Sensor Networks Routing via Bayesian Exploration

Shuang Hao
haos05@cs.ubc.ca

Ting Wang
guasars@cs.ubc.ca

Department of Computer Science, University of British Columbia,
Vancouver, B.C. V6T 1Z4

Abstract

There is increasing research interest in solving routing problems in sensor networks subject to constraints such as data correlation, link reliability and energy conservation. Since information concerning these constraints are unknown in an environment, a reinforcement learning approach is proposed to solve this problem. To this end, we deploy a Bayesian method to offer good balance between exploitation and exploration. It estimates the benefit of exploration by value of information therefore avoids the error-prone process of parameter tuning which usually requires human intervention. Experimental results have shown that this approach outperforms the widely-used Q-routing method.

1. Introduction

A wireless sensor network (WSN) [1] usually contains hundreds or thousands of sensing devices, which collect the environmental information and manage transmitting them to the user via other sensors. Unlike network with powerful nodes and stable links, e.g. LAN, sensor networks have various concerns which are unknown in advance. So a reinforcement learning is practical in the routing scenery. In the present work, we propose to use the model-based approach, Bayesian Exploration [3], to take account of all dynamic features and achieve good routing schemes through trial-and-error interactions with the environment.

2. Routing Schemes

A typical reinforcement learning problem is to control a *Markov decision process* (MDP), a tuple $< S, A, D, R >$, with finite state and action sets S, A, reward function R and transition dynamics D. Finding the optimal strategy is equivalent to maximize the q-value $Q^*(s, a)$, which represents the expected reward by starting from the state s, taking action a and following the optimal strategy after. The recursive solution of q-value is well defined by *Bellman equation*. In our approach, a state is that data pass through a sensor node and selecting an action means sending packets to a neighboring node.

Q-routing [2] directly updates estimates of q-values by Q-learning, which uses ϵ-greedy selection mechanism to address the trade-off between exploration and exploitation.

In Bayesian exploration [3], we focus on maintaining an updated model of D and R based on the observation history. If we assume conjugate prior, update of the model is simply counting. Because of the uncertainty about the environment, we treat the q-value as a random variable $q_{s,a}$. The finite number of MDPs are sampled from current distribution of the model to approximate the expectation of q-value, $\mathbb{E}(q_{s,a})$.

In order to find an efficient exploration policy (rather than the ϵ-greedy selection), we consider what can be gained by learning the true value $q^*_{s,a}$ of $q_{s,a}$. That is the value of perfect information (VPI) which represents the expected reward gain for taking an action. Thus in each state, our strategy is to choose the action that maximizes $\mathbb{E}(q_{s,a}) + \mathrm{VPI}(s, a)$ rather than merely $\mathbb{E}(q_{s,a})$. It shows a reasonable tradeoff between exploitation and exploration. Furthermore, the operator has no need to set the parameters as in Q-routing. This advantage is very important in real applications because the process of parameter setting may produce bad performance in an on-line system.

In order to calculate the immediate reward, each sensor node stores three types of information: (1) its Manhattan distance (L1 norm) to the final destination, $d(s)$; (2) its residual energy $e(s)$; (3) the number of previous aggregated transmissions at the node, $c(s)$.The immediate reward is a linear combination of the metrics mentioned below.

- Latency: When the packets are from s to s', the reward gains $d(s)-d(s')$. The routing prefer the shortest path.
- Data Correlation: The immediate reward coupled with data correlation is specified by $c(s)$. The more data aggregate, the more rewards are granted.
- Residual Energy: If the traversed sensor's energy $e(s)$ is lower than a threshold, a negative value is punished.

1-4244-0419-3/06/$20.00 ©2006 IEEE

- Link Reliability: If a transmission is failed, all the above mentioned immediate rewards are dismissed.

3. Experimental Results

We run the experiments on a $n \times n$ grid sensor network, where a sensor node can communicate with its 8 neighbors. At each *round*, a batch of data are generated in the data collectors. A round will end when either all packets reach the base or this round time-stamp goes over the threshold. At each *step*, the nodes holding the data will choose the successors and forward the packets.

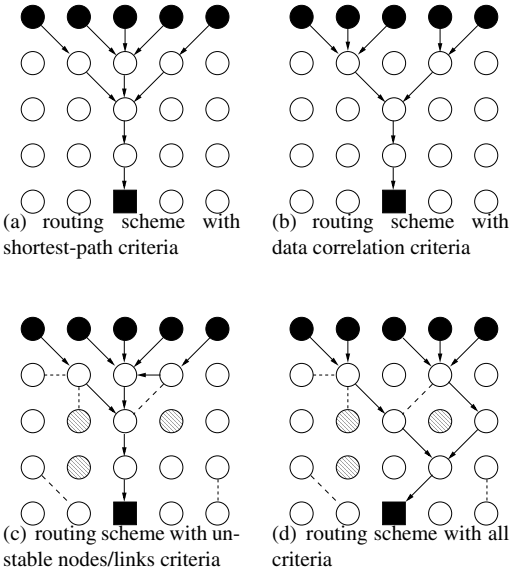

Figure 1. Routing schemes

To demonstrate the learnt routing scheme, we first run Bayesian exploration with different criteria in a 5×5 sensor network for 20 rounds. In Fig. 1, the shaded nodes are the data collectors that gather environment information and try to deliver them to the base represented by the square node; the striped nodes mean their energy is low; the dashed lines indicate the links are not reliable and the arrowed lines show the routing pathes.

1) If the only concern is to find the shortest path and the data correlation is small, the direct route to the base is optimal as shown in Fig. 1(a).

2) If the data correlation is high, the result routing pathes prefer to aggregate earlier as in Fig. 1(b).

3) When the data have very low correlation, but we care about unreliable links and low-energy nodes in the network, the routing scheme is in Fig. 1(c). It is based on the shortest path and keeps away from the malfunctioned nodes and links.

4) Finally, when we consider all metrics (more realistic case), the routing scheme is demonstrated in Fig. 1(d).

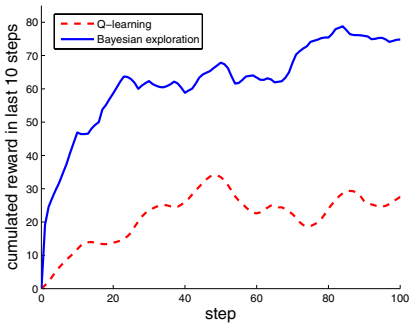

Figure 2. Cumulated reward by different reinforcement learning methods

Now we compare the performances of different reinforcement learning approaches on a more complex setting. The sensor network contains 9×9 nodes. The unstable links are randomly picked with probability 0.05 and the nodes are assigned low battery power with probability 0.15. We set Q-learning's learning rate $\alpha = 0.5$ and exploration rate $\epsilon = 0.3$ (it is noticed that different settings produce similar results) . The cumulated rewards in the last ten steps are shown in Fig. 2. Because Bayesian exploration makes a good tradeoff between exploitation and exploration, it gains more reward during the learning process.

4 Conclusion

In this paper, we try to take account of different features to optimize the routing scheme in a WSN. Users can adjust their preferences on those measures according to the real situation and their requirements. The Bayesian exploration framework smoothly integrates model extraction with prior knowledge and optimal exploration. We save the efforts of explicitly tuning the parameters and the performance of online routing scheme is greatly improved as a result.

References

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

[2] P. Beyens, M. Peeters, K Steenhaut and A. Nowe. "Routing with compression in wireless sensor networks: a Q-learning approach". In the Fourth International Joint Conference on Autonomous Agents and Multi Agent Systems, 2005.

[3] R. Dearden, N. Friedman, and D. Andre. "Model-based Bayesian exploration". In Proc. Fifteenth Conf. on Uncertainty in Artificial Intelligence, pp.150–159, 1999.