

Towards Active Crowdsourcing for Smart Cities

François Schnitzler*, Jia Yuan Yu†, Shie Mannor*

We present a new application for reinforcement learning: the monitoring of large dynamic processes (traffic, flood level...) through active crowdsourcing. Today, radio broadcasts and/or government organisations rely on people to report on unusual events, such as traffic jams. Our goal is to develop and test an active setting, where voluntary individuals are queried and compensated for their answers.

Requesting a continuous report from each volunteer, while very informative, is not feasible in practice. Only a subset of people can be queried at each time step, because of resources availability and avoiding pestering participants. Moreover, the responses are noisy. The noise associated to each participant varies and is a priori unknown.

The question we want to answer is which sensors should be queried at each time-step, in order to maintain a “good” (this is application dependent) estimate of the system state. We have identified two subtasks for which reinforcement learning plays a key role.

The first subtask is an exploration-exploitation tradeoff between the uncertainty of the participants’ noise and the uncertainty of the system. Preestimating this noise is not possible: we expect people to join and leave the program continuously. This is close to a multi-armed bandit problem, where each participant is a complex arm, but with a key difference. The reward received by querying a (combination of) participant(s) is not known directly. Rather, it depends on the quality of the arm, and must be estimated by the learning algorithm based on its estimate of the system state. When a new observation is received, former system state estimates and rewards can be re-evaluated.

The second subtask is the computation of a querying policy to optimise the rewards. Even if the noise is known, selecting the sensors can be nontrivial for some reward functions, due to the complexity of the system.

Formalisation: Because it is tractable (and has been used for traffic monitoring [1]), we model (for now) the process as a linear system with hidden variables x , that can be monitored using the sensors y_i (the participants, assumed fixed, $i \in \{1, \dots, I\}$, I large):

$$\begin{aligned} x[t+1] &= A_t x[t] + w[t] & x[1] &\sim \mathcal{N}(\bar{x}_1, \Sigma_1) & w[t] &\sim \mathcal{N}(0, \Sigma_{w,t}) \\ y_i[t] &= C_{i,t} x[t] + v_i[t] & v_i[t] &\sim \mathcal{N}(0, \sigma_i^2) \quad \forall i \end{aligned} .$$

Only $K < I$ sensors y_i can be used at each time step. Each y_i is perturbed by white gaussian noise $\mathcal{N}(0, \sigma_i^2)$ of variance σ_i^2 . $A_t, C_{i,t}, \bar{x}_1, \Sigma_1, \Sigma_{w,t}$ are supposed known $\forall i, t$ but the parameters $\Theta \equiv \{1/\sigma_i^2\}_i$ are unknown. $u_t, \hat{x}_{t|t',u}$ and $\hat{\Sigma}_{t|t',u}$ respectively denote the sensors used at time t , the system state estimate and the covariance of the estimate error at time t , given a selection of sensors $u_{1:t'}$ from time 1 to t' .

Our objective is to select the sensors that minimise the expected quadratic estimation error over time while satisfying the constraints:

$$\arg \min_{u'_{1:\infty}: |u'_t| \leq K \quad \forall t} \sum_{t=1}^{\infty} \gamma^t \text{trace}(\hat{\Sigma}_{t|t,u'}) \quad \gamma \in]0, 1] .$$

*Department of Electrical Engineering, Technion, Israel

†IBM Research - Ireland

Algorithm 1

for $t = 1 \rightarrow \infty$ **do**
 Compute $\hat{\Theta}_t \in \max_{\Theta} P(\Theta | y_{1:t-1})$ ▷ EM algorithm
 $u_t = \arg_{u_t} \min_{u_{t:\infty}: |u_t| \leq K \forall t} \sum_{t'=1}^{\infty} \gamma^{t'} \text{trace}(\hat{\Sigma}_{t'|t', [u_{1:t-1} u_{t:\infty}]})$ ▷ classical planning
end for

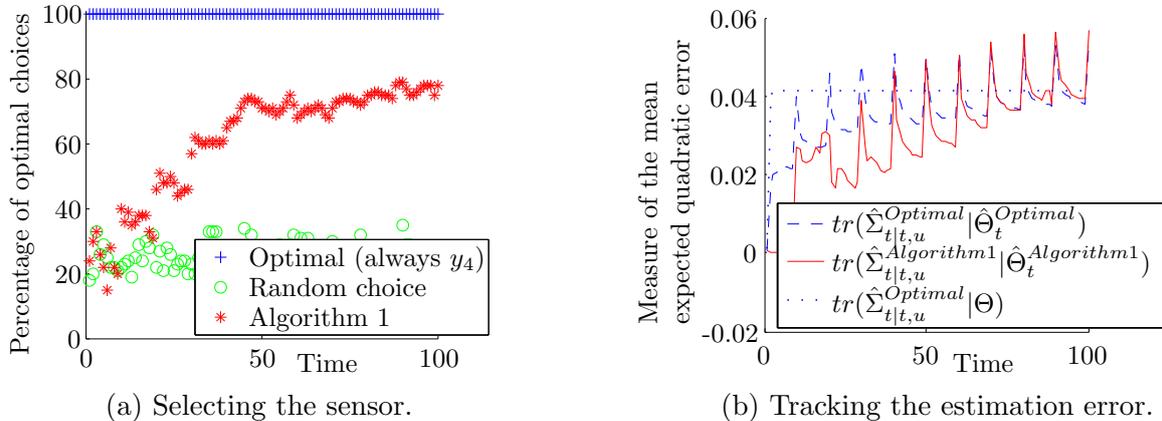


Figure 1: Averages over 100 experiments on a toy problem: one variable x , 4 sensors and $K = 1$; $A_t = [1]$; $C_{i,t} = [1]$; $\Sigma_{w,t} = [0.2] \forall t, i$; $[\sigma_1 \sigma_2 \sigma_3 \sigma_4] = [1 \ 0.5 \ 0.1 \ 0.05]$, $\gamma = 1$.

First algorithm: Algorithm 1 tackles a simple form of the first subtask. For each time step, it computes a maximum likelihood estimate $\hat{\Theta}_t$ of Θ using the observations collected so far, reestimates the state uncertainty and optimises u_t based on $\hat{\Theta}_t$. The latter step is a classical planning problem: the evolution of $\hat{\Sigma}_{t|t-1,u}$ is deterministic in a Kalman filter.

The sensors selected and the reward estimated by Algorithm 1 on a toy problem are illustrated in Figure 1. Note that $\hat{\Theta}_{t-1}$ is used as the initial value of Θ in the EM algorithm at time t , except when $t \bmod 10 = 0$, where $\hat{\Theta}_t$ is reinitialised.

Future research: We would like to derive convergence rates, evaluate scalability and, if necessary, develop approximation schemes for Algorithm 1 (and a similar algorithm, based on Thompson sampling and not detailed here). The formalisation can also be refined by considering uncertain dynamics, alternative models or the participants' lassitude.

Our medium-term goal is to deploy and test the algorithms developed in real large-scale settings, and in particular for traffic estimation in the city of Dublin (120 km², 500 000 inhabitants). Active crowdsourcing would complement a thousand inductive loop detectors used to measure traffic and adjust traffic light timings.

Our current plan, in collaboration with Dublin city council, is to query volunteers through smartphone applications. To make the system engaging, relevant compensations for participants must be designed. Possible schemes include parking vouchers, currently used by Dublin radio to reward voluntary reports, the possibility to select a song to played on the radio, public recognition through a system of points.

We hope these experiments will drive further research in reinforcement learning.

References:

- [1] Y. Wang and M. Papageorgiou. Real-time freeway traffic state estimation based on extended kalman filter: A general approach. *Transportation Research Part B: Methodological*, 39(2):141–167, 2005.