Wireless sensor networks for monitoring the environmental activities

Conference Paper · January 2011

DOI: 10.1109/ICCIC.2010.5705791 · Source: IEEE Xplore

2 authors, including:

Ruchi Mittal
Netaji Subhas Institute of Technology
25 PUBLICATIONS **161** CITATIONS

Wireless Sensor Networks for Monitoring the Environmental Activities

Ruchi Mittal[1], M.P.S Bhatia[2]
[1]Department of Computer Science and Engineering, Netaji Subhash Institute of Technology, New Delhi, India
[2]Department of Computer Science and Engineering, Netaji Subhash Institute of Technology, New Delhi, India
(ruchi.mittal138@gmail.com, bhatia.mps@gmail.com)

Abstract - **The area of sensor network has a long history and many kind of sensor devices are used in various real life applications. Here, we introduce Wireless sensor network which when combine with other areas then plays an important role in analyzing the data of forest temperature, bioinformatics, water contamination, traffic control, telecommunication etc. Due to the advancement in the area of wireless sensor network and their ability to generate large amount of spatial/temporal data, always attract researchers for applying data mining techniques and getting interesting results. Wireless sensor networks in monitoring the environmental activities grows and this attract greater interest and challenge for finding out the patterns from large amount of spatial/temporal datasets. These datasets are generated by sensor nodes which are deployed in some tropical regions or from some wearable sensor nodes which are attached with wild animals in wild life centuries. Sensor networks generate continuous stream of data over time. So, Data mining techniques always plays a vital role for extracting the knowledge form large wireless sensor network data. In this paper, we present the detection of sensor data irregularities, Sensor data clustering, Pattern matching and their interesting results and with these results we can analyze the sensor node data in different ways.**

Keywords **- wireless sensor network, clustering, pattern matching, mining, sensor node.**

I. INTRODUCTION

Wireless sensor network grows and rapidly improves, this enable the new communication services. Sensor networks are the most useful way to collect the various parameters and information needed by environments like industrial, home, shipboard, buildings, utilities, transportation systems automation etc. Recent various areas like terrorist and animal warfare countermeasures require distributed sensor networks that can be deployed easily using aircraft, and have self-organizing capabilities etc. with the advancement in the recent technology development of small, battery-powered, wireless sensor nodes [2] [5] [6] is enabled. But, sensors have serious resource constraints including battery lifetime, communication bandwidth, CPU capacity and storage [16].also, due to increase in mobility of the sensor node, may increase the complexity of sensor data because a sensor may be in a different neighborhood at any point of time [17]. Whenever we need to install a sensor network then it must be fast, easy to install and maintain. Sensor network basically consist of large amount of sensor nodes that are deployed to large physical area to monitor and detect the real time environmental activities. These sensor nodes works together to collect the data like temperature, humidity, acceleration etc from surroundings. As, wireless sensor networks consisting of such sensors nodes those create exciting opportunities for large scale, data intensive measurement and surveillance applications. In such type of application, in order to take intelligent decision mine sensor data or readings for pattern. In this paper, we study the challenges, problems and the solution comes out to be from mining data from live sensor nodes.

As sensor network is useful in application like in habitat monitoring, health monitoring, traffic, weather, pollution etc and in all such real life application sensor nodes generate large amount of data so mining data is really a fruitful task. Also, wild life habitat surveillance [11], battlefield troop coordination, and traffic monitoring are the initial application of data mining.

Due to advancement in the wireless sensor networks the networks have ability to generate a large amount of data, and to find out the useful knowledge regarding the sensor network we apply data mining techniques.

Here we link up the wireless sensor network with environment monitoring and this link helps in various areas like in fire detection in forest areas, saving wild life, and in other tropical conditions by analyzing temperature, humidity etc.

In this paper, we present the few important data mining techniques like data pre-processing, cluster analysis and pattern matching. These are the techniques which we applied to the senor data so that we got the appropriate result and we analyze the results.

We organize this paper as follows. Section 2 reviews the related works in wireless sensor network field and briefly discusses the background knowledge related to sensor network and data mining techniques. Section 3 presents the dataset that we had taken for the experiment. Section 4 presents the implementation strategies of the data mining techniques which we applied to sensor data. Section 5 shows the results. Section 6 concludes this paper. Section 7 presents the future scope.

II. LITRATURE SURVEY

Area of sensor network is very wide and used in various applications. There has been much work done in data mining, sensor network, data stream etc. but very less work has been done at the combination of these areas. So we introduce the combination of wireless sensor network and data mining to get some interesting and fresh results.

Research in WSNs area has focused on two separate aspects of such networks, namely networking issues, such as capacity, delay, and routing strategies; and application

-

978-1-4244-5967-4/10/$26.00 ©2010 IEEE

issues [15]. The beauty of sensor networking protocols is that they attracted a tremendous amount of research effort [5]. For large sensor network the management of sensor database is itself a big task. Query processing techniques have been proposed for acquiring and managing sensor data [11]. One major research goal of this problem in the database community is to efficiently detect outliers in a large-scale database [12]. To collect the data we treat the Sensor networks as databases and SQL queries [13]. Clustering of sensor network [3] is related with network topology, not on the sensor data.

In our study we are basically concentrating towards Spatial-Temporal data. Traditional data mining works on Association Rule [7], looking for patterns of the form: "Customers who buy bread, also buy milk, with probability x%." There is significant literature [8]. The area of spatial-temporal data is limited and always looking for rules of the form "customers who buy cars now, will buy tires in 2 years."[9], with such kind of rules we can extend our research area. We will use some of the matching algorithms to detect correlated sequence.

There is much literature regarding irregularity detection [14]. But, none of these approaches are directly applicable to a wireless sensor network environment.

In this paper, we are focusing towards the area of detection of irregular pattern. Either detect irregular sensory data of a single attribute with respect to time [10] and space or detect the irregular patterns of multiple sensory attributes.

The main problem with detecting irregularities is to find those sensory values that deviate significantly from the norm [4]. This problem is very important because if we are getting irregular data continuously then we can detect weather the tropical condition around the sensor node are not well or we have faulty sensor node . So, Detection of irregularities in the data strongly related to modelling of sensor data.

To find out the irregularities in the multi-attribute pattern detection problem we assume that that there are some normal patterns among multiple sensory attributes, which are true in some natural phenomena. Hence, whenever some of these normal patterns are broken somewhere, then the irregularity in the data is detected and reported. Similarly, in finding the irregularities in the single attribute sensor data problem, we examines the temporal and spatial characteristics of a sensor node and find out the irregularity in comparison with the node's previous data or the data of the neighbor node.

III. DATASET

We download the dataset for our work from http://daac.ornl.gov/LBA/guides/CD04_Meteorology_Fluxes.html, this data is presented as value measured at 30 minute interval over 3.5 years and compiled at the km 83 tower site. This data includes the variables relate to meteorology, soil moisture, fluxes of momentum, heat, water vapor, and carbon dioxide beneath the flux sensors.

III. IMPLEMENTATION STRATEGY

The goal of the work is the development of new data-mining techniques for knowledge discovery in monitoring environmental activity like detecting forest fire databases and a decision framework is used to inform the development and placement of the sensor networks. Thus, our goal is not restricted to one environmental problem, but it will produce a generalizable solution to the issue of interpreting and using environmental data. Hence we need huge amount of data for the well defined and accurate results.

A. Sensor Data Pre-processing

Data Pre-processing involves cleaning the data by putting in missing values and removing uninteresting data. It may also include summarization and aggregation of the data. This step basically involves preparing the data for analysis [2].

So, firstly we can detect the irregularities in the sensor data and apply pre-processing technique. Pre-processing involves cleaning the data by putting in missing values and removing uninteresting data. It may also include summarization and aggregation of the data. This step basically involves preparing the data for analysis [2]. Hence with the help of weka's filtering tool we pre-process our dataset.

B. Sensor data Clustering

The term Data Clustering means the process of organizing or arranging a set of objects into groups or clusters in such a manner so that objects within a cluster have the most similarity to one another and the most dissimilarity to objects in other clusters.

Clustering is a useful and basic task in data mining. We generally used it as initial step in data mining for future data analysis. So, here we purpose clustering of sensor data along with their sensor attributes. Here, we use the simple of K-Means clustering algorithm for making clusters of sensor node data in a wireless sensor network.

TABLE I. USEFUL DATA FOR MINING

Attribute	**Description**
Solar_in	mean daily downward flux of solar radiation
Solar_out	mean daily upward flux of solar radiation
T64	mean daily air temperature at 64 m height
T40	mean daily air temperature at 40 m height
T10	mean daily air temperature at 10 m height
T2	mean daily air temperature at 2 m height
Press	mean daily air pressure measured at 64m height
WS_64	mean daily wind speed measured on the sonic anemometer at 64 m height
WD_64	mean daily wind direction measured on the sonic anemometer at 64 m height
Rain	total daily rainfall measured at 64 m height using a

	tipping bucket rain gauge
Tsoil_10	mean daily soil temperature at 10 cm depth
Tsoil_20	mean daily soil temperature at 20 cm depth
Tsoil_50	mean daily soil temperature at 50 cm depth
h2o_soil_10	mean daily soil moisture at 10 cm depth
h2o_soil_20	mean daily soil moisture at 20 cm depth
h2o_soil_40	mean daily soil moisture at 40 cm depth
Co2_64	mean daily atmospheric CO2 concentration at 64 m height
Fco2	mean daily CO2 flux at 64 m

These are the attributes to which we are focusing and do our research work.

IV. RESULTS

Here, we firstly show the result of the clustering by applying the k-mean algorithm, where we take the value of k=6 because with this value number of squared errors within the cluster is reduced. Also, when we apply the k-mean algorithm to our dataset then it creates 1354 instances and missing values were replaced with mean/mode. Now, for each cluster number of instances was divided as shown in table :

TABLE II. CLUSTER INSTANCES

Cluster ID	Clusters
	Cluster Instance
0	303 (22%)
1	256 (19%)
2	79 (6%)
3	154(11%)
4	359 (27%)
5	203 (15%)

The cluster centroids for all the clusters and for each attribute can be given as follow:

TABLE IIII. CLUSTERING RESULTS OF THE SELECTED ATTRIBUTES

Attribute	Full Data (1354)	0 (116)	1 (102)	2 (493)	3 (151)	4 (350)	5 (142)
year	2001.8331	2002.2328	2001.9118	2002.0832	2003.2649	2000.6029	2002.0915
month	6.6713	6.5862	5.0196	6.499	7.6623	6.8029	7.1479
day	15.664	15.6207	14.7549	16.0588	14.6954	15.8257	15.6127
solar_in	-4.7675	198.5737	-648.218	175.4889	176.9272	165.8195	-948.1711
solar_out	-451.5165	-1.8369	-909.4881	23.1958	-280.725	-999	-970.2004
T64	-96.7335	8.4834	15.0824	21.8444	12.2302	-12.1647	-999
T40	-97.0646	8.3029	14.6759	21.7249	12.087	-13.0373	-999
T10	-97.7013	7.4788	14.2245	20.8989	11.3784	-13.6268	-999
T2	-37.4048	24.8921	24.2897	20.6637	-2.4604	24.2503	-523.3404
press	-207.8238	-33.8519	-945.1995	-26.0778	-112.2577	-39.3747	-968.0839
WS_64	-191.5058	-6.2504	-763.5053	-28.1206	-77.3834	-63.7173	-935.5408
WD_64	-132.9028	41.8833	-676.3456	59.1769	89.3266	-66.607	-951.9135
rain	3.8651	1.8106	8.3048	3.2205	4.3746	4.1979	3.2302
Tsoil_10	-114.7901	25.3259	-35.1826	25.2691	-1.9386	-235.5064	-595.16
Tsoil_20	-35.3804	25.35	25.093	25.288	-1.9149	24.8426	-523.084
Tsoil_50	-35.2962	25.4324	25.2413	25.3692	-1.7895	24.9087	-523.0332
h2o_soil_10	-58.7034	0.2711	0.3717	0.3008	-26.1011	0.396	-534.5023
h2o_soil_20	-196.6384	0.4056	-195.5056	0.4306	-999	0.4736	-675.2321
h2o_soil_40	-196.6476	0.4157	-195.5243	0.4306	-999	0.4427	-675.2379
co2_64	90.2457	-999	-54.5746	395.3443	68.3337	279.6214	-418.6432
fco2	-237.7173	-104.438	-969.6114	-71.6449	-74.5962	-78.2857	-963.8668

The above results have more accuracy because as we increase the value of k then the squared error decrease and we get more valuable result. Now when we want to figure out the different climate attributes for different aspects these results really help. For example let we want to determine the forest fire in a particular region then using the above clustering result we will easily figure out the area.

Here, we plot the graph for some attributes. The analysis graphs are done by using weka tool. In these graphs the x-axis denotes the time slice and the y-axis denotes the various attributes.

- Wind Analysis

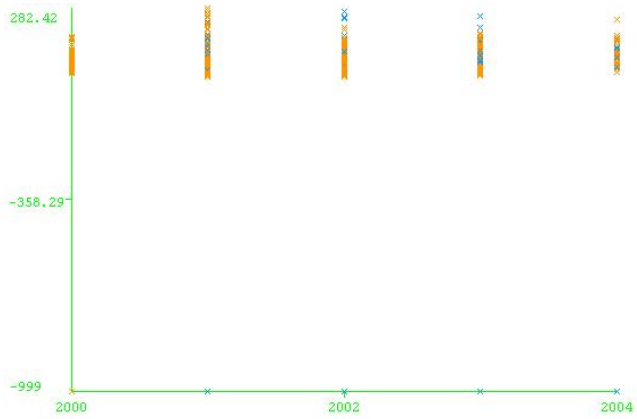

Fig 1. Graph soil moisture at 40 cm

- Temperature (64 m) Analysis

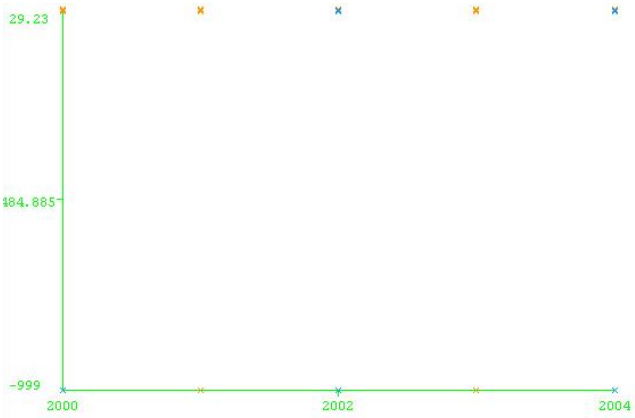

Fig 2. Graph Temperature at 64 m

- Temperature (2 m) Analysis

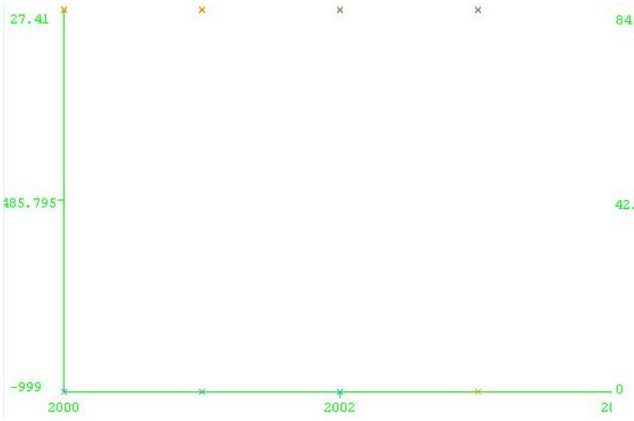

Fig 3. Graph temperature at 2 meter

- Downward flux of solar radiation Analysis

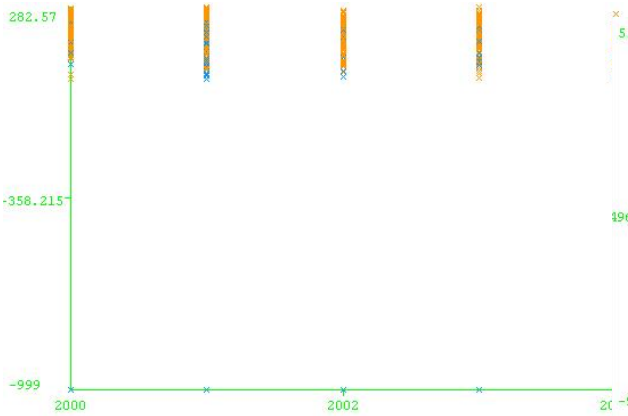

Fig 4. Graph for downward flux of solar radiation

- Upward flux of solar radiation Analysis

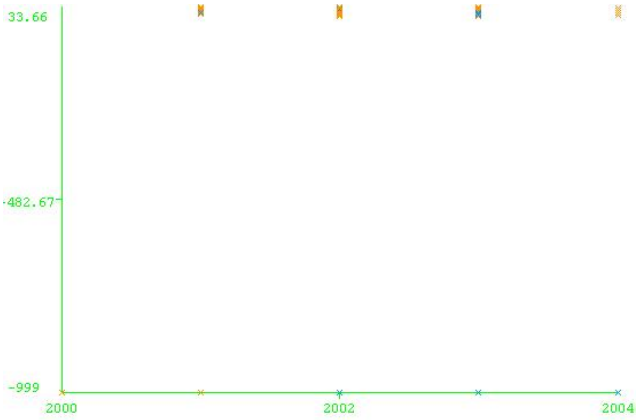

Fig 5. Graph upward solar flux

- Rain Analysis

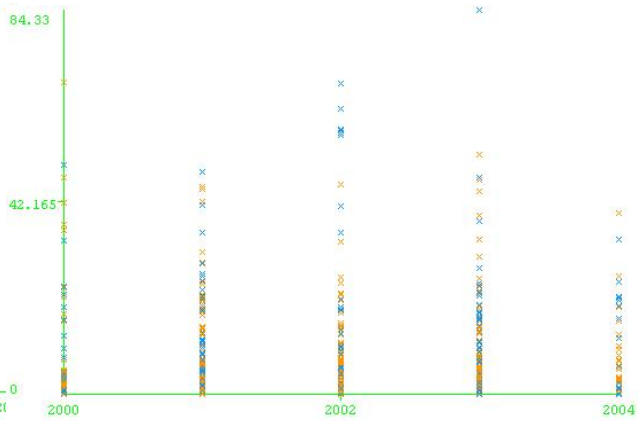

Fig 6. Graph for rain

- CO2 Flux Analysis

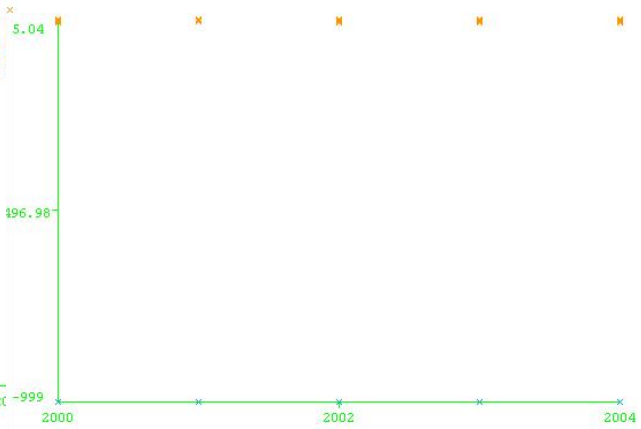

Fig 7. Graph for CO2 Flux

- Soil Moisture(40 cm) Analysis

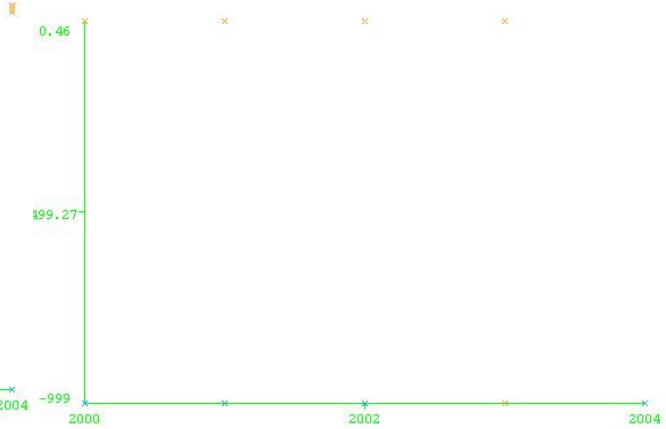

Fig 8. Graph for soil moisture at 10 cm

In the above graphs, blue denotes for small values and orange denotes higher values. Now, these graphs shows the variation in values of different attributes with respect to time and whenever we want to match these pattern according to time then these graphs are really helpful.

So, whenever we require pattern matching to analyse the environment activities at different time and in different weather conditions then we use the above graphs. For this we define the reference frame for normal pattern for different attributes then discover reference frame that results the patterns with multiple irregularities and finally if there is change in data values then automatically adjust for normal patterns and discover irregularities. For example in winters, from each sensor node we calculate the minimum and maximum temperature and plot the graphs to analyse the data. If there is fire in forest then to detect this we introduce the concept of pattern matching strategy and match the graphs of different nodes. This results the graph which shows irregular points and different from the graphs of other node then from this we quickly find the fire area.

IV. DISCUSSION

As future work, we will consider more about fault-tolerance, power consumption, scalability, inter sensor node communication and most important is security. We use the sensor network and data mining concept in other application areas like detecting tsunami, detecting heart attack etc. We also try to introduce the concepts of machine learning for further analysis and try to find out more interesting and correct results.

V. CONCLUSION

As, the sensor network is used at large scale in various applications. These sensor networks produce large datasets, which will require sophisticated data mining techniques and also, modeling to enable optimal decision-making. In this paper, we identify the major challenges for mining data from large sensor data. Here, various useful techniques of data mining are discussed which we are applied to wireless sensor network data for monitoring the environmental activities. The main area which we focus is accuracy and enhancement in the performance of getting the correct and fruitful results. We identify some of the research problems to work on sensor data irregularities detection, sensor data clustering and pattern matching. The work described in this paper link up the development of new knowledge discovery and decision making research designed to evaluate and optimize sensor network.

ACKNOWLEDGMENT

We worked on open source data mining tool for our analysis and this can be downloaded from site: http://www.cs.waikato.ac.nz/~ml/weka/. And the dataset repository site from where we download the dataset of sensor network for our work http://daac.ornl.gov/LBA/guides/CD04_Meteorology_Fluxes.html .

REFERENCES

[1] C.C.Aggarwal, J.Han, J.Wang, P.S.Yu. A Framework for Clustering Evolving Data Streams. VLDB2003. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] I.F.Akyildiz, W.Su, Y.Sankarasubramaniam, and E.Cayirci. Wireless Sensor Networks: A Survey. Computer Networks, Vol.38, No.4, pp.393-422, 2002.

[3] S.Bandyopadhyay and E.J.Coyle. An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks. IEEE INFOCOM 2003.

[4] D.Barbara, W.Dumouchel, C.Faoutsos and P.Haas. The New Jersey Data Reduction Re-port. IEEE Data Engineering Bulletin, Vol.20, No.4, pp.3-45, 1997.

[5] D.Estrin, R.Govindan, J.Heidemann, and S.Kumar. Next Century Challenges: Scalable Coordination in Sensor Networks. MobiCOM 1999

[6] G.J.Pottie and W.J.Kaiser. Wireless Integrated Network Sensors. Communications of the ACM. Vol.43, No.5, pp.51-58, May 2000.

[7] Agrawal, R.; Imielinski, T.; Swami, A. *IEEE\z* Transactionon Knowledge and Data Engineering; 1993; p 6.

[8] Han, J.; Kamber, M. Data Mining: Concepts and Techniques; Morgan Kaufmann: 2000.

[9] Dongqing Yang, Shiwei Tang, Qiong Luo ,Xiuli Ma, Dehui Zhang, Shuangfeng Li. Online Mining in Sensor Networks. NPC 2004, LNCS 3222, pp. 544-550, 2004.

[10] Agrawal, R.; Srikant, R. Mining Sequential Patterns.*ICDE* **1995**, 3-14.

[11] Madden, M.J.Franklin, J.M.Hellerstein, and W.Hong. TAG: A Tiny Aggregation Service for Ad-hoc Sensor Networks. In Symposium on OSDI, 2002.

[12] A. Lazarevic and V. Kumar. Feature bagging for outlier detection. In KDD '05, New York, NY, USA, 2005.I.F.Akyildiz,W.Su, Y.Sankarasubramaniam, and E.Cayirci. Wireless Sensor Networks: A Survey. Computer Networks, Vol.38, No.4, pp.393-422, 2002.

[13] S. R. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. TinyDB: an acquisitional query processing system for sensor networks. ACM Trans. Database Syst., 30(1), 2005.

[14] E.M.Knorr and R.T.Ng. Algorithms for Mining Distance-based Outliers in Large Datasets. VLDB 1998.

[15] S. R. K. Joel B. Predd, and H. Vincent Poor, "Distributed Learning in Wireless Sensor Networks - application issues and the problem of distributed inference," *IEEE Signal Processing Magazine*, 2006.

[16] S. Madden, M.J.Franklin, J.M.Hellerstein, and W.Hong. The Design of an Acquisitional Query Processor for Sensor Networks. SIGMOD 2003.

[17] M.J.Franklin. Challenges in Ubiquitous Data Management. Informatics 2001.