OVERCOMPLETE REPRESENTATION OF SENSOR NETWORK DATA BY DISTRIBUTED SYSTEM MODELING

Károly MOLNÁR and Gábor PÉCELI
Department of Measurement and Information Systems
Budapest University of Technology and Economics, and
Embedded Information Technology Research Group
Hungarian Academy of Sciences
Budapest, Hungary,
{karoly.molnar, gabor.peceli}@mit.bme.hu

Abstract: This paper presents a technique for representing distributed data of sensor networks. The approach is based on a general distributed regression framework that models sensor data by fitting a global function to each of the local measurements. The presented method explores the possible extensions of this distributed regression, by using more complex signal representations. In order to reduce the amount of processed data and the required communication, the signal model has to be as compact as possible, with the ability to restore the arbitrary measurements. To achieve this, the regression is followed by a data compression step, where the basis function set is changed to an overcomplete set, as these have special advantages in cases of nonstationary signal modeling than complete base representations.

Key words: sensor networks, distributed systems, parameter learning, overcomplete signal representation, adaptive signal processing

1 Introduction

Recently, the use of large scale sensor networks consisting of numerous independent nodes is more and more widespread. These systems are mainly used for environmental monitoring, but many other promising uses are feasible. An example system is presented in [1].

Distributed sensor networks process large amounts of data, therefore the compression of the signals are a key issue, as the cost of communication is relatively high in terms of battery energy. This is particularly important in case of wireless sensor networks. A possible power-aware solution is the use of appropriate signal preprocessing and compression in the signal data representation.

The proposed approach is a novel method of representing sensor network data, based on the distributed regression framework presented in [2]. The method is fundamentally a two-step data reduction process, where the data model calculated by the distributed regression is further compressed by using an overcomplete basis.

The overcomplete basis representation is a technique where the signal is decomposed using more signal components than it is necessary. The basis generally consists of functions that form one or more orthogonal bases and some extra functions. As the overcomplete basis contains more basis functions than the minimum number to represent the signal, the signal representation is not unique. The overcomplete signal representation is a well explored area, numerous bases are used and several algorithms are developed [3] [4]. The goal of these signal decomposition techniques is to overcome the difficulties of the most common orthogonal basis expansions (e.g. Fourier and wavelet bases) for modeling arbitrary signals. However, most algorithms based on overcomplete basis representation are not suitable for sensor network applications due to the excessive amount of required computation power. In [5], the authors present a more realizable technique that uses a weakly overcomplete basis, where a complete orthogonal basis is merged with one extra basis function. For this basis, the optimal solution in the L_1 norm sense can be calculated without vast amount of computation, making the algorithm realizable in distributed systems.

In Section 2 the distributed regression is outlined briefly. Based on this framework, in Section 3, the implementation of the overcomplete basis approach is shown. Simulation results are presented in Section 4.

2 Distributed regression in sensor networks

The distributed regression proposed in [2] is a framework for in-network modeling sensor network data. The nodes collaborate to optimally fit a global function to each of their local measurements. In the framework, a set of basis functions h_i is given and their parameters w_i are continuously fitted to the measured data. This modeling process is achieved using regression, where $f(t)$ is the measured signal.

$$\hat{f}(t) = \sum_i w_i h_i(t) \approx f(t)$$

Least squares (LS) optimization can be done by linear regression:

$$\mathbf{w} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{f} = \mathbf{A}^{-1}\mathbf{b}$$

where **H** is the basis matrix with one column for each basis function and one row for each measurement, and **f** is the measurement vector with one element for each measurement. The measurements are functions of time and spatial location, $f(x, y, t)$, therefore the basis functions are also a function of time and space. However, for sake of simplicity, the space-dependent and the time-dependent functions are separated, and the space-dependent ones are linear:

$$\sum_i h_i(x, y, t) = x + y + \sum_{i=1}^{N} h_i(t)$$

The linear regression is made computationally feasible and distributed by the use of kernel regression, where the model takes the form of a weighted sum of local basis functions, making the **A** matrix sparse. The algorithm solves the global linear system **Aw** = **b** by a distributed application of Gaussian elimination.

The complete algorithm is realized by sending regression messages between neighboring nodes. The message routing is not trivial, as before computing the local linear system, all the other information of the neighboring nodes has to be known. This forms a nonserial dynamic programming problem that can be solved by several algorithms [6].

3 Overcomplete basis in the distributed regression framework

In case of overcomplete basis representation, the signals are decomposed into a number of optimal basis components that are found from an overcomplete basis dictionary by some optimization method. As the computation cost of such an algorithm has to be low, a relatively simple method needs to be chosen. A feasible choice is the use of weakly overcomplete representation [5] based on the concept of the basis pursuit method [4].

The overcomplete base consists of a complete base, $[h_1, h_1, \dots h_N]$ and one additional function (h_0), that can be expressed by the same base: $h_0 = [g_1, g_2, \dots g_N]$. The input of the method is time-windowed data, that is transformed to the base $[h_1, h_2, \dots h_N]$ as a first step. If the extra basis function is taken into consideration by an unknown weight factor c, than the input signal can be represented as: $[ch_0, h_1 - cg_1, h_2 - cg_2, \dots h_N - cg_N]$. Since it is known that the L_1 norm minimum ensures the minimum number of nonzero coefficients, at least one of the above coefficients has to be zero, which means that the value of the c factor is $\{0, h_1/g_1, h_2/g_2, \dots, h_N/g_N\}$. Substituting c with these values, the one which has the minimum norm provides the most compact representation. This results in a complete, (not necessarily orthogonal) basis representation.

The method is implemented in the distributed regression framework in two steps. The first step is to calculate the coefficients $w_1, w_2, \dots, w_{N+2}$ of the basis functions $h_1, h_2, \dots, h_{N+2}$ by the Gaussian elimination method. There are $N+2$ basis functions, $h_1, h_2, \dots, h_N$ being time-dependent, while $h_{N+1} = x$ and $h_{N+2} = y$ are not time-dependent. The functions $h_1, h_2, \dots, h_N$ form an orthogonal base, e.g. Gabor functions (Gaussian windowed sinusoids) or wavelets.

The second step is to calculate the overcomplete solution based by the above presented method. This is done by adding the extra basis function h_0 to the $h_1, h_2, \dots, h_N$ orthogonal base, and recalculate the coefficients $w_0, w_1, \dots, w_N$.

The important choice of the extra basis function strongly depends on the application of the sensor network. In general, a priori information should be used for the choice, e.g. the signal can be derived from the expected ideal signal of the monitored phenomena. In this case, the overcomplete representation can be interpreted as a separation of the measured signal to ideal signal and noise, where the noise (or the difference from the ideal) is represented by the orthogonal basis.

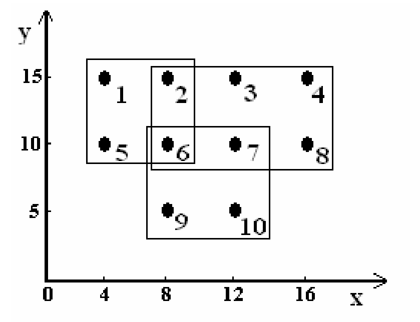

Fig. 1. Simulated sensor field

If there is no a-priori information available on the signals, then the extra basis function is suggested to be an element of another basis, or it can be derived from previous measurements of the same node.

The overcomplete signal representation can be used for signal compression if some of the calculated coefficients of the basis functions are less significant than others. In this case, the less significant coefficients are considered to be zeros, and the signal is represented only by the significant basis functions. The compression is ideal if there is only one significant coefficient (i.e. the measured signal is identical to one of the basis functions).

4 Simulation results

The above presented solution is realized in a Matlab environment, where a sensor-network system is simulated. The field layout is shown in Fig. 1. The simulated field consists of three overlapping kernels with 10 sensor nodes. The chosen time-dependent basis consists of an orthogonal base of N discrete cosine functions, where N being the length of the time-window used for the computation, and one extra function.

(Note: in this case, the linear regression step of the calculation is equivalent to the Discrete Cosine Transform of the measured signal.)

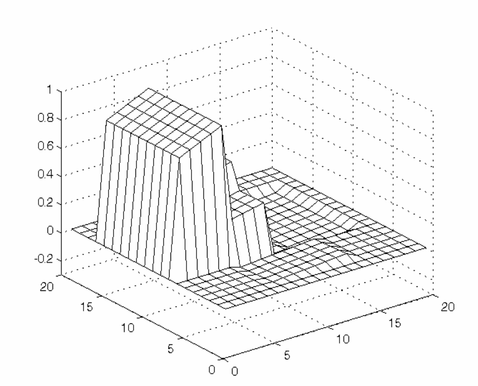

Fig.2. The reconstructed signal $f(x, y, t_1)$

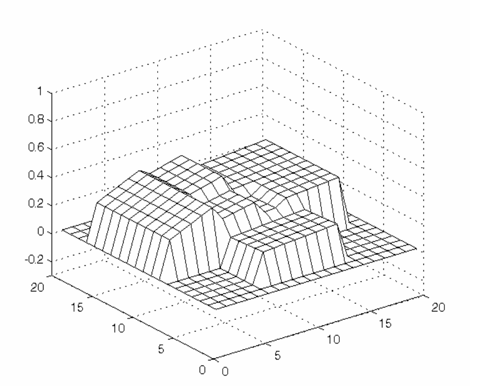

Fig.3. The reconstructed signal $f(x, y, t_2)$

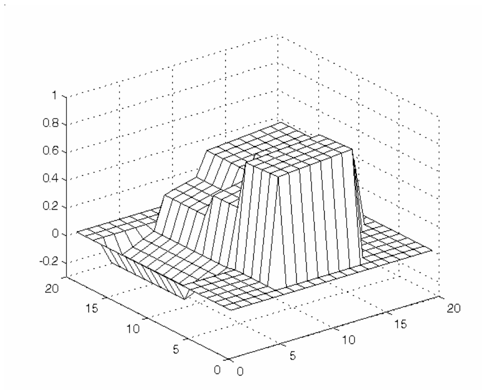

Fig.4. The reconstructed signal $f(x, y, t_3)$

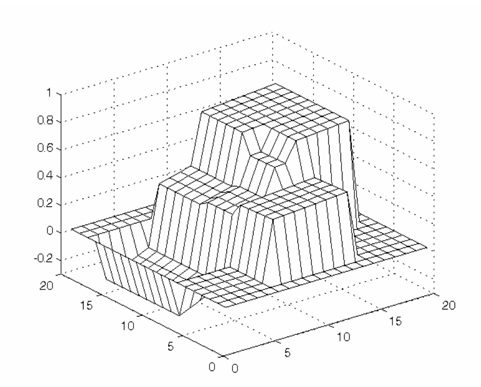

Fig.5. The reconstructed signal $f(x, y, t_4)$

The Dirac-impulse is chosen to be the extra function in the overcomplete base. This choice follows the a-priori assumption that the measured signal is impulse-like. The simulation presents an event when an impulse-shaped wavefront $f(x, y, t)$ is traveling over the sensor field along the x axis:

$$f(x, y, t) = \delta(x - x_0 + ct) + N$$

where c is the speed of the wavefront and N is white Gaussian noise. The impulse reaches nodes 1, 5 at time instant t_1, nodes 2, 6, 9 at t_2, nodes 3, 7, 10 at t_3, and finally nodes 4, 8 at

t_4. This measurement is a simplified simulation of an acoustic impulse (e.g. a blast) traveling above a network of acoustic sensor nodes. In Fig. 2 – 5. the calculated $f(x, y, t)$ is shown for t_1, t_2, t_3, t_4 respectively.

The compression ratio is close to ideal in this case, as the measured time-dependent signal is similar to the extra basis function.

5 Conclusions

In this paper, an expansion of the distributed regression framework is presented. The solution uses a weakly overcomplete basis to represent the signals measured by the sensors. This signal representation is inherently a compressed form of the measurements, making the solution appropriate for power-aware sensor networks.

The required amount of communication is fairly reduced by this solution, but the individual nodes still has to perform a fairly high amount of computation (e.g. matrix inversion calculations for the regression). Possible future directions of research are the use of recursive structures and to evaluate a technique that uses overcomplete signal representation before the regression phase in order to further decrease the required computation and communication cost.

References

[1] SIMON, Gy. et al. Sensor Network-Based Countersniper System. In *SenSys 04,* Baltimore, November 2004.

[2] GUESTRIN, C. et al. Distributed Regression: an Efficient Framework for Modeling Sensor Network Data. In *Proceedings of the Third International Symposium on Information Processing in Sensor Networks 2004 (IPSN-04)*.

[3] MALLAT, S., ZHANG, Z. Matching pursuit in a time-frequency dictionary. In *IEEE Trans. on Signal Processing*, vol. 41., pp. 3397 – 3415., Dec 1993.

[4] CHEN, S. S., DONOHO, D. L., SAUNDERS, M. A.. Atomic Decomposition by Basis Pursuit. In *SIAM Journal on Scientific Computing*. vol. 20., Issue 1, Aug 1998.

[5] VÁRKONYI-KÓCZY, A. R., FÉK, M. Recursive Overcomplete Signal Representations. In *IEEE Trans. on Instrumentation and Measurement*, vol. 50., NO. 6, Dec 2001.

[6] PASKIN, M. A., LAWRENCE, G.D. Junction Tree Algorithms for Solving Sparse Linear Systems. *Technical Report UCB/CSD-03-1271*, University of California, Berkeley.