

Redes Neurais Artificiais para o Preenchimento de Falhas em Séries Temporais Meteorológicas

Ferreira, J.A., Tapajós, R. P. e Conde, G. A. B.

Resumo— A maioria das séries temporais instrumentais são afetadas por uma certa proporção de dados ausentes devidos a problemas diversos que vão desde falhas de equipamentos até mesmo problemas com o fornecimento de energia. Este trabalho aborda a aplicação de técnicas baseadas em Redes Neurais Artificiais para a previsão de séries temporais meteorológicas como alternativa para o preenchimento de falhas nestas séries de forma robusta e eficiente no tratamento desses dados faltosos. Os resultados deste trabalho apontam para a aplicabilidade da utilização das técnicas de Redes Neurais como ferramentas eficientes para a previsão de séries temporais meteorológicas.

Palavras-Chave: *Séries Temporais Meteorológicas, Preenchimento de Falhas, Redes Neurais Artificiais.*

Abstract— Most instrumental time series are affected by a certain proportion of missing data due to various problems ranging from equipment failures even problems with the power supply. This paper discusses the application of techniques based on Artificial Neural Networks for forecasting meteorological time series as an alternative to the gap filling these robust and efficient series in the treatment of these faulty data. The results of this study point to the applicability of the use of neural network techniques as efficient tools for predicting meteorological time series..

Keywords: *Meteorological Time Series, Gap Filling, Artificial Neural Networks.*

I. INTRODUÇÃO

O estudo e análise da temperatura do ar é de grande relevância para o entendimento dos demais fenômenos climatológicos de uma região. Os elementos do clima estão diretamente relacionados com a formação e mudanças na flora e na fauna de determinada região, além de influenciar nas práticas do cotidiano das pessoas. As temperaturas máximas e mínimas estão associadas a outras variáveis meteorológicas, como: disponibilidade de energia solar, nebulosidade, umidade do ar e do solo, vento (direção e velocidade) e a parâmetros geográficos como topografia, altitude e latitude do local, além da cobertura e tipo de solo [10]. Os dados são coletados em estações meteorológicas equipadas com diversos sensores que coletam as mais diversas variáveis, como por exemplo a temperatura, que são armazenados ao longo de anos para que possam ser analisados posteriormente e passam a ser chamados de series temporais, pois mostram o comportamento dessas variáveis ao longo de um determinado período de tempo. Entretanto, é comum que essas series temporais de dados apresentem falhas, ou seja, períodos em que faltam

dados ou os dados não são validos. Pode-se apontar diversas razões para a leitura de dados com falhas, como por exemplo, interrupções ocasionais de estações automáticas, mau funcionamento dos instrumentos de medição, reorganização das redes de estações, falta de energia, etc. Uma forma de lidar com esse tipo de dificuldade é simplesmente excluir os períodos com dados faltantes das análises ou ignorar o problema, se a quantidade de falhas não for muito grande. Entretanto, esse tipo de abordagem desconsidera informações que podem ser relevantes para a análise dos dados e pode induzir vies no resultado final. Várias técnicas têm sido desenvolvidas nas últimas décadas para tratar do problema de dados faltantes em séries temporais meteorológicas mensais ou anuais. Métodos para manipular dados faltantes na escala diária são, de fato, escassos na literatura, além de pouco conhecidos pela maioria de pesquisadores de áreas mais aplicadas.

As Redes Neurais Artificiais (RNAs) tem sido empregadas com sucesso para resolver problemas diferentes e de características gerais, tendo como área de aplicação uma gama de bastante extensa. Entre as aplicações usuais das RNAs têm-se: reconhecimento e classificação de padrões, controle, agrupamento, aproximação de funções, otimização, processamento de sinais, análise de imagens, bioinformática, mineração de dados e, o foco deste trabalho, previsão de séries temporais [6], [7], [9].

Séries temporais são representadas por uma sequência de medições relativas a um determinado evento e organizadas cronologicamente, que aparentemente não seguem nenhuma lei ou tendência. No entanto, em séries temporais relacionadas a fenômenos naturais, econômicos ou físicos, pode-se notar certas características que se repetem após determinado período de tempo (sazonalidade) e outras que se mantêm durante o intervalo considerado (tendências), mesmo que não obedeçam a padrões lineares [2].

O objetivo deste trabalho é aplicar técnicas baseadas em RNAs para o preenchimento de falhas em bancos de dados meteorológicos diários, usando dados reais de temperatura e pressão atmosférica.

II. METODOLOGIA

2.1 OS DADOS

Os dados usados neste trabalho foram cedidos pelo laboratório de física da atmosfera da universidade federal do oeste do Pará – Ufopa, que realiza pesquisas a mais de uma

década dentro do Programa de Grande Escala da Biosfera-Atmosfera na Amazônia – LBA, que é um programa multidisciplinar que busca entender o funcionamento dos ecossistemas amazônicos em todas as suas vertentes e estudar o sistema amazônico como uma entidade regional no sistema Terra, assim como as causas e efeitos das mudanças em curso nesta região. Foram utilizados os dados da estação meteorológica localizada na cidade de Belterra – Pará, situada a beira da rodovia Santarém- Cuiabá, figura 1, que coleta dentre outras variáveis, a temperatura (coletada pelo sensor VAISALA modelo HMP45C) e pressão atmosférica (coletada por um Barômetro modelo VAISALA, PTB 101). Para este trabalho, foram selecionados os dados de temperatura e pressão durante os anos de 2000 a 2011. Para efeito de melhoria do processo de previsão a série temporal modelada teve como foco as médias diárias de temperatura e pressão durante os meses do período de seca amazônico compreendido durante os meses de agosto a dezembro dos anos selecionados para este trabalho.



Figura 1. Cidade de Belterra e sua região entorno.

Os dados utilizados neste trabalho, apresentaram um elevado índice de falhas para as variáveis temperatura e pressão atmosférica para o intervalo de tempo analisado, como pode ser visualizado na figura 2. Nesta figura, são mostrados os percentuais de falhas para os meses analisados durante o período de estudo. É importante perceber que a frequência com que ocorreram falhas nas leituras das variáveis temperatura e pressão é bastante elevada, ocorrendo em todos os meses e chegando a 100% no mês de agosto de 2008.

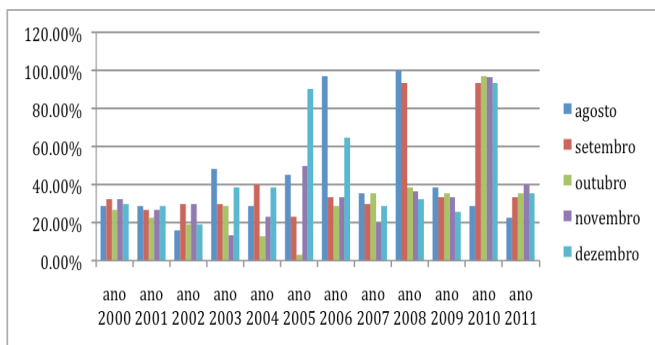


Figura 2. Percentual de falhas nos dados da série temporal analisada.

2.2 Redes Neurais Artificiais

Diversos métodos de inteligência artificial têm sido empregados em problemas de previsão de séries temporais, com diferentes níveis de sucesso. Entretanto, as RNAs têm sido muito aceitas como técnicas adotadas em sistemas de previsão [1],[9].

Para a construção do modelo de previsão é necessária a definição de três etapas principais: (a) identificação e definição da informação necessária para representar a série e o horizonte de previsão; (b) modelagem da RNA; (c) escolha do algoritmo de aprendizado da RNA capaz de aprender as características da série e realizar a previsão com sucesso.

2.2.1. Seleção das entradas

Durante o processo de identificação e definição das variáveis de entrada para o processo de previsão, a variável temperatura e pressão atmosférica foram escolhidas pelos especialistas do domínio devido a correlação entre essas. Além da variável temperatura e pressão, foram utilizadas as variáveis ano e dia, onde dia representa um determinado dia compreendido entre 1 e 366. Como os meses utilizados para a modelagem da série temporal foram os de agosto à dezembro o domínio da variável dia ficou delimitado entre os dias 212 (ago) e 366 (dez).

Partindo dos estudos realizados na série temporal aqui analisada, empregou-se a técnica de “janelamento” sobre a variável temperatura para realizar a previsão com RNAs. Essa técnica consiste em usar seqüências de valores passados (janela de tempo) da variável sobre a qual se deseja realizar a previsão bem como o horizonte de previsão, ou seja, os valores a serem previstos. A Figura 3 ilustra as entradas e saídas selecionadas para modelar as camadas de entrada e saída das RNAs projetadas neste trabalho. Após diversos experimentos com diferentes seqüências dos valores passados da variável temperatura, a janela de tempo que apresentou melhores resultados foi a de tamanho 6, ou seja, é passado para a RNA o valor da temperatura atual e cinco valores históricos.

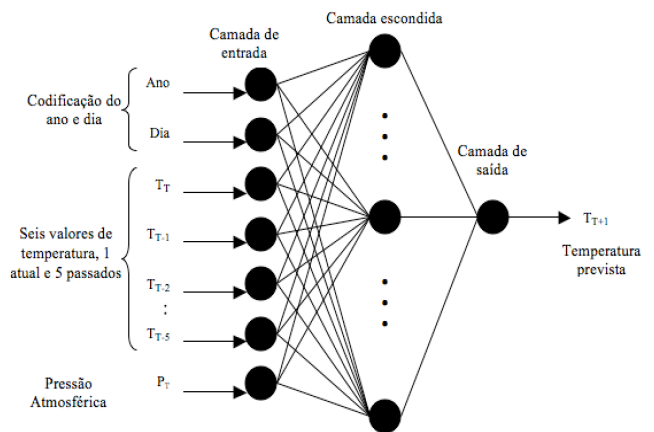


Figura 3. Arquitetura da RNA para a previsão de temperatura.

Após análise dos resultados iniciais com os especialistas, optou-se também em realizar as simulações com as RNAs sem levar em consideração a variável pressão.

2.2.2. Definição dos parâmetros das RNAs

Neste trabalho se optou por utilizar a rede perceptron multicamada feedforward (MLP), com apenas uma camada escondida, devido ao seu amplo emprego em sistemas de previsão [1].

Nesta etapa do processo de modelagem da rede MLP, o número de neurônios da camada de entrada e saída foram fixados de acordo com a seleção das entradas e saídas descritas no item 2.2.1. Entretanto, o número de neurônios na camada escondida é determinado dinamicamente durante o processo de treinamento da rede MLP, como será descrito no item 2.2.3.

2.2.3. Treinamento das redes MLP

O algoritmo de treinamento escolhido para o aprendizado das redes MLP foi o backpropagation com algumas modificações para o ajuste dinâmico dos parâmetros de configuração.

Durante a etapa de treinamento da rede MLP, o número de neurônios da camada escondida, a taxa de aprendizado da camada escondida e a taxa de aprendizado da camada de saída são variadas repetidamente, entre um mínimo e máximo definidos pelo usuário. Para cada combinação desses parâmetros é criada uma rede MLP e aplicado separadamente o algoritmo de treinamento backpropagation adaptativo. Desta forma, é identificada a melhor combinação de parâmetros (melhor rede MLP) para o conjunto de dados utilizado, que, por sua vez, provêem os resultados com os menores erros de previsão obtidos com durante o treinamento. Esse processo foi repetido para as simulações com as variáveis data, dia, temperatura e pressão e para as simulações com as variáveis data, dia e temperatura.

III. AVALIAÇÃO DOS RESULTADOS OBTIDOS

A avaliação dos resultados obtidos com a aplicação dos métodos de inteligência computacional foi realizada considerando a previsão de temperatura média para o instante T+1 através da aplicação de redes neurais, conforme detalhado a seguir, apontando também um comparativo para a aplicação das simulações considerando as informações de temperatura e pressão e para a consideração somente da temperatura. As performances dos modelos serão avaliadas de acordo com o erro percentual médio absoluto (MAPE) (equação 1).

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left(\frac{|y_i - \hat{y}_i|}{y_i} \right) \times 100\% \quad (1)$$

Onde N representa o número de exemplos, y o valor histórico real e \hat{y} o valor estimado.

De acordo com os parâmetros definidos para as RNAs na seção 2, foram realizadas simulações identificando e selecionando a melhor rede MLP. As simulações foram realizadas com dois conjuntos de dados, como descrito a seguir.

Primeiramente as redes MLP utilizaram para o conjunto de treino os valores dos dados históricos referentes ao intervalo de 2000 até 2010 com os valores de médias diárias de temperatura e pressão atmosférica para os meses de agosto à dezembro. O

ano de 2011 também fez parte do conjunto de treino, entretanto somente os meses de agosto à novembro compuseram os dados de treino, pois os dados referentes ao mês de dezembro de 2011 foram utilizados para o conjunto de teste. Para o mês de dezembro de 2011, dos 31 dias 10 (32.25%) não apresentam valores de temperatura nem pressão, ocasionando na série temporal falhas sobre essas informações. Após esse processo de treinamento, também foram realizadas simulações com as mesmas características em termos de intervalo da série temporal mas sem os dados referentes a variável pressão atmosférica. Assim, a composição dos conjuntos de treino e teste do segundo conjunto de simulações tiveram a mesma composição das primeiras mas sem os dados de pressão atmosférica. Nas figuras 4 e 5 são mostrados as previsões dos conjuntos de testes para as primeiras e segundas simulações, respectivamente, de acordo com as melhores RNAs encontradas durante o processo de aprendizado.

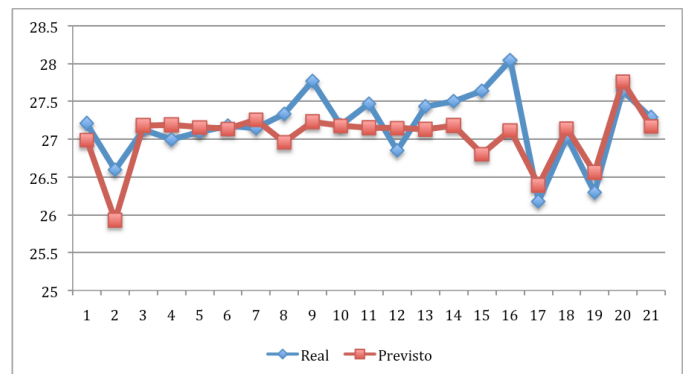


Figura 4. Valores reais e estimados de temperatura para o conjunto teste utilizando as variáveis de temperatura e pressão atmosférica.

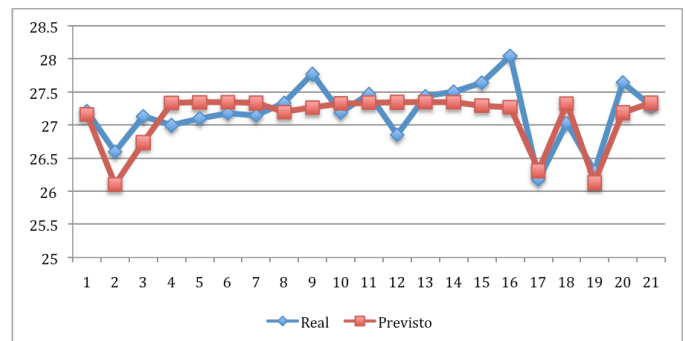


Figura 5. Valores reais e estimados de temperatura para o conjunto teste utilizando a variável de temperatura.

Como pôde ser observado, os resultados obtidos pelas estimações com e sem a utilização dos dados de pressão atmosférica apresentaram erros MAPE de aproximadamente 6.82% e 6.46% respectivamente, para os arquivos testes. Embora os resultados mostram que os erros obtidos em ambos os conjuntos simulações não foram estatisticamente diferentes, nota-se que o processo de previsão de temperatura sem a utilização dos valores de pressão atmosférica indicam que as RNAs conseguem realizar a previsão da série temporal das médias diárias de temperatura sem a variável pressão atmosférica, haja vista o desempenho equiparável dos dois conjuntos de simulações.

Após a finalização dos processos de aprendizagem das RNAs e obtidas as medidas de desempenho baseadas no erro MAPE das melhores redes para os conjuntos de teste, como mostrado nas figuras 4 e 5, foi realizada a estimação dos valores de médias diárias de temperaturas para todo o mês de dezembro de 2011 com o objetivo de realizar a estimação de dados ausentes na série temporal de temperatura para esse mês de dezembro. Desta forma, na figura 6 são mostrados as previsões para dezembro de 2011 baseadas nas RNAs avaliadas nas figuras 4 e 5.

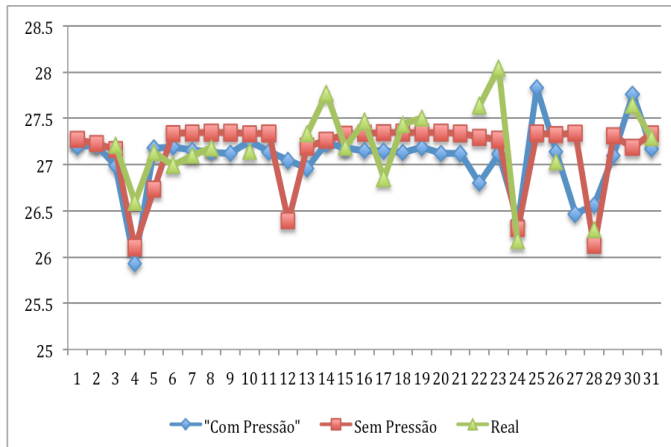


Figura 6. Valores reais (com falhas) e estimados de temperatura para dezembro de 2011 utilizando os dados de temperatura com pressão e temperatura sem pressão.

Na figura 6, pode ser observado três séries temporais. A série denominada “Real” representa os dados de médias diárias de temperatura para o mês de dezembro de 2011. Observa-se que esta série possui algumas falhas representadas na figura 6 pelos “buracos” no desenho da série. Também pode ser observado na figura 6 a série “Com Pressão” que representa os dados previstos para médias diárias de temperaturas para o mês de dezembro de 2011 utilizando a RNA treinada com os dados de temperatura e pressão atmosférica, enquanto que a série “Sem Pressão” também estima os valores de médias diárias de temperaturas para o mês de dezembro mas sem a utilização dos valores dos dados de pressão atmosférica.

IV. CONCLUSÕES

Neste trabalho foram aplicadas as técnicas de RNAs para a predição de valores de médias diárias de temperaturas através da utilização de janelas de previsão utilizando dados de data (ano e dia), seis valores de temperatura, sendo uma medida no instante atual e cinco valores históricos, além de uma medida de pressão atmosférica no instante atual. Outro conjunto de simulações também usou a janela de previsão do parágrafo anterior sem os dados de pressão atmosférica. Nas simulações realizadas foi observada, para os conjuntos de teste, uma boa capacidade de predição, apresentando erros muito semelhantes sem diferenciação estatística entre eles.

Os resultados deste trabalho apontaram a aplicabilidade da utilização das técnicas de RNAs como ferramentas eficientes para a previsão de séries temporais meteorológicas e por conseguinte o preenchimento de falhas existentes nestas séries através de um método robusto e com alta capacidade de tratar estes dados ruidosos com um erro consideravelmente baixo.

AGRADECIMENTOS

Os autores agradecem ao laboratório de física da atmosfera da universidade federal do oeste do Pará – Ufopa pela cessão dos dados utilizados neste trabalho e aos financiadores que fundamentam este artigo: CAPES, CNPq e FAPESPA.

REFERÊNCIAS

- [1] Adya, M. e Collopy, F.. “How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation”, In: *Journal of Forecasting*, vol. 17, 1998, pp. 481-495.
- [2] Box, G. E. P.; Jenkins, G. M.; Reinsel, G. C. R.. *Time Series Analysis: Forecasting and Control*. Prentice Hall, New Jersey, 3. ed. (1994)
- [3] Dillon, W. R. e Goldstein, M.. “*Multivariate Analysis - Methods and Applications*”, John Wiley & Sons. 1984.
- [4] Douglas, A.P., Breipohl, A.M., Lee, F.N. e Adapa, R.. “The impacts of temperature forecast uncertainty on Bayesian loadforecasting”. *IEEE Transactions on Power Systems*, 1998, vol. 13.
- [5] Hair, J. F. JR., Anderson, R. E., Tatham, R. L. e Black, W. C.. “*Multivariate data analysis*”. Prentice-Hall. 1998.
- [6] Rezende, S. O. *Sistemas Inteligentes – Fundamentos e Aplicações*. Ed. Manole, Rio de Janeiro (2003).
- [7] Russel, S. e Norvig, P.. “*Artificial Intelligence – A Modern Approach*”. Prentice Hall. 2003.
- [8] Senjyu, T., Takara, H., Uezato, K. e Funabashi, T.. “One-hour-ahead Load Forecasting Using Neural Network”. *IEEE Transactions on Power Systems*, 2002, vol. 17, no. 1.
- [9] Haykin, S.. *Neural Networks: a comprehensive Foundation*, Prentice Hall, 2nd Ed. 1998.
- [10] OMETTO, J. C. *Bioclimatologia vegetal*. São Paulo: Agronômica Ceres 1981. 440 p