

Detecção de Anomalias em Redes de Sensores Sem Fios Utilizando o Modelo *Moving Z-Score*

Rychelly Glenneson da Silva Ramos, Paulo Ribeiro Lins Júnior
Instituto Federal de Educação, Ciência e Tecnologia da Paraíba
Campus Campina Grande

Capítulo Técnico Estudantil da ComSoc - Ramo Estudantil IEEE IFPB – Campina Grande
Email: rychelly.g.ramos@ieee.org, paulo.ribeiro.lins.jr@gmail.com

Resumo—Redes de Sensores sem Fios (RSSFs) tem ganhado ampla aceitação como um *framework* para telemetria e monitoramento remoto para aplicações como telemedicina, agricultura de precisão e monitoramento do clima. As características complexas e dinâmicas dessas redes as torna vulneráveis a anomalias, ou seja, observações que não correspondem ao comportamento real e natural das medições realizadas. Este trabalho, têm por objetivo a avaliação do modelo *Z-Score* para detecção de anomalias em redes de sensores, com intuito de determinar a parametrização sob as quais esta técnica leva minimização de falsos-positivos.

Index Terms—Redes de sensores sem fios, anomalias, moving z-score, detecção.

I. INTRODUÇÃO

Redes de Sensores sem Fios (RSSFs), infra-estrutura básica da *Internet of Things* (IoT), podem apresentar anomalias devido ao mau funcionamento de *hardware* ou *software*, que causam falhas em nós sensores [1], [2]. Tais anomalias podem gerar o desligamento do nó sensor, quando relacionadas ao gerenciamento de energia dos mesmos, mas podem causar erros de funcionamento que não levam à interrupção da transmissão, mas comprometem a confiabilidade dos dados transmitidos, resultando em erros na informação, diminuindo a confiabilidade dos dados gerados pelo nó defeituoso, podendo até mesmo afetar a operação da rede a partir de um ponto de vista da segurança [3]. Portanto, ferramentas de gerenciamento de rede devem ser capazes de detectar não só as condições de desligamento de nós sensores, mas também o seu mau funcionamento [4].

Esse trabalho avalia o modelo *Moving Z-Score* aplicado a detecção de anomalias no tráfego de dados recebidos no principal de uma RSSF. O restante do artigo se apresenta da seguinte forma: na Seção II é apresentado o modelo *Z-Score*, na Seção III são apresentados os resultados e as discussões do uso desse modelo e na Seção IV, as considerações finais.

II. MOVING Z-SCORE

O *Moving Z-Score* é um modelo simples para medição de anomalias de cada *datapoint* em um *dataset* sequencial, i.e., uma série temporal. Dado um intervalo w , o *moving Z-score* é o quanto uma medida se afasta da média em termos de desvios padrão, onde a média e o desvio padrão são computados somente através de prévias observações de w [5].

O método *Z-score* não necessita dos valores de outras variáveis para realizar as estimativas, ou seja, esse método leva

em consideração apenas o valor da própria variável que está sendo tratada (w). Este método irá atribuir uma pontuação para cada dado avaliado. Quanto mais esta pontuação distancia de zero, maior a probabilidade desse *datapoint* ser uma anomalia [6]. Esta pontuação é calculada por meio da Eq. (1).

$$Z(x_i) = \frac{x_i - \bar{x}_i}{s_i} \quad (1)$$

Onde a média móvel é definida pela Eq. (2) e o desvio padrão móvel é definido pela Eq. (3).

$$\bar{x}_i = \frac{1}{w} \sum_{j=i-w}^{i-1} x_j \quad (2)$$

$$s_i = \sqrt{\frac{1}{w} \sum_{j=i-w}^{i-1} (x_j - \bar{x}_i)^2} \quad (3)$$

Então, para calcular a pontuação de cada *datapoint*, primeiramente é necessário calcular a média (\bar{x}) e o desvio padrão (s_i) da variável que está sendo tratada. Depois, é calculado a diferença entre o *datapoint* (x_i) e a média, e depois dividido pelo desvio padrão [6]. O resultado da equação (1) é a pontuação que irá sugerir se aquele dado é uma anomalia ou não.

Após calcular a pontuação para todos os dados da variável que está sendo tratada, é possível ordená-los de acordo com sua pontuação. Para tanto, o módulo da pontuação é utilizado.

Quando o valor de $Z(x_i)$ é positivo isto indica que o dado está acima da média e quando o mesmo é negativo significa que o dado está abaixo da média. Nesse método é possível escolher que sejam indicados como anomalias os N dados com maiores pontuações ou atribuir um limiar da pontuação, no qual os dados com pontuações maiores que o limite serão considerados como anomalias[6].

III. RESULTADOS E DISCUSSÕES

Nesse trabalho, afim de avaliar a técnica considerada, usou-se um *dataset* de RSSF rotulado. Os dados, disponíveis em [7], foram coletados a partir de uma rede de sensores sem fios *multi-hop* implantada utilizando módulos TelosB. Os dados consistem em medições de umidade coletados durante um período de 6 horas, com frequência de amostragem 0,2 Hertz. O rótulo '0' denota dados normais, e dados rotulados com '1'

indicam um evento introduzido. Neste caso, o vapor de água quente foi introduzido para aumentar a umidade, gerando a anomalia.

A implementação do arcabouço de detecção foi feito usando Python e Graphlab Create™, um *framework* de aprendizado de máquina extensível que a criação de aplicações e serviços inteligentes em escala [8].

A análise consiste em transformar o *dataset* em uma série temporal e usa-la para criar um modelo de detecção de anomalia baseado em *Moving Z-Score*. Após a criação do modelo, é determinado um conjunto de valores chamado *anomaly_score*, que consistem em valores que, em uma distribuição normal, variam de -3 a 3 , em 95% das amostras, em dados com distribuição não-gaussiana os valores não obedecem um intervalo bem definido [6].

É necessário observar que este modelo de detecção somente deverá ser utilizado se a variável sensoriada, em condições normais, não sofrer constante variação no tempo [5].

Na Figura 1 a curva azul representa o *dataset* utilizado para fazer as análises, onde é possível observar picos de *datapoint*, que conseqüentemente, consiste em anomalias. Na curva em vermelho, por outro lado, temos os valores de *anomaly_score* dos *datapoints* no mesmo instante de tempo, indicando o grau que esse *datapoint* é uma possível anomalia.

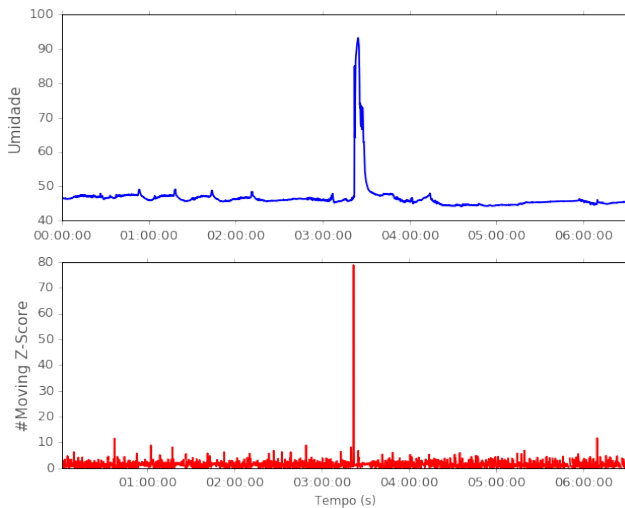


Figura 1. Série temporal de representação do *dataset* (curva azul) e *#Moving Z-Score* (curva vermelha).

Nesta coleção de dados, 100 entradas estão incorretas e rotuladas, tornando possível a avaliação do melhor tamanho da janela de análise de dados [7].

Na Figura 2 é possível observar que, quando a janela de análise do *dataset* aumenta, a acurácia do sistema aumenta e converge para um patamar de aproximadamente 97% e é possível observar que o índice de *true_negatives* converge para quase a totalidade do número de amostras presentes no *dataset*, assim como, nota-se a convergência dos *false_positives* e *false_negatives* para um número que tende a 0, assim como a ascensão do índice de *true_positives* para um número que tende a 100, valor referente ao número total de anomalias presente no *dataset*.

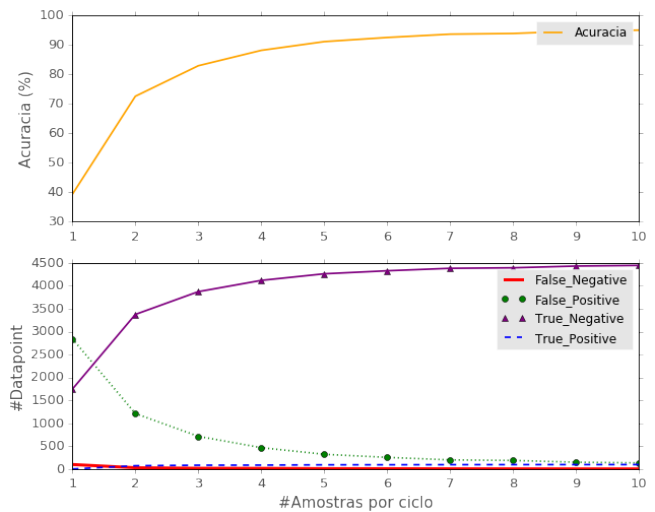


Figura 2. Resultados da análise dos dados utilizando o modelo de *Moving Z-Score* e acurácia das análises em relação as anomalias previamente rotuladas.

IV. CONCLUSÕES

Neste trabalho, foi investigado o uso do modelo *Moving Z-Scores* para para detecção de anomalias no tráfego de dados em redes de sensores sem fios. A eficiência técnica da aplicação do algoritmo de detecção de anomalia em redes de sensores sem fios foi testada em um *dataset*, e foi possível observar a influência do tamanho da janela de análise na eficiência da abordagem, considerando a incidência de falsos alarmes na métrica utilizada.

AGRADECIMENTOS

Os autores agradecem ao IFPB, *Campus* Campina Grande e ao Ramo Estudantil do IEEE IFPB-CG pelo apoio institucional. Este trabalho foi financiado pelo CNPq.

REFERÊNCIAS

- [1] M. A. Rassam, A. Zainal, and M. A. Maarof, "Advancements of Data Anomaly Detection Research in Wireless Sensor Networks: A Survey and Open Issues," *Sensors*, vol. 13, no. 8, pp. 10 087–10 122, 2013.
- [2] SBC, "Grandes Desafios da Pesquisa em Computação no Brasil – 2006-2016," Sociedade Brasileira de Computação, Tech. Rep., 2006.
- [3] H. Sagha, J. d. R. Mill, R. Chavarriga *et al.*, "Detecting and Rectifying Anomalies in Body Sensor Networks," in *2011 International Conference on Body Sensor Networks*. IEEE, 2011, pp. 162–167.
- [4] R. G. S. Ramos, P. R. Lins Júnior, and J. V. M. Cardoso, "Anomalies Detection in Wireless Sensor Networks Using Bayesian Changepoints," in *13th International Conference on Mobile Ad hoc and Sensor Systems*. IEEE, 2016.
- [5] "Dato Machine Learning Platform User Guide - Moving Z-Scores," https://dato.com/learn/userguide/anomaly_detection/, accessed: 2016-08-18.
- [6] T. M. Ventura, "Criação de um Ambiente Computacional para Detecção de Outliers e Preenchimento de Falhas em Dados Meteorológicos," Ph.D. dissertation, Universidade Federal de Mato Grosso, 2015.
- [7] S. Suthaharan, M. Alzahrani, S. Rajasegarar, C. Leckie, and M. Palaniswami, "Labelled Data Collection for Anomaly Detection in Wireless Sensor Networks," in *Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*. IEEE, 2010, pp. 269–274.
- [8] "GraphLab Create," <https://dato.com/products/create/>, accessed: 2016-06-15.