Managing energy and data quality in large sensor swarms

Jinseok Yang, Sameer Tilak, Tajana Simunic Rosing

UCSD

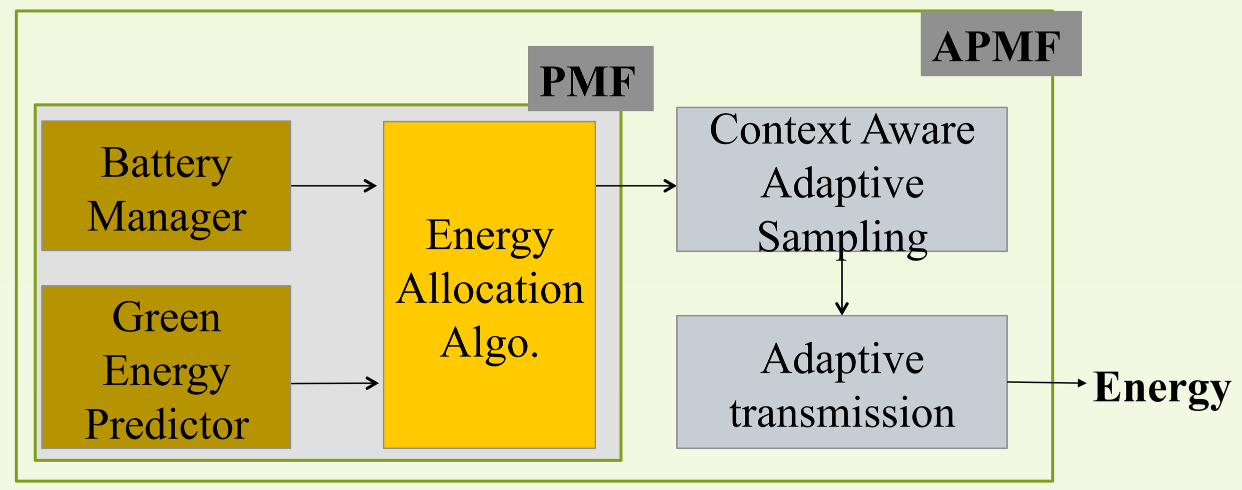
Background & Objectives

- *Modern applications of WSNs measure several variables (such as temperature, humidity, etc.) for extended periods of time over a large area.*
- *To meet the application requirements, the design and deployment of a WSNs has to carefully balance between two competing goals: (1) high spatio-temporal resolution to ensure the accuracy of the collected data (2) minimum energy consumption to maximize the network lifetime and limit node maintenance*

Challenges

- ***Reducing total number of samples collected by heterogeneous sensors****: To maintain an acceptable amount of total measurements, the missing data is inferred with statistical models that capture how the data evolves. In addition to enhancing the battery life, these approaches can also estimate any lost or corrupted data.*
- ***Integrated energy efficiency****: Low powered embedded sensor nodes consist of multiple heterogeneous components such as battery, CPU, and transceiver. Each component has its own power management technique, leading to suboptimal energy management. We need to define a method to combine these components into a unified whole.*
- ***Scalability****: Distributed and low complexity energy management techniques are essential for thousands of nodes in swarms.*

System Architecture



System adapts sampling and transmission rates as a function of the battery level, harvested energy availability and application context.

Methods

1. Estimating missing data:

- *We develop a latent factorization model for multivariate spatio-temporal data by applying tensor decomposition.*
- *Extract K dimension unobserved vectors (called latent factors) from training set*
- *Given T different time instances (a_t), N different nodes (c_n), and S different sensor types (b_s), sensor reading x can be estimated as:*

$$x_{tsn} = \sum_{k=1}^{K} a_{tk} b_{sk} c_{nk} + \varepsilon, \ \varepsilon \sim N(0, \sigma^2)$$

2. Power management framework:

- *Adaptive sampling module calculates the sampling rate by capturing variations in the phenomena.*
- *Transmission policy module estimates an optimal transmission instance by calculating reward in terms of energy gain and freshness loss of buffered data:*

$$\text{reward} = gain(\text{delay}) \times loss(\text{delay})$$

Results

Figure 1. Comparison of the temperature signal collected from the CIMIS network (top) and the 3ENCULT network (bottom)

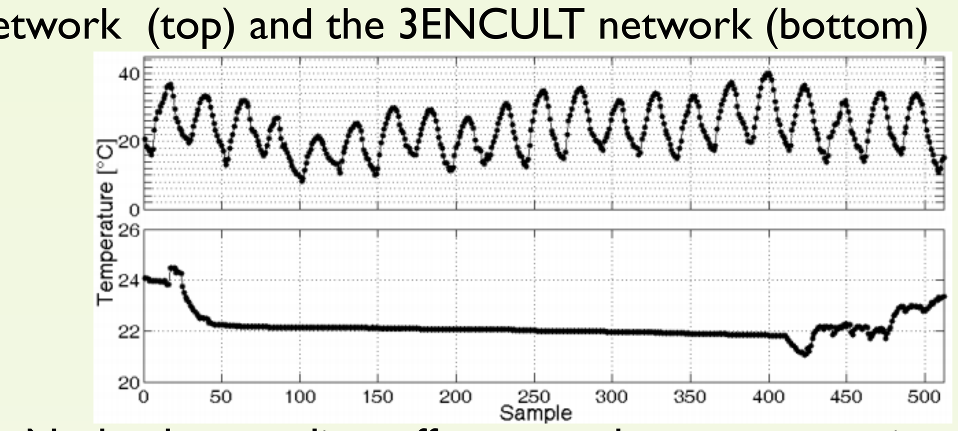


Figure 2. Node duty cycling effects on the reconstruction error as computed from the standard tensor factorization approach (CP) and our technique (MAP): (a) CIMIS dataset and (b) 3ENCULT.

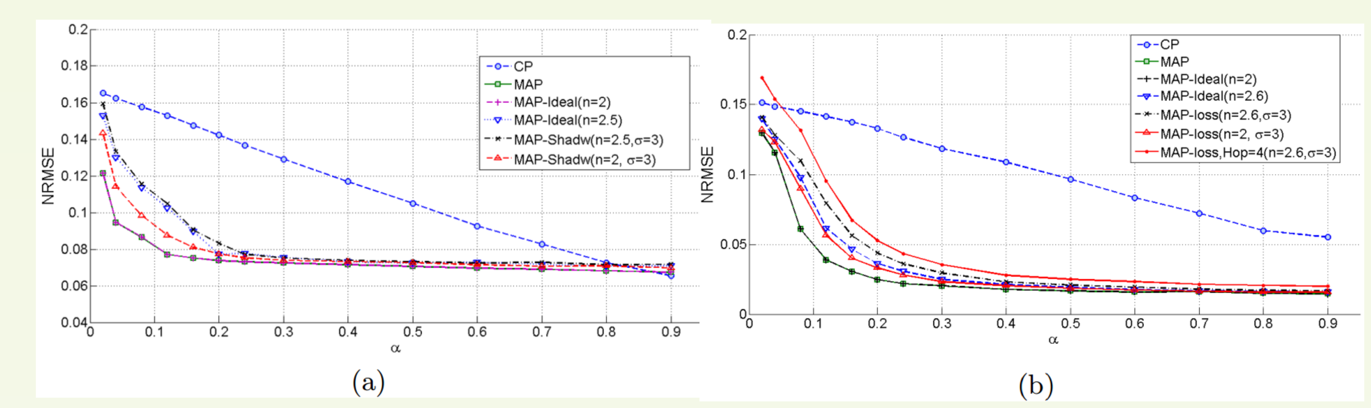


Figure 3. Comparison of our technique (MAP) with a distributed compressive sampling (DCS) technique for the different variables in the 3encult dataset: (a) temperature, (b) humidity and (c) light.

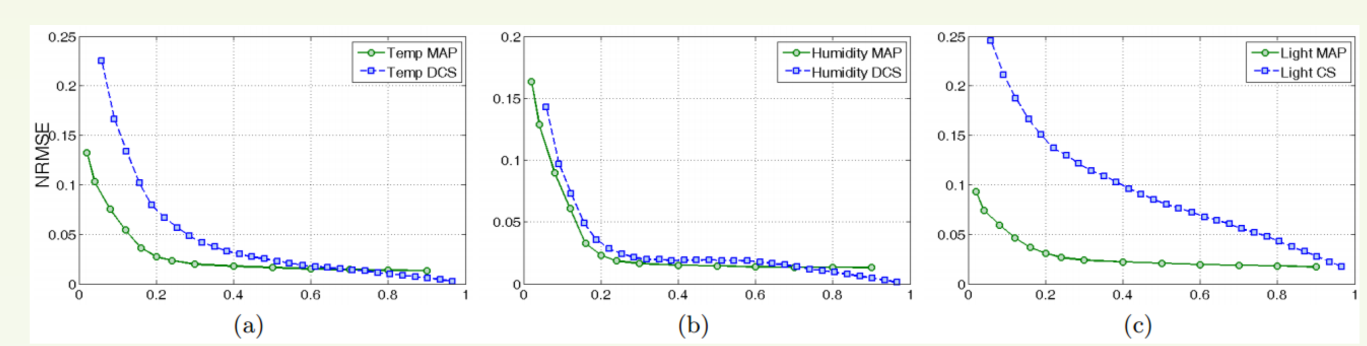


Table 1. Relative energy in comparison with PMF using buoy data

vs. PMF	Humidity	Temperature
APMF	27%	32%
APMF w/o TPM	62%	72%
APMF w/o ASM (5min)	48%	53%
APMF w/o ASM (10min)	46%	42%

Table 2. Estimation error in comparison with PMF with buoy data

vs. PMF	Humidity	Temperature
APMF	0.015	0.005
APMF w/o TPM	0.012	0.003
APMF w/o ASM (5min)	0.009	0.003
APMF w/o ASM (10min)	0.013	0.005

Table 3. Relative energy consumption in comparison with PMF with north temperate lake data

	Energy consmp. (%)
APMF	62%
APMF w/o TPM	88%
APMF w/o ASM (5min)	74%
APMF w/o ASM (10min)	69%

Real-world data sets

We use real-world data from multiple different types of sensor networks:

1. Estimating missing data

- *CIMIS: data set from California Irrigation Management Information Systems(i.e. air/soil temperature, humidity, pressure, precipitation)*
- *3EN-CULT: structural health monitoring data (i.e. air temperature, humidity)*

2. APMF

- *Multiple types of data sets gathered with a buoy deployed on a lake*
- *North temperate lake ecological study*

Conclusion & Future work

The key advantage of using a latent variable model is that it provides a compact representation of the gathered data that can be used to recover the missing samples. APMF provides applications a finer control over delay-energy tradeoff.

The next step in our studies is to develop a technique to decide how often latent variables need to be recomputed online to adapt to ever-changing environmental conditions. We will apply our strategy to a large, swarm size deployment, that covers most of San Diego County. The data from the deployment has 1000s of sensors over an area size of 100x100 sq miles, and has been collected over the period of last 10 years [3].

References

[1] B. Milosevic et. al, "Efficient Energy Management and Data Recovery in Sensor Networks using Latent Variables Based Tensor Factorization" IEEE/ ACM MsWiM 2013

[2] JS.Yang et. al, "Leveraging application context and detailed battery model for efficient sensing", submitted to DATE 2014

[3] hpwren.ucsd.edu

NSF TerraSwarm