Managing energy and data quality in large sensor swarms

Jinseok Yang, Sameer Tilak, Tajana Simunic Rosing

UCSD

Modern applications typically require measuring several variables for extended periods of time over a large area. To meet the application requirements, the design and deployment has to carefully balance two competing goals: (1) high spatio-temporal resolution to ensure the accuracy of the collected data, and (2) minimal energy consumption to maximize the network lifetime and limit node maintenance. The amount of data that each node collects and processes directly affects both its power consumption and the accuracy of the information obtained.

Extending the uptime of a swarm node - especially when they are deployed in difficult to access locations – is an active topic of research. Common approaches try to enhance the battery life directly by harvesting energy from the environment and employing low-power hardware, or using improved wireless protocols and distributed computation for data processing. More recently, researchers are optimizing the battery life indirectly by reducing the overall amount of sensed data. Here, the data is selectively sampled according to a predetermined protocol, reducing the total amount of samples collected by the individual sensor nodes, thus minimizing the energy consumption. To maintain an acceptable amount of total measurements, the missing data is inferred according to statistical models that capture how the data evolves. In addition to enhancing the battery life, these approaches are also able to estimate any lost or corrupted data, making them a popular choice.

We recently proposed an energy efficient, data driven technique to estimate missing data within a heterogeneous sensor network [1]. The latent variable factorization model which typically considers only dyadic interactions in data has been extended to multivariate spatio-temporal data, by applying tensor decomposition techniques. 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. In order to perform well under extreme sampling conditions, we explicitly incorporate the spatial, temporal, and inter-sensor correlations. The study focuses on the trade-off between the accuracy in recovering the missing data and the energy consumption when sensor nodes duty cycle to save energy. The proposed technique drastically reduces the amount of sampled data at each node, thus allowing the nodes to spend more time in a low-power sleep mode and save energy. The lower amount of sampled data implies a lower amount of data to transmit from the node to a central gathering station, reducing also the power consumptions associated with the radio communications. Our experiments with the OMNeT++ network simulator using realistic wireless channel conditions, on data collected from two real-world sensor networks, show that we can sample just 20% of the data and can reconstruct the remaining 80% of the data with less than 9% mean error, outperforming similar state-of-the-art techniques such is distributed compressive sampling. In addition, energy savings ranging up to 76%, depending on the sampling rate and the hardware configuration of the node. 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, and to then 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 [2].

[1] B. Milosevic, JS. Yang, N. Verma, S. Tilak, P. Zappi, E. Farella, L. Benini, and T. S. Rosing, "Efficient Energy Management and Data Recovery in Sensor Networks using Latent Variables Based Tensor Factorization" to appear in IEEE/ACM MsWiM 2013.

[2] hpwren.ucsd.edu