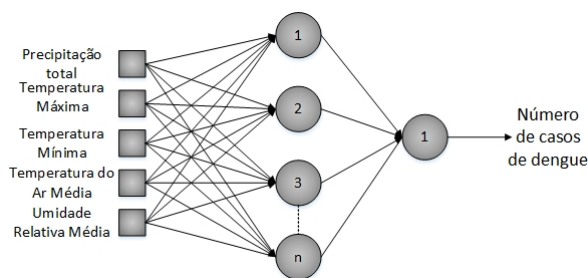


Figura 1: Modelo de Rede MLP

Nas redes NARX modeladas, optou-se pela utilização de dois atrasos nos valores de entrada e dois atrasos de realimentação, após a realização de alguns testes preliminares com atrasos maiores houve a constatação de que nestes não havia melhorias significativas nos resultados das redes.

A Figura 2 apresenta o modelo de rede NARX adotado, com dois atrasos de entrada e dois atrasos de realimentação, nos quais a saída estimada da rede nos momentos anteriores também é utilizada como parâmetro de entrada da rede no momento atual.

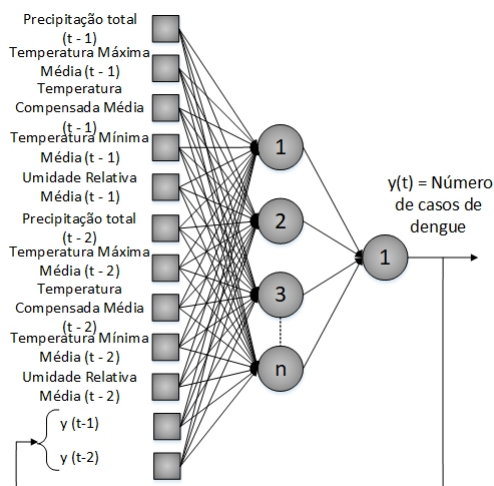


Figura 2: Modelo de Rede NARX

Além das redes descritas, foram modeladas RNAs com 15 variações do conjunto de entrada. Por exemplo, algumas redes não incluíram como entrada dados de todas as variáveis ou adicionaram informações referentes ao número de casos de dengue dos meses anteriores.

6 Resultados

Para a previsão de casos de dengue em Guarulhos, foram treinadas e testadas um total de 704 redes MLP e a mesma quantidade de redes NARX. Além do conjunto de entrada das RNAs, nestes treinamentos e testes foram alterados os seguintes parâmetros das redes: a função de ativação da camada oculta, o algoritmo de treinamento e a quantidade de neurônios na camada intermediária (5 a 15 neurônios). Alguns testes foram realizados com menos de cinco e mais de quinze neurônios na camada oculta, entretanto, a rede não produziu melhorias significativas nos resultados, apenas aumentava consideravelmente o tempo de treinamento.

Além disso, cada uma destas configurações de RNA, tanto MLP como NARX, foram treinadas 50 vezes, com o objetivo de minimizar os efeitos causados pela aleatoriedade das condições iniciais.

Em relação às redes MLP, a rede que obteve o melhor RMSE no conjunto de testes ($4,11 \times 10^{-3}$) foi treinada com o algoritmo de Regularização Bayesiana, função de ativação logística e 10 neurônios na camada oculta. Os sinais de entrada desta rede foram compostos apenas pela quantidade de casos de dengue dos dois meses anteriores. Esta rede obteve MAPE de 0,8613%. A Figura 3 apresenta a comparação entre a saída calculada pela RNA e o número de casos mensais observados.

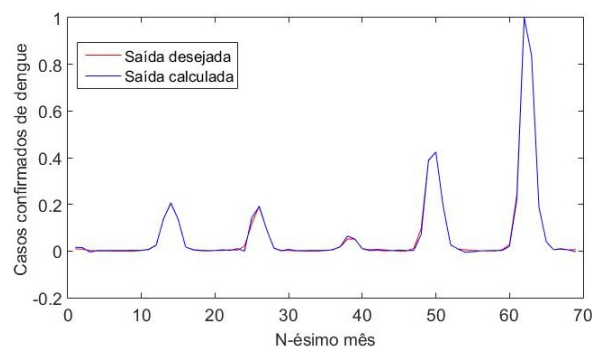


Figura 3: Rede MLP que obteve melhor RMSE

Em relação as redes NARX, o melhor

RMSE no conjunto de testes foi de $4,74 \times 10^{-3}$. A rede que obteve este RMSE foi treinada com 8 neurônios na camada oculta, função de ativação logística e com o algoritmo Levenberg-Marquardt. O conjunto de entrada desta rede incluiu os dados de precipitação, temperatura máxima, temperatura compensada, temperatura mínima e umidade. As entradas da rede possuíam um mês de atraso em relação a saída (número de casos de dengue de um mês a frente). Esta rede obteve MAPE de 1,1925%. A Figura 4 apresenta a saída calculada por esta RNA em relação a saída desejada. Pela observação gráfica, é possível perceber uma maior taxa de erros apenas quando o número de casos de dengue era baixo.

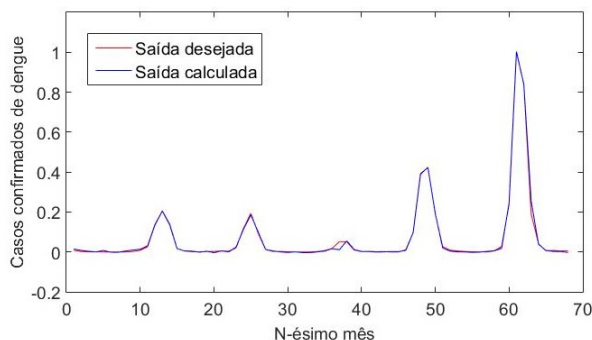


Figura 4: Rede NARX que obteve melhor RMSE

Comparando-se as redes MLP e NARX, pode-se observar que a rede MLP obteve um desempenho superior à rede NARX em relação ao RMSE e ao MAPE. Ambas as alternativas de rede neural não tiveram erros significativos na previsão, entretanto a RNA mais aderente à área de estudo foi a rede MLP cujo conjunto de entrada foi composto pela quantidade de casos de dengue dos dois meses anteriores.

7 Considerações

Este trabalho descreve a modelagem de RNAs para o problema de previsão de casos de dengue em Guarulhos - SP. Além disso, este trabalho compara a utilização de duas arquiteturas distintas de RNAs, as redes Perceptron Multicamadas e as Redes Neurais Auto-Regressivas

com Entradas Exógenas. As redes MLPs e NARXs foram comparadas entre si para eleição da rede com melhor desempenho na previsão de dengue na área de estudo. Assim, o melhor desempenho foi obtido pela MLP modelada com 10 neurônios na camada oculta, função de ativação logística e algoritmo de treinamento de Regularização Bayesiana. O conjunto de entrada que obteve este resultado foi composto apenas pela quantidade de casos de dengue dos dois meses anteriores ao mês em que se realizava a previsão. Nota-se que as variáveis climáticas não foram as mais adequadas para estimar a quantidade de casos de dengue, pois a rede treinada com informações sobre casos passados de dengue teve um desempenho superior.

Os resultados apresentados pelas redes mostram que é possível fazer a previsão de casos de dengue na área de estudo, utilizando RNAs com dados meteorológicos e/ou históricos da quantidade de casos de dengue de meses anteriores. No entanto, cabe salientar que como toda modelagem matemática, a generalização dos resultados obtidos em um estudo de caso específico não pode ser utilizada diretamente para outros casos, pois se deve levar em consideração que as características de cada região são únicas, sendo a utilização dos modelos construídos neste trabalho restrita a área de estudo selecionada.

Para trabalhos futuros, sugere-se o desenvolvimento de um portal que aplique o modelo, visando disponibilizar informações que possam auxiliar a população e os órgãos públicos na preparação e controle de uma possível epidemia de dengue.

Referências

- ABURAS, H. M.; CETINER, B. G.; SARI, M. Dengue confirmed-cases prediction: A neural network model. *Expert Systems with Applications*, Elsevier Ltd, v. 37, n. 6, p. 4256–4260, 2010. ISSN 09574174. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2009.11.077>>.
- ADNAN, R. et al. 5 hours flood prediction modeling using improved NNARX structure:

- Case study Kuala Lumpur. *Proceedings of the 2014 IEEE 4th International Conference on System Engineering and Technology, ICSET 2014*, n. NOVEMBER, 2014.
- BRAGA, A. d. P.; CARVALHO, A. P. d. L. F. d.; LUDERMIR, T. B. *Redes Neurais Artificiais: Teoria e Aplicações*. 2. ed. Rio de Janeiro: LTC, 2011. 226 p.
- CVE. *Centro de Vigilância Epidemiológica "Prof. Alexandre Vranjac"*. 2016. Disponível em: <<http://www.cve.saude.sp.gov.br>>.
- DOM, N. C. et al. Generating temporal model using climate variables for the prediction of dengue cases in Subang Jaya, Malaysia. *Asian Pacific Journal of Tropical Disease*, v. 3, n. 5, p. 352–361, 2013. ISSN 22221808.
- EARNEST, A. et al. Comparing statistical models to predict dengue fever notifications. *Computational and Mathematical Methods in Medicine*, v. 2012, 2012. ISSN 1748670X.
- GHARBI, M. et al. Time series analysis of dengue incidence in Guadeloupe, French West Indies: forecasting models using climate variables as predictors. *BMC infectious diseases*, v. 11, n. August 2015, p. 166, 2011. ISSN 1471-2334.
- HAYKIN, S. *Redes neurais: princípios e práticas*. 2. ed. Porto Alegre: Bookmann, 2001.
- IBGE. *Guarulhos*. 2016. Disponível em: <<http://cidades.ibge.gov.br/xtras/perfil.php?lang=%5C&codmun=351880>>.
- INMET. *Dados Históricos*. 2016. Disponível em: <<http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep>>.
- LIMA, E. D. A.; FIRMINO, J. L. D. N.; FILHO, M. F. G. A relação da previsão da precipitação pluviométrica e casos de dengue nos estados de Alagoas e Paraíba nordeste do Brasil. *Revista Brasileira de Meteorologia*, v. 23, n. 3, p. 264–269, 2008.
- LOESCH, C.; SARI, S. T. *Redes Neurais Artificiais: Fundamentos e Modelos*. 1. ed. Blumenau: Editora da FURB, 1996. 166 p.
- MANSO, W. *Estudo epidemiológico da dengue no período de 2000 a 2005 no município de Guarulhos (SP)*. [S.l.]: Dissertação (Mestrado em Medicina Veterinária). Departamento de Medicina Veterinária Preventiva e Saúde Animal, Universidade de São Paulo., 2006. 85 p.
- MITTELMANN, M.; SOARES, D. G. Previsão de Casos de Dengue em Itajaí - SC por meio de Redes Neurais Artificiais Multicamadas e Recorrentes. In: *Anais do Computer on the Beach [recurso eletrônico]*. Florianópolis: [s.n.], 2017. p. 130–139.
- PHUNG, D. et al. Identification of the prediction model for dengue incidence in Can Tho city, a Mekong Delta area in Vietnam. *Acta Tropica*, v. 141, n. August 2015, p. 88–96, 2015. ISSN 0001706X. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0001706X14003179>>.
- PROMPROU, S.; JAROENSUTASINEE, M.; JAROENSUTASINEE, K. Forecasting dengue haemorrhagic fever cases in Southern Thailand using ARIMA Models. *Dengue Bulletin*, v. 30, n. August, p. 99–106, 2006. ISSN 1020895X.
- RIBEIRO, A. F. et al. Associação entre incidência de dengue e variáveis climáticas. *Revista de Saude Publica*, v. 40, n. 4, p. 671–676, 2006. ISSN 00348910.
- SILVA, I. N. d.; SPATTI, D. H.; FLAUZINO, R. A. *Redes Neurais Artificiais para engenharia e ciências aplicadas*. São Paulo: Artliber, 2010. 399 p. ISBN 9788588098534.