Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence

A Market-Based Coordination Mechanism for Resource Planning Under Uncertainty

Hadi Hosseini and **Jesse Hoey** and **Robin Cohen**
David R. Cheriton School of Computer Science
University of Waterloo
200 University Avenue West
Waterloo, ON N2L 3G1
{h5hossei,jhoey,rcohen}@uwaterloo.ca

Introduction

Multiagent Resource Allocation (MARA) distributes a set of resources among a set of intelligent agents in order to respect the preferences of the agents and to maximize some measure of global utility, which may include minimizing total costs or maximizing total return. We are interested in MARA solutions that provide optimal or close-to-optimal allocation of resources in terms of maximizing a global welfare function with low communication and computation cost, with respect to the priority of agents, and temporal dependencies between resources. We propose an MDP approach for resource planning in multiagent environments. Our approach formulates internal preference modeling and success of each individual agent as a single MDP and then to optimize global utility, we apply a market-based solution to coordinate these decentralized MDPs.

Background

Resource planning under uncertainty in multiagent environments scales exponentially with the number of agents and resources. Therefore, in the presence of multiple agents with various possible actions and state, the size of a Markov model increases exponentially. Consequently, solving large MDPs in multiagent systems becomes NEXP-complete and computationally expensive, and in fact, MMDP and Dec-MDP approaches become infeasible (Bernstein et al. 2002) when dealing with a large number of agents and resources. One rational way of breaking down a multiagent system is to consider each agent as a single MDP, responsible for maintaining its local state and actions. This will decrease the number of states dramatically. If A is a set of permissible actions, and S is a set of possible states for each agent (assuming all patients have the same set of actions and states), and N is the set of agents in a multiagent setting, the number of state-action pair using a large MDP is equal to $|S|^{|N|}|A|^{|N|}$ while Dec-MDP decreases this number to $|S||N||A|$ in the best case, i.e., fully independent Dec-MDPs. Decentralized MDPs make the computation easier; however, there is still a need for global coordination of the local MDPs within the system in order to find an overall optimal policy. Coordinating many agents can add communication surplus to the system, and hence, it compensates the reduced computation costs. Therefore, there is always a tradeoff between communication level, interdependence, and finding optimal policy (Capitán et al. 2011).

We propose an auction-based mechanism for coordinating the local MDPs and maximizing the global utility function. Our approach consists of one part of planning and higher level of coordination between agents. In fact, coordination can be sought of as part of the planning environment, granting permissions to agents to perform their optimal actions, only if it maximizes the global social welfare function.

MDP representation of agents

Our model is a factored MDP represented as a set of variables and functions $\langle \mathbf{R}, \psi, P_T, \Phi, A \rangle$ where $\mathbf{R}$ is a finite set of resource variables, each of which is $R \in \{n, v, d\}$ representing that the agent $n = needs$, $v = has$, or has $d = had$ the resource in question. ψ is a variable measuring agent success (e.g. patient health when resources are doctors/equipment). We use $S = \{\mathbf{R}, \psi\}$ to denote the complete set of **state variables**. $\Phi(S, S')$ is a **reward function**, P_T is a **transition model** that gives the probability of reaching state S' after having health state H and resources $\mathbf{R}$, and A is a set of **actions**. There is one action for each resource, plus a null action. The resource actions represent bids for the corresponding resource, as detailed below.

The policy can be obtained in a number of ways, including by computing a value function $V^*(s)$ for each state $s \in S$, that is maximal for each state (i.e. that satisfies the Bellman equation (Bellman 2003)).

$$V^*(s) = \underset{a}{max}\, \gamma \sum_{s' \in S} [\Phi(s, s') + P_T(s'|s, a)V^*(s')] \quad (1)$$

The policy is then given by the actions at each state that are the arguments of the maximization in Equation 1.

The agent success (ψ) is conditionally independent of the agent action (e.g. bid) given the current resources and the previous success. Second, the agent actions only influence the resource allocation, since the agent can only influence success indirectly by bidding for resources. Third, the reward function is only dependent on the agent success, ψ. Thus, $P(\mathbf{r}', \psi' | \mathbf{r}, \psi, a) = P(\mathbf{r}' | \mathbf{r}, \psi, a) P(\psi' | \mathbf{r}, \psi)$, where we define $\Lambda_R \equiv P(\mathbf{r}' | \mathbf{r}, \psi, a)$ is the probability of getting the next set of resources given the current success, resources, and action, and $\Omega_\psi \equiv P(\psi'|\mathbf{r}, \psi)$ is a dynamic model for the agent's success rate. We will refer to Λ_R as the *resource obtention* model and to Ω_ψ as the *succession* model.

Bidding Mechanism

The bidding mechanism is based on the expected *regret* of not obtaining a given resource. The expected value, $Q_i(\psi, \mathbf{r}, a_i)$ for being in success state ψ with resources $\mathbf{r}$ at time t, bidding for (denoted a_i) and receiving resource r_i at time $t + 1$ is: $Q_i \equiv \sum_{\mathbf{r}'_{-i}} \sum_{\psi'} P(\psi'|\psi, \mathbf{r}) V_i(r'_i, \mathbf{r}'_{-i}, \psi') \delta(\mathbf{r}_{-i}, \mathbf{r}'_{-i})$, where $\mathbf{r_{-i}}$ is the set of all resources except r_i and $\delta(x, y) = 1 \leftrightarrow x = y$ and 0 otherwise. Similarly this is calculated for value for not receiving the resource, $\bar{Q}_i(\psi, \mathbf{r}, a_i)$, Thus, the expected regret for not receiving resource r_i when in success state ψ with resources $\mathbf{r}$ and taking action a_i is:

$$R_i(\psi, \mathbf{r}, a_i) = Q_i - \bar{Q}_i \qquad (2)$$

Note that Q is an optimistic estimate, since the expected value assumes the optimal policy can be followed after a single time step (which is untrue). This myopic approximation enables us to compute on-line allocations of resources in the complete multiagent problem, as described later in the paper.

Market-Based Coordination of MDPs

We propose a simple auction-based mechanism where agents send their current estimates of regret to resources that allocate their timeslots in an iterative auction (Algorithm 1).

Algorithm 1: Resource agents

```
Input: Set of resources $R$, set of bids
Output: Mapping of agents to timeslots
initialization;
foreach Timeslot $t$ do
    resource: open up auction for $t$;
    $Bid_t \leftarrow$ receive($bid_j$) ;        // bid from agent $j$
    $j_t = \arg\max_{j \in N}\{bid_j\}$ ;        // awarding phase
    alloc($A_j$, $t$);
```

The auction proceeds in an iterative fashion, and each consumer agent bids on the resource with highest regret first (Algorithm 2). The highest bidder is considered to have a higher expected regret for not getting a resource. If an agent does not win the auction for its highest regret resource, it waits until the auction for its second highest regret resource becomes available. It can also decide to give up, and is then resigned to not having any resources in the next time step (which may be a better option than taking a resource ahead of time). While this iterative auction mechanism is not optimal, it is demonstrated to work well in simulated environments.

Validation and Future Direction

To validate this approach, we have adopted a healthcare resource allocation problem. Since the medical procedures and uncertainty in patient behaviors require stochastic analysis, the healthcare domain is a prime candidate for our analysis and evaluation purposes. We have shown, in simulation, that our approach provides an effective allocation of resources to the agents by increasing the overall social welfare (health state of the patients) while coping with stochastic, temporal, and dynamic elements (beyond models like (Paulussen et al. 2003)).

Algorithm 2: Consumer agents: bidding mechanism

```
Input: A condition profile including a set of needed resources
Output: Bid values, schedule
// Initialization
begin
    $\Lambda_R \sim Dir(\boldsymbol{\alpha_r})$;        // resource obtention
    $\Omega_\psi \sim Dir(\boldsymbol{\alpha}_\psi)$;          // succession model
    Solve MDP;
while $\mathbf{r}$ is nonempty do
    forall the $r_i \in \mathbf{r}$, $a$, $\psi$ do
        $R_i(\psi, \mathbf{r}, a_i) = Q_i - \bar{Q}_i$
    foreach Timeslot $t$ do
        $\mathbf{r}^t \leftarrow \mathbf{r}$;
        while $\mathbf{r}^t$ is nonempty $\wedge$ $schedule^t_j$ is empty do
            $i \leftarrow \arg\max_{i \in \mathbf{r}}\{R_i\}$;
            submit $bid^i_j$ to resource $i$;
            if $j$ is winner then
                $update(r_i, t)$;
                remove $r_i$ from $\mathbf{r}$;
            else remove $r_i$ from $\mathbf{r}^t$
```

We are planning to illustrate the effectiveness of our approach from a theoretical point of view by proposing mathematical proofs for the lower and upper bounds of our approximate solution. In future work, we would like to refine our model to incorporate more sophisticated bidding languages (Nisan 2000) to better cope with combinatorial problems where agents need to expose a richer preference model in order to optimize the allocation process.

References

Bellman, R. 2003. *Dynamic programming*. Courier Dover Publications.

Bernstein, D.; Givan, R.; Immerman, N.; and Zilberstein, S. 2002. The complexity of decentralized control of Markov decision processes. *Mathematics of operations research* 27(4):819–840.

Capitán, J.; Spaan, M.; Merino, L.; and Ollero, A. 2011. Decentralized multi-robot cooperation with auctioned pomdps. In *Sixth Annual Workshop on Multiagent Sequential Decision Making in Uncertain Domains (MSDM-2011)*, 24.

Nisan, N. 2000. Bidding and allocation in combinatorial auctions. In *Proceedings of the 2nd ACM conference on Electronic commerce*, 1–12. ACM.

Paulussen, T.; Jennings, N.; Decker, K.; and Heinzl, A. 2003. Distributed patient scheduling in hospitals. In *International Joint Conference on Artificial Intelligence*, volume 18, 1224–1232. Citeseer.