

# Predição Casos de Dengue em Séries Temporais

Antonio José Pinheiro Prado, Juliano Siloto Assine e Luiz Eduardo Pita Mercês Almeida

Faculdade de Engenharia Elétrica  
e de Computação, Unicamp,

## Introdução

Para o projeto da disciplina de “Dados, Inferência e Aprendizagem” ministrada no 2o Semestre de 2019 pelos professores José Cândido Silveira Santos Filho e Flavio du Pin Calmon, escolhemos um desafio público de predição do número de casos de dengue baseado na série histórica das cidades de San Juan, Porto Rico e Iquitos, Peru. O desafio é oferecido pela plataforma [DrivenData\[1\]](#) e a descrição completa do desafio está disponível no seu [site oficial](#).

Consideramos que apesar de se tratar de uma competição, em que não temos acesso aos dados de saída de teste e que estamos sujeitos aos [regulamentos da competição](#), a escolha deste desafio está de acordo com a filosofia da disciplina, pois provê aos alunos uma ótima oportunidade de aplicar ferramentas de inferência em um problema real e de relação direta às suas vidas cotidianas, visto que de janeiro a junho de 2019 já foram confirmados [mais de 18,000 casos de dengue na cidade de Campinas](#).

## Dados

Os dados são representados por uma série temporal (Figura 1) com resolução semanal dos casos reportados de dengue, além de 21 atributos contendo localização, dados meteorológicos e de vegetação, agregando 4 fontes distintas. A descrição detalhada está disponível na [página do desafio](#)

## Modelo

Tradicionalmente a análise de séries temporais utiliza modelos auto-regressivos da família [ARMA](#) (Autoregressive–moving-average), a partir de uma pesquisa preliminar de estudos preditivos de séries temporais sobre epidemias (Tabela 1), temos que os modelos mais comuns são os modelos (S)ARIMA com método de otimização [Box-Jenkins](#), porém, estes modelos dependem de suposições fortes de estacionariedade dos dados e requerem estratégias manuais de agregação dos atributos, o que cresce combinatorialmente em complexidade com o número de atributos (21 no nosso caso).

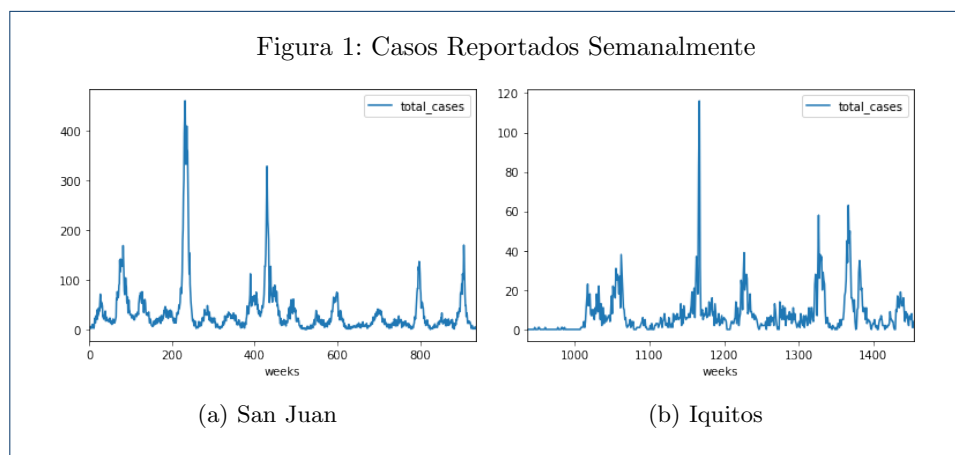
Também é do interesse do grupo a exploração de modelos baseados em redes neurais, trabalhos mais recentes apontam o uso dessas técnicas, em particular o trabalho em [\[2\]](#) utiliza redes neurais para o nosso dataset em específico, mas não parece existir consenso sobre as melhores escolhas de arquiteturas e hiperparâmetros e portanto uma análise mais profunda é necessária.

Por fim, o critério de otimização é um dado do problema. O modelo deve minimizar a métrica de erro absoluto médio (norma  $l_1$ )

## Referências

1. Bull, P., Slavitt, I., Lipstein, G.: Harnessing the power of the crowd to increase capacity for data science in the social sector. arXiv preprint arXiv:1606.07781 (2016)
2. Islas Abud, K., Emmanuel Vallejo Clemente, E., Sánchez Castellanos, H.: A novel deep recurrent neural network architecture for time series forecasting of mosquito-borne disease case counts. Master's thesis (May 2019)
3. Cortes, F., Martelli, C.M.T., de Alencar Ximenes, R.A., Montarroyos, U.R., Junior, J.B.S., Cruz, O.G., Alexander, N., de Souza, W.V.: Time series analysis of dengue surveillance data in two brazilian cities. *Acta tropica* **182**, 190–197 (2018)
4. Wu, Y., Yang, Y., Nishiura, H., Saitoh, M.: Deep learning for epidemiological predictions. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 1085–1088 (2018). ACM
5. Guo, P., Liu, T., Zhang, Q., Wang, L., Xiao, J., Zhang, Q., Luo, G., Li, Z., He, J., Zhang, Y., *et al.*: Developing a dengue forecast model using machine learning: A case study in china. *PLoS neglected tropical diseases* **11**(10), 0005973 (2017)
6. Anwar, M.Y., Lewnard, J.A., Parikh, S., Pitzer, V.E.: Time series analysis of malaria in afghanistan: using arima models to predict future trends in incidence. *Malaria journal* **15**(1), 566 (2016)
7. Midekisa, A., Senay, G., Henebry, G.M., Semuniguse, P., Wimberly, M.C.: Remote sensing-based time series models for malaria early warning in the highlands of ethiopia. *Malaria journal* **11**(1), 165 (2012)
8. Gharbi, M., Quenel, P., Gustave, J., Cassadou, S., La Ruche, G., Girdary, L., Marrama, L.: Time series analysis of dengue incidence in guadeloupe, french west indies: forecasting models using climate variables as predictors. *BMC infectious diseases* **11**(1), 166 (2011)
9. Wangdi, K., Singhasivanon, P., Silawan, T., Lawpoolsri, S., White, N.J., Kaewkungwal, J.: Development of temporal modelling for forecasting and prediction of malaria infections using time-series and arimax analyses: a case study in endemic districts of bhutan. *Malaria Journal* **9**(1), 251 (2010)
10. Lu, L., Lin, H., Tian, L., Yang, W., Sun, J., Liu, Q.: Time series analysis of dengue fever and weather in guangzhou, china. *BMC Public Health* **9**(1), 395 (2009)
11. Choudhury, Z.M., Banu, S., Islam, A.M.: Forecasting dengue incidence in dhaka, bangladesh: A time series analysis. (2008)
12. Luz, P.M., Mendes, B.V., Codeço, C.T., Struchiner, C.J., Galvani, A.P.: Time series analysis of dengue incidence in rio de janeiro, brazil. *The American journal of tropical medicine and hygiene* **79**(6), 933–939 (2008)

## Figuras



## Tabelas

Tabela 1: Relação de trabalhos similares e características dos modelos utilizados

Doença	Outros Atributos	Modelo	Algoritmo	Obs	Trabalho
Dengue, Zica	Não	Vários	Vários	Tese de Mestrado com nosso dataset	[2]
Dengue	Não	ARIMA	Box Jenkins		[3]
Influenza	Não	ARMA-like, Processos Gaussianos, Deep Learning	N/A		[4]
Dengue	Temperatura, Precipitação, Humidade Relativa, Pesquisas no Baidu	Vários	Vários	Support Vector Regression teve os melhores resultados	[5]
Malaria	Temperatura, Precipitação, Humidade Relativa, Índice de Vegetação	ARIMA	Box Jenkins	Trabalha com escala logaritimica dos dados.	[6]
Malaria	Temperatura do chão, Precipitação, Índice de Vegetação, EvapoTranspiração	SARIMA	N/A	Trabalha com escala logaritimica dos dados.	[7]
Dengue	Temperatura Max/Min/Avg, Precipitação, Humidade Relativa	SARIMA	Box Jenkins	Variavel com lag de 3 meses e Temperatura são os atributos mais preditivo.	[8]
Malaria	Temperatura Max/Min/Avg, Precipitação, Humidade Relativa	ARIMAX	N/A	Variáveis não eram transferíveis para diferentes localizações	[9]
Dengue	Temperatura Max/Min, Precipitação, Humidade relativa, Vento	Regressão de Poisson	GEE/QICu	Temperatura minima/ Humidade são positivamente correlacionadas, vento é negativamente correlacionado	[10]
Dengue	Não	SARIMA	Normalized Bayesian Information Criteria		[11]
Dengue	Temperatura Max/Min, Precipitação diaria e anual	ARIMA	Box Jenkins	Trabalha com escala logaritimica dos dados	[12]