

# OPTIMAL ACTUATION STRATEGIES FOR SENSOR/ACTUATOR NETWORKS

*M.J. Coates, F. Thouin and R. Thommes*

Department of Electrical and Computer Engineering  
3480 University St, Montreal, Quebec, Canada H3A 2A7.  
{coates,fthoui,rthommes}@ece.mcgill.ca

## Introduction

Sensor networks have been successfully deployed for passively monitoring environments, for example [1], but there has been much less progress towards developing networks that interact with the environment. Wireless sensor and actuator networks (SANETs) represent an important extension, allowing nodes to perform actions (actuation) in response to sensor measurements and shared information. The potential applications of such wireless SANETs are widespread, including agricultural maintenance and localized delivery of medication.

A dynamic actuation strategy is a set of decision rules, one per time-step, that determine the nature of the actuation performed throughout the operation of the SANET. In this paper, we focus on the problem of designing actuation strategies to achieve a desired response from the environment. We consider the scenario where no such knowledge exists, but measurement data are available from previous actuation epochs. The goal is then to use the experimental data to identify an actuation strategy that maximizes an estimated marginal mean response (or reward). Our problem formulation and methodology are based on batch Q-learning approaches [2]. We depart from this work in adopting a semi-parametric model for the expected marginal response to a given actuation strategy. We use a penalized likelihood formulation and estimate the Q-functions using distributed learning of piecewise-constant dyadic regression trees [3].

## Problem Statement

We consider a wireless SANET comprised of joint sensor-actuator nodes labelled  $i = 1, \dots, N$ . At the beginning of the discrete time interval  $t$ , node  $i$  measures a set of environmental variables  $V_t^{(i)}$ , and decides the level of actuation (indicated by the random variable  $A_t^{(i)}$ ) according to  $g(V_t, A_{0:t-1})$ , a known, probabilistic function of the measured variables and actuation history. At the end of a measurement (actuation) epoch, each wireless SANET node senses the environment to record a local response variable  $Y_T^{(i)}$ . Measurements are made at the discrete time-intervals  $t = 0 : T$ . We denote the observed data at the end of the epoch as  $\{X_{0:T}, A_T, Y_T\}$ , where  $X_{0:t} = \{A_{0:t-1}^{(i)}, V_{0:t}^{(i)}\}_{i=1}^N$ . We assume that there are no *unmeasured direct confounders* in the observed data [2]; that is, the actuation decision  $A_t$  is independent of *potential* outcomes given the actuation and measurement history  $X_{0:t}$ . Furthermore, we assume that measurements are independent and actuations have a local effect; they do not affect measurements at other nodes.

Define a dynamic actuation strategy as a collection of functions  $d_t(X_{0:t})$ , each of which at time  $t$  generates an actuation  $A_t$  based on the observed data and actuation history. We represent this dynamic actuation strategy as a vector  $d = (d_0, \dots, d_T)$ . Denote the class of actuation strategies that have non-zero probability of appearing in the observed data as  $D_P$ , where  $P$  is the distribution of the observed data. Our objective

---

This research was funded by NSERC.

is to identify the dynamic actuation strategy  $d^* \in D_P$  that maximizes the marginal mean response  $E_P[Y_T|d]$ .

### Methodology

Define the  $Q$ -functions and *optimal benefit-to-go functions*  $J_0, \dots, J_{T-1}$  with the recursive relationships:

$$Q_0(X_{0:T}, a_T) = E[Y|X_{0:T}, A_T = a_T] \quad (1)$$

$$J_0(X_{0:T}) = \sup_{a_T: p_T(a_T|X_{0:T}) > 0} Q_0(X_{0:T}, a_T) \quad (2)$$

$$Q_{T-j}(X_{0:j}, a_j) = E[J_{T-j-1}(X_{0:j+1})|X_{0:j}, A_j = a_j] \quad j = 0, \dots, T-1 \quad (3)$$

$$J_{T-j}(X_{0:j}) = \sup_{a_j: p_j(a_j|X_{0:j}) > 0} Q_{T-j}(X_{0:j}, a_j) \quad j = 0, \dots, T-1 \quad (4)$$

Iteratively determining each optimal benefit-to-go function describes a dynamic program that identifies  $d^*$ :

$$d_j^*(V_{0:j}, a_{0:j-1}) = \arg \sup_{a_j: p_j(a_j|X_{0:j}) > 0} Q_{T-j}(V_{0:j}, a_{0:j}) \quad j = 0, \dots, T \quad (5)$$

This procedure cannot be applied directly because the probability distributions of the response functions,  $p(Y|X_{0:T}, A_t)$ , are not specified; the expectations cannot be evaluated explicitly. We complete the approach by iteratively performing a semi-parametric estimation of  $Q_j$  for  $j = 0 : T$ . We use a complexity-penalized regression tree procedure to form estimates  $\tilde{Q}_{T-j}(X_{0:j}, a_j)$  [3]. The set of functions,  $\tilde{d}_j(V_{0:j}, a_{0:j}) = \arg \max_{a_j} \tilde{Q}_{T-j}(V_{0:j}, a_{0:j})$ , where the maximization is over actuations with non-zero probability, forms our actuation strategy.

### Distributed Evaluation

We consider scenarios where the dimension of the data collected at the sensor/actuator nodes is large. In such cases, transporting raw (or even compressed) measurements to a central fusion site is energy-expensive. We propose and examine two alternative algorithms. In the first approach, each tree grows a full-depth tree based on its data (the maximum depth is specified), and sends the mean value and sum-of-squares error of each leaf node to the fusion centre. These form sufficient statistics for construction of a piecewise-constant tree (the same tree as if the fusion centre had access to all measurements). In a second approach, each sensor node grows and prunes its own tree, and sends a description of its tree structure to the fusion centre. The fusion centre applies a heuristic to merge the individual trees. The second approach can substantially reduce communication requirements, but the tree derived at the fusion centre generally differs from the full-data tree.

### References

- [1] D. Myung, B. Duncan, D. Malan, M. Welsh, M. Gaynor, and S. Moulton, "Vital dust: Wireless sensors and a sensor network for real-time patient monitoring," in *Proc. New Eng. Reg. Trauma Conf.*, Burl., MA, Nov. 2003.
- [2] S.A. Murphy, "Optimal dynamic treatment regimes," *J. Roy. Stat. Soc., Ser. B*, vol. 65, no. 2, pp. 331–366, 2003.
- [3] C. Scott, R. Willett, and R. Nowak, "CORT: Classification or regression trees," in *Proc. IEEE ICASSP*, Hong Kong, China, April 2003.