

Uso de Redes Neurais para Previsão da Temperatura da Superfície do Mar do Oceano Atlântico Tropical

Patrícia Mattos, Ana-Maria Dubois Iorgulescu, Rebecca Salles
Eduardo Bezerra, Eduardo Ogasawara, Leonardo Lima

¹Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ

{ebezerra,leolima}@cefet-rj.br, eogasawara@ieee.org

Resumo. *Eventos ambientais extremos afetam a vida de milhões de pessoas. As secas encontram-se nesta categoria. Embora não se possa impedir que este tipo de evento ocorra, a previsão de sua ocorrência possibilita mitigar parte do dano resultante. Uma das variáveis importantes para identificação de ocorrências de secas é a temperatura de superfície do mar (TSM). Este trabalho apresenta um modelo baseado em Redes Neurais para previsão de TSM do Oceano Atlântico Tropical. O modelo é constituído por dois workflows. O primeiro workflow realiza o ajuste do modelo de previsão, enquanto que o segundo realiza a previsão propriamente dita. Os experimentos preliminares apresentaram resultados positivos e apontam para oportunidades de refinamento de modelos tanto para horizontes de previsão de curto prazo como de longo prazo.*

1. Introdução

Alguns fenômenos da natureza, como os eventos extremos, afetam a vida de milhões de pessoas. Um evento natural é caracterizado como extremo normalmente a partir de três critérios: raridade, intensidade e severidade [Beniston et al., 2007]. Dentre os eventos extremos, as secas se distinguem por causarem impactos mais severos, uma vez que ocasionam um número maior de mortes quando comparadas aos demais desastres naturais [EM-DAT, 2015]. Este tipo de evento extremo pode ser definido como uma condição de umidade insuficiente causada por uma redução da quantidade de precipitação recebida durante um determinado período de tempo [Mishra and Singh, 2011]. Além disto, o crescimento da população mundial e o conseqüente aumento da demanda por água têm tornado a seca um problema ainda mais preocupante.

Um agravante a ser considerado quando se trata de secas é a dificuldade de avaliá-las e prevê-las, dado que a avaliação e previsão de secas, segundo Svoboda et al. [2002], podem ser consideradas mais difíceis do que as de outros desastres naturais. Essa maior dificuldade resulta no fato de que métodos tradicionais de previsão ou de avaliação podem não servir para o caso das secas. Isso motiva estudos que procurem contribuir para melhorar as técnicas de monitoramento e previsão desses eventos. O propósito é reduzir, dentro do que for possível, a vulnerabilidade da sociedade como um todo perante as secas e os seus efeitos indesejáveis.

Existem muitos modelos de previsão propostos na literatura [De Gooijer and Hyndman, 2006]. Em termos de desempenho de previsão, nenhum destes modelos é notoriamente superior aos demais. Sendo assim, o processo de previsão é exploratório e demanda experimentação. É necessário aplicar diversas técnicas de transformação de

dados. A gerência de todos estes dados acaba se tornando uma tarefa árdua. Aspectos como proveniência [Marinho et al., 2012] passam a ser fundamentais para se garantir a reprodutibilidade.

Algumas outras iniciativas centradas em dados procuram introduzir primitivas para se realizar previsão diretamente nas bases de dados temporais [Parisi et al., 2013]. Estas iniciativas estão alinhadas ao objetivo deste trabalho, com a diferença de que o nosso foco visa à elaboração de modelos centrados em dados de previsão baseados em *workflows* e não à incorporação de tais primitivas em sistemas de banco de dados.

O presente trabalho contribui ao estudar a temperatura da superfície do mar usando Redes Neurais (RN). O modelo descrito neste trabalho inclui a previsão dos dados de TSM em diferentes periodicidades, *i.e.*, a partir de séries diárias e mensais. Durante a pesquisa bibliográfica, não foi detectado nenhum estudo de previsão de TSM do Oceano Atlântico Tropical que contemplasse diferentes periodicidades.

2. Temperatura de Superfície do Mar

A temperatura de superfície do mar (TSM) é a temperatura da água do oceano em sua superfície, sendo considerada um indicador de acúmulo de calor do oceano [Lins et al., 2013]. Em diversas partes do mundo, a frequência e quantidade de precipitação tem sido associadas a fenômenos atmosféricos, como o El Niño Oscilação Sul (ENOS) e a TSM, na qual tem se observado um forte relacionamento estatístico entre esses fenômenos e as chuvas [Mishra and Singh, 2011].

A relação entre as secas na região nordeste do Brasil e a TSM no Atlântico tropical já foi abordada por alguns autores. Segundo Hastenrath [2011], a região Nordeste apresenta uma curta estação chuvosa entre março e abril, momento no qual o gradiente inter-hemisférico da TSM no Atlântico Tropical é mais fraco. Eventos de grande impacto socioeconômicos acontecem nesta região, mas há carência de estudos [Good et al., 2008].

A fim de tentar suprir essa lacuna de informação, foi proposto o projeto *Pilot Research Moored Array in the Tropical Atlantic* (PIRATA). O PIRATA é uma rede de observação composta por boias espalhadas pelo Oceano Atlântico Tropical, planejada para monitorar uma série de variáveis dos processos de interação oceano-atmosfera [Bourlès et al., 2008]. As boias utilizadas pelo programa são conhecidas como *Autonomous Temperature Line Acquisition System* (ATLAS) e foram projetadas para medir variáveis meteorológicas de superfície através de sensores hidrológicos entre a superfície e quinhentos metros de profundidade.

3. Modelos de Previsão de TSM Usando Redes Neurais

Para abordar a previsão de TSM do Atlântico Tropical usando RN foram modelados dois workflows: um de ajustamento e outro de previsão. O primeiro, correspondente ao ajustamento do modelo, é indicado na Figura 1.a e contém cinco atividades. A primeira atividade refere-se à preparação da série estacionária. A partir de uma série temporal X , analisa-se a sua estacionariedade via *unit root test* [Shumway and Stoffer, 2010] e, caso necessário, obtém-se uma série transformada Y com propriedades estacionárias via diferenciação [Shumway and Stoffer, 2010]. A segunda atividade consiste em separar Y em duas séries Y_{tm} e Y_{am} , que são utilizadas, respectivamente, para treinamento e

avaliação do modelo. A terceira atividade consiste em converter as séries Y_{tm} e Y_{am} em janelas deslizantes J_{tm} e J_{am} . Cada janela contém um conjunto defasado de observações formando uma matriz. Cada linha contém um número de observações defasadas que são usadas para formar as entradas e o valor alvo da RN. A quarta atividade contempla o treinamento da RN propriamente dito usando J_{tm} , que produz o modelo ajustado \bar{RN} . Finalmente, a quinta atividade faz a avaliação, usando os dados de J_{am} , para medir o erro de previsão da rede. O segundo workflow, indicado na Figura 1.b, utiliza o modelo ajustado \bar{RN} e realiza a previsão com os dados de teste.

4. Avaliação Experimental e Considerações Finais

A fim de realizar uma análise preliminar da abordagem, realizou-se um experimento utilizando dados obtidos do PIRATA [GOOS-Brasil, 2015]. Tanto o *workflow* de ajustamento de modelo quanto o *workflow* de previsão foram modelados utilizando um sistema de workflow relacional [Ogasawara et al., 2013]. Os experimentos utilizaram como entrada séries de TSM com duas periodicidades distintas: diária e mensal.

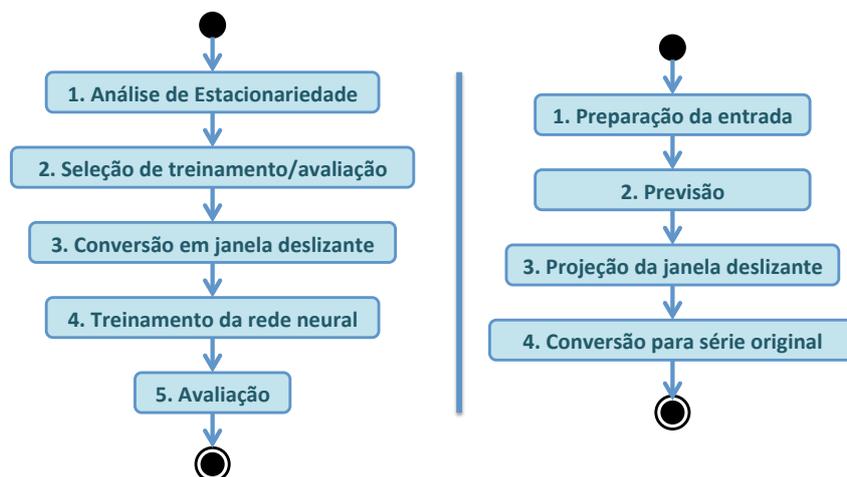


Figura 1. Workflow de ajustamento (a); Workflow de previsão (b)

Tabela 1. Resumo dos resultados das previsões

Erro	Modelo	Previsão Diária (7 dias)	Previsão Mensal (3 meses)
EMA	RW	0,24	1,38
	RN	<u>0,20</u>	<u>0,70</u>
EMPA	RW	1,01%	5,22%
	RN	<u>0,82%</u>	<u>2,66%</u>
SEQ	RW	0,48	8,09
	RN	<u>0,33%</u>	<u>0,67%</u>

A Tabela 1 apresenta os resultados da previsão das RN e sua comparação com *Random Walk*, que corresponde a um indicador de qualidade na área de séries temporais [Diebold and Mariano, 2002]. Com o intuito de avaliar os resultados de previsão obtidos foram utilizadas três medidas de desempenho: erro médio absoluto (EMA), erro médio percentual absoluto (EMPA) e soma dos erros quadráticos (SEQ).

Os resultados conduzem ao entendimento que as RN apresentaram bom desempenho. Na série diária, as RN apresentaram um erro de previsão inferior a 1% em 7 passos a frente e, no caso da previsão mensal, este valor chegou a 2,66%, em 3 passos a frente. Os resultados indicam potencial de previsão de longo prazo usando séries de menor periodicidade, o que é muito importante para o contexto das secas.

Referências

- Beniston, M., Stephenson, D. B., Christensen, O. B., Ferro, C. A. T., Frei, C., Goyette, S., Halsnaes, K., Holt, T., Jylhä, K., Koffi, B., Palutikof, J., Schöll, R., Semmler, T., and Woth, K. (2007). Future extreme events in European climate: an exploration of regional climate model projections. *Climatic Change*, 81(1):71–95.
- Bourlès, B., Lumpkin, R., McPhaden, M. J., Hernandez, F., Nobre, P., Campos, E., Yu, L., Planton, S., Busalacchi, A., Moura, A. D., Servain, J., and Trotte, J. (2008). The Pirata Program: History, Accomplishments, and Future Directions. *Bulletin of the American Meteorological Society*, 89(8):1111–1125.
- De Gooijer, J. G. and Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22(3):443–473.
- Diebold, F. X. and Mariano, R. S. (2002). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 20(1):134–144.
- EM-DAT (2015). The OFDA/CRED International Disaster Data Base. Centre for Research on the Epidemiology of Disasters. Technical report, <http://www.emdat.be/database>.
- Good, P., Lowe, J. A., and Rowell, D. P. (2008). Understanding uncertainty in future projections for the tropical Atlantic: relationships with the unforced climate. *Climate Dynamics*, 32(2-3):205–218.
- GOOS-Brasil (2015). PIRATA Dataset. Technical report, <http://www.goosbrasil.org>.
- Hastenrath, S. (2011). Exploring the climate problems of Brazil's Nordeste: a review. *Climatic Change*, 112(2):243–251.
- Lins, I. D., Araujo, M., Moura, M. d. C., Silva, M. A., and Droguett, E. L. (2013). Prediction of sea surface temperature in the tropical Atlantic by support vector machines. *Computational Statistics & Data Analysis*, 61:187–198.
- Marinho, A., Murta, L., Werner, C., Braganholo, V., Cruz, S. M. S. d., Ogasawara, E., and Mattoso, M. (2012). ProvManager: a provenance management system for scientific workflows. *Concurrency and Computation: Practice and Experience*, 24(13):1513–1530.
- Mishra, A. K. and Singh, V. P. (2011). Drought modeling – A review. *Journal of Hydrology*, 403(1–2):157–175.
- Ogasawara, E., Dias, J., Silva, V., Chirigati, F., de Oliveira, D., Porto, F., Valduriez, P., and Mattoso, M. (2013). Chiron: a parallel engine for algebraic scientific workflows. *Concurrency and Computation: Practice and Experience*, 25(16):2327–2341. 00015.
- Parisi, F., Sliva, A., and Subrahmanian, V. S. (2013). A temporal database forecasting algebra. *International Journal of Approximate Reasoning*, 54(7):827–860.
- Shumway, R. H. and Stoffer, D. S. (2010). *Time Series Analysis and Its Applications: With R Examples*. Springer, New York, 3rd ed. 2011 edition edition. 01259.
- Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., Miskus, D., and Stephens, S. (2002). The Drought Monitor. *Bulletin of the American Meteorological Society*, 83(8):1181–1190.