

Um estudo sobre estratégias para compactação de Séries Temporais para aplicações de IoT

Marcos A. de Oliveira Jr.^{1,2}, Gerson Geraldo H. Cavalheiro²

¹Instituto Federal de Educação, Ciência e Tecnologia Farroupilha
Santa Maria/RS – Brasil

²Centro de Desenvolvimento Tecnológico – Programa de Pós-Graduação em Computação
Universidade Federal de Pelotas (UFPEL) – Pelotas, RS – Brasil

{marcos.oliveira, gerson.cavalheiro}@inf.ufpel.edu.br

***Resumo.** Séries temporais são amplamente utilizadas para registro de informações em aplicações baseadas em Internet das Coisas, onde sensores são utilizados para coleta de dados em quantidades massivas e existe uma significativa limitação tecnológica dos dispositivos. O presente trabalho propõe a investigação de alternativas para a representação de séries temporais, de forma a reduzir a necessidade de recursos para a manipulação desses dados.*

1. Introdução

Com a evolução dos dispositivos baseados no conceito de Internet das Coisas e a crescente utilização de recursos tecnológicos para coleta de dados, séries temporais tem sido amplamente utilizadas para o armazenamento de dados observados por sensores e outros mecanismos. No entanto, ainda que seja notável o recente progresso tecnológico desses dispositivos, o grande volume de dados coletados se apresenta como um desafio a ser contornado pelas aplicações que manipulam esses dados.

Na grande maioria das aplicações, os dados são coletados na camada de borda (*edge*), transmitidos pela camada de neblina (*fog*), enviados para a nuvem (*cloud*) para algum processamento mais pesado e, posteriormente, devolvidos para as camadas mais abaixo. Durante esse ciclo é necessário, principalmente, armazenar e transmitir os dados respeitando a limitação de recursos. Para isso, é importante que existam formas alternativas para representar os dados coletados, geralmente em quantidades massivas, de forma a minimizar a demanda de recursos, porém sem perda de informação. Assim, é fundamental a investigação de alternativas, como [Azar et al. 2020], que viabilizem uma eficiente coleta e utilização desses dados.

Este trabalho tem como objetivo propor uma investigação de alternativas de representação de séries temporais, com o intuito de identificar meios de compactação desses dados, a fim de reduzir a necessidade de recursos, proporcionando, assim, maior eficiência na utilização desse formato de informação.

2. Estratégias e Experimentos

O ponto de partida da investigação será a seleção de um benchmark de algoritmos de classificação de séries temporais [Bagnall et al. 2017], bem como a reprodução dos

experimentos com os conjuntos de dados originais, a fim de se obter o resultado das classificações com as séries temporais não compactadas. Em seguida, será realizada a busca e classificação dos algoritmos de compressão de séries temporais para aplicações de Internet das Coisas, como [Blalock et al. 2018], [Eichinger et al. 2015] e [Chirikhin and Ryabko 2019], a fim de explorar algoritmos de compressão e suas características e identificar quais aqueles que se aplicam ao estudo em questão.

Posteriormente, serão escolhidos alguns algoritmos de compactação de séries temporais, de acordo com a sua adequação após a análise da etapa anterior, para implementação e aplicação na compactação dos dados utilizados nos experimentos iniciais. Dessa forma, as séries temporais originais serão compactadas para realização de nova rodada de experimentos.

Com os dados compactados, serão novamente reproduzidos os experimentos com os algoritmos de classificação selecionados no início do processo, a fim de identificar quais algoritmos são capazes de classificar as séries temporais de forma correta, mesmo em seu formato compactado. A partir desses testes, pressupõe-se que seja possível identificar a eficiência e a viabilidade da utilização de séries temporais compactadas em aplicações de classificação.

3. Trabalhos Futuros

A partir da identificação de estratégias eficientes para a representação de séries temporais de forma reduzida, ou seja, com uma menor utilização de recursos, a continuidade do trabalho se dará na aplicação dessas estratégias a problemas com viés agrícola. Em especial, serão utilizadas as melhores alternativas de forma combinada a um protótipo desenvolvido por pesquisadores da mesma instituição, que utiliza sensores eletrodos para coletar dados de plantas (como salinidade e nível de recursos hídricos) e registra as informações na forma de séries temporais. O objetivo futuro é contribuir com o aperfeiçoamento desse protótipo, contornando principalmente o gargalo de armazenamento, proporcionado a coleta de um maior número de informações em um dispositivo com recursos limitados.

References

- Azar, J., Makhoul, A., Couturier, R., and Demerjian, J. (2020). Robust iot time series classification with data compression and deep learning. *Neurocomputing*, 398:222–234.
- Bagnall, A., Lines, J., Bostrom, A., Large, J., and Keogh, E. (2017). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. In *Data Mining and Knowledge Discovery*, page 606–660. Springer.
- Blalock, D., Madden, S., and Guttag, J. (2018). Sprintz: Time series compression for the internet of things. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(3).
- Chirikhin, K. S. and Ryabko, B. Y. (2019). Application of data compression techniques to time series forecasting.
- Eichinger, F., Efron, P., Karnouskos, S., and Böhm, K. (2015). A time-series compression technique and its application to the smart grid. *VLDB Journal*, 24(2):193–218.