Social Behaviour Modeling and Optimization through Big Data and Reinforcement Learning

Alina Vereshchaka and Wen Dong

Department of Computer Science and Engineering,
State University of New York at Buffalo, Buffalo, USA
{avereshc, wendong}@buffalo.edu

1 Introduction

Solving complex real-world modeling and optimization problems presents several key challenges, including balancing the huge number of agents and system states, considering the complicated interactions between agents, and accounting for constant change in both the environment and the specification of optimality. Existing algorithms generally cannot cope with these challenges. In this abstract, we propose two approaches to this problem that utilize big data and the reinforcement learning framework. These approaches can be applied to complex social behaviour problems, such as allocating limited resources during natural disaster events and optimizing dynamics in critical infrastructure.

2 Background

In this section we will introduce the notation we will use for the standard optimal control or reinforcement learning formulation. We consider innite-horizon partially observable Markov decision process (POMDP), defined by the tuple $\langle S, A, O, P, r, \rho_0, \gamma \rangle$, where $s \in S$ denotes states, describing the possible configurations of all agents; $a \in A$ denotes actions, which can be discrete or continuous; $P : S \times A \times S \to \mathbb{R}$ is the states transition probability distribution, where states evolve according to the stochastic dynamics $p(s_{t+1}|s_t, a_t)$, which are in general unknown; O is a set of observations for each agents; $r : S \to \mathbb{R}$ is the reward function; $\rho_0 : S \to [0, 1]$ is the distribution of the initial state s_0; $\gamma \in [0, 1]$ is a discount factor. For our research we make assumptions about the stochastic action choices, in which actions may be chosen with any probabilities to facilitate more robust prediction and planning. To choose an action, each agent uses a stochastic policy $\pi_\theta : O \times A \to [0, 1]$, which produces the next state according to the states transition probability. Each agent obtains rewards as a function of the state and agents action $r : S \times A \to \mathbb{R}$, and receives a private observation correlated with the state $\mathbf{o} : S \to O$.

Markov property:

$$\mathbb{P}[S_{t+1}|S_1, S_2, \ldots, S_t] = \mathbb{P}[S_{t+1}|S_t]$$

Solving a MDP means finding a policy π^* that maximizes the expected long-term reward $R = \sum_{t=0}^{T} \gamma^t r^t$, where T is the time horizon.

3 Social Behaviour Modeling during Natural Disasters

Natural disasters are important and devastating events for a country, which makes it difficult to make informed decisions in terms of allocating limited resources for mitigation efforts. According to the United Nations Office for Disaster Risk Reduction natural disasters are happening more frequently, which has caused a rise of 151% over the last twenty years in economic losses from climate-related disasters. During 1998-2017, disaster-hit countries reported direct economic losses from natural disasters of US$2,245 billion and these hazards are responsible for 77% of the total losses in these countries. The USA has the greatest economic losses among other countries of US$ 944.8 billion. Over the last twenty years, disasters claimed more than 1.3 million lives, and more than 4 billion people were injured, rendered homeless or in need of assistance.

We introduce the novel approach of reinforcement learning framework to model optimal resource allocation for natural disasters. Our hybrid model is a combination of social behavioural modeling and deep reinforcement learning. We propose a multi-agent real-time resource allocation framework to respond to the disaster scenarios. The model can be used by experts, decision makers, disaster managers and emergency personnel to assess the critical event and respond appropriately in order to mitigate the disaster effects in real-time. This framework can be applicable to various scenarios in response to the natural hazards, like allocating firetrucks in case of wildfire, allocating snow plows during the winter storms or allocating rescue crews during flooding. [1]

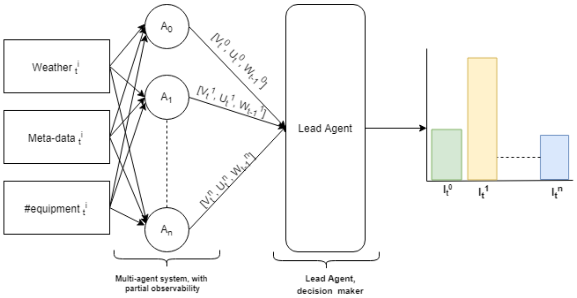
Weather $_t^i$
Meta-data $_t^i$
#equipment $_t^i$
A_0
A_1
A_n
$[V_t^0, U_t^0, W_{t-1}^0]$
$[V_t^1, U_t^1, W_{t-1}^1]$
$[V_t^n, U_t^n, W_{t-1}^n]$
Lead Agent
I_t^0
I_t^1
I_t^n
Multi-agent system, with partial observability
Lead Agent, decision maker

Fig. 1: An overview of our framework. Left: a view of the module with a bundle of independent agents. Right: a lead agent that makes an allocation decision based on the data broadcasted by each of the subagents.

As shown on Fig. 1 our framework consists of two parts. On the lower level (left side) there is a set of multi-agents $(A_0, A_1, ..., A_n)$ that corresponds to the number of sub-regions. Each agent predicts the volume of hazards and the importance, taking as input an observation from the time-series environment. The total number of agents is bounded by the number of sub-regions. On the higher level (right side) there is a lead agent, which is trained to make resource allocation decisions for assigning volume distribution of involved resources.

4 Social Behaviour Modeling in Critical Infrastructure

We formulate a stochastic kinetic process to model the diverse dynamics of agent interactions and decision-making in a complex network as a sequence of atomic events that individually induce minimal changes to the network but together show diverse behavior. With this formulation, learning and control in a network involve learning to set how fast these interaction events happen in response to the noisy signals from within and outside the network. Such networked and event-based control often appears in networked social, biological, and engineered systems, which need to solve different optimization problems at different times.

To solve the learning and control problems of a stochastic kinetic model, we reduce these problems to parameter-learning and inference problems in a mixture of dynamic Bayesian networks. With this reduction we can bring in many existing parameter-learning and statistical inference techniques for optimizing the interactions of a networked system based on partial observations about the complex environment, and integrate signal processing and decision-making into a holistic framework. Specifically, we develop a particle filter algorithm to model how the networked system continually tracks the current state of itself and the environment using noisy observation streams, and to learn how to make near-optimal plans by adjusting the rates at which interaction events happen (Fig. 2). In this sense, our learning algorithm works through policy search that maximizes the expected log-likelihood over agent policies in a mixture of dynamic Bayesian networks. [2].

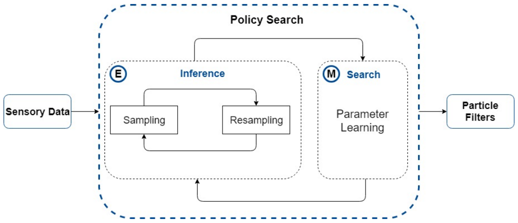

Fig. 2: Our policy search algorithm is done in two parts. During the Inference we estimate the state density and reward associated with each particle. During the Search step we update policy parameters according to the Inference results.

References

1. A. Vereshchaka, W. Dong, Dynamic resource allocation during natural disasters using multi-agent environment, in: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, Springer, 2019.
2. F. Yang, A. Vereshchaka, W. Dong, Predicting and optimizing city-scale road traffic dynamics using trajectories of individual vehicles, in: 2018 IEEE International Conference on Big Data (Big Data), IEEE, 2018, pp. 173–180.