

Detecção de *Outliers* Multivariados em Redes de Sensores Sem Fio

Disciplina de Projeto e Análise de Algoritmos

Fabício G. Valadares, André L. L. Aquino, Ricardo A. Rabelo
PPGCC - Programa de Pós-Graduação em Ciência da Computação
UFOP - Universidade Federal de Ouro Preto
Ouro Preto, Minas Gerais, Brasil
email: {fabicio.valadares, alla.lins, rrabelo}@gmail.com

Resumo—No campo das redes de sensores sem fio, algumas medições podem desviar do padrão de um conjunto de dados monitorado. Esses desvios são conhecidos como *outliers*, e podem ser causados por ruídos ou erros, ataques maliciosos ou, um evento. Apresentaremos neste artigo um comparativo entre alguns algoritmos para detecção de *outliers* multivariados, aplicados à detecção de eventos em redes de sensores sem fio. Serão realizadas análises de complexidade de tempo e espaço, considerando-se as restrições de uma rede de sensores.

Keywords—Redes de sensores sem fio, detecção de eventos, *outliers*, algoritmos

I. INTRODUÇÃO

O mundo ao nosso redor possui uma variedade de fenômenos que podem ser descritos por algumas grandezas como temperatura, pressão e umidade, e podem ser monitorados por dispositivos com poder de sensoriamento, processamento e comunicação [1]. Trabalhando de forma cooperativa, esses dispositivos constituem as redes de sensores sem fio [2], [3], [4], que são um tipo especial de rede *ad hoc* [5], formadas por elementos conhecidos como nós sensores. Outros elementos que também fazem parte das redes de sensores são os nós atuadores, que possuem a capacidade de interferir no ambiente monitorado, e podem ser fisicamente os próprios nós. Além deles, temos também os sorvedouros, que possuem capacidade computacional superior a dos demais nós, e são responsáveis pelo recebimento e processamento das informações provenientes do sensoriamento. Existem também os *gateways*, dispositivos responsáveis por conectar a rede de sensores a outras redes, como a Internet. Os *gateways* e sorvedouros podem ser o mesmo elemento.

Os nós sensores são dispositivos compactos e autônomos, com capacidade de processamento local e sensoriamento. Devido ao seu tamanho reduzido, possuem uma série de restrições, como, capacidade computacional limitada, largura de banda, memória e principalmente energia, que é fornecida por bateria. Como estes dispositivos são geralmente depositados em áreas remotas e/ou de difícil acesso, a substituição da bateria é normalmente inviável. De acordo com [2], [6], o rádio é o componente de maior consumo energético, muito superior ao processamento, desta forma, o envio de grandes quantidades de dados pode reduzir a vida útil de

uma rede de sensores. Conforme pesquisa citada em [3], o consumo energético para a execução de 3000 instruções é equivalente ao envio de um *bit* por 100 metros via rádio. Sendo assim, devemos favorecer o pré-processamento dos dados diretamente na rede.

O principal objetivo de uma rede de sensores sem fio é o envio de informações a um observador externo. Porém, a transmissão de dados está sujeita a interferências do ambiente, falhas na comunicação, ruídos, o que pode causar a perda de valores, dados duplicados ou inconsistência nas informações. Para muitas aplicações, como por exemplo, detecção de incêndio em florestas e sensores de inundação, não é possível realizar a tomada de decisões baseadas em dados incorretos. Estas anomalias podem ser encontradas na literatura como *outliers*, termo que tem origem na estatística [7].

Em [7] encontramos algumas definições para os *outliers*. Definições clássicas: Segundo Hawkins [8], “um *outlier* é uma observação, que desvia muito de outras observações que despertam suspeitas que são gerados por um mecanismo diferente.” Segundo Barnett and Lewis [9], “um *outlier* é uma observação (ou um subconjunto de observações) que parece ser inconsistente comparado ao restante do conjunto de dados.” No campo das redes de sensores, podemos definir os *outliers* como as medições que desviam significativamente do padrão normal dos dados sensorizados.

Existem três fontes de *outliers* nas redes de sensores, os ruídos e os erros, os ataques maliciosos e os eventos [7]. Neste trabalho estamos interessados na detecção de eventos, que segundo [2], é a principal tarefa de um nó no campo de monitoramento. Esses eventos são detectados a partir dos dados do ambiente, que são coletados por sensores presentes em cada nó. É importante destacar que os nós nas redes de sensores sem fio podem possuir um ou mais tipos de sensores, o que permite monitorar um ou mais tipos de fenômenos. Desta forma, os dados que trafegam pela rede podem ser classificados como univariados ou multivariados [10].

Dados univariados representam um único conjunto de valores de um mesmo tipo de fenômeno, como por exemplo, um nó com capacidade para monitorar apenas a temperatura

ambiente. Por sua vez, os dados multivariados são representados por um conjunto de valores de diferentes tipos, como os gerados por um nó que possui sensores que monitoram temperatura, pressão e umidade simultaneamente, ou ainda os dados recebidos por um nó, que é o responsável pelo processamento dos dados enviados por um conjunto de nós capazes de monitorar um único tipo de fenômeno, como temperatura, por exemplo [1]. Neste trabalho estamos interessados na utilização de técnicas para detecção de *outliers* multivariados na detecção de eventos.

O artigo está organizado como segue. A Seção II apresenta a justificativa e relevância do trabalho. Os objetivos são apresentados na Seção III. A metodologia é mostrada na Seção IV e a Seção V discorre sobre os resultados esperados para o trabalho.

II. JUSTIFICATIVAS

As técnicas para detecção de *outliers* são utilizadas em diversas áreas de pesquisa, como aprendizagem de máquina, mineração de dados, estatística, teoria da informação, entre outras, onde são aplicadas por exemplo, para detecção de intrusos, detecção de fraudes [11]. Conforme relatado em [7], apesar destas técnicas ainda não serem encontradas na literatura de detecção de eventos, elas podem ser adaptadas para tal fim. Neste trabalho, realizaremos a análise de alguns algoritmos para detecção de *outliers* multivariados que podem ser adequados para a detecção de eventos, avaliaremos sua complexidade de tempo e espaço considerando a fragilidade dos nós sensores.

III. OBJETIVOS

O principal objetivo deste trabalho será o estudo teórico de algoritmos para detecção de *outliers* multivariados, onde avaliaremos sua complexidade de tempo e espaço considerando as restrições de uma rede de sensores sem fio. Buscaremos também implementar estes algoritmos em um cenário *offline*, no programa estatístico R¹.

IV. METODOLOGIA

Conforme citado na Seção III, a metodologia empregada consistirá do estudo teórico dos algoritmos para detecção de *outliers*, e sua classificação para redes de sensores sem fio. Buscaremos também a implementação *offline* dos algoritmos no R.

V. RESULTADOS ESPERADOS

Ao final deste trabalho esperamos classificar prováveis algoritmos para detecção de *outliers*, que serão candidatos a aplicação nas redes de sensores sem fio, um ambiente bastante restritivo. Esperamos que estes algoritmos possam ser aplicados de forma eficiente na detecção de eventos. Ilustraremos também a análise de complexidade de tempo e espaço dos algoritmos selecionados.

¹The R Project for Statistical Computing. Disponível em: <http://www.r-project.org/>

REFERÊNCIAS

- [1] A. L. L. Aquino, "Redução de dados em redes de sensores sem fio baseada em stream de dados," Ph.D. dissertation, Universidade Federal de Minas Gerais, February 2008.
- [2] I. F. Akyildiz, W. Su, Y. Sankarubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, August 2002.
- [3] S. Tilak, N. B. Abu-Ghazaleh, and W. Heizelman, "A taxonomy of wireless micro-sensor network models," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 6, no. 2, pp. 28–36, April 2002.
- [4] T. Arampatzis, J. Lygeros, and S. Manesis, "A survey of applications of wireless sensors and wireless sensor networks," *In 13th IEEE Mediterranean Conference on Control and Automation (MED'05)*, pp. 719–724, June 2005.
- [5] E. M. Royer and C. K. Toh, "A review of current routing protocols for ad-hoc mobile wireless networks," *IEEE Personal Communications*, vol. 6, no. 2, pp. 46–55, April 1999.
- [6] A. A. F. Loureiro, J. M. F. Nogueira, L. B. Ruiz, R. A. de Freitas Mini, E. F. Nakamura, and C. M. S. Figueiredo, "Redes de sensores sem fio (tutorial)," *XXI Simpósio Brasileiro de Redes de Computadores (SBRC'03)*, pp. 179–226, May 2003.
- [7] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *Communications Surveys Tutorials, IEEE*, vol. 12, no. 2, pp. 159–170, quarter 2010.
- [8] D. Hawkins, "Identification of outliers," *Capman and Hall*, 1980.
- [9] V. Barnett and T. Lewis, "Outliers in statistical data," *John Wiley Sons*, 1994.
- [10] O. S. Junior, A. L. L. Aquino, C. Figueiredo, and R. A. F. Mini, "Multivariate reduction in wireless sensor networks," *14th IEEE Symposium on Computers and Communications (ISCC'09)*, pp. 726–729, July 2009.
- [11] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, pp. 15:1–15:58, July 2009. [Online]. Available: <http://doi.acm.org/10.1145/1541880.1541882>

APÊNDICE

Todos os arquivos de testes, assim como os algoritmos estão disponíveis no endereço <https://sites.google.com/site/fabricioppgccufop/outliers>.