

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

Fabício G. Valadares

Orientadores:

Álvaro R. Pereira Jr.(UFOP)

André L. L. Aquino (UFAL)



UFOP



UFOP

Universidade Federal
de Ouro Preto

22 de julho de 2011



Sumário

- 1 Redes de Sensores Sem Fio
- 2 Outliers
- 3 Análise de Complexidade
- 4 Teste
- 5 Conclusões



Componentes das redes de sensores sem fio

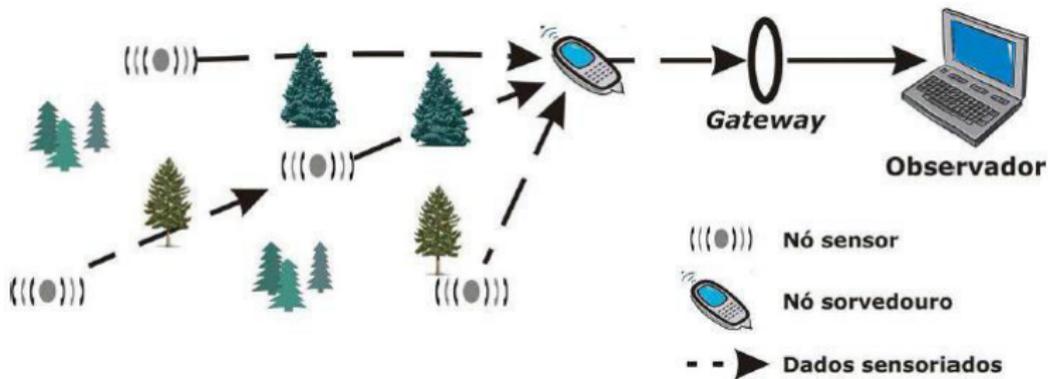


Figura adaptada de [1]



Estrutura básica de um nó sensor

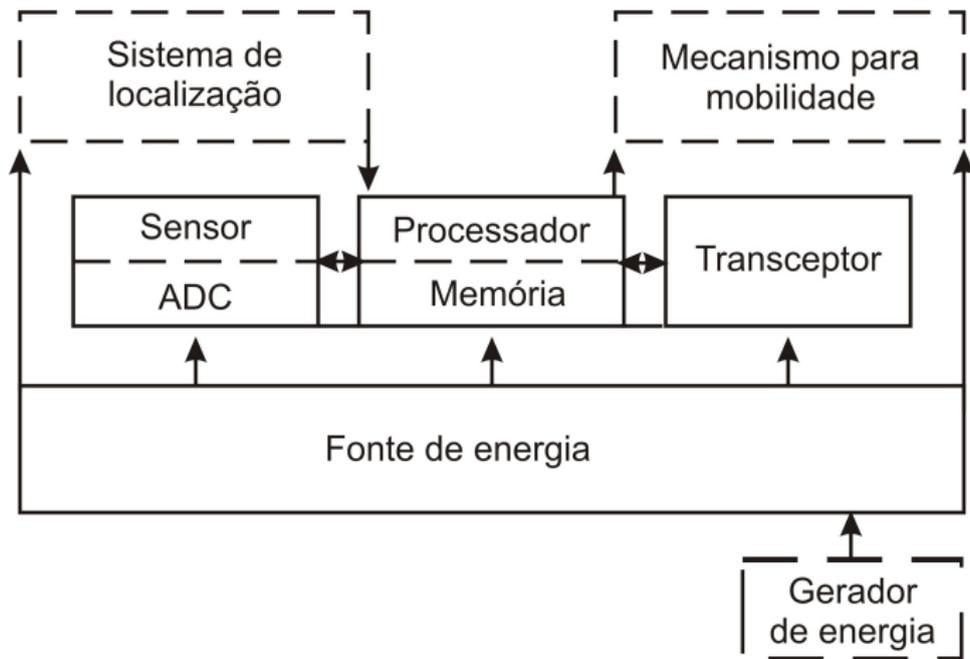


Figura adaptada de [2]



UFOP

Universidade Federal
de Ouro Preto

Algumas características das RSSF

- Tipos de dados
 - ▶ Univariados
 - ▶ Multivariados
- Detecção de eventos no campo de monitoramento
- Fragilidade da comunicação sem fio
- Anomalias nos dados



Definição de Outliers

- Pesquisa em áreas como estatística, mineração de dados
 - ▶ Identificação de fraudes, intrusão na rede, previsão do tempo[3].



Definição de Outliers

- Pesquisa em áreas como estatística, mineração de dados
 - ▶ Identificação de fraudes, intrusão na rede, previsão do tempo[3].
- Barnett and Lewis [4], “*um outlier é uma observação (ou um subconjunto de observações) que parece ser inconsistente comparado ao restante do conjunto de dados.*”



Definição de Outliers

- Pesquisa em áreas como estatística, mineração de dados
 - ▶ Identificação de fraudes, intrusão na rede, previsão do tempo[3].
- Barnett and Lewis [4], “*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 sensoriados [5].



Fontes de *outliers* em RSSF



Figura adaptada de [5]



Fontes de *outliers* em RSSF

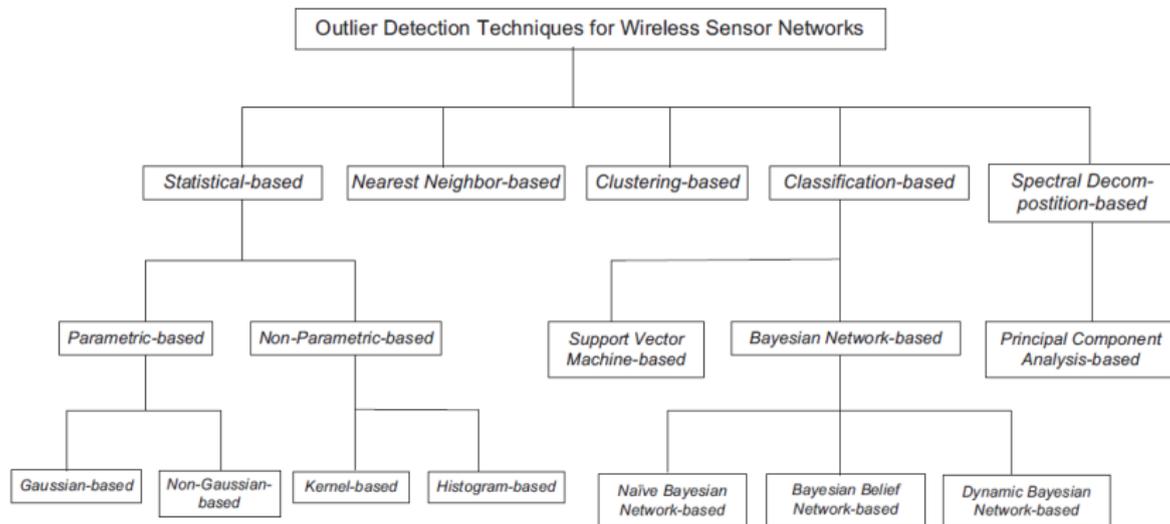


Figura adaptada de [5]



UFOP

Universidade Federal
de Ouro Preto

Cálculo de *outliers* univariados

- Distância Euclidiana



$$ED_i = \sqrt{(x_i - \bar{x})^2}$$

- Distância de Mahalanobis



$$MD_i = \sqrt{(x_i - \bar{x})^T S_x^{-1} (x_i - \bar{x})}$$



UFOP

Universidade Federal
de Ouro Preto

MVE - *Minimum Volume Ellipsoid*

- Distância de Mahalanobis robusta
- Estimadores são $T(X)$ e $C(X)$, respectivamente um vetor de tamanho p e uma matriz de tamanho $p \times p$, onde $T(X)$ é o centro do elipsoide e $C(X)$ é a estrutura de covariância do mesmo elipsoide.



MCD - *Minimum Covariance Determinant*

- Distância de Mahalanobis robusta
- São calculados estimadores $T(X)$ e $C(X)$, que permitem traçar um elipsoide, e um limiar, a partir de onde as amostras serão consideradas *outliers*.



UFOP

Universidade Federal
de Ouro Preto

Métodos utilizados

MED - *Max-eigen difference*

- Estimadores calculados através dos autovetores e autovalores da matriz de covariância, aos quais são aplicados a norma euclidiana



UFOP

Universidade Federal
de Ouro Preto

Análise de Complexidade

- MVE
 - ▶ NP -completo
- MCD
 - ▶ Quando univariado $O(n \lg n)$
 - ▶ Quando multivariado NP -completo
- MED
 - ▶ $O(k^3)$, onde k representa a quantidade de elementos de uma matriz $n \times p$.



Teste

- 5 sensores, 64 amostras
- Média 25, desvio padrão 5
- Linhas 62, 63, 64 alteradas, com dois desvios padrões

Todas as técnicas identificaram os *outliers*, porém apenas o MCD, através da função *dd.plot* alcançou 100% de acerto, sem falsos positivos. O MVE e o MED, encontraram respectivamente 21 e 13 falsos positivos. O segundo teste do MCD identificou dois falsos positivos.



UFOP

Universidade Federal
de Ouro Preto

Conclusões

- Identificadas algumas técnicas para a detecção de *outliers* multivariados.
- Os algoritmos analisados são em sua maioria aproximados.
- Teste não conclusivo, serviu para demonstrar o funcionamento das técnicas.
- Acreditamos, que através da utilização das ferramentas para detecção de *outliers*, possamos identificar eventos de forma eficiente e dinâmica.
- Os próximos passos da pesquisa são um estudo mais profundo das técnicas aqui discriminadas e um estudo de complexidade mais avançado.
- Uma observação importante, é que todas as técnicas implementadas neste trabalho são gráficas.



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

Fabício G. Valadares

Orientadores:

Álvaro R. Pereira Jr.(UFOP)

André L. L. Aquino (UFAL)



UFOP



Universidade Federal
de Ouro Preto

22 de julho de 2011



REFERÊNCIAS I

- [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. Sankarusubramaniam, and E. Cayirci, “A survey on sensor networks,” *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, August 2002.
- [3] F. R. de Santana Giroldo and L. P. Barroso, “Alguns métodos robustos para detectar outliers multivariados,” Master’s thesis, Instituto de Matemática e Estatística - Universidade de São Paulo - IME-USP, 2008.
- [4] V. Barnett and T. Lewis, “Outliers in statistica data,” *John Wiley Sons*, 1994.
- [5] 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.

