Proactive Distributed Constraint Optimization Problems

Doctoral Mentoring Program

Khoi Hoang
Washington University in St. Louis
St. Louis, MO
khoi.hoang@wustl.edu

ABSTRACT
Current approaches that model dynamism in DCOPs solve a sequence of static problems, reacting to changes in the environment as the agents observe them. Such approaches thus ignore possible predictions on future changes. To overcome this limitation, we introduce (finite-horizon) Proactive Dynamic DCOPs (PD-DCOPs) and Infinite-Horizon PD-DCOPs (IPD-DCOPs) to model dynamic DCOPs in the presence of exogenous uncertainty. In contrast to reactive approaches, PD-DCOPs and IPD-DCOPs are able to explicitly model the possible changes to the problem, and take such information into account proactively, when solving the dynamically changing problem. The additional expressivity of these formalisms allows them to model a wider variety of distributed optimization problems. Our work presents both theoretical and practical contributions that advance current dynamic DCOP models.

KEYWORDS
Distributed Problem Solving; Distributed Constraint Optimization; DCOP

2 PROPOSED MODELS
2.1 PD-DCOPs
A Proactive Dynamic DCOP (PD-DCOP) is a tuple \((A, X, D, F, h, T, c, y, p, \alpha, \beta)\), where:
- \(A = \{a_i\}_{i=1}^{p}\) is a set of agents.
- \(X = \{x_i\}_{i=1}^{n}\) is a mixed set of decision and random variables. To differentiate between decision variables and random variables, we use \(Y \subseteq X\) to denote the set of random variables that model uncontrollable stochastic events.
- \(D = \{D_x\}_{x \in X}\) is a set of finite domains. Each variable \(x \in X\) takes values from the set \(D_x \in D\). We also use \(\Omega = \{\Omega_{y}\}_{y \in Y} \subseteq D\) to denote the set of event spaces for the random variables such that each \(y \in Y\) takes values in \(\Omega_y\).
- \(F = \{f_i\}_{i=1}^{m}\) is a set of reward functions, each defined over a mixed set of decision variables and random variables: \(f_i : \prod_{x \in x^i} D_x \rightarrow \mathbb{R}^+ \cup \{\bot\}\), where \(x^i \subseteq X\) is scope of \(f_i\) and \(\bot\) is a special element used to denote that a given combination of values for the variables in \(x^i\) is not allowed.
- \(h \in \mathbb{N}\) is a finite horizon in which the agents can change the values of their variables.
- \(T = (T_y)_{y \in Y}\) is the set of transition functions \(T_y : \Omega_y \times \Omega_y \rightarrow [0, 1] \subseteq \mathbb{R}\) for the random variables \(y \in Y\), describing the probability for a random variable to change its value in successive time steps. For a time step \(t > 0\), and values \(\omega_i \in \Omega_y, \omega_j \in \Omega_y\), \(T_y(\omega_i, \omega_j) = P(y^t = \omega_j | y^{t-1} = \omega_i)\), where \(y^t\) denotes the value of the variable \(y\) at time step \(t\), and \(P\) is a probability measure. Thus, \(T_y(\omega_i, \omega_j)\) describes the probability for the random variable \(y\) to change its value from \(\omega_i\) at a time step \(t - 1\) to \(\omega_j\) at a time step \(t\). Finally, \(\sum_{\omega_j \in \Omega_y} T_y(\omega_i, \omega_j) = 1\) for all \(\omega_i \in \Omega_y\).
\[ \mathbf{c} \in \mathbb{R}^+ \text{ is a switching cost, which is the cost associated with the change in the value of a decision variable between time steps.} \]

\[ \mathbf{y} \in [0, 1) \text{ is a discount factor, which represents the decrease in the importance of rewards/costs over time.} \]

\[ P^0_Y = \{p^0_y \mid y \in Y \} \text{ is a set of initial probability distributions for the random variables } y \in Y. \]

\[ \alpha : X_Y \rightarrow A \text{ is a function that associates each decision variable to one agent. We assume that the random variables are not under the control of the agents and are independent of decision variables.} \]

The goal of a PD-DCOP is to find a sequence of \( h + 1 \) assignments \( x^* \) for all the decision variables in \( X \setminus Y \):

\[
\mathbf{x}^* = \underset{x=(x^1,\ldots,x^h)}{\text{argmax}} \mathcal{F}^h(x) \tag{1}
\]

\[
\mathcal{F}^h(x) = \sum_{t=0}^{h-1} \mathcal{F}_x^t(x^t) + \mathcal{F}_y^t(x^t) - \sum_{t=0}^{h-1} \mathcal{F}_x^t(x^t) \mathcal{F}_y^t(x^t) + \mathcal{F}_x(x^h) - \mathcal{F}_y(x^h) \tag{2}
\]

\[
\mathcal{F}_x(x) = \sum_{i \in [1, \ldots, n]} c \cdot (x^i, x^{i+1}) \tag{3}
\]

\[
\mathcal{F}_y(x) = \sum_{j \in [1, \ldots, m]} \mathcal{F}_y^t(x^j) \tag{4}
\]

where \( \Sigma \) is the assignment space for the decision variables of the PD-DCOP, at each time step. These assignments to values of decision variables maximizes the sum of two terms. The first term maximizes the discounted net utility, that is, the discounted rewards for the functions that do not involve exogenous factors (\( \mathcal{F}_x \)) and the expected discounted random rewards (\( \mathcal{F}_y \)) minus the discounted penalties over the first \( h \) time steps. The second term maximizes the discounted future rewards for the problem.

### 2.2 IPD-DCOPs

At a high level, the Infinite-Horizon Proactive Dynamic DCOP (IPD-DCOP) model is a straightforward, but essential and significant, extension of the PD-DCOP model. Both IPD-DCOP and PD-DCOP models assume that a finite horizon \( h \) is given. However, PD-DCOPs ignore all changes to the problem after its finite horizon \( h \) and, thus, do not optimize for them, while IPD-DCOPs don’t take into account the discount factor and assume that the Markov chains will converge to stationary distributions after that horizon and an optimal solution for those stationary distributions should be adopted after the horizon.

As agents will keep their optimal solution from horizon \( h \) onwards, the goal of an IPD-DCOP is to find a sequence of \( h + 1 \) assignments \( \mathbf{x}^* \) for all the decision variables in \( X \):

\[
\mathbf{x}^* = \underset{x=(x^1,\ldots,x^h)}{\text{argmax}} \mathcal{F}(\mathbf{x}) \tag{5}
\]

\[
\mathcal{F}(\mathbf{x}) = \sum_{t=0}^{h-1} \mathcal{F}_x^t(x^t) - \sum_{t=0}^{h-1} \mathcal{F}_x^t(x^t) \mathcal{F}_y^t(x^t) + \mathcal{F}_x(x^h) - \mathcal{F}_y(x^h) \tag{6}
\]

where \( \Sigma \) is the assignment space for the decision variables of the IPD-DCOP at each time step. \( \Delta : \Sigma \times \Sigma \rightarrow \{0\} \cup \mathbb{N} \) is a function counting the number of assignments to decision variables that differs from one time step to the next.

Note that at horizon \( h \), agents solve the problem with the stationary distribution of random variables and keep this optimal solution onwards. This would maximize the expected reward at each time step after the distributions have converged to the stationary distribution. If choosing any other assignment, then, it will result in a smaller reward.

### 3 CONCLUSIONS

In real-world applications, agents often act in dynamic environments. Thus, the Dynamic DCOP formulation is attractive to model such problems. Current research has focused on solving such problems reactively, thus discarding the information on possible future changes, which is often available in many applications. To cope with this limitation, we (i) introduce Proactive Dynamic DCOPs (PD-DCOPs), which model the dynamism in Dynamic DCOPs; (ii) provide theoretical results on the complexity class of PD-DCOPs; and (iii) develop an exact PD-DCOP algorithm that solves the problem proactively as well as an approximation algorithm with quality guarantees that can scale to larger and more complex problems.

Moreover, we also propose the Infinite-Horizon PD-DCOP (IPD-DCOP) model, which extends PD-DCOPs to optimize the cumulative reward obtained across an infinite number of time steps. It exploits the convergence properties of Markov chains and assumes that the underlying Markov chain in the problem is guaranteed to converge to the unique stationary distribution.

### ACKNOWLEDGEMENTS

This work is supervised by Dr. William Yeoh at Washington University in St. Louis, and supported by NSF grants 1345232, 1540168, 1550662 and JSPS KAKENHI Grant Number 24220003.

### REFERENCES


