[International Journal of Computer and Communication Technology](https://www.interscience.in/ijcct)

Volume 2 | Issue 2 Article 10

April 2011

# Mining Wireless Sensor Network Data: an adaptive approach based on artificial neuralnetworks algorithm

Lambodar Jena
*Dept. of CSE,C.V.Raman College of Engineering, Bhubaneswar, India*, lmjena@yahoo.com

Ramakrushna Swain
*Department of CSE and IT, , C.V.Raman College of Engineering, Bhubaneswar, India;*, rkswain1@gmail.com

N.K. kamila
*C.V.Raman College of Engineering and Technology*, nkamila@yahoo.com

# Mining Wireless Sensor Network Data: an adaptive approach based on artificial neural-networks algorithm

Lambodar Jena[1], Ramakrushna Swain[2], Narendra K. Kamila[3]
[1,3] Dept. of CSE,C.V.Raman College of Engineering, Bhubaneswar, India
[2]Dept. of CSE and IT,C.V.Raman College of Engineering, Bhubaneswar, India
[1]lmjena@yahoo.com, [2]rkswain1@gmail.com, [3]nkamila@yahoo.com

**Abstract- This paper proposes a layered modular architecture to adaptively perform data mining tasks in large sensor networks. The architecture consists in a lower layer which performs data aggregation in a modular fashion and in an upper layer which employs an adaptive local learning technique to extract a prediction model from the aggregated information. The rationale of the approach is that a modular aggregation of sensor data can serve jointly two purposes: first, the organization of sensors in clusters, then reducing the communication effort, second, the dimensionality reduction of the data mining task, then improving the accuracy of the sensing task . Here we show that some of the algorithms developed within the artificial neural-networks tradition can be easily adopted to wireless sensor-network platforms and will meet several aspects of the constraints for data mining in sensor networks like: limited communication bandwidth, limited computing resources, limited power supply, and the need for fault-tolerance. The analysis of the dimensionality reduction obtained from the outputs of the neural-networks clustering algorithms shows that the communication costs of the proposed approach are significantly smaller, which is an important consideration in sensor-networks due to limited power supply. In this paper we will present two possible implementations of the ART and FuzzyART neural-networks algorithms, which are unsupervised learning methods for categorization of the sensory inputs. They are tested on a data obtained from a set of several nodes, equipped with several sensors each.**

***Keywords : Sensor Networks, Data Mining,Artificial neural network.***

## I. INTRODUCTION

Intelligent sensor networks has numerous applications for distributed information gathering and processing, monitoring, supervision of hazardous environments, intrusion detection, cooperative sensing, tracking. The increasing use of sensing units asks for the development of specific data mining architectures. There is a need for minimization of the communication and computational effort demanded to each single sensor unit. The simplest approach to the analysis of sensor network data makes use of a centralized architecture where a central server maintains a database of readings from all the sensors. The whole analysis effort is localized in the server, whose mission is to extract from the flow of data the high-level information expected to be returned by the monitoring system. If we assume that reasonable-size sensor networks will be made of thousands of nodes, the limitation of this approach is strikingly evident: the number of messages sent in the system as well as the number of variables of the data mining task are too large to be managed efficiently[7]. A centralized data clustering in wireless sensor networks is difficult and often not scalable because of various reasons such as limited communication bandwidth and limited power supply for running the sensor nodes. It is also inefficient given that sensor data has significant redundancy both in time and in space. In cases when the application demands compressed summaries of large spatio-temporal sensor data and similarity queries, such as detecting correlations and finding similar patterns, the use of a neural-network algorithm is a reasonable choice. The development of the wireless sensor networks is accompanied by several algorithms for data processing which are modified regression techniques from the field of multidimensional data series analysis in other scientific fields, with examples like nearest neighbor search, principal component analysis and multidimensional scaling (e.g. [11], [13]).

At the same time techniques of data compression, like Principal Component analysis (PCA) or Independent Component Analysis (ICA) [8], are often used in data mining to reduce the complexity of modeling tasks with a very large number of variables. It is well known in the data mining literature that methods for reducing complexity are beneficial for several reasons: improvement of the accuracy and intelligibility of the model, reduced storage and time requirements.A modular organization of the sensor network can be used to jointly address the two main issues in mining sensor network data : the minimization of the communication effort and the accurate extraction of high-level information from massive and streaming datasets.

It is proposed[7] a data driven procedure to configure a two-layer topology of a sensor network (Figure 1) made of

1. a lower level whose task is to organize the sensors in clusters, compress their signals and transmit the aggregate information to the upper level,

2. an upper level playing the role of a data mining server which uses the aggregate information to carry out the required sensing task.

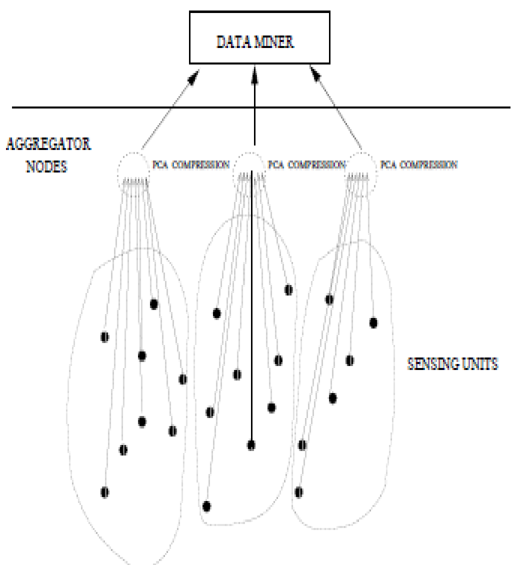


Figure 1: The black dots represent the sensing units. The dotted circles represent the aggregator nodes which carry out the fusion of data coming from neighboring sensors before sending the aggregated signals up to the data mining server.

Auto-classification of the sensor readings is important in sensor networks since the data obtained with them is with high dimensionality and very immense, which could easily overwhelm the processing and storage capacity of a centralized database system[7]. On the other hand, the data obtained by the sensor networks are often self-correlated over time, over space and over different sensor inputs, due to the nature of the phenomena being sensed which is often slowly changing, due to the redundant sensor nodes dispersed near each other, and due to the fact that often the sensor readings are correlated over different modalities sensed at one node (e.g. sound and light from cars in traffic control application).

Neural-networks algorithms, on the other hand, use simple computations and do not represent big burden to memory. The proposed adaptations of the ART neural networks models can be easily parameterized according to user needs for greater or lower level of details of the sensor data. Unsupervised learning Artificial Neural Networks typically perform dimensionality reduction or pattern clustering. They are able to discover both regularities and irregularities in the redundant input data by iterative process of adjusting weights of interconnections between a large numbers of simple computational units (called artificial neurons). As a result of the dimensionality reduction obtained easily from the outputs of these algorithms, lower communication costs and thus bigger energy savings can also be obtained.

A neural network algorithm can be implemented in the tiny platform of Smart-It units, which are kind of sensor nodes or motes. Thus instead of reporting the raw-data, each Smart-It unit can send only the cluster number where the current sensory input pattern has been classified. In that way a huge dimensionality reduction can be achieved depending on the number of sensor inputs in each unit. In the same time communication savings will benefit from the fact that the cluster number is a small binary number unlike sensory readings which can be several bytes long real numbers converted from the analog inputs. Since the communication is the biggest consumer of the energy in the units, this leads to bigger energy savings as well.

## II. Related work

According to [10] Kohonen Self Organizing Maps (SOMs) model and ART models are similar in a way that they are both prototype-based networks where they both create a set of prototypes and then compare an unknown input vector with the stored prototypes in order to implement the mapping or clustering. The advantages of SOMs over other Artificial Neural Network models include the ability to provide real-time nearest-neighbor response as well as topology-preserving mapping of the input data. Still, the limitations are extensive off-line learning and most importantly, the need of a predefined map size, i.e. a fixed number of output clusters or categories. In many real-world situations, there is no a priori information on variability present in the data stream, so we can not determine in advance the required number of output clusters in which the input patterns will fit. Thus this straightforward implementation of the Kohonen neural network seems rather rudimentary and the only justification for it can be the mere possibility to apply some principles of Artificial Neural Networks for data processing in wireless sensor networks. DIMENSIONS [11] is another model where they treat the problems of data storage and handling in sensor systems. DIMENSIONS incorporates spatio-temporal data reduction to distributed storage architectures, introduces local cost functions to data compression techniques, and adds distributed decision making and communication cost to data mining paradigms. It provides unified view of data handling in sensor networks incorporating long-term and short-term storage with increasing lossy compression over time, multiresolution data access using different wavelet parameters at different hierarchical levels, and spatio-temporal pattern mining. Several aspects of our approach are common to DIMENSIONS such as spatio-temporal data reduction, limited communication costs, long-term and short-term storage and hierarchical access with different level of details, although these aspects are achieved by completely different algorithms.

Consider a sensor network S made of S sensors where P is a [S, 3] matrix containing the three-dimensional coordinates of the S sensors and

(2.1) $x(t) = \{s_1(t), s_2(t), \ldots, s_S(t)\}$

is the state (or snapshot) of the sensor network at time t. Suppose we intend to employ S to perform a supervised learning task, for example a regression problem

(2.2) $y(t) = f(x(t)) + \varepsilon(t)$

where y is the variable to be predicted at time t on the basis of the state x(t) of the network S and ε is usually thought as the term including modeling error, disturbances and noise.

If we have available a finite dataset $D_N = \{ [x(t_i), y(t_i)], i = 1, \ldots, N\}$ of N input-output observations, this problem can be tackled as a conventional regression problem, by first estimating an approximator of f on the basis of $D_N$ and then using this estimator as a predictor of y. However, if, like in the case of sensor networks, the number S is huge, the mapping f is non-stationary and the data are collected sequentially, conventional techniques reach rapidly their limits. In particular, the large dimensionality of the problem asks for feature selection problem as well as the streaming aspect of the problem requires sequential (also called recursive) estimation approaches.

This paper proposes an approach to the problem of data mining in sensor networks which tries to conciliate the needs for an accurate prediction of the output y with the constraints related to energy reserves, communication bandwidth and sensor computational power.

The following subsections will rapidly sketch the two computational modules used in our approach: the recursive PCA and the ART and FuzzyART. Both will describe how these modules are combined in our architecture for mining sensor networks.

## III. ART AND FUZZYART ALGORITHMS

There are numerous models of unsupervised Artificial Neural Networks have been proposed like Multi-layer Perceptron (MLP), Self-Organizing Maps (SOMs), and Adaptive Resonance Theory (ART) ([12] and [13]). Out of these we have chosen the ART models for implementation in the field of sensor networks because they do not constrain the number of different categories in which the input data will be clustered.Although the later extensions of MLP and SOMs involve the principle of incrementally growing structure, their topological self-organization is possible only with so called off-line learning cycle separate from the classification cycle. Having two separate cycles is inconvenient in the presence of potentially unlimited stream of input data with no reliable method of choosing the suitably representative subset for a learning cycle. ART algorithms offer another example of topological self-organization of data but they can adapt structure quickly in the fast-learning mode. Adaptive Resonance Theory (ART) has been developed by Grossberg and Carpenter for pattern recognition primarily. Models of unsupervised learning include ART1 [1] for binary input patterns and FuzzyART [2] for analog input patterns.

ART networks develop stable recognition codes by self organization in response to arbitrary sequences of input patterns. They were designed to solve the so called stability-plasticity dilemma: how to continue to learn from new events without forgetting previously learned information. ART networks model several features such as robustness to variations in intensity), detection of signals mixed with noise, and both short- and long-term memory to accommodate variable rates of change in the environment. There are several variations of ART-based networks: ART1 (three-layer network with binary inputs), Fuzzy ART (with analog inputs, representing neuro-fuzzy hybrids which inherit all key features of ART), their supervised versions ARTMAP and FuzzyARTMAP and many others. ARTMAP models [9], for example, combine two unsupervised modules to carry out supervised learning.

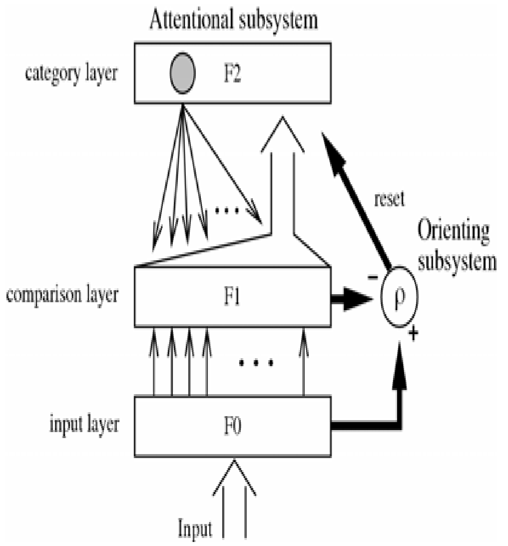


Figure 2: Architecture of the ART network.

In Figure 2 typical representation of an ART Artificial Neural Network is given. Winning F2 category nodes are selected by the attentional subsystem. Category search is controlled by the orienting subsystem. If the degree of category match at the F1 layer is lower than the so called vigilance level ρ, a reset signal will be triggered, which will deactivate the current winning F2 node for the period of presentation of the current input. An ART network is built up of three layers: the input layer (F0), the comparison layer (F1) and the recognition layer (F2) with N, N and M neurons, respectively. The input layer stores the input pattern, and each neuron in the input layer is connected to its corresponding node in the comparison layer via one-to-one, non-modifiable links.

Nodes in the F2 layer represent input categories. The F1 and F2 layers interact with each other through weighted bottom-up and top-down connections that are modified when the network learns. The learning process of the network can be described as follows: At each presentation of a non-zero binary input pattern x ( $x_j$ ε {0,1};j=1,2,., N), the network attempts to classify it into one of its existing categories based on its similarity to the stored prototype of each category node. More precisely, for each node i in the F2 layer, the bottom-up activation Ti is calculated, which can be expressed as

$$T_i = \frac{|W_i \cap X|}{\beta + |W_i|} \quad i = 1,\dots, M \tag{1}$$

Where | . | is the norm operator ( for a vector u it is : $|u| \equiv \Sigma_{j=1}^{N} u_j$ ), $w_i$ is the (binary) weight vector or prototype of category i, and β > 0 is a parameter. Then the F2 node I that has the highest bottom-up activation, i.e $T_1$ =max{ Ti| i = 1,…, M}, is selected ( realizing so called winner node ($W_I$) will then be compared to the current input at the comparison layer. If they are similar enough, i.e. if they satisfy the matching condition:

$$\frac{|W_I \cap X|}{|X|} \geq \rho \tag{2}$$

Where ρ is a system parameter called vigilance( 0 < ρ ≤ 1), then the F2 node I will capture the current input and the network learns by modifying Wi.

$$W_I^{new} = \eta(W_I^{old} \cap X) + (1 - \eta)W_I^{old} \tag{3}$$

Where η is the learning rate (0 < η ≤ 1) (the case when η=1 is called "fast learning"). All other weights in the network remain unchanged.

If, however, the stored prototype WI does not match the input sufficiently, i.e. if the condition (2) is not met, the winning F2 node will be reset (by activating the reset signal in figure 1) for the period of presentation of the current input. Then another F2 node (or category) is selected with the highest Ti, whose prototype will be matched against the input, and so on. This "hypothesis-testing" cycle is repeated until the network either finds a stored category whose prototype matches the input well enough, or allocates a new F2 node in which case learning takes place according to (3).

As a consequence of its stability-plasticity property, the network is capable of learning "on-line", i.e. refining its learned categories in response to a stream of new input patterns, as opposed to being trained "off-line" on a finite training set.

The number of developed categories can be controlled by setting the vigilance ρ: the higher the vigilance level, the larger number of more specific categories will be created. At its extreme, if ρ=1, the network will create a new category for every unique input pattern.

FuzzyART is an analog version of the ART1 algorithm, which takes analog inputs and classifies them in a similar way as ART1. The main ART1 operations of category choice (1), match (2), and learning (3) translate into Fuzzy ART operations by replacing the ordinary set theory intersection operator

∩ of ART1 with the fuzzy set theory conjunction MIN operator ∧.

In FuzzyART (but as well in ART1), complement coding of the input vector prevents a type of category proliferation that could otherwise occur when weights erode. Complement coding doubles the dimensionality of an input vector b≡ $(b_1,\ldots b_N)$ by concatenating the vector b with its complement $b^c$. The input to the FuzzyART network (F0 in Figure 2) is then a 2N-demensional vector: I=B≡(b, $b^c$), where $(b^c)_i \equiv (1-b_i)$. If b represents input features, then complement coding allows a learned category representation to encode the degree to which each feature is consistently absent from the input vector, as well as the degree to which it is consistently present in the input vector, when that category is active. Because of its computational advantages, complement coding is used in nearly all ART applications, and we have used it in our models as well.

The strengths of the ART models include its unique ability to solve a stability-plasticity dilemma, extremely short training times in the fast-learning mode, and an incrementally growing number of clusters based on the variations in the input data. The network runs entirely autonomously; it dose not need any outside control, it can learn and classify at the same time, provides fast access to match results, and is designed to work with infinite stream of data. All these features make it an excellent choice for application in wireless sensor networks.

## IV. PCA COMPRESSION TECHNIQUES

As discussed above, each data mining problem in the context of sensor network data with large S has to face the problem of reducing dimensionality. Existing techniques for feature selection can be grouped into two main approaches: the wrapper and the filter approach.

In the wrapper approach [5] the feature subset selection algorithm exists as a wrapper around the learning algorithm, which is often considered as a black box able to return (e.g. via cross-validation) an evaluation of the quality of a feature subset. On the contrary, the filter approach selects features using a preprocessing step independently of the learning algorithm.

In this paper we will adopt the Principal Component analysis (PCA) technique, an instance of the filter approach. PCA is a classic technique in statistical data analysis, feature extraction and data compression [8]. Given a set of multivariate measurements, its goal is to find a smaller set of variables with less redundancy, that would give as good a representation as possible. In PCA the redundancy is measured by computing linear correlations between variables. PCA entails transforming the n original variables $x_1, \ldots, x_n$ into m new variables $z_1, \ldots, z_m$ (called principal components) such that the new variables are uncorrelated with each other and account for decreasing portions of the variance of the original variables. Consider a vector x of size n and a matrix x containing n measures of the vector x. The m principal components

$$z_k = \sum_{j=1}^{n} W_{jk} X_j = W_k^T X, \quad k=1,\ldots, m \tag{4.1}$$

are defined as weighted sums of the elements of x with maximal variance, under the constraints that the weights are normalized and the principal components are uncorrelated with each other. It is well known from basic linear algebra that the solution to the PCA problem is given in terms of the unit-length eigenvectors $e_1, e_2, \ldots, e_n$ of the correlation matrix of x. Once ordered the eigenvectors such that the corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, the principal components $z_k$ is given by

$$z_k = e_k^T x.$$

It can be shown that the PCA problem can be also put in the form of a minimum mean-square error compression of x. This means that the computation of the wk for the orthonormal basis $w_1, \ldots, w_m$ that minimizes

$$J_{PCA} = \frac{1}{N}\sum_{t=1}^{N} \left\| x(t) - \sum_{k=1}^{m} \left(w_k^T x(t)\right) w_k \right\|^2 \tag{4.2}$$

If we denote W=(w1, …., wm )T where W is a matrix of size[m,n] we have

$$J_{PCA} = \frac{1}{N}\sum_{t=1}^{N} \left\| x(t) - W^T W_x(t) \right\|^2 \tag{4.3}$$

What is appealing in this formulation is that a recursive formulation of this least-squares problem is provided by the Projection Approximation Subspace Tracking (PAST) algorithm proposed by [6]. This algorithm, based on the recursive formulation of the least squares problem, has low computational cost and makes possible an updating of the principal components as new observations become available.

Once the matrix W is computed a reduction of the input dimensionally is obtained by transforming the input matrix X into the matrix $Z = XW^T$ and by transforming the regression problem of dimensionality n into a problem of dimensionality m in the space of principal components.

At this step the question arises of how to choose m. The techniques more commonly used rely either on the absolute values of the eigenvalues or on procedures of cross-validation [4].

## V. Discussion

In the proposed architecture the cluster heads collecting all sensor data from its cluster units. One cluster head at the upper level collects only clustering outputs from the other units which can be generalized to a hierarchical cascade classification scheme and transmitted to the data mining server which performs the regression modeling by using adaptive lazy learning algorithm.

The sensor units at the lowest level will be grouped in to small groups having one cluster head, a recursive PCA is performed on each cluster to return the principal components. Several cluster heads can be grouped and their outputs can be classified at a cluster head one level higher using ART1 classifier and so on.

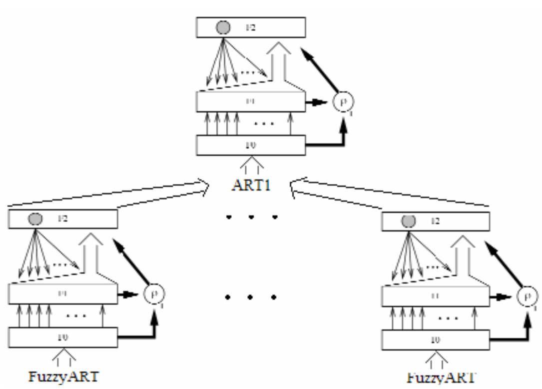


Figure 3: Clusterhead collecting and classifying the data after they are once classified at the lower level.

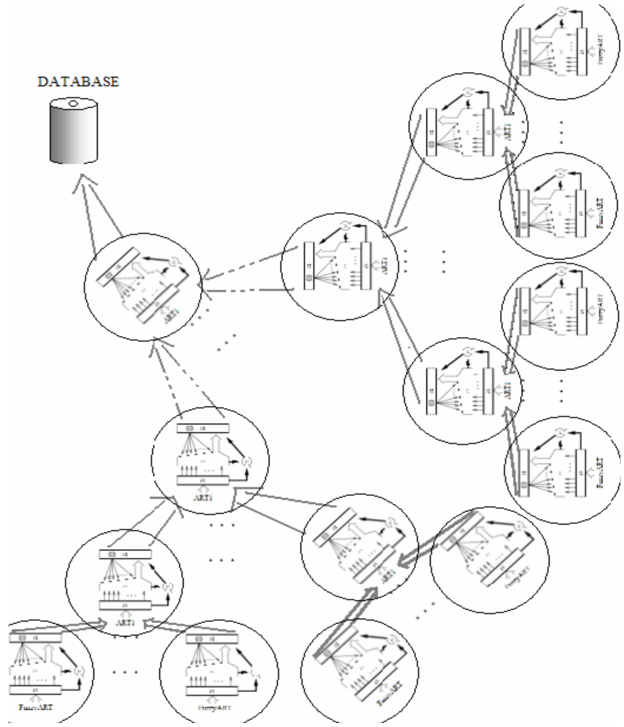


Figure 4: Hierarchical cascades of ART neural-network classifiers implemented in units of a sensor network.

## VI. Conclusion

This paper proposes possible adaptive methodology like ART model and PCA technique to mine data in large sensor networks. The structure of the processing architecture of a sensor network must be taken into account for data mining task. Data clustering algorithms for data spread over a sensor network are necessary in many applications based on sensor-networks. The use of limited resources together with the distributed nature of the sensor-networks demands a fundamentally distributed algorithmic solution for data clustering. According to the sensing task like classification or prediction the organization of the sensor network may change , thus the accuracy and quality of the data mining task must be taken in to account.

## VII. REFERENCES

[1] M. Birattari, G. Bontempi, and H. Bersini. Lazy learning meets the recursive least-squares algorithm. In M. S. Kearns, S. A. Solla, and D. A. Cohn, editors,NIPS 11, pages 375–381, Cambridge, 1999. MIT Press.

[2] C. M. Bishop. Neural Networks for Statistical Pattern Recognition. Oxford University Press, Oxford, UK, 1994.

[3] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. Journal of Machine Learning Research, 3:1157–1182, 2003.

[4] A. Hyvarinen, J. Karhunen, and E. Oja. Independent Component Analysis. Wiley, 2001.

[5] R. Kohavi and G. H. John. Wrappers for feature subset selection. Artificial Intelligence, 97(1-2):273–324, 1997.

[6] B. Yang. Projection approximation subspace tracking. IEEE Transactions on Signal Processing, 43(1):95–107,1995.

[7] Gianluca Bontempi, Yann-A¨el Le Borgne ., U An adaptive modular approach to the mining of sensor network data. LB Machine Learning Group Universit´e Libre de Bruxelles, Belgium

[8] Andrea Kulakov, Computer Science Department Faculty of Electrical Engineering Skopje Macedonia, Danco Davcev.,Data mining in wireless sensor networks based on artificial neural-networks algorithms. Computer Science Department Faculty of Electrical Engineering Skopje, Macedonia

[9] CARPENTER, G.A., GROSSBERG, S., MARKUZON, N., REYNOLDS, J.H., AND ROSEN, D.B., Fuzzy ARTMAP:A neural network architecture for incremental supervised learning of analog ultidimensional maps,IEEE Transactions on Neural Networks, vol. 3, pp. 698-713, 1992.

[10] CATTERALL, E., VAN LAERHOVEN, K., AND STROHBACH, M., Self-Organization in Ad Hoc Sensor Networks: An Empirical Study, in Proc. of Artificial Life VIII, The 8th Int. Conf. on the Simulation and Synthesis of Living Systems, Sydney, NSW, Australia,2002.

[11] GANESAN, D., ESTRIN, D., HEIDEMANN, J.,DIMENSIONS: Why do we need a new Data Handling architecture for Sensor Networks?, in Proc.Information Processing in Sensor Networks, Berkeley,California, USA, 2004.

[12] GROSSBERG, S. Adaptive Resonance Theory in Encyclopedia Of Cognitive Science, Macmillan Reference Ltd, 2000.

[13] GUESTRIN, C., BODIK, P., THIBAUX, R., PASKIN M.,AND MADDEN, S., Distributed Regression: an Efficient Framework for Modeling Sensor Network Data, in Proceedings of IPSN'04, April 26– 27, 2004, Berkeley,California, USA, 2004.