Balancing Lifetime and Classification Accuracy of Wireless Sensor Networks

Kush R. Varshney and Peter M. van de Ven
Business Analytics and Mathematical Sciences Department
IBM Thomas J. Watson Research Center
1101 Kitchawan Road, Route 134
Yorktown Heights, New York
{krvarshn,pmvandev}@us.ibm.com

ABSTRACT

Wireless sensor networks are composed of distributed sensors that can be used for signal detection or classification. The likelihood functions of the hypotheses are often not known in advance, and decision rules have to be learned via supervised learning. A specific learning algorithm is Fisher discriminant analysis (FDA), the classification accuracy of which has been previously studied in the context of wireless sensor networks. Previous work, however, does not take into account the communication protocol or battery lifetime; in this paper we extend existing studies by proposing a model that captures the relationship between battery lifetime and classification accuracy. To do so, we combine the FDA with a model that captures the dynamics of the carrier-sense multiple-access (CSMA) algorithm, the random-access algorithm used to regulate communications in sensor networks. This allows us to study the interaction between the classification accuracy, battery lifetime and effort put towards learning, as well as the impact of the back-off rates of CSMA on the accuracy. We characterize the tradeoff between the length of the training stage and accuracy, and show that accuracy is non-monotone in the back-off rate due to changes in the training sample size and overfitting.

Categories and Subject Descriptors

C.2 [**Computer Systems Organization**]: Computer-Communication Networks

General Terms

Performance

Keywords

carrier-sense multiple-access, linear discriminant analysis, sensor networks

1. INTRODUCTION

Wireless sensor networks are used for detection or classification, whether for surveillance, environmental monitoring, or any of the myriad other application domains that are emerging in the age of big data. In many such applications, the likelihood functions of the hypotheses, e.g., the presence or absence of a particular physical phenomenon, are not known before the sensor network is deployed; in these applications, the sensor network requires training prior to operation via supervised learning [13, 18]. The resulting classification accuracy improves with the number of measurements taken during training [17], but increasing the length of the training stage further reduces the limited battery capacity for the operational stage. Therefore, the amount of resources expended during training mediates operational lifetime and accuracy of the sensor network.

The energy consumption of sensor nodes, and thus the lifetime of the network, is dominated by energy expended on communication. Node transmissions in wireless sensor networks are commonly regulated by the carrier-sense multiple-access (CSMA) algorithm [1, 16]. This algorithm is implemented in TinyOS, a popular open source operating system for wireless sensor networks, and is part of the IEEE 802.15.4 standard for wireless sensor network communication. Nodes using CSMA access the medium in a distributed manner, and wait some random back-off time between successive transmissions.

In this paper we consider a scenario where a set of measurements and classification is required every time unit. Only nodes that are active at that time perform a measurement and transmit the result, so the number of measurements collected varies over time. We develop and analyze a model of sensor networks that perform supervised classification *in situ*, using the Fisher discriminant analysis (FDA) learning algorithm, with a training stage and an operational stage enabled by CSMA. The specific analysis of focus is the relationship between operational accuracy and lifetime, which we show to be of a fundamentally different character than for the case of detection with known likelihood functions, due to overfitting. In characterizing operational classification accuracy (in contrast to classification accuracy on training samples), we make use of generalization approximations for FDA developed by Raudys et al. [15].

Battery capacity is characterized by the number of transmissions that can be performed, whether they be during training or operation. As every measurement corresponds to one transmission, the expected network lifetime is inversely

proportional to the node throughput in our model. The performance measures of interest are the classification accuracy and operational lifetime, which is the lifetime spent in the operational stage, not in the training stage. The two main parameters available for configuring the sensor network are the CSMA back-off rates (the reciprocal of the mean back-off time), and the fraction of the lifetime spent in training.

As the back-off rates of the nodes increase, states with many actively transmitting nodes are more likely. This requires more energy consumption, and also affects classification accuracy. Classification accuracy is not monotonically increasing in the number of active nodes due to the phenomenon of overfitting [17]. We also show that operational accuracy as a function of back-off rate exhibits the hallmarks of overfitting in one regime, but in another regime, has a behavior quite different than any behavior usually encountered in statistical learning.

The analysis of supervised classification for sensor networks in the literature is limited [13]: investigations have been predominantly concerned with the detection case where the likelihood functions are known. Moreover, sensor network research tends to separate learning issues from the communication aspect. There are several works that model CSMA communication in sensor networks generally, e.g. [8] and references therein, but not with the supervised classification application as part of the formulation. Cross-layer work that does consider the networking issues together with a detection or estimation application, e.g., the correlation-based collaborative MAC protocol [20], is again focused on the case with known likelihoods. So although FDA and the performance of CSMA-like algorithms have been widely studied, we are the first to jointly consider classification accuracy and communication aspects of wireless sensor networks.

We consider both the case of statistically independent and identically distributed (i.i.d.) measurements from different nodes, and the case of measurements exhibiting correlation that depends on the spatial distance between the nodes. Having i.i.d. measurements is a common simplifying assumption in wireless sensor network detection [6]. A model with spatially-correlated measurements is much closer to reality in most applications [10]. The spatial correlation is encoded via a Gauss–Markov random field (GMRF) model.

The CSMA model under consideration was first introduced in the 1980s in the context of packet radio networks [5] and was later applied to networks based on the IEEE 802.11 standard [7, 19]. More recently, it has been used to study so-called adaptive CSMA algorithms, where the back-off rate of the nodes changes with their congestion level [14, 4, 9]. Although the representation of binary exponential back-off mechanism in the above-mentioned models is far less detailed than in the landmark work of Bianchi [3] and similar results focusing on sensor networks, e.g., [11], the general interference graph offers greater versatility and covers a broad range of topologies.

The remainder of the paper is organized as follows. In Section 2, we describe the setup of the sensor network system from the FDA supervised classification perspective and in Section 3, we describe the setup of the sensor network system from the CSMA communication perspective. In Section 4 we derive the relationship between operational lifetime and accuracy, and Section 5 presents numerical results of lifetime and accuracy for two special cases, illustrating

the complicated balancing act that is involved. Section 6 provides a discussion and several ideas for future research.

2. FISHER DISCRIMINANT ANALYSIS

Consider a sensor network consisting of n sensor nodes each taking a scalar measurement combined into a joint measurement vector $\mathbf{x}_j \in \mathbb{R}^n$. In the general supervised classification problem, we are given m sample pairs $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$ known as the training set, with measurement $\mathbf{x}_j$ and class label or hypothesis $y_j \in \{0, 1\}$. The training samples are acquired by the network after deployment and before the operational stage. Once the training set is acquired, the samples are used to learn a classification function or decision rule $\hat{y}(\cdot)$ that will accurately classify new unseen and unlabeled samples $\mathbf{x}$ from the same distribution from which the training set was drawn.

In this paper, we focus on a simple, classical decision rule $\hat{y}$, the Fisher discriminant analysis classifier:

$$\hat{y}(\mathbf{x}) = \text{step}(\mathbf{w}^T \mathbf{x} + \theta), \tag{1}$$

where $\mathbf{w} = \left(\hat{\boldsymbol{\Sigma}}_0 + \hat{\boldsymbol{\Sigma}}_1\right)^{-1} (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_0)$, $\theta = -\frac{1}{2}\mathbf{w}^T (\hat{\boldsymbol{\mu}}_0 + \hat{\boldsymbol{\mu}}_1)$, and $\hat{\boldsymbol{\mu}}_0$, $\hat{\boldsymbol{\mu}}_1$, $\hat{\boldsymbol{\Sigma}}_0$, and $\hat{\boldsymbol{\Sigma}}_1$ are the conditional sample means and covariances of the m training samples. The FDA rule is a plug-in classifier that follows from the likelihood ratio test for optimal signal detection between Gaussian signals with the same covariance and different means.

Given the FDA decision rule (1), we would like to characterize its performance, specifically its classification accuracy as it generalizes to new unseen samples in the operational stage. Generalization accuracy, however, is a functional of the underlying data distribution $f_{\mathbf{x},y}(\mathbf{x}, y)$, and we must first specify a probability distribution of the sensor measurements. We employ the same GMRF statistical model for sensor measurements as [17, 2]. That is, the n sensor nodes are deployed on the plane with spatial locations $\mathbf{v}_i \in \mathbb{R}^2$, $i = 1, \ldots, n$. The likelihoods of the two hypotheses are Gaussian: $f_{\mathbf{x}|y}(\mathbf{x}|y = 0) \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ and $f_{\mathbf{x}|y}(\mathbf{x}|y = 1) \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$. The prior probabilities of the hypotheses are equal: $\Pr[y = 0] = \Pr[y = 1] = 1/2$. For simplicity of exposition $\boldsymbol{\mu}_0 = \mathbf{0}$ (the vector of all zeroes) and $\boldsymbol{\mu}_1 = \mathbf{1}$ (the vector of all ones).

The covariance structure is based on the Euclidean nearest neighbor graph of the sensors: the (undirected) nearest neighbor graph contains an edge between sensor i and sensor i' if sensor i is the nearest neighbor of sensor i' or if sensor i' is the nearest neighbor of sensor i. The set of edges in the nearest neighbor graph is denoted $\mathcal{E}$. The diagonal elements of $\boldsymbol{\Sigma}$ are all equal to σ^2. The elements of $\boldsymbol{\Sigma}$ corresponding to edges in the nearest neighbor graph are:

$$\{\boldsymbol{\Sigma}\}_{ii'} = \sigma^2 g(d(\mathbf{v}_i, \mathbf{v}_{i'})), \quad (i, i') \in \mathcal{E}, \tag{2}$$

where $g(\cdot) : \mathbb{R}^+ \to (0, 1)$ is a decreasing function that encodes correlation decay with distance. The inverse covariance matrix $\mathbf{J} = \boldsymbol{\Sigma}^{-1}$ is used to specify the remaining elements. The off-diagonal elements of $\mathbf{J}$ corresponding to sensor pairs (i, i') that do not have an edge in the nearest neighbor graph are zero, i.e.

$$\{\mathbf{J}\}_{ii'} = 0, \quad i \neq i', (i, i') \notin \mathcal{E}. \tag{3}$$

We also consider the case of i.i.d. observations in the paper, in which case $g(d) = 1$ for $d = 0$ and $g(d) = 0$ otherwise.

A highly accurate approximation of generalization accuracy $A = \Pr[\hat{y}(\mathbf{x}) = y]$ for FDA with GMRF is found in [17]. Based on [15], this approximation is

$$A \approx \Phi\left(\frac{\delta}{2}\left[\left(1 + \frac{4n}{m\delta^2}\right)\frac{m}{m-n}\right]^{-\frac{1}{2}}\right), \quad m > n, \quad (4)$$

where $\Phi(\cdot)$ is the Gaussian cumulative distribution function and δ is known as the Mahalanobis distance

$$\delta^2 = \frac{n}{\sigma^2} - \frac{2}{\sigma^2}\sum_{(i,i')\in\mathcal{E}}\frac{g(d(\mathbf{v}_i, \mathbf{v}_{i'}))}{1 + g(d(\mathbf{v}_i, \mathbf{v}_{i'}))}. \quad (5)$$

In case $m \leq n$ there are insufficient training samples for accurate classification and we have $A = 0.5$. In the i.i.d. case, δ^2 simplifies to $\frac{n}{\sigma^2}$.

3. CARRIER-SENSE MULTIPLE-ACCESS

The CSMA algorithm is a random-access algorithm where nodes decide for themselves when to transmit based only on local information. We assume that the n nodes share the wireless medium according to a CSMA-type protocol.

The network is described by an undirected conflict graph $(\mathcal{V}, \mathcal{E})$, where the set of vertices $\mathcal{V} = \{1, \ldots, n\}$ represents the nodes of the network and the set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ indicates which pairs of nodes cannot activate simultaneously. For ease of presentation we assume that the conflict graph is the same as the nearest neighbor graph introduced in Section 2. Nodes that are neighbors in the conflict graph are prevented from simultaneous activity by the carrier-sensing mechanism. An inactive node is said to be blocked whenever any of its neighbors is active, and unblocked otherwise.

The nodes continually monitor the phenomenon of interest, i.e. they are not event-triggered. The transmission times of node i are independent and exponentially distributed with unit mean. When node i is blocked, it remains silent until all its neighbors are inactive, when it tries to activate after an exponentially distributed back-off time with mean $1/\nu_i$.

The set Ω of all feasible joint activity states of the network in this case corresponds to the incidence vectors of all independent sets of the conflict graph. Let the network state at time t be denoted by $\mathbf{Y}(t) = (Y_1(t), Y_2(t), \ldots, Y_n(t)) \in \Omega$, with $Y_i(t)$ indicating whether node i is active at time t ($Y_i(t) = 1$) or not ($Y_i(t) = 0$). Then $\{\mathbf{Y}(t)\}_{t\geq 0}$ is a Markov process which is fully specified by the state space Ω and the transition rates

$$r(\omega, \omega') = \begin{cases} \nu_i, & \text{if } \boldsymbol{\omega}' = \boldsymbol{\omega} + \boldsymbol{e}_i \in \Omega, \\ 1, & \text{if } \boldsymbol{\omega}' = \boldsymbol{\omega} - \boldsymbol{e}_i \in \Omega, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

Since $\mathbf{Y}(t)$ is reversible (see [5]), the following product-form stationary distribution π exists:

$$\pi(\boldsymbol{\omega}) = \begin{cases} Z^{-1}\prod_{i=1}^{n}\nu_i^{\omega_i}, & \text{if } \boldsymbol{\omega} \in \Omega, \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

where Z is the normalization constant that makes π a probability measure.

We assume that nodes attempt to send out data constantly for streaming classification, so the output of a node is limited only by its back-off process. The rate θ_i at which sensor node i makes and transmits an observation is referred to as the throughput of this node, and may be written as

$$\theta_i = \sum_{\boldsymbol{\omega}\in\Omega}\pi(\boldsymbol{\omega})\mathbb{I}_{\{\omega_i=1\}}. \quad (8)$$

Sensor nodes rely on batteries for energy, and we assume that all nodes have a battery that allows them to make l transmissions each before their battery is drained. Consequently, the expected lifetime of a node can be written as

$$T_i = \frac{l}{\theta_i}. \quad (9)$$

The activity process in the training stage is the same as in the operational stage. We denote by $0 \leq \alpha \leq 1$ the fraction of the battery capacity that is dedicated to training the sensor network. So the testing lifetime of node i is αT_i, and the operational lifetime is:

$$U_i = (1-\alpha)T_i = (1-\alpha)\frac{l}{\theta_i}. \quad (10)$$

The model we have specified is fully general for any n-node conflict graph. We work with this general model throughout the remainder of the paper, but also focus on two illustrative special cases. The two special cases of the CSMA network we consider are an n-node network where all networks are disjoint and a three-node linear network.

3.1 Independent Nodes

First, consider an n-node network where all nodes can be active simultaneously. This corresponds to an interference graph with an empty edge set $E = \phi$ and $\Omega = \{0,1\}^n$. Note that although in this case the lack of interference eliminates the need for a MAC protocol like CSMA, it is nevertheless interesting because it allows us to derive closed-form results. Moreover, deployment of sensor networks is typically random, and such disjoint network may arise in practice. We set $\nu_i \equiv \nu$ so the stationary distribution (7) simplifies to

$$\pi(\boldsymbol{\omega}) = \frac{1}{(\nu+1)^n}\nu^{\|\boldsymbol{\omega}\|_1}. \quad (11)$$

The stationary probability of any particular state only depends on the number of active nodes in that state and on the back-off rate ν. Thus, for notational convenience, we introduce $\pi(k)$ as the stationary probability of being in any state with k active nodes, and we write

$$\pi(k) = \frac{1}{(\nu+1)^n}\binom{n}{k}\nu^k, \quad (12)$$

which follows since there are $\binom{n}{k}$ different activity states with k nodes transmitting.

With equal back-off rates and disjoint nodes, the stationary throughput (8) is the same for all nodes

$$\theta_i \equiv \theta = \frac{\nu}{\nu+1}. \quad (13)$$

Moreover, all nodes have the same lifetime, and the operational lifetime of the network may be written as

$$U_i \equiv U = (1-\alpha)l\frac{\nu+1}{\nu}. \quad (14)$$

3.2 A Three-Node Linear Network

Consider the three-node network where the nodes are positioned such that the carrier-sensing mechanism prevents

node 2 from activating while either node 1 or node 3 is active. Nodes 1 and 3 can be active simultaneously, but their observations are correlated. The network can take five possible states

$$\Omega = \{\mathbf{0}, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_1 + \boldsymbol{e}_3\}. \tag{15}$$

Using (7) we compute the following stationary probabilities:

$$\pi(\mathbf{0}) = Z^{-1},$$
$$\pi(\boldsymbol{e}_i) = Z^{-1}\nu_i, \quad i = 1, 2, 3,$$
$$\pi(\boldsymbol{e}_1 + \boldsymbol{e}_3) = Z^{-1}\nu_1\nu_3. \tag{16}$$

In order to make sure that all nodes have the same throughput and lifetime, we fix some parameter $\eta > 0$ and choose $\nu_1 = \nu_3 = \eta$ and $\nu_2 = \eta(\eta + 1)$. So node 2 has a shorter mean back-off time in order to compensate for its disadvantageous position in the network, and all nodes have throughput (see [19])

$$\theta_i \equiv \theta = \frac{\eta}{2\eta + 1} \tag{17}$$

and operational lifetime

$$U_i \equiv U = (1 - \alpha)l\frac{2\eta + 1}{\eta}. \tag{18}$$

The normalization constant with these back-off rates is

$$Z = 2\eta^2 + 3\eta + 1. \tag{19}$$

4. RELATIONSHIP BETWEEN LIFETIME AND ACCURACY

We are now in position to combine the FDA model from Section 2 and the CSMA model presented in Section 3 to derive the relationship between generalization accuracy and operational lifetime. This is mediated by two parameters: the back-off rate ν or η and the fraction of the lifetime spent in the training stage α.

Due to the interference constraints and the intermittent nature of CSMA communications, not all nodes produce and validly communicate measurements at all times. So the training samples are acquired under different activity states $\boldsymbol{\omega} \in \Omega$. Thus studying the relationship between accuracy and lifetime is not simply a matter of joining the corresponding expressions (4) and (10).

This issue of incomplete data due to the activity process can be addressed in several ways, including data imputation. Although various elaborate schemes are available, they come at the cost of additional computation, communication, and coordination that are at a premium in the sensor network setting. Instead, we choose to model the classification by having separately learned classifiers for different activity states. In the operational stage the appropriate classifier is used for prediction based on the activity state of the measurements. In this setup, we associate with each state $\boldsymbol{\omega}$ a number of training samples

$$m_{\boldsymbol{\omega}} = \alpha T \pi(\boldsymbol{\omega}). \tag{20}$$

Then we compute the overall generalization accuracy as the weighted sum of the individual generalization accuracies for each pattern according to their stationary probabilities:

$$A \approx \sum_{\boldsymbol{\omega} \in \Omega} \pi(\boldsymbol{\omega}) \Phi \left(\frac{\delta}{2} \left[\left(1 + \frac{4\|\boldsymbol{\omega}\|_1}{m_{\boldsymbol{\omega}}\delta^2} \right) \frac{m_{\boldsymbol{\omega}}}{m_{\boldsymbol{\omega}} - \|\boldsymbol{\omega}\|_1} \right]^{-\frac{1}{2}} \right), \tag{21}$$

with π the stationary distribution (7) and $m_{\boldsymbol{\omega}}$ as in (20).

We now compute the generalization accuracies for the two special cases introduced in Section 3 with the GMRF of the measurements having the same graph structure as the CSMA network.

4.1 Independent Nodes

As discussed in Section 3 for a set of n disjoint nodes, all patterns with k active nodes have the same stationary probability $\pi(k)$ given in (12), and all nodes have equal throughput (13) and lifetime (14). We denote by m_k the number of training samples for patterns with k active nodes, and by summing (20) over all states with k active nodes, we write

$$m_k = \alpha l \binom{n}{k} \frac{\nu^{k-1}}{(\nu + 1)^{n-1}}. \tag{22}$$

As discussed in Section 2, with i.i.d. measurements from n sensors, the squared Mahalanobis distance is $\frac{n}{\sigma^2}$. Thus, with k active sensors, the squared Mahalanobis distance is $\frac{k}{\sigma^2}$. Substituting the expression for the stationary distribution (12) and the number of training samples (22) into the expression for the generalization accuracy (21) we obtain

$$A \approx \frac{1}{(\nu + 1)^n} \sum_{k=0}^{n} \binom{n}{k} \nu^k \Phi \left(\frac{\sqrt{k}}{2\sigma} \left[\left(1 + \frac{4\sigma^2}{m_k} \right) \frac{m_k}{m_k - k} \right]^{-\frac{1}{2}} \right). \tag{23}$$

4.2 A Three-Node Linear Network

Recall from Section 3.2 that the three-node network has 5 feasible states. The four non-empty states have squared Mahalanobis distance

$$\delta^2_{\boldsymbol{e}_i} = \frac{1}{\sigma^2}, \quad i = 1, 2, 3,$$
$$\delta^2_{\boldsymbol{e}_1 + \boldsymbol{e}_3} = \frac{2}{\sigma^2} \cdot \frac{1}{1 + g(d(\mathbf{v}_1, \mathbf{v}_2))g(d(\mathbf{v}_2, \mathbf{v}_3))}. \tag{24}$$

Note that since $g < 1$, the Mahalanobis distance of the larger state is larger than that of the states with only one node active, and is more valuable.

Evaluating (20) we obtain an expression for the number of training samples for each state:

$$m_{\mathbf{0}} = \alpha l \frac{1}{\eta^2 + \eta}$$
$$m_{\boldsymbol{e}_i} = \alpha l \frac{1}{\eta + 1}, \quad i = 1, 3,$$
$$m_{\boldsymbol{e}_2} = \alpha l,$$
$$m_{\boldsymbol{e}_1 + \boldsymbol{e}_3} = \alpha l \frac{\eta}{\eta + 1}. \tag{25}$$
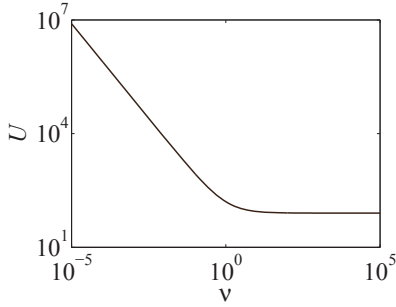
Figure 1: Operational lifetime as function of back-off rate.

By weighting the individual generalization accuracies (21),

$$A \approx$$

$$\frac{1}{2\eta^2 + 3\eta + 1} \left[\frac{1}{2} + \eta\Phi\left(\frac{1}{2\sigma} \left[\left(1 + \frac{4\sigma^2}{m_{e_1}}\right) \frac{m_{e_1}}{m_{e_1} - 1} \right]^{-\frac{1}{2}} \right) + \right.$$

$$(\eta^2 + \eta)\Phi\left(\frac{1}{2\sigma} \left[\left(1 + \frac{4\sigma^2}{m_{e_2}}\right) \frac{m_{e_2}}{m_{e_2} - 1} \right]^{-\frac{1}{2}} \right) +$$

$$\eta\Phi\left(\frac{1}{2\sigma} \left[\left(1 + \frac{4\sigma^2}{m_{e_3}}\right) \frac{m_{e_3}}{m_{e_3} - 1} \right]^{-\frac{1}{2}} \right) +$$

$$\left. \eta^2\Phi\left(\frac{\delta_{e_1+e_3}}{2} \left[\left(1 + \frac{8}{m_{e_1+e_3}\delta_{e_1+e_3}^2}\right) \frac{m_{e_1+e_3}}{m_{e_1+e_3} - 2} \right]^{-\frac{1}{2}} \right) \right].$$
(26)

5. EXAMPLES

In Section 4 we derived the operational lifetime U, the number of training samples m_{ω} and the operational classification accuracy A for a wireless sensor network with random-access communication as a function of the back-off rate ν and the fraction of the lifetime spent in training α. Here we numerically evaluate these quantities for the special cases of independent nodes and the three-node linear network. We include a comparison to the Bayes optimal detector with known likelihood functions and see that the accuracy behavior is markedly different. Additionally, we see that there are two different regimes in the accuracy behavior as a function of the back-off rate, the second regime different than that usually seen in statistical learning. The overall behavior is unique due to the combination of CSMA and FDA.

5.1 Independent Nodes

We consider a network of $n = 8$ independent nodes with $l = 100$ transmissions allowed by the battery per node. The sensor measurement noise variance is set to $\sigma^2 = 1$. Other parameter settings produce qualitatively similar results. First, in Fig. 1, we plot the operational lifetime U as a function of the back-off rate ν for a fixed lifetime fraction devoted to training: $\alpha = 0.2$. The operational lifetime is very high with low back-off rate because the system is mostly in the state with $k = 0$ active sensors, which does not drain sensor batteries at all. Once states with more active sensors become more probable with increasing back-off rate, the lifetime drops rapidly to $U = (1 - \alpha)l$, the lifetime
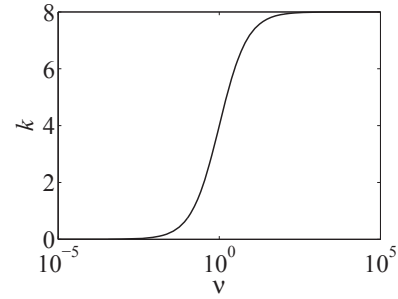

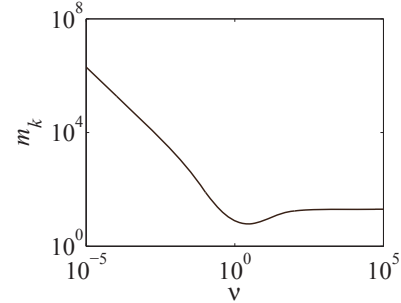Figure 2: Expected number of active sensors as function of back-off rate.

Figure 3: Expected number of training samples per state classifier as function of back-off rate.

in the case where all nodes are always active. Fig. 2 shows the expected number of active sensors $\bar{k}$ as a function of ν.

One of the components of the generalization accuracy expression (23) is the number of active sensors k; the other is m_k, the number of training samples. In Fig. 3 we set $\alpha = 0.2$, and plot $\bar{m}_k$, the weighted average over k of m_k:

$$\bar{m}_k = \sum_{k=0}^{n} \pi(k)m_k.$$
(27)

Interestingly, this number is not monotonically decreasing as a function of ν like we see with the operational lifetime. This is because when several different states all have non-negligible probability, the acquired training samples get divided to all of the different states. Initially the number of training samples is very high because almost all of the training samples are for the state with no active sensors. For large ν the number of training samples approaches αl.

Now that we have looked at $\bar{k}$ and $\bar{m}_k$, we now examine the accuracy A as a function of ν, plotted in Fig. 4 for $\alpha = 0.2$. The figure also shows the detection accuracy of the Bayes optimal decision rule with known likelihood functions. The Bayes optimal accuracy is monotonically increasing in ν, following the expected value of k. On the other hand, the FDA classification accuracy first increases in ν, starts decreasing with local bumps, and then increases. The local bumps arise from the generalization accuracy behavior for different states, which are shown in Fig. 5. Specifically, the figure shows the different $\Phi(\cdot)$ components of A given in (23); these are functions of ν because the m_k are.

The phenomenon of overfitting, demonstrated in [17], is that for a fixed number of training samples and an increasing number of sensors, the generalization accuracy first increases and then decreases. Conversely, for a fixed number of sensors and an increasing number of training samples, the
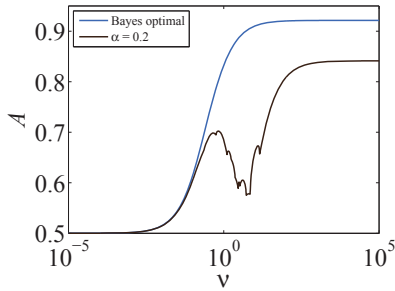
Figure 4: Operational classification accuracy as function of back-off rate.
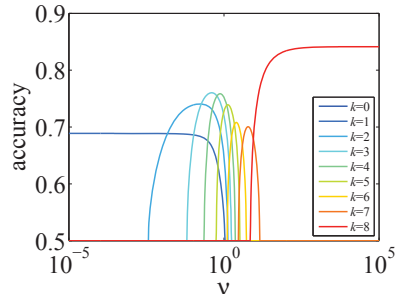

Figure 6: Relationship between operational lifetime and operational classification accuracy.

Figure 5: Operational classification accuracy of different state classifiers as function of back-off rate.

Figure 7: Operational lifetime as function of back-off rate.

generalization accuracy monotonically increases. With the wireless sensor network with CSMA communication, both of these effects intermingle as a function of ν because both the number of active sensors and the number of training samples changes. The initial increase and decrease in A is the manifestation of overfitting, where the generalization accuracy is best around $k = 3$ and $k = 4$. For large ν, the number of active sensors is essentially fixed at $k = n$ (seen in Fig. 2) and the number of training samples increases (seen in Fig. 3), resulting in improving classification accuracy.

Finally, we examine the relationship between lifetime and accuracy in Fig. 6. For comparison, the figure shows the relationship for the Bayes optimal decision rule, in which there is no lifetime devoted to training, only to operation. All curves represent parametric functions of ν, and correspond to different values of α ranging from 0.1 to 0.9. Different values of α contribute to the frontier of the relationship, the parts of the curve closest to the Bayes decision rule and closest to the top right corner of the plot. At the extreme of random guessing, i.e. $A = 0.5$, lifetime is maximized by not doing any training, i.e. $\alpha = 0$. Small but increasing values of α then contribute to the frontier until a point when smaller values of α abruptly again become part of the frontier. Very large values of α contribute to the frontier only when the very best accuracies possible are desired.

5.2   A Three-Node Linear Network

Having seen quite interesting behaviors for independent nodes, we now turn to a three-node linear network with correlated measurements and conflict graph preventing sensors 1 and 2, and sensors 2 and 3 from transmitting simultaneously. We present similar plots as in Section 5.1, with $l = 10$ and $\sigma^2 = 1$. We present results for $g(d(\mathbf{v}_1, \mathbf{v}_2)) = g(d(\mathbf{v}_2, \mathbf{v}_3)) = \frac{1}{4}$ as the distance-based correlations. For the
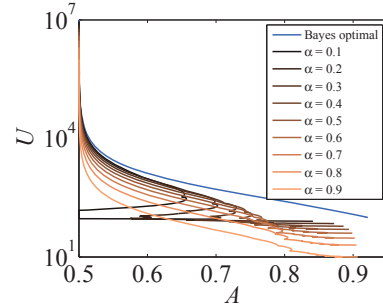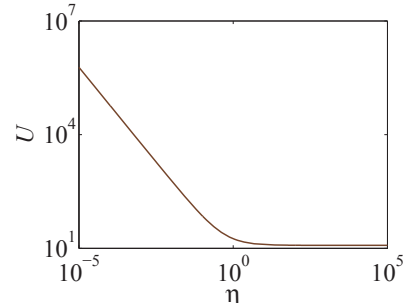
dependent linear network case, we see more or less the same behavior as for the independent nodes in Fig. 7–Fig. 12. Fig. 7, Fig. 9, and Fig. 11 are given for $\alpha = 0.4$. One difference from the independent nodes case is that for the $\mathbf{e}_2$ state, $m_{\boldsymbol{\omega}}$ is constant and not a function of η.

6.   CONCLUSION AND OUTLOOK

In this paper we proposed a model to investigate the interaction between generalization accuracy and operational lifetime in wireless sensor networks. We demonstrate that this relationship is highly nontrivial, due to the joint effects of overfitting, the number of training samples and the changing weights of the various states. The two special cases for which we provide result plots are qualitatively similar, and changing the conflict graph and spatial correlation does not affect the general behavior.

For small increasing back-off rates, the accuracy improves until peaking. At intermediate back-off rates, the accuracy gets worse due to overfitting, and then improves again for large back-off rates due to increasing training samples per state. Due to these different regimes of increasing and decreasing accuracy in the back-off rate, along with different values of the fraction of lifetime to spend in training affecting both the operational lifetime and the accuracy, setting the parameters ν and α to achieve certain target performance is not straightforward. The parameterized curves in Fig. 6 and Fig. 12 give (not necessarily intuitive) recommendations for balancing lifetime and accuracy as a function of the back-off rate and training fraction parameters.

The classification and communication models we have used, i.e. FDA with GMRF-dependent sensor measurements and binary exponential back-off mechanism, are certainly simplified, but are general and amenable to analysis. The guidelines and behaviors we see will transfer over in a broad sense
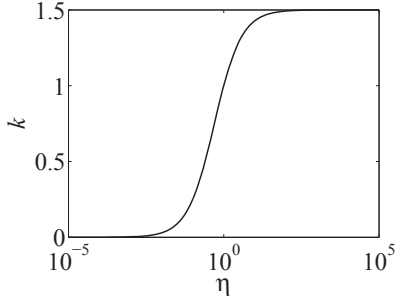
Figure 8: Expected number of active sensors as function of back-off rate.
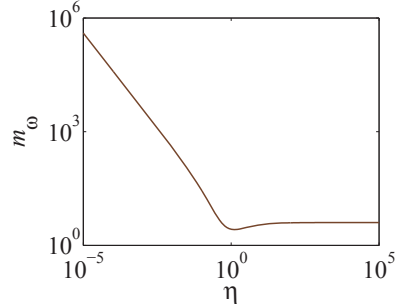

Figure 9: Expected number of training samples per state classifier as function of back-off rate.
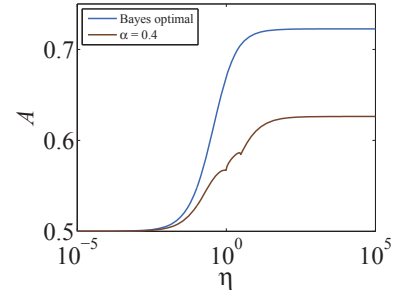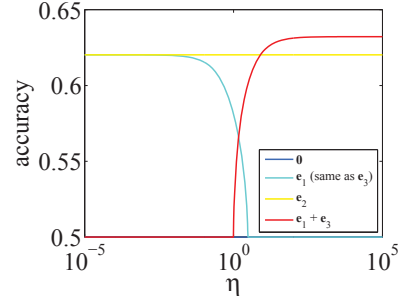

Figure 10: Operational classification accuracy as function of back-off rate.

Figure 11: Operational classification accuracy of different state classifiers as function of back-off rate.

to other classifiers and other similar random-access communication protocols.

6.1 Outlook

Having made the connection between lifetime and accuracy in Fig. 6 and Fig. 12, the next step is to find the values of α and ν_i that achieve a certain target performance. For example, we may want to maximize the lifetime of the network, subject to certain accuracy constraints $\beta \in (0, 1)$:

$$(\alpha^*, \nu^*) = \arg\max U(\alpha, \nu_1, \ldots, \nu_n)$$
$$\text{s.t. } A(\alpha, \nu_1, \ldots, \nu_n) \geq \beta. \qquad (28)$$

The optimization problem (28) is non-convex (as illustrated in Fig. 6), and we may approximate its solution using numerical methods. Some preliminary results are shown in Figs. 13 and 14, where we plot the solution to (28) for increasing β, in the model with n independent nodes with the parameters as in Section 5. Fig. 13 shows that the ν increases almost monotonically, and jumps to infinity around $\beta = 0.772$. In practice we see that the back-off rate is constrained by physical limitations and by the communication protocol, so ν is bounded from above. Fig. 14 shows a more irregular behavior for α, with a sharp drop when ν jumps to infinity.

The non-monotonicity of the classification accuracy in the back-off rates makes an analytic approach to optimization difficult, and an alternative solution would be to approximate the expression for the detection accuracy (21) with some convex function. This would reduce the complexity of numerical optimization, and may even allow for analytical results.

The effect of overfitting for medium back-off rates can be mitigated by choosing different back-off rates of the training stage and operational stage. For example, choosing larger back-off rates during training should increase the number of samples for states with many active nodes, thus reducing the risk of overfitting. Although this would simultaneously reduce the number of samples for smaller states, the risk of overfitting is not as high there due to the smaller number of active nodes.

Another direction for future research is to model temporal correlation in the sensor measurements in addition to spatial correlation. In the present work, successive measurements in time are assumed independent, but including temporal correlation is more realistic. If temporal correlation is part of the sensing and classification model, its interaction with the temporal back-off mechanism may produce quite interesting phenomena. Moreover, many other properties and parameters of the network such as link quality, physical properties measured in joules, watts and volts, network diameter, and multihop routing can be analyzed.

Finally, we also mention that asymptotic analysis is of interest in the future study of this cross-layer supervised learning and random-access communication setup. Developing expressions for the three-node dependent network, e.g., (25) and (26), requires us to keep track of many details; larger networks will require us to keep track of many more. By performing an asymptotic analysis of an increasing number of randomly placed sensor nodes with constant density we are able to eliminate many such details in the sensor network generalization error using geometric probability [12]. Having now set forth this extended model with CSMA communication, similar asymptotic analysis using geometric probability is certainly warranted. Such asymptotics can also inform classifier design in that a separate classifier for each specific state not be required.
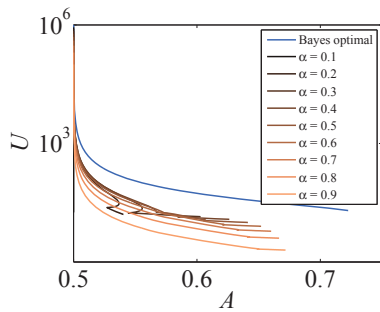
Figure 12: Relationship between operational lifetime and operational classification accuracy.
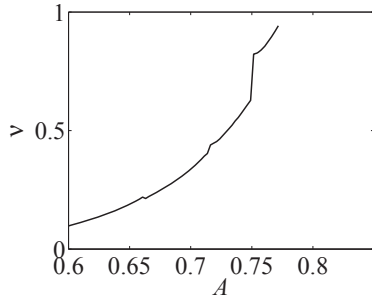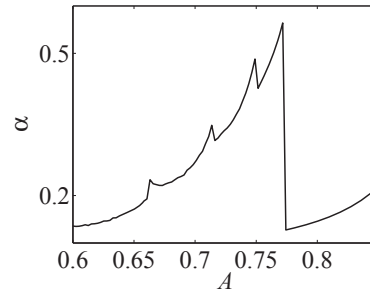

Figure 13: ν^* as function of desired accuracy.

Figure 14: α^* as function of desired accuracy.

7. REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *IEEE Commun. Mag.*, 40(8):102–114, Aug. 2002.

[2] A. Anandkumar, L. Tong, and A. Swami. Detection of Gauss–Markov random fields with nearest-neighbor dependency. *IEEE Trans. Inf. Theory*, 55(2):816–827, Feb. 2009.

[3] G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE J. Sel. Areas Commun.*, 18(3):535–547, Mar. 2000.

[4] T. Bonald and M. Feuillet. On the stability of flow-aware CSMA. *Perform. Evaluation*, 67(11):1219–1229, Nov. 2010.

[5] R. Boorstyn, A. Kershenbaum, B. Maglaris, and V. Sahin. Throughput analysis in multihop CSMA packet radio networks. *IEEE Trans. Commun.*, COM-35(3):267–274, Mar. 1987.

[6] J.-F. Chamberland and V. V. Veeravalli. Decentralized detection in sensor networks. *IEEE Trans. Signal Process.*, 51(2):407–416, Feb. 2003.

[7] M. Durvy, O. Dousse, and P. Thiran. Self-organization properties of CSMA/CA systems and their consequences on fairness. *IEEE Trans. Inf. Theory*, 55(3):931–943, Mar. 2009.

[8] E. Feo and G. A. Di Caro. An analytical model for IEEE 802.15.4 non-beacon enabled CSMA/CA in multihop wireless sensor networks. Technical Report 05-11, Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, Lugano, Switzerland, May 2011.

[9] L. Jiang and J. Walrand. A distributed CSMA algorithm for throughput and utility maximization in wireless networks. *IEEE/ACM Trans. Netw.*, 18(3):960–972, June 2010.

[10] A. Jindal and K. Psounis. Modeling spatially correlated data in sensor networks. *ACM Trans. Sensor Netw.*, 2(4):466–499, Nov. 2006.

[11] T. Park, T. Kim, J. Choi, S. Choi, and W. Kwon. Throughput and energy consumption analysis of IEEE 802.15.4 slotted CSMA/CA. *Electron. Lett.*, 41(18):1017–1019, Sept. 2005.

[12] M. D. Penrose and J. E. Yukich. Weak laws of large numbers in geometric probability. *Ann. Appl. Prob.*, 13(1):277–303, Jan. 2003.

[13] J. B. Predd, S. R. Kulkarni, and H. V. Poor. Distributed learning in wireless sensor networks. *IEEE Signal Process. Mag.*, 23(4):56–69, July 2006.

[14] S. Rajagopalan, D. Shah, and J. Shin. Network adiabatic theorem: An efficient randomized protocol for content resolution. In *Proc. ACM SIGMETRICS/Performance*, pages 133–144, Seattle, WA, June 2009.

[15] Š. Raudys and D. M. Young. Results in statistical discriminant analysis: A review of the former Soviet Union literature. *J. Multivariate Anal.*, 89(1):1–35, Apr. 2004.

[16] Y. C. Tay, K. Jamieson, and H. Balakrishnan. Collision-minimizing CSMA and its applications to wireless sensor networks. *IEEE J. Sel. Areas Commun.*, 22(6):1048–1057, Aug. 2004.

[17] K. R. Varshney. Generalization error of linear discriminant analysis in spatially-correlated sensor networks. *IEEE Trans. Signal Process.*, 60(6):3295–3301, June 2012.

[18] K. R. Varshney and A. S. Willsky. Linear dimensionality reduction for margin-based classification: High-dimensional data and sensor networks. *IEEE Trans. Signal Process.*, 59(6):2496–2512, June 2011.

[19] P. M. van de Ven, J. S. H. van Leeuwaarden, D. Denteneer, and A. J. E. M. Janssen. Spatial fairness in linear random-access networks. *Perform. Evaluation*, 69(3–4):121–134, Mar.–Apr. 2012.

[20] M. C. Vuran and I. F. Akyildiz. Spatial correlation-based collaborative medium access control in wireless sensor networks. *IEEE/ACM Trans. Netw.*, 14(2):316–329, Apr. 2006.