An Incentive-Based, Multi-Period Decision Model for Hierarchical Systems

Christian Wernz (Dept. of Industrial and Systems Engineering, Virginia Tech)
Abhijit Deshmukh (Dept. of Industrial and Systems Engineering, Texas A&M)

Abstract

This paper derives multi-period decision strategies for agents in hierarchical systems. We contribute to the multiscale decision theory by extending a one-period, base model to a multi-period framework. By combining game theory and Markov decision processes, we enable strategically interacting agents to derive optimal decision strategies. The model allows supremal agents to determine incentive levels that encourage cooperative behavior from their subordinates. Furthermore, our results show that limited information about the current and next period is sufficient for both agents to make optimal decisions for their entire decision horizon.

1. Introduction

In business and governmental organizations, decision-makers (agents) need to consider the choices made by other agents in the system to determine optimal decision strategies. When agents consider only immediate and local effects of their interaction with other agents, their decisions can become suboptimal. Effective decision-makers take into account the multi-period consequences of their actions on other agents – and the effect of other agents on them.

In this paper, we contribute to the multiscale decision theory [3, 5] by extending a one-period, two-agent model – first presented by Wernz and Deshmukh [4] – to a multi-period model. Multiscale decision theory recognizes and incorporates in its models that decisions occur on and interact across multiple organizational, temporal and informational levels, or scales. We model the agent interaction by combining Markov decision processes and game theory and derive optimal decision strategies that account for all data in all periods, but only require a small subset of information.
2. Model

We consider two agents in a hierarchical superior-subordinate relationship, which we refer to as agent SUP (supremal) and agent INF (infimal). The agents’ hierarchical interaction during one period can be described as follows: agent INF performs work for agent SUP that affects the chances of success of agent SUP. To motivate good work, and thereby increase the likelihood of success for agent SUP, agent SUP rewards agent INF a share of its payoff. Hence, agent INF has an incentive to support agent SUP given the right influence and incentive structures. Good decisions in one period result in higher transition probabilities to preferred states, which put both agents in an advantageous position for the next period.

Decisions are made at the beginning of a period; this point in time is referred to as the decision epoch \( t \in \mathbb{T} \). At every decision epoch \( t = \{1,...,N-1\} \), each agent chooses an action \( a_t \). Decisions are not made at the final decision epoch \( N \); however, agents receive their final reward during this period. Depending on their current state and decision, each agent moves to a state \( s_{t+1} \in S \) with probability \( p_t(s_{t+1} | s_t, a_t) \). Each agent receives a reward that depends on the state at the next decision epoch \( r_t(s_{t+1}) \). This process repeats for every period \( t = \{1,...,N-1\} \).

The action spaces for agents SUP and INF are denoted as \( A^{SUP} = \{a_{t,1}^{SUP}, a_{t,2}^{SUP}\} \), \( A^{INF} = \{a_{t,1}^{INF}, a_{t,2}^{INF}\} \) and their state spaces by \( S^{SUP} = \{s_{t,1}^{SUP}, s_{t,2}^{SUP}\} \), \( S^{INF} = \{s_{t,1}^{INF}, s_{t,2}^{INF}\} \) for \( t = 1,...,N \). Note that each agent has a distinct set of actions and states during each period.

The rewards for agent SUP in each period are \( r_t^{SUP}(s_{t+1}^{SUP}) = p_t^{SUP}, r_t^{SUP}(s_{t+1}^{SUP}) = p_t^{SUP} \) for \( t = 1,...,N-1 \).

The rewards for agent INF are defined similarly by replacing superscript SUP with INF. The initial transition probabilities for agent SUP, without agent INF’s influence, given that agent SUP is in state \( s_{t,i}^{SUP} \), with \( i = 1,2 \) and \( t = 1,...,N-1 \) are

\[
p_t^{SUP}(s_{t+1}^{SUP} | s_{t,i}^{SUP}, a_{t,i}^{SUP}) = \alpha_{1,1,i}^{SUP}, \quad p_t^{SUP}(s_{t+1}^{SUP} | s_{t,1}^{SUP}, a_{t,1}^{SUP}) = 1 - \alpha_{1,1,i}^{SUP},
\]

\[
p_t^{SUP}(s_{t+1}^{SUP} | s_{t,i}^{SUP}, a_{t,2}^{SUP}) = 1 - \alpha_{1,2,i}^{SUP}, \quad p_t^{SUP}(s_{t+1}^{SUP} | s_{t,1}^{SUP}, a_{t,2}^{SUP}) = \alpha_{1,2,i}^{SUP},
\]
with $0 \leq \alpha_{i,m,t}^{SUP} \leq 1$, $m = 1,2$. Again, the transition probabilities for agent INF are denoted similarly by replacing the superscript SUP with INF.

The magnitude of the influence is determined by the state to which an agent moves. We assume that the transition probabilities of agent SUP are affected by an additive influence function $f$ such that

$$
P_{\text{final},t}^{SUP} \left( s_{k,t+1}^{SUP}, s_{j,t}^{INF}, a_{m,t}^{SUP} \right) = p_t^{SUP} \left( s_{k,t+1}^{SUP}, s_{j,t}^{INF}, a_{m,t}^{SUP} \right) + f_t^{SUP} \left( s_{k,t+1}^{SUP}, s_{j,t}^{INF} \right) \quad \text{for } i, j, k, l, m = 1,2, \ t = 1,\ldots,N-1.
$$

We choose the influence function to be a constant and define it as

$$
f_t^{SUP} \left( s_{k,t+1}^{SUP}, s_{j,t}^{INF} \right) = \begin{cases} c_i & \text{if } k = l, \\ -c_i & \text{if } k \neq l, \end{cases} \quad \text{with } c_i > 0.
$$

Constants $c_i$ are referred to as change coefficients. Since probabilities cannot be negative nor exceed unity, condition $0 \leq P_{\text{final},t}^{SUP} (\cdot) \leq 1$ must hold in general.

The meaning and impact of the chosen change coefficient structure is as follows: for $t = 2,\ldots,N$, state $s_{i,t}^{INF}$ increases the probability of state $s_{i,t}^{SUP}$ and consequentially reduces the probability of state $s_{2,t}^{SUP}$.

The probabilities change in opposite direction for state $s_{2,t}^{INF}$: state $s_{2,t}^{SUP}$ becomes more likely and state $s_{1,t}^{SUP}$ less likely. This effect on transition probabilities applies to situations where agent INF’s state supports or hinders agent SUP reaching a specific state.

In a final step, we mathematically formulate the reward influence. The rewards for agent INF are affected by agent SUP’s state dependent rewards, of which agent INF receives proportional shares $b_l$; we call $b_l$ share coefficients. The final rewards per period for agent INF are

$$
r_{\text{final},t}^{INF} \left( s_{k,t+1}^{INF}, s_{l,t+1}^{INF} \right) = r_t^{INF} \left( s_{l,t+1}^{INF} \right) + b_{l,t} r_t^{SUP} \left( s_{k,t+1}^{SUP} \right) = \rho_{l,t}^{INF} + b_{l,t} \rho_{k,t}^{SUP} \quad \text{for } k, l = 1,2, \ t = 1,\ldots,N-1.
$$

Agent SUP’s initial rewards are reduced by the reward shares given to agent INF, resulting in the final rewards

$$
r_{\text{final},t}^{SUP} \left( s_{k,t+1}^{SUP} \right) = (1 - b_l) \rho_{k,t}^{SUP} \quad \text{for } k = 1,2, \ t = 1,\ldots,N-1.
$$

The following assumptions are made in order to create an interesting agent interaction that requires the agents to reason about a non-trivial decision strategy:
\[
\rho_{1,t}^{\text{SUP}} > \rho_{2,t}^{\text{SUP}}, \quad \rho_{1,t}^{\text{INF}} < \rho_{2,t}^{\text{INF}},
\]
\[
\alpha_{m,t}^{\text{SUP}}, \alpha_{n,t}^{\text{INF}} > \frac{1}{2} \quad \text{for} \quad m, n = 1, 2, \quad t = 1, \ldots, N - 1.
\]

Inequalities in (1) express that agent SUP prefers states \(s_{1,t+1}^{\text{SUP}}\) over \(s_{2,t+1}^{\text{SUP}}\) and agent INF reversely prefers \(s_{2,t+1}^{\text{INF}}\) over \(s_{1,t+1}^{\text{INF}}\), at least initially. Expression (2) states that an action is linked to the state with the same index; in other words, there is a corresponding action for every state, which is the most likely consequence of the respective action. This restriction circumvents redundant cases in the analysis, but does not limit the generality of the model.

3. Analysis

We assume that agents are risk-neutral and rational, i.e., agents maximize their expected utilities, or equivalently their expected rewards. As information asymmetry is commonplace in organizations, we assume that agents have only private information and need to communicate with other agents to elicit their information. Rational agents are able to calculate both their own and the other party’s expected rewards, and thus can decide which decisions yield the highest expected rewards for themselves. Hence, agents will engage in a game-theoretic reasoning process, recognizing the dependency of each other’s decisions. The expected reward for agent INF for a given period \(t\) is calculated as follows:

\[
E(r_{n,1}^{\text{INF}} | s_{i,1}^{\text{SUP}}, s_{j,1}^{\text{INF}}, a_{m,n}, a_{n,n}) = \sum_{k=1}^{2} \sum_{l=1}^{2} r_{n,1}^{\text{INF}}(s_{k,t+1}^{\text{SUP}}, s_{l,t+1}^{\text{INF}}) \cdot p_{n,1}^{\text{INF}}(s_{l,t+1}^{\text{INF}} | s_{j,t}^{\text{INF}}, a_{n,n}) \cdot p_{n,1}^{\text{SUP}}(s_{k,t+1}^{\text{SUP}} | s_{i,t}^{\text{SUP}}, s_{j,t}^{\text{INF}}, a_{m,n}).
\]

The expected reward for agent SUP is calculated similarly by replacing \(r_{n,1}^{\text{INF}}(\cdot)\) with \(r_{n,1}^{\text{SUP}}(\cdot)\).

Agents seek to maximize their cumulative rewards \(r_{n,1}^{\text{final}(N-1)}\), i.e., the sum of all \(r_{n,1}^{\text{final}(t)}\) over the time horizon, given agents’ initial states. The cumulative reward from period \(t\) to period \(N-1\) for agent INF – and accordingly for agent SUP – is:

\[
r_{n,1}^{\text{INF}}(s_{i,1}^{\text{SUP}}, s_{j,1}^{\text{INF}}) = \sum_{t=1}^{N-1} r_{n,1}^{\text{INF}}(s_{i,t+1}^{\text{SUP}}, s_{j,t+1}^{\text{INF}}) \quad \text{for} \quad t = 1, \ldots, N - 1
\]

To calculate the expected cumulative reward, which agents seek to maximize, the backward induction principle [1, 2] is applied, starting in the last period of the time horizon and working backwards to period 1:
\[ E\left( r_{\text{final}}^{\text{INF}} \mid s_{1,t}^{\text{SUP}}, s_{j,t}^{\text{INF}}, a_{m,t}^{\text{SUP}}, a_{n,t}^{\text{INF}} \right) = E\left( r_{\text{final}}^{\text{INF}} \mid s_{1,t}^{\text{SUP}}, s_{j,t}^{\text{INF}}, a_{m,t}^{\text{SUP}}, a_{n,t}^{\text{INF}} \right) + \sum_{k=1}^{2} \sum_{l=1}^{2} p_{l} \left( s_{k,t+1}^{\text{SUP}}, s_{l,t+1}^{\text{INF}} \mid s_{1,t}^{\text{SUP}}, s_{j,t}^{\text{INF}}, a_{m,t}^{\text{SUP}}, a_{n,t}^{\text{INF}} \right) \cdot E\left( r_{\text{final(t+1)}}^{\text{INF}} \mid s_{k,t+1}^{\text{SUP}}, s_{l,t+1}^{\text{INF}} \right) \]

with

\[ p_{l} \left( s_{k,t+1}^{\text{SUP}}, s_{l,t+1}^{\text{INF}} \mid s_{1,t}^{\text{SUP}}, s_{j,t}^{\text{INF}}, a_{m,t}^{\text{SUP}}, a_{n,t}^{\text{INF}} \right) = P_{l}^{\text{INF}} \left( s_{l,t}^{\text{INF}} \mid s_{1,t}^{\text{SUP}}, a_{m,t}^{\text{INF}} \right) \cdot p_{l}^{\text{SUP}} \left( s_{k,t+1}^{\text{SUP}} \mid s_{1,t}^{\text{SUP}}, s_{l,t+1}^{\text{INF}}, a_{m,t}^{\text{SUP}} \right) \]

and \( E\left( r_{\text{final(N)}}^{\text{INF}} \mid s_{k,N}^{\text{SUP}}, s_{l,N}^{\text{INF}} \right) = 0 \) for \( i, j, m, n = 1, 2, t = 1, \ldots, N - 1 \).

To derive \( E\left( r_{\text{final(t)}}^{\text{INF}} \mid s_{1,t}^{\text{SUP}}, s_{j,t}^{\text{INF}} \right) \) from \( E\left( r_{\text{final(t)}}^{\text{INF}} \mid s_{1,t}^{\text{SUP}}, s_{j,t}^{\text{INF}}, a_{m,t}^{\text{SUP}}, a_{n,t}^{\text{INF}} \right) \), agents determine their optimal action pair \( \left( a_{m,t}^{\text{SUP}}, a_{n,t}^{\text{INF}} \right) \). Agent SUP wants agent INF to switch from its initially preferred actions \( a_{2,j}^{\text{INF}} \) to \( a_{1,t}^{\text{INF}} \). Agent SUP can incentivize agent INF by choosing sufficiently large \( b_{t} \).

Expressed mathematically, agent SUP choose share coefficients \( b_{t} \) such that

\[ E\left( r_{\text{final(t)}}^{\text{INF}} \mid s_{1,t}^{\text{SUP}}, s_{j,t}^{\text{INF}}, a_{1,t}^{\text{INF}}, a_{2,t}^{\text{INF}} \right) \geq E\left( r_{\text{final(t)}}^{\text{INF}} \mid s_{1,t}^{\text{SUP}}, s_{j,t}^{\text{INF}}, a_{m,t}^{\text{SUP}}, a_{n,t}^{\text{INF}} \right) \] for \( i, j = 1, 2, t = 1, \ldots, N - 1 \).

Starting in period \( N - 1 \), we determine

\[ E\left( r_{\text{final(N-1)}}^{\text{INF}} \mid s_{1,N-1}^{\text{SUP}}, s_{j,N-1}^{\text{INF}}, a_{1,N-1}^{\text{SUP}}, a_{2,N-1}^{\text{INF}} \right) \geq E\left( r_{\text{final(N-1)}}^{\text{INF}} \mid s_{1,N-1}^{\text{SUP}}, s_{j,N-1}^{\text{INF}}, a_{1,N-1}^{\text{INF}}, a_{2,N-1}^{\text{INF}} \right) \]

\[ b_{N-1} \geq \frac{\rho_{2,N-1}^{\text{INF}} - \rho_{1,N-1}^{\text{INF}}}{2c_{N-1}^{\text{SUP}} \left( \rho_{1,N-1}^{\text{SUP}} - \rho_{2,N-1}^{\text{SUP}} \right)} \] for \( i, j = 1, 2 \).

Notice that inequality (7) merely depends on reward differences and is independent of transition probabilities and current states.

Agent SUP chooses the smallest \( b_{N-1} \), i.e., chooses \( b_{N-1} \) such that (7) is an equality, since this maximizes its expected reward. We assume that data are such that paying the reward share and encouraging cooperative behavior is more profitable than not paying an incentive. This participation constraint can be expressed mathematically:

\[ E\left( r_{\text{final(t)}}^{\text{SUP}} \mid s_{1,t}^{\text{SUP}}, s_{j,t}^{\text{INF}}, a_{1,t}^{\text{INF}}, a_{2,t}^{\text{INF}} \right) \bigg| b_{t} \geq E\left( r_{\text{final(t)}}^{\text{SUP}} \mid s_{1,t}^{\text{SUP}}, s_{j,t}^{\text{INF}}, a_{1,t}^{\text{INF}}, a_{2,t}^{\text{INF}} \right) \bigg| b_{t} = 0 \] for \( i, j = 1, 2, t = 1, \ldots, N - 1 \).

Alternatively, we could consider a model variation where the reward share is paid by a third party interested in agent SUP attaining states \( s_{1,t}^{\text{SUP}} \).
We continue with the backward induction and evaluate (6). We get the same result for every period, regardless of the initial state:

\[ b_t \geq \frac{\rho_{2,t}^{\text{INF}} - \rho_{2,t}^{\text{INF}}}{2c_t (\rho_{1,t+1}^{\text{SUP}} - \rho_{1,t+1}^{\text{SUP}})} \left( \alpha_{1,1,t+1}^{\text{SUP}} + \alpha_{2,1,t+1}^{\text{SUP}} \right) \left( \rho_{2,t+1}^{\text{INF}} - \rho_{1,t+1}^{\text{INF}} \right) \text{ for } t = 1, \ldots, N - 2. \] (8)

Agent SUP chooses \( b_t \) such that (8) become an equality, and the optimal strategy for agents is to choose \( (\alpha_{1,t}^{\text{SUP}}, \alpha_{2,t}^{\text{INF}}) \) for \( t = 1, \ldots, N - 1 \). Agents merely have to look ahead one period to determine their optimal decisions. Periods further into the future do not affect the current decision. This fact can be explained by the choices of share coefficients \( b_t^* \) which make agent INF’s expected rewards for both of its actions equal. The rewards of periods beyond the next period do not play a role in the decision process at the current epoch. Further analyses show that if agent SUP chose \( b_t > b_t^* \), the current decision would be influenced by all future periods.

4. Conclusions

We extended the hierarchical agent interaction model presented in [4] from one to multiple periods and thereby advanced the multiscale decision-making framework introduced by Wernz and Deshmukh [5]. For the described influence and incentive structures, we showed that agents can determine their optimal decision strategies for a multi-period horizon by looking ahead only one period. Furthermore, agents need only limited and aggregate information about their and the other agent’s parameters, which helps to keep communication costs low and shows the model’s robustness towards data uncertainty. These results inform managers how to design incentive and influence structures and data management systems.

References