Constrained Multiagent MDPs: A Taxonomy of Problems and Algorithms

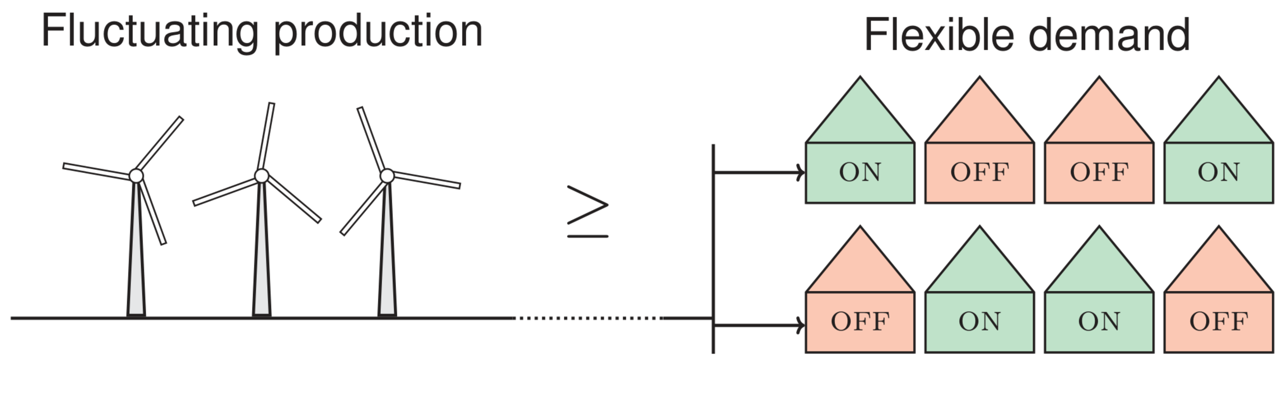
Frits de Nijs, Erwin Walraven, Mathijs de Weerdt, Matthijs Spaan

Abstract

In domains such as electric vehicle charging, smart distribution grids and autonomous warehouses, multiple agents share the same resources. When planning the use of these resources, agents need to deal with the uncertainty in these domains. Although several models and algorithms for such constrained multiagent planning problems under uncertainty have been proposed in the literature, it remains unclear when which algorithm can be applied.

In this survey we conceptualize these domains and establish a generic problem class based on Markov decision processes. We identify and compare the conditions under which algorithms from the planning literature for problems in this class can be applied: whether constraints are soft or hard, whether agents are continuously connected, whether the domain is fully observable, whether a constraint is momentarily (instantaneous) or on a budget, and whether the constraint is on a single resource or on multiple. Further we discuss the advantages and disadvantages of these algorithms. We conclude by identifying open problems that are directly related to the conceptualized domains, as well as in adjacent research areas.

1. Constrained Multiagent MDP Example



Example setting

- Neighborhood of houses with flexible loads (electric vehicle, washing machine, heating)
- Powered by a renewable source like wind farm.
- Goal:
  Maximize collective comfort within production.
- Under uncertainty:
  power production, occupancy, temperature.

Single-agent MDP model

States $s =$ occupancy, temperature, wind speed
Actions $a =$ which loads to activate
$T(s, a, s') =$ given state transition model
Reward $R(s, a) =$ comfort level
Usage $C_t(s, a) =$ load of action a at time t

Constrained Multiagent MDP model

Joint state	$S = \langle s_1, s_2, \ldots, s_n \rangle$
Joint action	$A = \langle a_1, a_2, \ldots, a_n \rangle$
Transition	$T(\mathbf{s}, \mathbf{a}, \mathbf{s}') = \prod_i T_i(s_i, a_i, s_i')$
Rewards	$R(\mathbf{s}, \mathbf{a}) = \sum_i R_i(s_i, a_i)$
Consumption	$C_j(\mathbf{s}, \mathbf{a}) = \sum_i C_j(s_i, a_i)$
Shared limits	$\sum_t C_j(\mathbf{s}, \mathbf{a}) \leq L_j, \; \forall j$

Goal: optimal policy π max. $\mathbb{E}[\sum R(s, \pi(s))]$,
s.t. $\sum_t C_j(s, \pi(s)) \leq L_j$

The problem: Constrained multiagent MDP couples agent MDPs through shared limits: exponential $|S|$, $|A|$.

2. The taxonomy

Communication

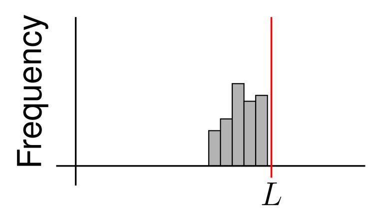
- *Disconnected* problems: unreliable or no communication,
  - Centralized planning, decentralized execution
- *Connected* problems: communication possible,
  - Centralized planning and on-line re-planning or adjustment of policy

State observability

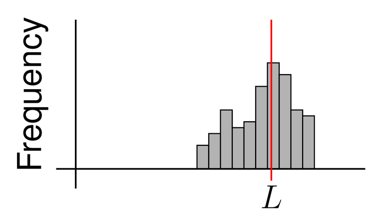
- *Fully observable*, observation o perfectly identifies state s
  - Condition policy $\pi(s) = a$ on current state
- *Partially observable* problems, observation function $O(s, a, o) = \Pr(o \mid s, a)$
  - Condition policy $\pi(b(s)) = a$ on belief $b(s) = \Pr(s)$ over current state
  - Worst-case resource use $\max_{s, b(s) > 0} C(s, a)$

Constraint strictness

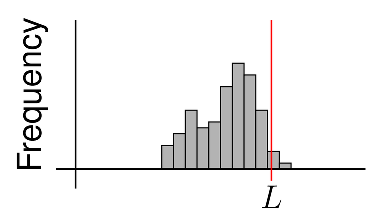
- *Hard constraints*, no reachable outcome may exceed limits, $C_j \leq L_j$



- *Soft constraints*, policy may not exceed limits in expectation, $\mathbb{E}[C_j] \leq L_j$



- *Bounded constraints*, fewer than ϵ constraint violations, $\Pr(C_j > L_j) \leq \epsilon$



Constraint time span

- *Budget constraints*, cumulative consumption over time is bounded, $\sum_t C_j \leq L_j$
- *Instant constraints*, every time step, consumption is limited, $C_t \leq L_t, \forall t$

3. Overview of Algorithms

Reference	Conn.	Strictness	PO	Timespan			Impl.*
				Budget	Inst.	Multi	
Altman, 1999		soft		✓	✓	m	✓
Yost and Washburn, 2000		soft	✓	✓	✓	m	✓
Isom et al., 2008		soft	✓	✓		1	
Kim et al., 2011		soft	✓	✓		m	
Poupart et al., 2015		soft	✓	✓		m	✓
Walraven and Spaan, 2018		soft	✓	✓	✓	m	✓
Dolgov and Durfee, 2003		soft w. bound		✓		m	
De Nijs et al., 2017		soft w. bound		✓	✓	m	✓
Wu and Durfee, 2010		hard		✓	✓	m	✓
Agrawal et al., 2016		hard		✓		m	
Meuleau et al., 1998	✓	hard		✓	✓	m	
Boutilier and Lu, 2016	✓	soft		✓		1	
De Nijs et al., 2015	✓	hard			✓	h	
Undurti and How, 2010	✓	hard	✓	✓		1	
Lee et al., 2018	✓	soft	✓	✓		m	✓

*Open source implementation of algorithm available as part of our Constrained Planning Toolbox. QR code links to the GitHub repository:

`https://github.com/AlgTUDelft/ConstrainedPlanningToolbox`

4. Observations

- Significant body of work on Constrained Multi-agent MDPs,
- Hard constraints received limited attention,
  - PSPACE-hard, requires approximation
  - Soft constraints give poly-time relaxation (but not safe)
- Hard constraints especially problematic when considering Partial Observability

5. Extensions and Challenges

- **Safe RL**: can successful solution methods like preallocations be employed in the learning setting?
- **Partial observability with hard constraints**: gap in existing work, under what conditions can we solve?
- **Partial communication**: communication models all-or-nothing; how to use intermittently available comms?
- **Agent interaction**: how can we relax the core assumption of independent agents?

MONASH University

TU Delft