Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 109C (2017) 210–216

Procedia Computer Science

www.elsevier.com/locate/procedia

The 8th International Conference on Ambient Systems, Networks and Technologies (ANT 2017)

Sparse Data Recovery using Optimized Orthogonal Matching Pursuit for WSNs

Vishal Krishna Singh[a,*], Ankur Kumar Rai[a], Manish Kumar[a]

[a]Indian Institute of Information Technology, 211012, Allahabad, India

Abstract

Compressed Sensing based recovery algorithms are bounded by limitations such as recovery error and high computational requirements. This work presents a novel algorithm based on Orthogonal Matching Pursuit algorithm for efficient and fast data recovery in wireless sensor networks. The proposed algorithm significantly reduces the number of iterations required to recover the original data in a relatively small interval of time. Simulations show that the proposed algorithm converges in a short interval of time with significantly better results as compared to the existing data recovery approaches in wireless sensor networks.

Keywords: Orthogonal Matching Pursuit; Sparse data recovery; Compressed Sensing; Wireless Sensor Networks.

1. Introduction

Compressed sensing (CS) framework is aimed at precisely obtaining a high dimensional signal x such that $x \in \mathbb{R}^{m\times n}$ from a relatively small sample set of linearly combined observations y such that $y \in \mathbb{R}^m$, given the linear measurements are the projections of the original signal which is sparse in some intitutively known domain. CS based compression is driven by the idea of obtaining sparse signals at a rate notably lesser than the Nyquist rate. The CS principle utilizes the sparse nature of the wireless sensor network (WSN) data for efficient compression and recovery, However, in practical scenarios the signal in question is not sparse itself but in most of the cases has a sparse representation in some $a \times b$ directory 'D' such that $(a < b)$. Given the inherent energy and bandwidth constraints of WSNs, CS based data gathering has gained the desired attention in recent years. The central idea is to compress the data at the deployed nodes before it is transmitted to the sink where the original data is recovered with almost no loss. Thus, the major energy consuming task is shifted from the nodes to the sink. However the data reconstruction problem is bounded by mainly two issues:

- The required number of measurements for reconstructing the signal.

* Corresponding author. Tel.: +91-955-962-6633
E-mail address: vashukrishna@gmail.com

10.1016/j.procs.2017.05.328

- The choice of the best suited algorithm for effective and accurate data recovery.

Accurate and fast recovery of the original data at the sink is one of the major issues for most of the data recovery algorithms[1]. Existing approaches tend to have almost full recovery but it is at the cost of added computational complexity. Hence, an accptable tradeoff is seen between fast recovery of the original data and complexity of the recovery algorithms[2,3]. The Orthogonal Matching Pursuit (OMP) [4] algorithm is one of the most promising work in this field. The basic idea behind OMP is trivial and spontaneous; i.e. to choose a column of measurement matrix Φ in every iteration whose correlation is maximum with the residual. The key for the selected column of the measurement matrix, is appended to the list. Also, the remnant of the selected columns is removed from the Φ, updating a residual for the further iterations. The steps are repeated until N columns are chosen from the measurement matrix (Φ). The approximation of 'x' is computed by the residual update by finding a solution of least squares (LS), which costs around 4Md flops and 2Md cost for updating the residual matrix. Considering that the OMP algorithm needs total M iterations, the total cost is around $2Mdn + 3M^2d$ where, M is the number of iterations, d is the number of measurements and n is total number of nodes. Thus, the sparsity 'S' has a major role in optimizing the algorithm's computational complexity. Recently, the TOMP [5], modified the OMP algorithm by introducing the threshold of recovery error, calculated by requirement i.e. how much fraction of original data needs to be recovered. Simultaneous Orthogonal Matching Pursuit algorithm (SOMP) [6] was used for inter-signal compressed sensing for clustering techniques like LEACH [11], HEED [12], and DEEC [13] etc. Another recovery algorithm is the Compressive Sampling Matching Pursuit (CoSaMP) [7] which is well matched for noisy measurements. The algorithm's output spectrum is sparse unlike a standard linear program solver techniques. A different approach of CS based schemes is the use of signal's support in the compressed observations which effectively reduces the number of samples required for accurate recovery. With the aim of reducing the samples required for reconstruction, recent works [8,9] proved that partially known supports can be used to accurately recover sparse signals with significantly reduced number of samples.

The rest of the paper is organised as follows, the section 2 presents the problem definition. In section 3, the preliminaries for the proposed scheme are discussed in detail followed by section 4 where the proposed scheme is explained in detail. The section 5 presents the results along with a detailed explanation. Finally in section 6, the conclusion and future work are presented.

2. Problem Definition

Accurate and fast recovery of the original data at the sink is one of the major issues for most of the data recovery algorithms in WSNs. Existing approaches tend to have almost full recovery but is at the cost of added computational complexity. In one iteration, a typical recovery algorithm selects a column from the received data matrix and compares it with the sparse representation of the original data. Therefore, minimizing the number of iterations for data recovery, can enhance the performance of algorithms significantly. However, efficient recovery of the compressed data, in WSNs, is bounded by three prominent issues, specifically:

- Computational complexity
- Recovery time and
- Accuracy.

An acceptable trade-off is seen between the recovery time and complexity of the recovery algorithms. This work aims at the fast and accurate recovery of sparse data without imposing any extra computational overhead.

3. Preliminaries for the Proposed Algorithm

The preliminaries for the proposed algorithm are:

- *Recovery Condition*
  The recovery condition is the number of non-zero coefficient of original signal (x) i.e. sparsity level (M). In order to identify the N correlated columns of the dictionary, N number of keys are chosen by finding the largest

correlation between residual of y and measurement matrix Φ. Such that, $y = \Phi x$ is the data received at the sink for reconstruction.

- *Column Appending*
  Selected columns are appended to the list to form a new dictionary. 'N' columns are added to the directory until the halting criteria is met.
- *Signal Estimation*
  The signal is estimated with the help of the appended dictionary and by solving the least square (LS) problem. LS is used find a unique solution as shown in equation given below:

$$x^k = argminx||y - \Phi_{\wedge^k} x||_2 \tag{1}$$

- *Residual Update* After each iteration the residual is updated to get the next maximum correlated 'N' columns as shown in equation given below:

$$r^k = y - \Phi_{\wedge^k} x^k \tag{2}$$

- *Halting Criteria*
  Compare the reconstruction error with threshold, if reconstruction error is less than threshold or number of selected columns are equal to 'M', then no further iterations are required and we jump to the output. The halting criteria is given by:

$$If \;\; ||r^k||_2 < \delta \; else \; N \times k \geq M \tag{3}$$

4. Proposed Approach

Based on the well known OMP algorithm, we propose a relatively fast and accurate version of the OMP and is termed as Optimized OMP (OOMP) here onwards. Multiple indices are chosen in each iteration for reducing the complexity and the execution time of the algorithm. The comparison of the reconstructed result with the original data is necessary for every iteration in the process of reconstruction of data. If $||r^k||_2$, then the reconstruction can be stopped and we can jump to output. That means, if the demand is met, the iteration can be halted earlier, rather than after M iterations. To reduce the complexity and speeding up the execution time, multiple indices are chosen in each iteration. The selection of multiple indices results in multiple keys to be added to list. Thus, the requirement of selection of M indices, is achieved with small number of iteration as compared to OMP. A reconstruction error threshold is incorporated, which is the necessary and sufficient condition for the convergence of algorithm. Reconstruction error is determined on the basis of the support of the similarity degree of original and recovered signal, i.e.

$$Reconstructed\ Error(\gamma) = \frac{||X - X^*||_2}{||X||_2} \tag{4}$$

Where the original signal is denoted by 'X' and the recovered signal by X^*. Thus, the reconstruction error is inversely proportional to the accuracy of the recovered data. The steps of the algorithm are as follows:

- Multiple columns 'N' of the measurement matrix 'Φ' are chosen which is maximally correlated to residual (identification).
- The key of the selected columns of the measurement matrix is appended to the list (augmentation).
- The remnant of selected columns is removed from the Φ as shown in 1.
- Updating a residual for the next iteration (residual update) as shown in 2.
- We check stopping criteria i.e. if reconstruction error is less than threshold or number of selected column is equal to M then jump to the output otherwise jump to step 1 as shown in 3.

The fig. 1 shows the uncompressed signal representation of the temperature data. The fig. 2 and fig. 3 shows the recovered signal using the proposed OOMP algorithm in noise and noise free environment respectively with sparsity value 10.

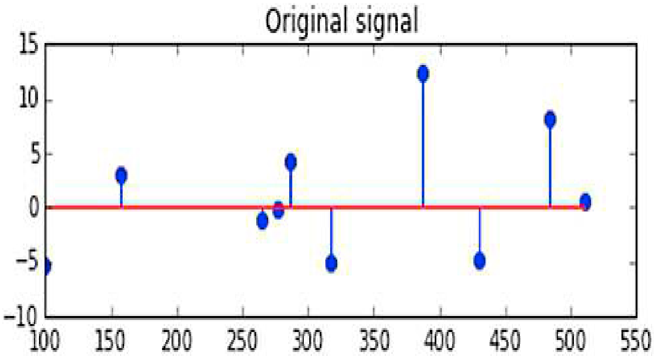

Fig. 1. Original Signal Representation of Temperature Data

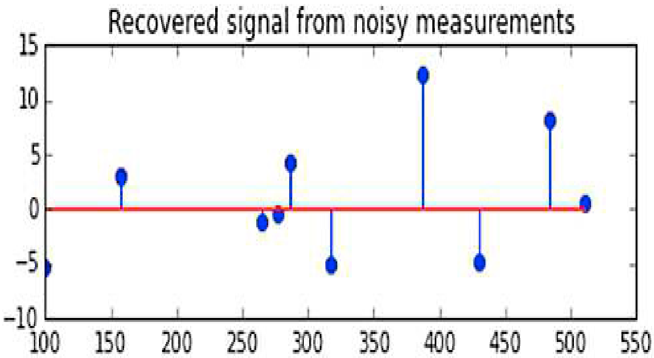

Fig. 2. Recovered Signal with Noise

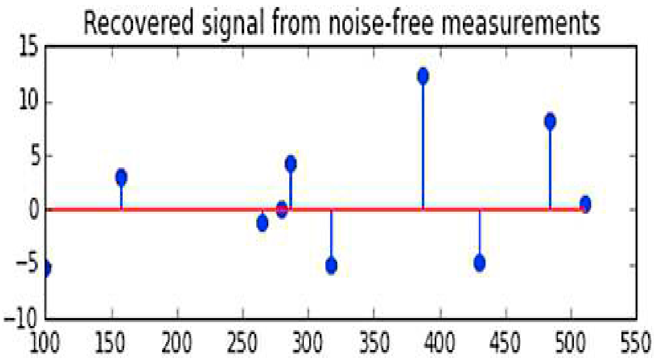

Fig. 3. Recovered Signal without Noise

5. Simulation and Results

5.1. Simulation Environment

Python was used for simulating the WSN environment with 1024 nodes randomly deployed in an area of $200 \times 200\ m^2$, The measurement matrix Φ is drawn by generating $n \times d$ Gaussian random variables. The simulations were performed for assessing the performance of the proposed OOMP with maximum number of selected columns in one

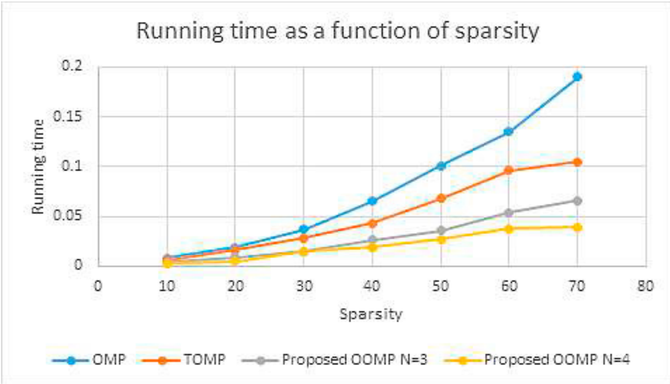

Fig. 4. Running Time as a Function of Sparsity

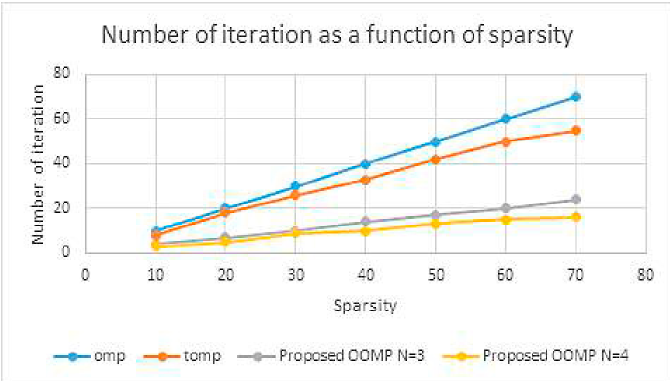

Fig. 5. Number of Iteration as a Function of Sparsity

iteration, $N = 3$ and 4. A comparative analysis of the proposed OOMP with the existing OMP[4] and TOMP[5] is presented in this section. The details of simulation parameters are presented in Table 1.

Table 1. Simulation Parameters

Parameter	Values
Area	$200 \times 200\ m^2$
Number of Nodes (n)	1024
Number of Measurement (d)	512
Number of Selected Columns (N)	3 and 4
Sparsity	10 – 70

5.2. Result and Analysis

Time Analysis: As shown in fig. 4 the recovery time of the proposed algorithm is minimum as compared to OMP[4] and TOMP[5]. This is due to fewer number of iterations for recovering the compressed data. The proposed algorithm outperforms the existing algorithm, with a fair margin, proving its efficiency in terms of recovery time.

Computational Complexity Analysis:Computational complexity for OOMP algorithm is around $2pdn + (2N^2 + N)p^2d)$ where 'p' denotes the number of total iterations required until the algorithm meets the halting criteria. As shown in fig. 5, the number of iterations of the proposed algorithm is minimum as compared to OMP and TOMP proving its efficiency over the two approaches.

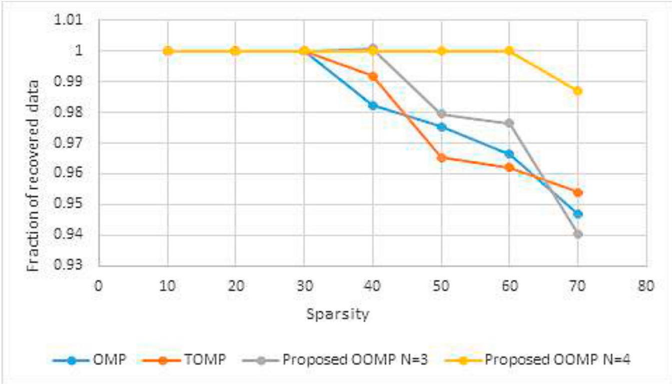

Fig. 6. Fraction of Recovered Data as a Function of Sparsity

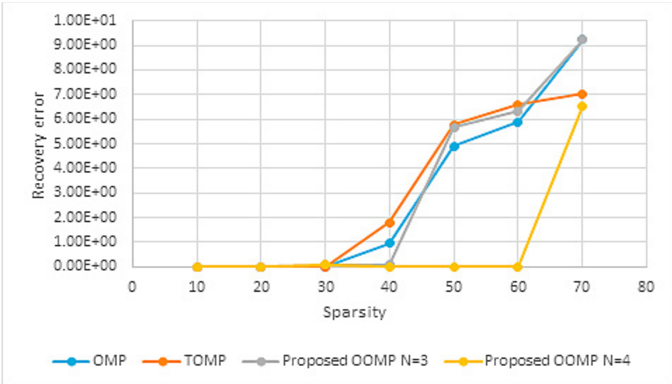

Fig. 7. Recovery Error as a Function of Sparsity

Reconstruction Analysis: The fig. 6 shows percentage recovery of the proposed algorithm as a function of sparsity. A comparative analysis of the percentage recovery of the proposed algorithm with existing approaches OMP and TOMP shows that critical sparsity for optimized OMP N=3 and N=4 is 40 and 60 respectively which is far better than OMP and TOMP. So higher critical sparsity inferred that, recovery performance of proposed OOMP is more efficient than existing OMP algorithms.

Error Analysis: Recovery error is determined by similarity degree of recovered and original signal i.e.

$$Recoveryerror = \| X - X^{*} \|_2 \tag{5}$$

Where X is original signal and X is recovered signal. As shown in fig. 7, the proposed algorithm outperforms existing recovery approaches based on CS with the error percentage as low as 0.01 for sparsity value of 40. Even in the worst case, the proposed algorithm has minimum error percentage of 0.02 for sparsity value of 70 as compared to the existing OMP which has the error percentage of 0.06 for the same value of sparsity. The proposed OOMP is able to give accurate results with large scale WSNs. Interestingly, the the error value decreases marginally with increase in the data correlation. Since, the data correlation is very weak in small sized WSN, the recovery accuracy is compromised as the size of the network is reduced (for detailed analysis of the effect of network size on the recovery accuracy, see our previous work on CS [10]).

6. Conclusion

This work presents an Optimized OMP algorithm for efficient and fast data recovery in WSNs. Selecting correlated columns helps to reduce the number of iterations required to recover the original data in a relatively small interval of time. Simulation results show that the proposed algorithm has better accuracy, computational complexity and fast recovery as compared to the existing recovery algorithm, specifically OMP and TOMP. The proposed OOMP converges quickly without affecting the quality of the recovered signal.

Future Work: We continue to extend our work in the direction of exploring recovery algorithms for small sized WSNs where the traditional recovery algorithms fail to give accurate results. Targeting the matrix completion problem, we aim to devise a reconstruction algorithm for such small scale WSNs where algorithms such as the proposed OOMP fail to achieve accurate results.

References

1. Singh, Vishal Krishna, Gajendra Sharma, Manish Kumar. Compressed sensing based acoustic event detection in protected area networks with wireless multimedia sensors. *Multimedia Tools and Applications* 2016; **1-25**.
2. Wu, Rui, Wei Huang, Di-Rong Chen. The exact support recovery of sparse signals with noise via orthogonal matching pursuit. *IEEE signal processing letters* 2013;**20(4)**:403-406.
3. Donoho, David L., Yaakov Tsaig, Iddo Drori, Jean-Luc Starck. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Transactions on information theory* 2012;**58(2)**:1094-1121.
4. Tropp, J.A. and Gilbert, A.C.. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on information theory*,**53(12)**:4655-66,2007.
5. Ji S, Huang L, Wang J, Shen J, Kim JU. An Improved Reconstruction methods of Compressive Sensing Data Recovery in Wireless Sensor Networks. *International Journal of Security and Its Applications* 2014;**8(1)**:1-8.
6. He B, Li Y, Huang H, Tang H. Spatialtemporal compression and recovery in a wireless sensor network in an underground tunnel environment. *Knowledge and Information Systems*, **41(2),449-65, 2014.**
7. Needell D, Tropp JA. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis* ;**26(3):301-21, 2009.**
8. Vaswani N, Lu W. Modified-CS: Modifying compressive sensing for problems with partially known support. *IEEE Transactions on Signal Processing.*, **58(9),4595-607, 2010.**
9. W. Lu and N. Vaswani: Exact reconstruction conditions and error bounds for regularized Modified Basis Pursuit (Reg-modified-BP). *Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, **763-767, 2010.**
10. Singh, Vishal Krishna, Manish Kumar. Hierarchical compressed sensing for cluster based wireless sensor networks. *International Journal of Advance Computer Science Applications,* **1(7),58-67, 2016.**
11. Tyagi, S. and Kumar, N., A systematic review on clustering and routing techniques based upon LEACH protocol for wireless sensor networks. *Journal of Network and Computer Applications*, **36(2),623-645, 2013.**
12. Chand, S., Singh, S. and Kumar, B., Heterogeneous HEED protocol for wireless sensor networks. *Wireless personal communications*, **77(3), 2117-2139, 2014.**
13. Sharma, A. and Goyal, V., Performance Analysis During Transmission In WSN with Distributed Energy Efficient Clustering by Modified Threshold Protocol. *International Journal of Electronics Communication and Computer Engineering,* **8(1), 1. 2017.**