Poster Abstract: Selective Gossiping for SVM Training in Wireless Sensor Networks

Kallirroe Flouri
Department of Computer Science
University of Crete
FORTH-ICS,Greece
Email: flouri@ics.forth.gr

Baltasar Beferull-Lozano
Instituto de Robótica
Universidad de Valencia
ETSI, Spain
Email: baltasar.beferull@uv.es

Panagiotis Tsakalides
Department of Computer Science
University of Crete
FORTH-ICS,Greece
Email: tsakalid@ics.forth.gr

***Abstract*—The design of an integrated distributed wireless network comprised of thousands of microsensors that can perform tasks with low energy consumption poses some interesting signal processing challenges. In this paper, we propose two "selective gossip algorithms" taking advantage of the sparse representation that Support Vector Machines (SVMs) provide for the decision boundaries, to perform classification in wireless sensor networks (WSNs). Through analytical studies and simulation experiments, we show that the proposed energy efficient, distributed selective gossip algorithms train SVMs equally well with energy-consuming centralized approaches that need to transmit the whole data to a processing node.**

I. Introduction

As the research field of mobile computing and communication advances, so does the idea of a distributed, ad-hoc wireless network of hundreds to thousands of microsensors, which can be randomly deployed in an area of interest. On the other hand, the resource-constraint devices that comprise a sensor network usually operate in environments that are prone to link and node failure. Hence, it is important for sensor networks to be robust against changes in the topology. Motivated by applications to sensor networks, Gossip Algorithms have recently been studied for computation and information exchange in an arbitrarily connected network of nodes. The innovation in the iterative algorithms is that consensus on common global parameters is reached through a totally decentralized process, [1], [2], [3].

In this paper, addressing the transmission constraints in a WSN, we use the ideal characteristic of SVM-training (*i.e.*,that the data of a node can be compressed to their corresponding support vectors) in order to communicate minimal information to neighboring sensors. We also show that this is a sub-optimal strategy but with low energy cost. Investigating the geometrical aspect of the SVM, we propose an alternative algorithm that uses partial information but guarantees consensus to the optimal hyperplane that would have been constructed if all sensors had access to the entire information. Section II introduces the two Selective Gossip Algorithms we propose for training a SVM as applied to the classification problem in a WSN. Section III presents a set of simulation experiments in order to assess the performance of our proposed approaches.

II. Selective Gossip Algorithm for SVM Training

Let us consider a deployment of n sensors taking measurements in a certain area. Our goal is to be able to train a SVM in an efficient and distributed fashion so that: a) we can get good classification results on test data and b) all sensors keep refining their estimate concurrently in order to reach convergence to the global estimate at the same time.

In this work, we use Gossip Algorithms in the context of a SVM. There is a successive refinement to the estimate of each sensor based on communicating information with one-hop neighbors only. Therefore, at each time slot the new estimate is diffused to the next-hop neighbors and finally, after several time slots, all sensors reach consensus. Hence, it is important to exchange the minimum amount of data in order to ensure optimality. The more data are being exchanged, the more energy is consumed. The trade-off between optimality and energy consumption led our research to the following algorithms.

A. The Minimum Selective Gossip Algorithm (MSG-SVM). Communication links in the WSN comprised of n sensors are presented by a graph whose vertices are the sensors and whose edges are formed by the available communication links. The set of sensors having an active link with the i-th sensor comprise the neighborhood N_i. The WSN is deployed to train the SVM using the distributed measurements $M_i(0) := \{\mathbf{x_{i,j}(0)}\}_{\mathbf{i=1,\ldots,n}}$, where $j = 1, \ldots, k$ is the number of the measurements of sensor i.

We begin by taking k measurements at each node and then we train the SVM locally (for each sensor). The first estimate of the hyperplane is denoted by $\mathbf{w}_i(0)$, $i = 1, \ldots, n$, for each node. When training a SVM, only the support vectors determine the discriminant that separates the data collected by each sensor in two classes [4]. Therefore, the data of each node can be compressed to their corresponding estimated hyperplane and hence to the support vectors $SV_i(0) = \{\mathbf{x}_i(0) : \sum_i \alpha_i y_i \mathbf{x}_i(0) = \mathbf{w}_i(0),\ y_i = class\{1, -1\},\ \alpha_i \neq 0\}$. In general, it holds that $| SV_i(0) | << | M_i(0) |$, where $| SV_i(0) |$ and $| M_i(0) |$ denote the cardinality of $SV_i(0)$ and $M_i(0)$, respectively [5].

The proposed algorithm MSG-SVM is a gossip-based algorithm, hence the support vectors $SV_i(0)$ are communicated between one-hop neighbors. Therefore, for each node i, at time $t+1$, we update its estimate $w_i(t+1)$ by using all the information available at that moment, namely, the previously estimated set of support vectors $SV_i(t)$ at node i as well as the union of the sets of support vectors $SV_{N(i)}(t)$ that have been previously estimated by the neighbour nodes.

MSG-SVM provides a sub-optimal discriminant hyperplane while communicating minimum information. As we already mentioned, the data of a node can be compressed to their corresponding support vectors. But it cannot be guaranteed that a vector $x \in M_i(0)$ and $x \notin SV_i(0)$, is not a support vector in $\{M_{N_i}(0) \cup M_i(0)\},\ i \neq j$.

B. The Sufficient Selective Gossip Algorithm (SSG-SVM). We examine the convex hull of the training data of each class, and construct the plane that bisects the two closest points of the convex hulls, [4]. This is an alternative equivalent perspective for training a SVM. The closest points can be found by solving the following quadratic problem:

$$\min_{\alpha} \frac{1}{2} \parallel \mathbf{c} - \mathbf{d} \parallel^2 \tag{1}$$

$$\mathbf{c} = \sum_{y_i \in class\ 1} \alpha_i x_i, \quad \mathbf{d} = \sum_{y_i \in class\ -1} \alpha_i x_i,$$

subject to

$$\begin{aligned} \textstyle\sum_{y_i \in class\ 1} \alpha_i &= 1, \ \textstyle\sum_{y_i \in class\ -1} \alpha_i = 1 \\ \alpha_i &\geq 0 \ \text{ for } i = 1, 2, ...n. \end{aligned}$$

SSG-SVM takes advantage of the geometrical property of the SVM discriminant hyperplane. The sufficient data for the hyperplane construction are the vectors that lie on the boundary of the convex hulls of the two classes. For each node, the SSG-SVM discards all the vectors of the WSN nodes, except those located at the boundary of the convex hulls. Thus, neighboring sensors exchange the sufficient data only. After some communication, all WSN nodes have the information to construct a plane identical to the plane that would have been constructed if all sensors had access to the entire information.

Both algorithms are energy efficient, since data need not be transmitted to a fusion center. Instead, WSN nodes diffuse partial information to neighboring sensors. Furthermore, each node communicates the data that had not been sent in previous iterations, thus the energy spent for transmission is reduced.

III. Results and Discussion

In this Section, we evaluate the performance of the two proposed distributed algorithms in terms of the average classification error rate and we compare them to the ideal case where WSN nodes have access to the entire information. We consider a sensor network composed of $n = 15$ nodes distributed in a grid topology, where each of the sensors collects $\mid M_i(0) \mid = 20,\ i = 1, \ldots, 15$ sample vectors from two classes. In our experiments, we generate the sample data of the two classes using two Gaussian distributions with two different means. Each node in the network is connected with two one-hop neighbors.

We simulated 100 Monte Carlo runs in order to test the performance of the two proposed Selective Gossip Algorithms. Figure 1 represents the average classification error rates (%) for a randomly chosen sensor, as a function of the iteration steps. After only a few iterations, both algorithms result in trained SVM classifiers which exhibit similar performance to an SVM trained using the entire data from all sensors. SSG-SVM (dashed line) gives an optimal estimate of the discriminant after at most 8 iterations using only partial data. On the other hand, even though MSG-SVM (dash-dotted line) is a sub-optimal solution, it gives a good approximation of the optimal plane. Most importantly, with both introduced distributed schemes, all n sensors reach an agreement on the near optimal discriminant function.

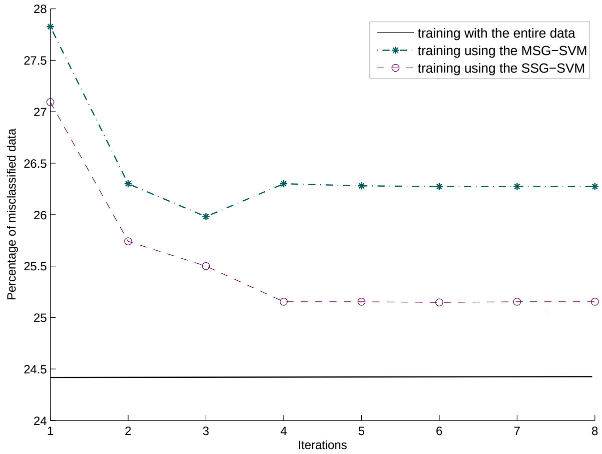

Fig. 1. The performance of the proposed algorithms.

IV. Conclusion

In this paper, we employed the concept of gossip algorithms for training a SVM in a Wireless Sensor Network. We introduced two distributed algorithms for training a SVM based on successive refinement of local estimates. In both cases, information is communicated to one-hop neighbors in order to update the estimate at each iteration. The sub-optimal algorithm MSG-SVM, uses only the support vectors of each node to reach an agreement. The SSG-SVM on the other hand, communicates larger amount of data, *i.e.* vectors lying on the convex hull boundaries, but converges to the optimal solution in a few iterations.

Acknowledgment

This work was funded by the Greek General Secretariat for Research and Technology under Program ΠΕΝΕΔ-Code 03ΕΔ69 and by the Marie Curie TOK-DEV "ASPIRE" grant (MTKD-CT-2005-029791) within the 6th European Community Framework Program.

References

[1] A. G. S. Boyd, B. Prabhakar, and D. Shah, "Gossip Algorithms: Design, Analysis and Applications," in *INFOCOM, 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, Miami, USA, March 2005, pp. 1653–1664.

[2] ——, "Randomized Gossip Algorithm," *IEEE/ACM Transactions on Networking*, vol. 52,6, pp. 2508–2530, 2006.

[3] S. B. L. Xiao and S. Lall, "A Scheme for Robust Distributed Sensor Fusion Based on Average Consensus," in *Proc. of 4th International Conference on Information Processing in Sensor Networks, IPSN'05*, Los Angeles, California, USA, April 2005.

[4] K. Bennett and C. Campbell, "Support vector machines: hype or hallelujah," *SIGKDD Explorations*, vol. 2,2, pp. 1–13, 2000.

[5] K. Flouri, B. Beferull-Lozano, and P. Tsakalides, "Training a SupportVector Machine- based Classifier in Distributed Sensor Networks," in *14nd European Signal Processing Conference (EUSIPCO '06)*, Florence, Italy, September 4-8, 2006.