

Vector based adaptive sampling in wireless sensor networks

Aravind Mohanoor, Sridhar Radhakrishnan

School of Computer Science, University of Oklahoma
Norman, OK 73019
{aravindmc, sridhar}@ou.edu

Thomas Hughes

School of Industrial Engineering, University of Oklahoma
Norman, OK 73019
tehughes84@ou.edu

Abstract—While in-network aggregation and query processing are the common forms of collecting data in a typical wireless sensor network, continuous monitoring allows for more complex querying of the data collected. The process of collecting this data subject to constraints on the network resources is termed adaptive sampling. In this paper we present a simple quantitative scheme for adaptive sampling by using a vector of suitably defined parameters. The adaptive sampling vector (ASV) specifies the baseline of the data which needs to be collected to meet information quality needs.

Keywords- wireless sensor network, adaptive sampling

I. INTRODUCTION

Data collection in sensor networks could be broadly classified into the following three categories: in-network aggregation, query processing and continuous monitoring [1]. While in-network aggregation and query processing are used in a typical wireless sensor network, continuous monitoring provides a much more complex and richer model for querying the sensor data [2]. Such a data collection method is well suited for ambient sensor networks [3, 4] which can help people monitor their personal living space and make decisions about their daily activities. However, the resource constraints in such battery operated networks require a careful adaptive sampling approach so that the network's resources are utilized properly.

In their paper describing a vision for the world wide sensor web, Balazinska et al. [5] discuss the problem of data ingest in sensor networks. The process of data ingest is of the steps required to “calibrate, gap-fill and regrid” necessary to perform more complex data analysis on the collected data. In this paper, we present a vector based adaptive sampling approach for sensor networks which can be very useful for the data ingest process described above. In this sense, the adaptive sampling vector is essentially a statistics-friendly, resource aware data collection tool.

II. ADAPTIVE SAMPLING VECTOR

Suppose we are interested in continuously monitoring a sensor network with N sensor nodes for a duration of T time slots. The entire operational lifetime of the sensor network is expected to be a fairly large multiple of T . During each time slot, each of the N sensor nodes may report their data to a designated sink node S . The duration of a single time slot is

assumed wide enough to allow each sensor node to report its data to S .

The adaptive sampling vector (ASV) is a vector of length 5 of the form $[u \ v \ w \ x \ y]$. All the parameters in the ASV are non-negative integers. The quantity u represents the number of time slots which we wish to sample. The quantity v represents the number of sensors in the network. Normally, u and v take the values T and N , respectively. As one would expect, in a sensor network with node failures, the value of u can be shortened to periodically reassess the duration of the data collection cycle and the value of v can be adjusted according to the number of sensors which are still operational. Incorporating this crucial feedback when performing multiple rounds of data collection makes our approach highly *adaptive*. The quantity w represents the maximum consecutive time slots in which a sensor may not report its data to sink S . If w is assigned a value of 1, then each sensor is expected to report its data at least every other time slot. The quantity x represents the minimum number of time slots that each sensor is expected to report its data within the T time slots and the quantity y represents the minimum number of sensors which are expected to report their data during each time slot. Clearly $x \leq u$ and $y \leq v$.

A. Consecutive missing data

Preventing consecutive missing data helps avoid ‘blocking’ of data – blocking here refers to too many consecutive readings. Errors from blocking of data can misrepresent the central tendency of a particular population. Columns 3, 4 and 5 in Table 1 are polynomial regression approximations obtained from the following data sets: (a) the complete data set with all 12 months of temperature data (b) the summer data set (6 months, Mar – August) and (c) the data set with alternate months (Jan, Mar, May, July, Sep, Nov). Both (b) and (c) contain temperature values for 6 out of the 12 months, but (c) is obviously better distributed, as we show. We can see that the trends predicted by taking every other reading are close to both the actual value and the trends predicted by taking the full data set. The summer data regression does a good job approximating the temperatures during the summer months, but if the data is extrapolated beyond those months the regression quickly fails. Looking at the temperature profile in December as an example, the actual temperature was 53 degrees. The regression value predicted by the full data set was 53.6 degrees and the regression value predicted by the alternating month data set was 53.4 degrees. Comparing these values to the 84 degree value predicted by the summer months’ regression

analysis makes the effect of extrapolating values beyond the data set obvious. Thus minimizing blocking of data into clusters helps reduce extrapolation error and provides a better representative sample of data.

TABLE I. TEMPERATURE READINGS AND REGRESSION APPROXIMATIONS FROM NORMAN, OKLAHOMA

Month	Actual Temp	2nd Order Poly. Regression Approx.		
		w/ Full Data Set	w/ Summer Data Set	w/ alt. month data set
Jan	50	43.51	40.17	46.25
Feb	56	58.16	53.79	60.07
Mar	66	70.07	65.49	71.26
Apr	75	79.23	75.26	79.82
May	82	85.64	83.11	85.74
Jun	89	89.30	89.02	89.03
Jul	95	90.22	93.01	89.68
Aug	94	88.40	95.06	87.70
Sep	86	83.82	95.19	83.09
Oct	76	76.50	93.39	75.84
Nov	63	66.43	89.66	65.96
Dec	53	53.62	84.01	53.44

III. DATA COLLECTION PATTERN

A data collection pattern P is a matrix representation of the data collected. The rows of P represent the sensor IDs and the columns of P represent the time slots. Each cell in the matrix has a binary value 0 or 1. If cell (r, c) has value 1, then the sensor ID corresponding to row r reports its data during the time slot corresponding to column c . If the value of (r, c) is 0, then no data is reported to the sink node. The matrix P with all cells having a value of 1 is called the identity data collection matrix I , and the matrix P with all cells having a value of 0 is called the null data collection matrix represented by ϕ .

IV. FEASIBLE PATTERNS

Given an adaptive sampling vector V_1 , a data collection pattern P_1 is said to be feasible for V_1 if it satisfies all the parameters in vector V_1 . While parameters u and v represent the size of the network and the duration of the sampling, which are usually fixed parameters, we are more interested in P_1 satisfying the parameters w , x and y in V_1 . Figure 1 shows an example of a data collection pattern matrix for a 5 sensor network over 6 time slots. This matrix is feasible for the vector $[6 \ 5 \ 1 \ 3 \ 2]$.

		Time slot					
		1	2	3	4	5	6
Sensor ID	21	1	0	1	0	1	0
	22	0	1	0	1	0	1
	23	1	0	1	1	0	1
	24	1	0	1	0	1	0
	25	0	1	0	1	0	1

Figure 1. Feasible data collection matrix for the adaptive sampling vector $[6 \ 5 \ 1 \ 3 \ 2]$

Lemma 1: The identity data collection matrix I is feasible for all vectors V (assuming N and T are satisfied).

Proof: Follows from definition of the ASV. ■

V. RESOURCE CONSTRAINTS

The essential reason for using the adaptive sampling vector is to ensure that the resources in a sensor network are well utilized. As shown in Lemma 1, any ASV is satisfied by I , which would deplete the energy of the network very quickly. The problem becomes more challenging as additional constraints are introduced. The objective is to define ASVs in such a way that we get good temporal and spatial resolutions of the data collected while finding ways to minimize the resource utilization of the data collection process. The temporal resolution is ensured by the quantities w and x , while y can help ensure good spatial resolution. Although y may not directly ensure good “coverage” of the sensor network, we could add additional parameters to our ASV to suit the requirements.

Problem MINSAMPLES: Given an ASV V_0 , we wish to find data collection pattern P_0 satisfying V_0 which minimizes the value of $\sum P(r, c)$ where $P(r, c) = 0$ or 1 is the value of the r^{th} row and c^{th} column of P_0 . (i.e. the fewest number of data samples which need to be reported to satisfy the ASV).

Theorem 1: MINSAMPLES is NP-Complete.

Proof: Omitted due to space constraints. ■

VI. RELATED WORK

The best performing adaptive sampling scheme in the literature is presented in [2], and its authors describe the major steps involved in the algorithm as follows: (i) sensing driven cluster-construction, (ii) correlation-based sampler selection and model derivation and (iii) adaptive data collection with model based prediction. While ASV performs adaptive data collection similar to step (iii), we avoid the energy expenditure involved in steps (i) and (ii). Besides, we do not require a priori assumptions about the existence of a correlation in the sampler which is selected. Hence our approach is more energy efficient and also tolerates varying degrees of correlation among the data while still permitting the larger goal of high quality data ingest.

REFERENCES

- [1] Chen D. and Varshney, P.K., “QoS support in wireless sensor networks: A survey,” *Proc. Intl. Conf. on Wireless Networks (ICWN 04)*, pp. 227-233, 2004.
- [2] Gedik, B., Ling Liu, Yu, P.S., “ASAP: An Adaptive Sampling Approach to Data Collection in Sensor Networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 18, no. 12, pp. 1766-1783, Dec. 2007.
- [3] Boekhorst, F., “Ambient intelligence, the next paradigm for consumer electronics: How will it affect silicon?,” *IEEE Intl. Solid State Circuits Conference (ISSCC)*, pp. 28-31, 2002.
- [4] Weber, W., “Ambient intelligence: industrial research on a visionary concept,” *Proc. of Intl. Symp. on Low Power Electronics and Design*, pp. 247-251, 2003.
- [5] Balazinska, M., et al. “Data Management in the Worldwide Sensor Web,” *IEEE Pervasive Computing*, vol. 6, no. 2, pp. 30-40, Apr-Jun, 2007.