Algoritmo para detección de fallas y suavizado de datos de una red inalámbrica de sensores para monitorización de variables ambientales

Juan E. Galvis-Bautista1, Christian D. Escobar-Amado1, Byron Medina1, Dinael Guevara1, Ángelo J. Soto-Vergel1, Jose M. Celis-Peñaranda1 and Brayan R. Acevedo J.2

1 Universidad Francisco de Paula Santander, Departamento de Electricidad y Electrónica, GIDET - Colombia. Av. Gran Col. # 12E-96

2 Universidade Federal de Minas Gerais, LITC - Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brazil

E-mail: juanernestogb@ufps.edu.co

**Abstract**. La presencia de valores atípicos extremos y enmascarados en las mediciones de variables físicas, afectan la toma de decisiones en el fenómeno de estudio. La investigación propone un algoritmo para la detección de fallas y suavizado en los valores medidos de manera secuencial por una red inalámbrica de sensores. El estudio utiliza datos de temperatura, humedad relativa, humedad de la tierra e intensidad lumínica; considerando la correlación entre éstas; donde, el ajuste de las series temporales se realiza con el criterio de Tukey, el rango de medición de los sensores y los cálculos de medianas móviles. Con esta propuesta se pretende mejorar la confiabilidad de la red inalámbrica de sensores y obtener la mejor representación del fenómeno con el propósito de contribuir a la toma de decisiones.

1. Introducción

La importancia del análisis de valores atípicos en series temporales radica en que a partir de éste se puede determinar la calidad y confiabilidad de los datos para su aplicación en la toma de decisiones o para entrenamiento de modelos de previsión[1]. La presencia de valores atípicos en conjuntos de datos obtenidos a partir de redes inalámbricas de sensores puede corresponder al comportamiento estocástico de las variables monitoreadas y a fallas en los sensores o en los nodos dispuestos para almacenar o transmitir la información[2].

Existen diferentes tipos de datos atípicos que pueden presentarse en series temporales y que se clasifican según tres criterios:

* Por su efecto sobre los datos, como atípicos aditivos, los cuales pueden afectar la serie en instantes de tiempo determinados y atípicos innovadores, cuyo efecto persiste sobre los datos siguientes a su ocurrencia[3].
* A partir de la descripción estadística de los datos, como valores atípicos extremos, que corresponden a registros con valores muy alejados de la media los datos, valores atípicos enmascarados, que presentan magnitudes cercanas a la media de los datos, pero no corresponden al comportamiento de la variable y atípicos empantanados, que ocurren cuando los valores normales son considerados como atípicos debido a la presencia de valores atípicos cercanos[4].
* Según la dimensión de los datos, como atípicos univariantes que afectan sólo una variable y atípicos multivariantes que ocurren simultáneamente en más de una variable[4].

En el estudio de series temporales multivariantes puede ser difícil determinar con exactitud cuando un valor corresponde realmente a una variación atípica que perjudique la integridad de los datos, esto ocurre cuando la presencia de valores atípicos afecta las relaciones de dependencia y de correlación temporal entre las variables[4][5].

Un enfoque utilizado para mejorar la calidad de los datos consiste en la aplicación de métodos de ajuste o suavizado de curvas, lo que permite mantener los registros de variables almacenados simultáneamente con valores atípicos, además el ajuste de curvas permite la reducción o eliminación de la cantidad de ruido en los registros y una representación de los datos en la que se identifican más fácilmente la tendencia, los patrones y la estacionalidad de las series temporales[6]. Sin embargo la aplicación de métodos de suavizado puede ser perjudicial ante la presencia de valores atípicos[7].

En el presente artículo se propone un algoritmo que permite detectar y suavizar valores atípicos extremos y enmascarados en conjuntos de datos multivariantes.

Falta hablar de:

* El objetivo del artículo (extenderlo)
* Hipótesis
* Ventanas deslizantes.
* Métodos robustos: medianas móviles.
* Desviación estándar.
* Variables agroclimáticas.
* Aplicaciones del análisis y limpieza de datos
* Beneficios debido a datos confiables.
* Tukey

1. Another section of your paper

The first paragraph after a heading is not indented (Bodytext style).

Other paragraphs are indented (BodytextIndented style).

## A subsection

Some text.

* + 1. A subsubsection. The paragraph text follows on from the subsubsection heading but should not be in italic.

References

[1] Y. Yu, Y. Zhu, S. Li, and D. Wan, “Time Series Outlier Detection Based on Sliding Window Prediction,” *Math. Probl. Eng.*, vol. 2014, p. 14 pages, 2014.

[2] Y. Zhang, N. Meratnia, and P. Havinga, “Outlier Detection Techniques for Wireless Sensor Networks : A Survey,” *IEEE Commun. Surv. Tutorials*, vol. 12, no. 2, pp. 159–170, 2010.

[3] H. Shang and B. Zhang, “Outliers detection in INAR (1) time series,” *J. Phys. Conf. Ser.*, vol. 1053, no. 1, 2018.

[4] Darmanto and S. Astutik, “The effectiveness of robust RMCD control chart as outliers’ detector,” *J. Phys. Conf. Ser.*, vol. 943, no. 1, 2017.

[5] G. Lim and S. Min, “Effect of Outliers and Non-consecutive Data Points on the Detrended Cross-Correlation Analysis,” *J. Korean Phys. Soc.*, vol. 72, no. 4, pp. 545–550, 2018.

[6] J. V Hansen and R. D. Nelson, “Neural Networks and Traditional Time Series Methods : A Synergistic Combination in State Economic Forecasts,” *IEEE Trans. Neural Networks*, vol. 8, no. 4, pp. 863–873, 1997.

[7] H. Hamid, “New Algorithm of Location Model based on Robust Estimators and Smoothing Approach,” *Int. Conf. Appl. Syst. Innov.*, vol. Sapporo, pp. 1091–1093, 2017.