Sensor Network deployment based on data variability

Roberto Henriques[†], Rui M. Rocha[¥]
[†] Instituto Superior de Estatística e Gestão de Informação da Universidade Nova de Lisboa,
Campus de Campolide, 1070-312 Lisboa, Portugal
Phone: +351 21 387 04 13, Fax: +351 21 387 21 40, e-mail: roberto@isegi.unl.pt
[¥] Instituto de Telecomunicações, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal
Phone: +351-218418455, Fax: +351-218418472, e-mail: rui.rocha@lx.it.pt

Abstract — **Deployment of sensor nodes is one of the key issues in wireless sensor networking. This issue is related with the fact that usually the number of sensor does not allow an individual deployment and, usually in these situations, sensors are randomly spread. Here we present a new method to evaluate the network sensing coverage, identifying those areas where more sensor nodes are needed. This method uses the sensed data variability to estimate a variability coverage index.**

I. INTRODUCTION

A sensor network is a collection of sensor nodes organized in a cooperative way. In a sensor network, each node has sensing, computation and communication capabilities, *i.e.* each sensor node is capable of sense its surroundings, process this data and communicate with other nodes forming a collaborative system capable of undertaking specific tasks. These inexpensive, low-power communication devices can be deployed throughout a physical space, providing dense sensing, quite close to physical phenomena [1]. There are several different applications where sensor networks are used such as environment monitoring [2], military surveillance or enemy tracking [3], health [4], education [5] and other commercial applications.

Depending on the number of sensor to be deployed, it is usually not possible to define each node position and, in many situations, sensors are randomly deployed [6]. In these situations, one of the fundamental issues in a wireless sensor network is the sensor nodes' deployment and consequently the coverage problem [7]. Some sensor nodes can possess locomotion capabilities forming what can be called a mobile sensor network. Nodes' mobility provides better coverage of the environment, improved response to changes and active information gathering capabilities. Another advantage of this mobility is the ability to self-deployment, *i.e.*, starting from some initial spatial arrangement the nodes in the network can spatially adapt themselves such that the sensed area is maximized [8].

Several coverage methods have been presented in the last years, from distributed algorithms to some others broadly inspired in facility location. However, to the best of our knowledge, there are few methods, which we will analyze in more detail, that take into account the sensed data variability.

Full coverage of a monitored area is a very important feature in a WSN, being usually associated with the detection of discrete phenomena, *e.g.* tracking people, vehicles, etc. In the case of continuous phenomena, such as temperature, when can one consider to have a full coverage? If the monitored phenomenon is continuous, we should possess an infinite number of nodes to fully cover the area. Physically, that is not possible, and the solution to cover this type of phenomena is to place nodes according to its variability. As an analogy, we can think in the work of cartographers that build soil maps. Several soil samples are collected from the field and the final map results from interpolating each sample data. Sensor nodes can act as sample suppliers, with the benefit of providing streams of data in temporal and spatial dimensions.

The goal of the paper is to present a new methodology to evaluate, in a decentralized way, the network sensing coverage, identifying those areas where a higher number of sensor nodes are needed. Based on the coverage of each sensor and on the sensed data variability, a new variability coverage index is calculated. In this calculation an artificial neural network is used, the self--organizing maps (SOM). This type of neural network is specialized in data clustering and visualization. In the particular case, we use SOM to process the sensed data and calculate the proposed index. Using this index, sensor nodes deployment may be adjusted (manually or automatically depending on the sensor network mobility) providing a better spatial and data variability coverage.

This paper is organized as follows. In section 2 we present the related work review along with a survey on self-organizing maps and sensor networks. In section 3 the proposed method is presented, with some simulation and tests presented in section 4. Finally, section 5 is devoted to the discussion of the main results obtained in this research work, followed by the appropriate conclusions.

II. COVERAGE OF SENSOR NETWORKS

One of the fundamental issues in sensor networks is the deployment, which will affect how well a sensor network is monitoring the environment. This is also called the sensing coverage of a network, which can be defined as the area that is being sensed by a group of nodes. Ideally, the cover area should be equal to the study area (except the obstacles) without holes. One of the first coverage definitions was the measure of quality of service of a sensor network [5]. Other authors define coverage, in a probabilistic view, as the probability that any target point is covered by the nodes detection range [9].

This work was supported in part by the Fundação para a Ciência e a Tecnologia under PhD grant SFRH/BD/30360/2006.

A. General Approach to WSN Deployment

To define which areas are insufficiently covered [7] proposes a method based on the cover of the perimeter of each sensor's sensing range. While each sensor perimeter is sufficiently covered, the whole area is sufficiently covered. Several authors have proposed different sensor networks covering methodologies in the past few years [6, 8, 10].

Howard et al. [10] uses a potential-field-based approach to spread sensor nodes throughout the environment, in which nodes are treated as virtual particles, subject to virtual forces. These virtual forces will motivate the nodes to avoid other nodes and obstacles, achieving an equilibrium state after a period of time. Lam et al. [8] use a regular metric grid (isogrid) to deploy the sensor network. Using this concept of isogrid the algorithm attempts to reposition the nodes on the grid vertex. This method seems to succeed in maintaining the network connectivity and topology, even in the presence of obstacles. Chellappan et al. [6] proposes a minimum-cost maximum-flow approach using a graph to model the problem. This graph is constructed using the nodes position (nodes) and possible paths (edges). Each node determines its position and region and sends this information to a central node. From this point, a central node processes the information and proposes an optimized movement plan which is sent and followed by all nodes.

B. SOM Approach to WSN Deployment

Teuvo Kohonen proposed the self-organizing maps (SOM) in the beginning of the 1980s [11] as a result of his work on associative memory and vector quantization. One of SOMs main objectives is to "extract and show" the essential structures in a dataset, through a map resulting from an unsupervised learning process. The SOM is normally used as a tool for mapping high-dimensional data into one, two, or three-dimensional feature maps. The basic idea of a SOM is to map the data patterns onto an n-dimensional grid of neurons or units. This mapping tries to preserve topological relations, i.e., patterns that are close in the input space will be mapped to units that are close in the output space, and vice-versa. The output of the SOM will be a set of neurons with weights. One way to visualize the distances between units is the use of the U-Matrices [12]. U-Matrices were originally proposed by Ultsch [12] and they are computed by finding the distances, in the input space, between neighboring units in the output space. These distances are represented using a color scheme.

The SOM algorithm has been used in several sensor networks applications. Some authors use variants of the SOM to create clusters of sensed data and obtain a topological SOM where similar data is closer [13]. Other authors present a SOM variant to perform a dynamic power management and thus saving energy [14]. SOM has also been used in the context of sensor networks deployment [15]. In this application the authors used the SOM as a tool for density discovery, placing the sensors in a way that minimizes the distance to all events detected. Some limitations on this work refer to its use before the sensor network is deployed defining each node best position. To predict the spatial pattern there is also a need to known in detail the phenomena to monitor. Finally this implementation uses a centralized algorithm to perform this operation.

III. Coverage Variability (KV) Index

The deployment problem is seen as a coverage problem. Thus, most of the methods algorithms try to cover the study area, with minimum overlaps. However, this is only possible to achieve if a necessary number of nodes exists. In the case of insufficient nodes to cover the whole area, holes will exist. In our case, we also try to cover the largest area possible, however if the number of nodes is insufficient, we propose the use of an index to maximize the deployment of nodes.

In this paper we proposed a new method, using SOM, to create a coverage variability index (*kv*), for each sensor node. This index will evaluate the sensed data variability in the area covered by the sensor network. Hence, higher values for this index represent areas where the number of sensor nodes is insufficient to cover the data variability. From this individual node index, it is also possible to calculate global variability coverage. The goal is to achieve a map where we present the coverage variability index distribution across the sensed area.

Let a sensor network S consisting of n nodes be $S = \{N'_0, N'_1, \ldots N'_n\}$. Each, node N'_i has a position defined by x_i and y_i and has the capability of capturing data D_i from the environment. Moreover, each node also has a communication range which intersects other nodes, being this neighborhood defined by $Vi = \{N'_0, N'_1, \ldots N'_m\}$ consisting of m neighbor nodes. Using each node communication capabilities, messages with the sensed data are exchanged across neighbors. Taking advantage of processing capabilities, a SOM is trained based on previous data. One dimension SOM is used, and the coverage variability (*kv*) index could be calculated using the neurons weights. Thus, the *kv* index is the average of the difference between neuron weights:

$$kv_{index} = \frac{\sum_{i=1}^{j-1} |u_i - u_{i+1}|}{j} \quad (1)$$

Where j is the number of neurons used in the SOM and u_i is the neuron i weight. From the analysis of the *kv* index, we can infer about the coverage on data variability on the area defined between the sensor node and its neighbors. From each node *kv* index we calculate a map with global coverage information, for the area of interest, as:

$$gkv_{x,y} = \frac{\sum_{i=1}^{j} kv_i}{j} \quad (2)$$

Where $gkv_{x,y}$ is the global coverage calculated for each point x, y, inside the area of interest, and j is the number of nodes with the point x, y inside its sense range.

To better explain the proposed method let us consider the following example. Fig. 1 shows an area with four nodes (s_1 to s_4). The solid isolines represent the phenomena to be monitored, temperature in this case. These lines result from the union of all the points where the temperature is equal. The dash line represents the s_1 communication range; therefore s_1 exchanges messages with s_2 and s_3.

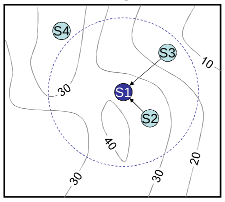
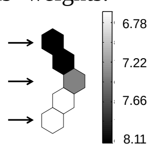


Fig. 1 – Sensor nodes in the area of interest.

Using the temperature data from its neighbors (aprox. 35ºC and 15ºC) and its own value (aprox. 35ºC), s_1 will train a SOM. A 3x1 SOM was used and the correspondent U-matrix is represented in Fig. 2. The arrows point to the units, while the hexagon between them represents their relative distance. Also in Fig. 2, we present the neurons' weights.

Neuron	Weight
1	33.43
2	26.65
3	18.53



a) b)

Fig. 2 – SOM outputs for s1; a) Neurons weights b) U-matrix

Using the SOM neurons weights, we calculate, for each node, the *kv* index which is shown in Table I (in this example, we assume that s_2 and s_3 have the same neighbor's range than s_1, producing the same *kv* index)

Table I
Sensor nodes' *Kv* indexes

Node	*Kv* index
s_1	1.33
s_2	1.33
s_3	1.33

Finally, from the set of calculated *kv* index it is also possible to derive a global coverage for the all area. Fig. 3 depicts the global coverage index, given by the average of the *kv* index in each point (in each point only the *kv* index belonging to the nodes that are in the communication range are used).

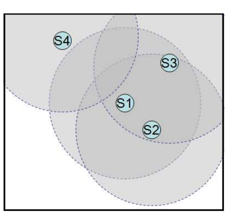
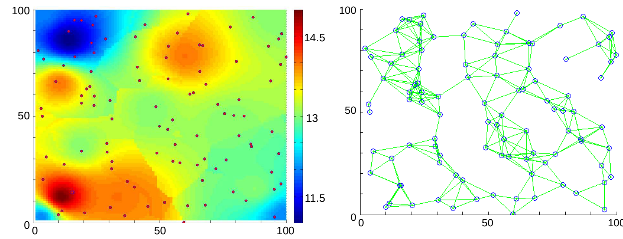


Fig. 3 – Global coverage index

IV. SIMULATIONS

To simulate the proposed method we have used Matlab®. In this simulation a set of sensor nodes (100 nodes) are randomly deployed over a certain area (100x100 units) and we continue to assume temperature as the variable of interest (which was also generated for this purpose).

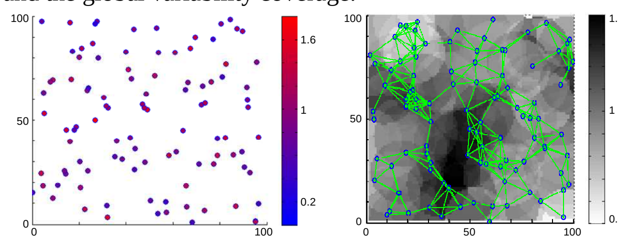
Sensors randomly displayed over this area are shown in Fig. 4 along with their communication connectivity. It is assumed that the communication range is homogenous for all the nodes. We also assume that for all the neighbors' nodes closer than a communication threshold, it is possible to exchange messages in both ways.



a) b)

Fig. 4 – Simulation; a) nodes generated (100) in a random position along with the temperature in the area of interest (100x100 units); b) nodes and their connectivity.

Fig. 5 shows the *kv* index calculated for each sensor node and the global variability coverage.



a) b)

Fig. 5 – Sensor nodes kv index; a) value of kv index for each node; b) global variability coverage and nodes' connectivity.

It is worthwhile to note that the regions with higher *kv* (darker areas) are those with a low number of nodes and, mostly, those where the communication between sensor nodes does not exist (Fig. 5b).

In order to evaluate how the proposed method reacts to changes in the initial parameters we calculate the index for the cases of increasing the number of sensors nodes and increasing the sensor communication range.

A. Changing the number of sensor nodes

In this test, the number of sensors is increased using two different approaches. In the first one, we random deployed the new added nodes, while in the second approach those nodes were added to the neighborhood of the nodes with higher *kv* index. Fig. 6 shows the mean global variability coverage along with the increase in the number of nodes. From Fig. 6 we can conclude that adding more sensor nodes produce similar results independent of the approach. We can also see that an increase in the number of sensors tends to produce a higher mean global variability index, although the opposite was expected. This could be explained by the fact that increasing the number of sensor nodes will reveal further holes. This holes detection will result in an increase on the *kv* index, since new uncovered data is added to this estimation.

B. *Changing the sensor nodes communication range*

In this second test, we calculate the mean global variability coverage increasing the sensor node communication range. From Fig. 7 we can conclude that, generally, adding more sensor nodes will produce a decrease in the mean global variability coverage index. This is an expected result, since the increase of the communication range will increase the number of neighbors exchanging messages. Using a higher number of temperature readings to calculate the *kv* index will tend to produce lower values, until a certain threshold. In this case, for a communication range of 40 meters the mean global coverage value will boost.

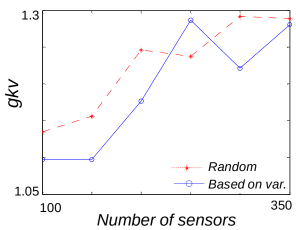


Fig. 6 – Mean global variability coverage changing the number of sensor nodes.

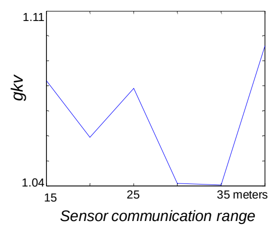


Fig. 7 – Change in the sensor nodes communication range

V. CONCLUSION

In this paper we have presented a new method to evaluate the variability on the sensed data, identifying those areas where a higher number of sensor nodes are needed. Simulation tests have shown that the calculated index has higher values for less covered areas and for areas where the connectivity between nodes is limited. For changes in the number of nodes and in the communication range, results are biased by the existence of holes in the area.

As future work we propose to adapt the method such that only clusterhead nodes have to process SOM. This is an important change, since computation and energy resources are very limited. Also as future work, we will implement the possibility of moving nodes towards the zones with higher coverage variability index.

REFERENCES

[1] D. Culler, E. Deborah, and M. Srivastava, "Overview of Sensor Networks," in *IEEE Computer, Special Issue in Sensor Networks*, 2004, pp. 41-49.

[2] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*. Atlanta, Georgia, USA: ACM Press, 2002.

[3] Á. Lédeczi, A. Nádas, P. Völgyesi, G. Balogh, B. Kusy, J. Sallai, G. Pap, S. Dóra, K. Molnár, M. Maróti, and G. Simon, "Countersniper system for urban warfare," *ACM Trans. Sen. Netw.*, vol. 1, pp. 153-177, 2005.

[4] J. A. Stankovic, Q. Cao, T. Doan, L. Fang, Z. He, R. Kiran, S. Lin, S. Son, R. Stoleru, and A. Wood, "Wireless Sensor Networks for In-Home Health-care: Potential and Challenges," presented at High Confidence Medical DeviceSoftware and Systems (HCMDSS) Workshop, Philadelphia, PA, 2005.

[5] S. Meguerdichian, F. Koushanfar, M. Potkonjak, and M. B. Srivastava, "Coverage Problems in Wireless Ad-hoc Sensor Networks," presented at IEEE Infocom, 2001.

[6] S. Chellappan, X. Bai, B. Ma, D. Xuan, and C. Xu, "Mobility Limited Flip-Based Sensor Networks Deployment," *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, pp. 199-211, 2007.

[7] C.-F. Huang and Y.-C. Tseng, "The Coverage Problem in a Wireless Sensor Network " *Mobile Networks and Applications*, vol. 10, pp. 519-528, 2005.

[8] M.-l. Lam and Y.-h. Liu, "ISOGRID: an Efficient Algorithm for Coverage Enhancement in Mobile Sensor Networks," presented at Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on, 2006.

[9] Y. Wang, X. Wang, D. P. Agrawal, and A. A. Minai, "Impact of Heterogeneity on Coverage and Broadcast Reachability in Wireless Sensor Networks," presented at Computer Communications and Networks, 2006. ICCCN 2006. Proceedings.15th International Conference on, 2006.

[10] A. Howard, M. J. Mataric, and G. S. Sukhatme, "Mobile Sensor Network Deployment using Potential Fields: A Distributed, Scalable Solution to the Area Coverage Problem," presented at DARS 02, Fukuoka, Japan, 2002.

[11] T. Kohonen, "Self-organizing formation of topologically correct feature maps," *RecMap: rectangular map approximations*, vol. 43 (1), pp. 59-69, 1982.

[12] A. Ultsch and H. P. Siemon, "Kohonen´s self-organizing neural networks for exploratory data analysis," presented at Proceedings of the International Neural Network Conference, Paris, 1990.

[13] E. Catterall, K. V. Laerhoven, and M. Strohbach, "Self Organization in Ad Hoc Sensor Networks: An Empirical Study," presented at Proc. of Alife VIII: the 8th International Conference on the Simulation and Synthesis of Living Systems, Sydney, Australia, 2002.

[14] Q. Huang, Y. Bai, and L. Chen, "SOM-based dynamic power management (SDPM) framework for wireless sensor networks " presented at Proceedings of the 2006 international conference on Wireless communications and mobile computing, Vancouver, British Columbia, Canada, 2006.

[15] C. Koutsougeras, Y. Liu, and R. Zheng, "Event-driven Sensor Deployment Using Self Organizing Maps," *International Journal of Sensor Networks* vol. 3, 2008.