Decentralized Decision support for an agent population in dynamic and uncertain domains

(Extended Abstract)

Pradeep Varakantham, Shih-Fen Cheng, Nguyen Thi Duon
School of Information Systems,
Singapore Management University,
{pradeepv,sfcheng,tdnguyen}@smu.edu.sg

ABSTRACT

This research is motivated by problems in urban transportation and labor mobility, where the agent flow is dynamic, non-deterministic and on a large scale. In such domains, even though the individual agents do not have an identity of their own and do not explicitly impact other agents, they have implicit interactions with other agents. While there has been much research in handling such implicit effects, it has primarily assumed controlled movements of agents in static environments. We address the issue of decision support for individual agents having involuntary movements in dynamic environments . For instance, in a taxi fleet serving a city: (i) Movements of a taxi are uncontrolled when it is hired by a customer. (ii) Depending on movements of other taxis in the fleet, the environment and hence the movement model for the current taxi changes. Towards addressing this problem, we make three key contributions: (a) A framework to represent the decision problem for individuals in a dynamic population, where there is uncertainty in movements; (b) A novel heuristic technique called Iterative Sampled OPtimization (ISOP) and greedy heuristics to solve large scale problems in domains of interest; and (c) Analyze the solutions provided by our techniques on problems inspired from a real world data set of a leading taxi company in Singapore. As shown in the experimental results, our techniques are able to provide strategies that outperform "driver" strategies with respect to: (i) overall availability of taxis; and (ii) the revenue obtained by the taxi drivers.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed AI

General Terms

Algorithms; Experimentation

Keywords

Multi-agent decision making, Uncertainty

1. INTRODUCTION

Research on understanding and controlling dynamic and large scale flow of agents (e.g., humans, industries, vehicles) between different states spans various domains such as urban transportation [2, 4] (e.g., movement of vehicles between different regions of an area), industry dynamics [3] (e.g., strategizing on marketing investments by different companies selling the same product), labor mobility between cities [1] (e.g., analyzing individuals search for jobs in new locations), advertising and others. The main challenge in these problems is accounting for the implicit interaction that exists between agents. For example, vehicles trying to get on the same road are implicitly competing. Existing literature has primarily focused on understanding behaviors and improving operational efficiency while accounting for the implicit interactions under the assumption that the agent movement is voluntary.

We are focused on similar problems, except in cases where there is involuntary (or forced) movement of agents. The first problem of interest is with respect to the operation of a taxi fleet. Taxi drivers are subject to both voluntary (at driver's own decision) and involuntary (when customers board taxis) movements. Different regions might have different demands for taxis (both in terms of numbers and revenues) and due to this an implicit competition exists between taxis. The goal here is to improve the operational efficiency of the fleet while improving the revenues obtained by taxi drivers. Secondly, in understanding labor mobility, which is governed by voluntary (quitting jobs and moving to other geographic location) and involuntary (getting laid off) movements. Different geographical regions might have different compensation levels, and individuals might need to invest beforehand in order to move from one region to another. Since the distribution of unemployed labor determines the chance of getting a job in a region, there is again implicit competition between individuals. Similarly, there are problems in analyzing industry dynamics, where different companies strategize to maintain their competitive advantage.

We were able to illustrate that our approach, ISOP and one of the greedy approaches provide solutions that improved significantly over real world taxi driver policies. This improvement was with respect to both the (a) operational inefficiency, characterized as congestion in our results; and (b) the minimum revenue obtained by any taxi driver and the average revenue of all the taxi drivers. These results emphasize the utility of our sampled optimization and greedy techniques in solving DDAP problems.

Cite as: (Extended Abstract), , *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. XXX-XXX.

2. MODEL

In this section, we describe the Decentralized decision model for Dynamic Agent Populations or DDAP. DDAP is a model to represent the decentralized decision problem for individual agents in a population operating in dynamic domains. It is represented using the tuple:
$\langle \mathcal{P}, \mathcal{S}, \mathcal{A}, \phi, \mathcal{R}i, \mathcal{R}p, H, D^0 \rangle$, where $\mathcal{P}$ represents the agent population. $\mathcal{S}$ corresponds to the set of states encountered by every agent in the population. $\mathcal{A}$ is the set of actions executed by each agent. ϕ represents the transition probability between agent states given the population distribution. $\mathcal{R}i^t(s, a, d)$ is the reward obtained by an agent due to its action alone, when in state s, taking action a and the state distribution is d at time t. $\mathcal{R}p^t(s, a, d)$ is the reward obtained due to implicit interaction with other agents in the population, when the state distribution is d at time t.

H is the time horizon for the decision process, with the underlying assumption that the distribution of agent states is available after every H time steps. D^0 represents the set of possible starting distributions. The objective is to compute a policy which maximizes social welfare without sacrificing agent interests.[1]

3. SOLVING A DDAP

3.1 ISOP

Algorithm 1 SolveDDAP()

$\pi_i \leftarrow \phi$
$\pi_{-i} \leftarrow$ InitializePolicy()
while *true* **do**
  $\pi_i \leftarrow Br(\pi_{-i})$
  if $\pi_i = \pi_{-i}$ **then**
    break while
  $\pi_{-i} \leftarrow \pi_i$
return π_i

We now introduce Iterative Sampled Optimization (ISOP), an approximate approach that scales to large DDAP problems. The overall idea of solving a DDAP is characterized by Algorithm 1. It provides two approximations to address the issues mentioned at the end of the previous section.

Firstly, we approximate the value function by making assumptions on the transition between distributions and the set of distributions. The set of distributions is obtained by sampling from the set of reachable distributions. The expression for the updated value function is as follows:

$$\mathcal{V}^t_{\pi_i,\pi_{-i}}(s, d) = \sum_{a \in \mathcal{A}} [\mathcal{R}p^t(s, a, d) + \pi^t_i(s, a) \cdot \{\mathcal{R}i^t(s, a, d) + \sum_{s'} \phi^t_d(s, a, s') \mathcal{V}^{t+1}_{\pi_i,\pi_{-i}}(s')\}] \quad (1)$$

$$\mathcal{V}^{t+1}_{\pi_i,\pi_{-i}}(s') = \sum_{d'} Pr^t(d'|D^0, \pi_i, \pi_{-i}) \mathcal{V}^{t+1}_{\pi_i,\pi_{-i}}(s', d') = \frac{\sum_{d' \in \tilde{D}} \mathcal{V}^{t+1}_{\pi_i,\pi_{-i}}(s', d')}{|\tilde{D}|} \quad (2)$$

Secondly, we approximate with respect to Algorithm 1. Algorithm 1 performs best response computation over the policy for the entire horizon at each iteration of the algorithm. We propose an approximation method inspired from best response computation in sequential games, where best response is computed for each decision epoch separately while backing up the value function. Instead of iterating until convergence, ISOP algorithm iterates until the time horizon and solves a linear optimization problem for computing one step best response at each iteration.

3.2 Greedy Approaches

The key approximation in greedy approaches is the assumption that no other agent is present in the environment, i.e. $D = \{d | d = \langle 0, 0, \cdots, 0 \rangle\}$. By substituting zero vector for d, we obtain the updated values for $\mathcal{R}p^t(s, a, d)$ and $\mathcal{R}i^t(s, a, d)$. These updated values of rewards are used to obtain greedy policies based on the parameter, g. When $g = 1$, the policy obtained is deterministic. When $g = 2$, the policy obtained is randomized over two actions for all the states and so on.

4. EXPERIMENTAL RESULTS

We compared the performance of ISOP and the suite of greedy approaches on a real world taxi data set of a cab company in Singapore. In the taxi domain, we were able to show (from a month of actual taxi data) that the taxi drivers adopt greedy policies, randomly choosing between the zones with the highest overall rewards $(\mathcal{R}i^t(s, a, \mathbf{0}) + \mathcal{R}p^t(s, a, \mathbf{0}))$ during that time step. The key evaluation metrics are: (a) The minimum revenue obtained by any taxi during the time horizon; (b) The average revenue obtained by all taxis; and (c) Overall congestion, which is the sum of the excess taxis and excess flow in all the zones. On each problem, values for these evaluation metrics are obtained by simulating the output policies of each of the approach on the customer flow model and revenues. We were able to show that when the number of zones is less than or equal to 20, ISOP is able to outperform the greedy approaches. However, as the number of zones increases above 20 the performance of ISOP degrades. We believe that this is due to the algorithm employing for obtaining the set of distributions.

5. REFERENCES

[1] J. R. Harris and M. P. Todaro. Migration, unemployment and development: A two-sector analysis. *The American Economic Review*, 60(1):126–142, 1970.

[2] J. G. Wardrop. Some theoretical aspects of road traffic research. In *Proceedings of Institute of Civil Engineers, Part II*, volume 1, pages 325–378, 1952.

[3] G. Y. Weintraub, L. Benkard, and B. V. Roy. Oblivious equilibrium: A mean field approximation for large-scale dynamic games. In *Neural Information Processing Systems (NIPS)*, 2006.

[4] H. Yang and S. C. Wong. A network model of urban taxi services. *Transportation Research Part B: Methodological*, 32(4):235–246, 1998.

[1]This optimization criterion can mean different things for different domains. In the taxi problem, this refers to minimizing starvation of taxis in all zones and maximizing revenue for taxi drivers.