Employment of Data Prediction Methods in Wireless Sensor Networks

Leonardo C. Monteiro
PPGI/DCC-IM
Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brazil
leonardo.monteiro@ufrj.br

***Abstract* — The purpose of the study presented in this paper is to identify the different themes in Wireless Sensor Networks (WSN) that employ data prediction methods. The study was undertaken as a Systematic Literature Review (SLR) with manual search in 5 well-known online digital libraries. A total of 49 themes with employment of data prediction methods were identified. These employments were also classified in 4 levels of processing in full WSN applications (from bottom hardware to top software). The results indicate that the use of data prediction methods in WSN is currently concentrated in solving their operation problems, like energy saving, than in top-level application models that effectively generate predictions to the end users of the sensor networks.**

Keywords — Data prediction; Systematic Literature Review; Wireless Sensor Networks

I. INTRODUCTION

A. About prediction

According to the Merriam-Webster dictionary, the word "prediction" means: "A statement about what will happen or might happen in the future". In this sense, a "prediction" is something that someone says now that will happen in the future.

In mathematical sciences, however, the use of the term "prediction" is more flexible about timing. A prediction (predicted data) is an expectation calculated by combining predictors (predictor data) through a model, employing one or more prediction methods. In this sense, we can say that a "data prediction method" employed at time t_0, using predictor data from time $t_p < t_0$, generates some predicted data that can be interpreted as an expectation about the time $t_f > t_0$. One important observation, it is possible that $t_f < t_{now}$, where t_{now} is the current time.

B. The study

The purpose of the study presented in this paper is to identify the different employments of data prediction methods in Wireless Sensor Networks (WSN). It was undertaken as a Systematic Literature Review (SLR) because this method allows systematic filtering (step-by-step) of a significant number of research works based in a strict set of criteria.

It is important to point out that the objective of the study is *not* to identify how data prediction methods are calculated, but how they are used in the context of WSN.

II. RELATED WORK

It seems that the study presented in this paper is unique. At least the author did not find any work in the literature with the same focus presented here. That is, to identify the different themes in Wireless Sensor Networks that employ data prediction methods.

Although there are a few works identifying the different data prediction methods used in WSN, like [1], none of them explores the different themes in WSN that use these methods.

III. METHOD

The study was undertaken as a Systematic Literature Review (SLR). The method employed was adapted from the guidelines proposed by da Silva et al. in [2].

A. Research Questions

The research questions addressed by this study are:

RQ1: What are the objectives of employing data prediction methods in WSN?

RQ2: What are the top-level application models employing data prediction methods in WSN?

The objective of the first question (RQ1) is to identify all kinds of data prediction uses related to WSN. It is important to point out that the purpose is to identify *where* data prediction methods are used and not *what* data prediction methods are used.

The objective of the second question (RQ2) is more restrictive, the purpose is to identify only data prediction methods used at the top-level of the WSN applications. That is, whose generated results are directly useful for the user of the WSN. This kind of prediction must necessarily use, at least as one of the inputs, the data measured by the sensors of the WSN.

B. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria used in the selection process of this study are:

RIEC1: Data prediction uses must be identified in all stages of the selection process. Works that do not describe any data predication use in WSN must be excluded.

RIEC2: Should only be included in the final stage of the selection process works that describe top-level uses of data prediction methods that effectively use the values measured by the sensors of a WSN.

C. Quality Assessment

Quality assessment was not employed in this study for two reasons:

- First, lack of peer-review. As only one person conducted this study, it was not possible to employ peer-review to qualify the works, which is a recognized good practice in SLR to avoid biased judgments [2].
- Second, the main objective of the study is to identify uses of data prediction in WSN and not to analyze in depth these uses. Therefore, to avoid exclude low-explored uses of data prediction, maybe only presented in a few low-quality works, all works were considered, without exclusion by quality criteria.

D. Search Process

The search process in this study refers to the task to judiciously search and download references to research works from a set of online digital libraries. Includes also the task to filter the downloaded references to exclude duplicates.

All search process was carried out manually, without the help of any special software code. However, to store and manage the references it was used the desktop version of the Mendeley system (http://www.mendeley.com). Mendeley was also used to store all data collected or extracted from the works along the study.

The parameters chosen as input to the search process are:

Document types: As the main purpose of the study is to identify all kinds of data prediction uses in WSN as possible, no restriction was imposed about the type of the documents.

Document languages: Only documents written in English.

Search terms: From the analysis of the research questions and the inclusion and exclusion criteria, the search terms chosen in this study are:

Terms that select the research field:

ST1.1: “Wireless Sensor Network”

ST1.2: “Internet of Things”

Term that restrict the search in the research field:

ST2.1: “Data Prediction”

Search string: Composed by the search terms, the search string used to search works in the digital libraries is:

SR1: (ST1.1 OR ST1.2) AND (ST2.1)

Search fields: The fields of the works used in the search process are:

SF1: Title

SF2: Abstract

SF3: Keywords

SF4: Main text

Search period: No restriction about the publication date of the works.

Search libraries: The following well-known online digital libraries were used as source for the works:

SL1: ACM Digital Library

SL2: IEEE Xplore

SL3: ScienceDirect

SL4: Engineering Village

SL5: Scopus

E. Selection Process

The selection process in this study refers to the task to check if the research works, previously selected in the search process, adhere to the inclusion and exclusion criteria. It was carried out in three stages:

- First, the title and abstract of the works were checked. Works that clearly did not adhered to the study criteria were excluded.
- Second, the introduction and conclusion section of the works were checked. Works that clearly did not adhered to the study criteria were excluded.
- In the last stage, the main text of the works were checked. Works that clearly did not adhered to study criteria were excluded.

During all stages it was also carried out for each work the identification of the themes in WSN that use data prediction methods.

F. Synthesis Process

The synthesis process in this study refers to the task to analyze and summarize all data collected or extracted from the research works.

G. Method Summary

The method employed in this study can be summarized in the Work Breakdown Structure (WBS) shown in Table I, with the respective inputs and outputs for each leaf activity.

IV. Results

This section summarizes the results of the study.

A. Search Results

The results of the search process are summarized in Table II. As shown, after exclude duplicates, 264 works were selected to the next stage of the study.

TABLE I. METHOD SUMMARY

Activity	Input	Output
Specification	-	-
Define research questions	Study objectives	RQ
Define inclusion and exclusion criteria	RQ	RIEC
Define quality assessment criteria	RQ	RQAC
Define document types	RQ	DT
Define document languages	RQ	DL
Define search terms	RQ, RIEC, DL	ST
Define search strings	RQ, ST	SR
Define search fields	RQ	SF
Define search period	RQ	SP
Define search libraries	RQ, DT, DL	SL
Search	-	-
Search in libraries	DT, DL, SR, SF, SP, SL	W1
Exclude duplicates	W1	W2
Selection	-	-
Selection by title and abstract	RIEC, W2	W3, RT1
Selection by introduction and conclusion	RIEC, W3, RT1	W4, RT2
Selection by main text	RIEC, W4, RT2	W5, RT3
Synthesis	RQ, W5, RT3	SYN
Documentation	Study objectives + all outputs	DOC

Abbreviations:

RQ: research questions
RIEC: inclusion and exclusion criteria
RQAC: quality assessment criteria
DT: document types
DL: document languages
ST: search terms
SR: search strings
SF: search fields
SP: search period
SL: search libraries
Wi: set of selected works i
RTi: set of research themes i
SYN: synthesis result
DOC: final study document

B. Selection Results

The results of the selection process are summarized in Table III. As shown, after the three stages of the process, only 14 works, fully adherent to all inclusion and exclusion criteria, were selected.

C. Data Prediction Themes

As mentioned above, during the selection process it was also carried out the identification of the themes described in each work that employ, *in one or more levels*, data prediction methods. This process identified 49 themes related to WSN, as shown in Figure 1, sorted by the number of works that describe the respective theme. Important: The author used the prefix "WSN" to emphasize that the theme is related to the internals of the WSN. For example, "WSN Energy Saving" means "saving the energy supply of some part of the WSN", and not "saving energy in homes or offices".

TABLE II. SEARCH RESULTS

ID	Source	# Works
SL1	ACM Digital Library	4
SL2	IEEE Xplore	144
SL3	ScienceDirect	24
SL4	Engineering Village	24
SL5	Scopus	143
W1	TOTAL	339
W2	After exclude duplicates	**264**

TABLE III. SELECTION RESULTS

ID	Selection by	# Works		
		# In	***# Out***	***# Excluded***
W3	Title and abstract	264	71	193
W4	Introduction and conclusion	71	19	52
W5	Main text	19	**14**	5

Note that two or more themes can occur in the same research work. For example, there is a strong relationship between "WSN Data Reduction" and "WSN Energy Saving", since reducing the data generated in the network usually imply saving energy to process these data.

It seems from Figure 1 that "WSN Energy Saving" is a big research issue. Which is reasonable due the nature of the WSN, where is common disconnected nodes from the power grid, especially when they are used in remote areas or in mobile scenarios.

The author identified at least four levels of employment of data prediction methods, as shown in Figure 2:

1) Node level: The nodes of the WSN use data prediction methods internally. For example, to adjust the sensors sample rate [3], reducing processing, communication, or storage demands.

2) Network level: Data prediction methods are used to improve the performance, lifetime, or reliability of the WSN. For example, data prediction at this level can be used to improve the positioning of the nodes of the WSN.

3) Data series level: Data prediction methods are used to correct or adjust the data series generated by the WSN and sent to the user.

4) Application level: The data that come from the WSN (previously adjusted or not) are used as input in top-level application models to generate some prediction directly useful for the end users of the WSN.

It is important to point out that the "user" of the WSN can be not only a human being (or another living creature) but also another system, like an actuators network.

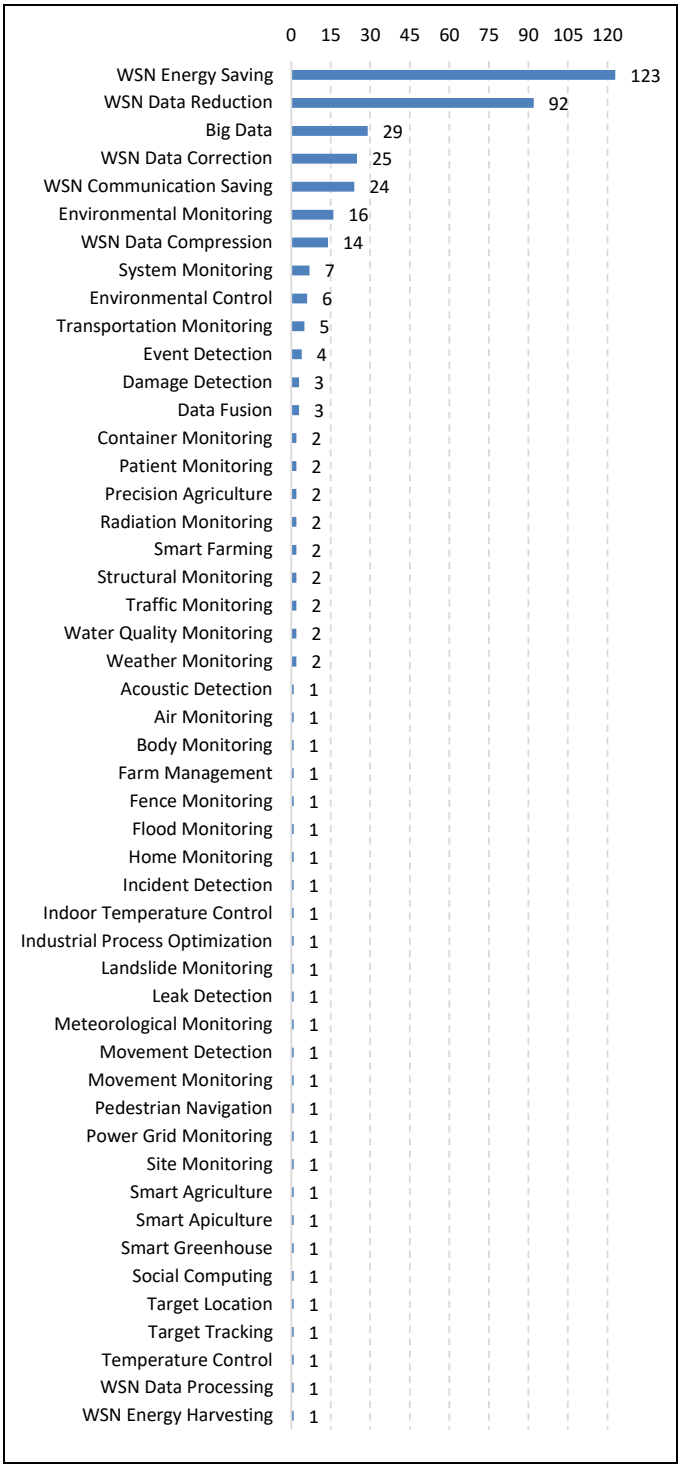

Figure 1. Data prediction themes in WSN sorted by number of works.

D. Top-level Data Prediction Themes

As mentioned above, only 14 works fully adherent to all inclusion and exclusion criteria were selected, especially to the criterion RIEC2, which filters only works that describe uses of data prediction methods at top-level application models.

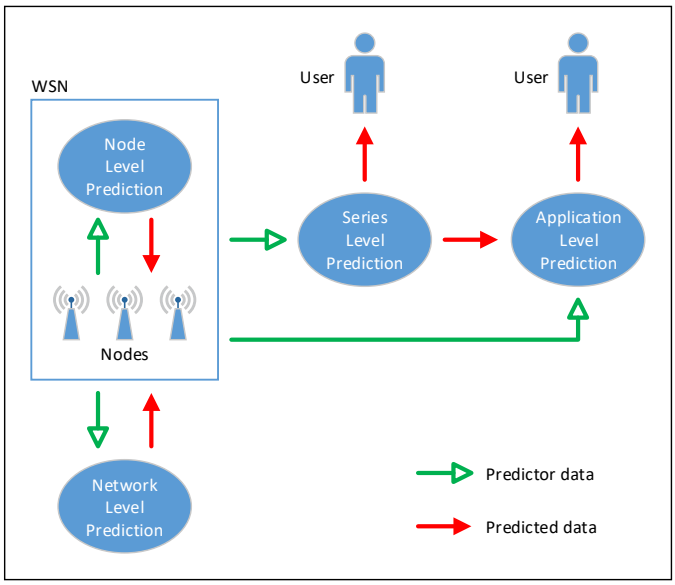

Figure 2. Data prediction levels in WSN.

From the 14 works, 13 themes with employment of data prediction methods at top-level were identified:

1) Fence Monitoring: Only one work was selected with this theme [5]. The purpose is to detect and classify the behavior of an agent at a fence, like kick, lean, peek, or climb. The author classified this theme as a subtheme of Event Detection.

2) Flood Prediction: Only one work was selected with this theme [15]. The purpose is to predict the occurrence of a flood in a given area. The author classified this theme as a subtheme of Weather Monitoring and Environmental Monitoring.

3) Landslide Monitoring: Only one work was selected with this theme [7]. The purpose is to predict the occurrence of a landslide in a given area. The author classified this theme as a subtheme of Environmental Monitoring.

4) Leak Detection: Only one work was selected with this theme [14]. The purpose is to detect the occurrence of a leak in a pipeline. The author classified this theme as a subtheme of Damage Detection and System Monitoring.

5) Movement Detection: Only one work was selected with this theme [16]. The purpose is to detect the presence of an agent that moves in a given area. The author classified this theme as a subtheme of Event Detection and Environmental Monitoring.

6) Pedestrian Navigation: Only one work was selected with this theme [4]. The purpose is to estimate the best comfort route for a pedestrian that walks in a given area, based on the environmental conditions. The author classified this theme as a subtheme of Environmental Monitoring.

7) Precision Agriculture: Only one work was selected with this theme [9]. The purpose is to monitor a crop area and

surroundings such that the farmers can be informed and control the best agricultural and environmental parameters for the crop. The author classified this theme as a subtheme of Environmental Monitoring. Here is important to point out that in the study was made a distinction between *Precision* Agriculture and *Smart* Agriculture. In Smart Agriculture the main objective is to build an agriculture system that auto-regulates, not necessarily in the best conditions, but more independent of the farmers work as possible.

8) Smart Apiculture: Only one work was selected with this theme [17]. The purpose is to build an auto-regulated system to beekeeping in a confined environment. The author classified this theme as a subtheme of Environmental Control.

9) Smart Greenhouse: Only one work was selected with this theme [8]. The purpose is to build an auto-regulated system to control the environment of a greenhouse. The author classified this theme as a subtheme of Environmental Control.

10)Structural Monitoring: Two works were selected with this theme [11] [12]. The purpose is to monitor a structure (for example, a bridge) and predict possible damages or faults.

11)Traffic Monitoring: Only one work was selected with this theme [10]. The purpose is to monitor the traffic of vehicles in a given area and assess the traffic conditions. The author classified this theme as a subtheme of Transportation Monitoring.

12)Water Quality Monitoring: Only one work was selected with this theme [6]. The purpose is to monitor a water body and estimate the water quality conditions. The author classified this theme as a subtheme of Environmental Monitoring.

13)Weather Monitoring: Only one work was selected with this theme [13]. The purpose is to monitor the weather in a given area and estimate the meteorological conditions. The author classified this theme as a subtheme of Environmental Monitoring.

It is important to point out that in this study the term "smart" is used to designate a system capable of self-regulate, composed by a combination of sensors and actuators and as independent as possible of a human operator.

V. Conclusion

As mentioned above, from 264 research works selected in the initial search of this study (Table II) only 14 (Table III) really use the data generated by the sensors of the WSN for top-level predictions. The significant difference between these numbers may indicate that, for the time being, the employment of data prediction methods in WSN is too focused in solve the operation problems of the networks than in generate direct results to the end users. At one side, this small number is disappointed, it was expected a bigger number of works. On the other hand, it can indicate that a lot of research still has to be done in this topic in WSN.

It was also identified that the themes in WSN employing data prediction methods can be classified in levels of processing of the full WSN applications (from bottom hardware to top software).

Acknowledgment

The author would like to thank the professors Flávia C. Delicato, Luci Pirmez, and Claudio M. de Farias for the valuable guidance.

References

[1] B. Stojkoska and K. Mahoski, "Comparison of different data prediction methods for wireless sensor networks," 10th Conf. Informatics, 2013.

[2] F. Q. B. da Silva, A. L. M. Santos, S. Soares, A. C. C. Frana, C. V. F. Monteiro, and F. F. MacIel, "Six years of systematic literature reviews in software engineering: An updated tertiary study," Information and Software Technology, vol. 53, no. 9. pp. 899–913, 2011.

[3] G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: A survey," Ad Hoc Networks, vol. 7, no. 3, pp. 537–568, 2009.

[4] C. Dang, M. Iwai, K. Umeda, Y. Tobe, and K. Sezaki, "NaviComf: Navigate pedestrians for comfort using multi-modal environmental sensors," in Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on, 2012, pp. 76–84.

[5] N. Dziengel, M. Seiffert, M. Ziegert, S. Adler, S. Pfeiffer, and J. Schiller, "Deployment and evaluation of a fully applicable distributed event detection system in Wireless Sensor Networks," Ad Hoc Networks, vol. 37, pp. 160–182, 2016.

[6] J. S. Horsburgh, A. Spackman Jones, D. K. Stevens, D. G. Tarboton, and N. O. Mesner, "A sensor network for high frequency estimation of water quality constituent fluxes using surrogates," Environ. Model. {&} Softw., vol. 25, no. 9, pp. 1031–1044, 2010.

[7] P. Kainthura, V. Singh, and S. Gupta, "Gis based model for monitoring and predition of landslide susceptibility," in 2015 1st International Conference on Next Generation Computing Technologies (NGCT), 2015, pp. 584–587.

[8] Y. Kaneda, H. Ibayashi, N. Oishi, and H. Mineno, "Greenhouse Environmental Control System Based on SW-SVR," Procedia Comput. Sci., vol. 60, pp. 860–869, 2015.

[9] T. D. Le, D. H. Tan, T. Dinh Le, and D. H. Tan, "Design and deploy a wireless sensor network for precision agriculture," in Information and Computer Science (NICS), 2015 2nd National Foundation for Science and Technology Development Conference on, 2015, pp. 294–299.

[10] Y. Ma, M. Chowdhury, A. Sadek, and M. Jeihani, "Real-Time Highway Traffic Condition Assessment Framework Using Vehicle-Infrastructure Integration (VII) With Artificial Intelligence (AI)," IEEE Trans. Intell. Transp. Syst., vol. 10, no. 4, pp. 615–627, 2009.

[11] K. Roy, H. Ogai, B. Bhattacharya, S. Ray-Chaudhuri, and J. Qin, "Damage Detection of Bridge using Wireless Sensors," IFAC Proc. Vol., vol. 45, no. 23, pp. 107–111, 2012.

[12] B. Samali, J. Li, and K. Aboura, "Monitoring Bridge Deterioration Using Sensors," in 2008 Third International Conference on Broadband Communications, Information Technology {&} Biomedical Applications, 2008, pp. 64–69.

[13] Y. Suzuki, H. Ibayashi, Y. Kaneda, and H. Mineno, "Proposal to Sliding Window-based Support Vector Regression," Procedia Comput. Sci., vol. 35, pp. 1615–1624, 2014.

[14] K. Wang, L. Zhuo, Y. Shao, D. Yue, and K. F. Tsang, "Towards Distributed Data Processing on Intelligent Leakpoints Prediction in Petrochemical Industries," IEEE Trans. Ind. Informatics, vol. PP, no. 99, p. 1, 2016.

[15] I. Wijesundera, M. N. Halgamuge, T. Nirmalathas, and T. Nanayakkara, "A Geographic Primitive-Based Bayesian Framework to Predict

Cyclone-Induced Flooding*," J. Hydrometeorol., vol. 14, no. 2, pp. 505–523, 2013.

[16] F. Xiao, X. Xie, H. Zhu, L. Sun, and R. Wang, "Invisible Cloak Fails: CSI-based Passive Human Detection," in Proceedings of the 1st Workshop on Context Sensing and Activity Recognition - CSAR '15, 2015, pp. 19–23.

[17] A. Zacepins, J. Meitalovs, V. Komasilovs, and E. Stalidzans, "Temperature sensor network for prediction of possible start of brood rearing by indoor wintered honey bees," in 2011 12th International Carpathian Control Conference (ICCC), 2011, pp. 465–468.