Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

On Fairness in Decision-Making under Uncertainty: Definitions, Computation, and Comparison

Chongjie Zhang and Julie A. Shah
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139
{chongjie,julie_a_shah}@csail.mit.edu

Abstract

The utilitarian solution criterion, which has been extensively studied in multi-agent decision making under uncertainty, aims to maximize the sum of individual utilities. However, as the utilitarian solution often discriminates against some agents, it is not desirable for many practical applications where agents have their own interests and fairness is expected. To address this issue, this paper introduces egalitarian solution criteria for sequential decision-making under uncertainty, which are based on the maximin principle. Motivated by different application domains, we propose four maximin fairness criteria and develop corresponding algorithms for computing their optimal policies. Furthermore, we analyze the connections between these criteria and discuss and compare their characteristics.

Introduction

Multi-agent Markov decision processes (MMDPs) provide a powerful framework for studying sequential decision making under uncertainty in the presence of multiple agents. Existing techniques for MMDPs (Guestrin 2003; Kok and Vlassis 2006; Zhang and Lesser 2011) have been focused on optimizing the utilitarian criterion, where the central decision maker aims to find a policy for maximizing the sum of individual utilities. The computed utilitarian solution is optimal from the perspective of the system where the performance is additive. However, as the utilitarian solution often discriminates against some agents, it is not desirable for many practical applications where agents have their own interests and fairness is expected. For example, in manufacturing plants, resources need to be fairly and dynamically allocated to work stations on assembly lines in order to maximize the throughput; in telecommunication systems, wireless bandwidth needs to be fairly allocated to avoid "unhappy" customers; in transportation systems, traffic lights are controlled so that traffic flow is balanced.

To tackle these decision-making problems, we need to introduce a notion of fairness in the computed policy. Fairness has been studied for resource division in economics (Andrews et al. 2001; Brams 1995), and primary goods distribution in social science (Rawls 1971) and bandwidth allocation in networking (Bonald and Massoulié 2001; Nace and Pióro 2008). These existing works on fairness have been focused on one-shot decision-making in deterministic settings. In this paper, we will study fairness in sequential decision-making under uncertainty. Our fairness notion is based on the maximin principle in Rawlsian theory of justice, maximizing the utility of the agents who are worse off. Motivated by different application domains, we propose four different maximin fairness solution criteria and develop corresponding algorithms for computing fair policies. In order to identify proper fairness criteria for different domains, we also formally analyze the connections between these four criteria and discuss and compare their characteristics.

Multi-Agent Decision Making Model

We are interested in multi-agent sequential decision-making problems, where agents have their own interests. We will focus on centralized policies. This focus is sensible because we assume that although agents have local interests, they are cooperative. Such cooperation can be proactive, e.g., sharing resources with other agents to sustain cooperation that benefits all agents, or can be passive, where agents' actions are controlled by a central decision maker, as in many practical resource allocation problems in manufacturing, cloud computing, networking, and transportation.

We use a multi-agent Markov decision processes (MDP) to model multi-agent sequential decision-making problems, defined by a tuple $\langle I, S, A, T, \{R_i\}_{i\in I}, s_1, H\rangle$, where

- I $= \{1, \ldots, n\}$ is a set of agent indices.
- S is a finite set of states.
- A $= \times_{i\in I} A_i$ is a finite set of joint actions, where A_i is a finite set of actions available for agent i.
- T: $S \times A \times S \to [0, 1]$ is the transition function. $T(s'|s, a)$ is the probability of transiting to the next state s' after a joint action $a \in A$ is taken by agents in state s.
- R_i: $S \times A \times S \to \Re$ is a reward function of agent i and provides agent i with an individual reward $R_i(s, a, s')$ after a joint action a taken in state s and resulting in state s'.
- s_1 is the initial state.
- H is the horizon.

In the multi-agent MDP model, each agent has its own reward function, representing its interest. In this paper, our discussion is based on the finite horizon case, but most fairness definitions and corresponding algorithms can be adapted for the infinite horizon case, as we will discuss later.

The objective of multi-agent MDPs is to find a centralized policy that optimizes a certain solution criterion. A policy can be deterministic or stochastic. A deterministic policy $\pi : S \to A$ is a function that returns action $a = \pi(s)$ for state s. A stochastic policy $\pi : S \times A \to [0, 1]$ is a function that returns the probability $\pi(s, a)$ of taking joint action $a \in A$ for any given state $s \in S$. Because we focus on the finite horizon, we will use non-stationary policies, whose action choices also depend on the time in addition to the state.

Solution Criteria and Algorithms

To define fairness criteria for multi-agent MDPs, we employ the maximin principle in the Rawlsian theory of justice (Rawls 1971), which states that social and economic inequalities should be arranged such that "they are to be of the greatest benefit to the least-advantaged members of society." In other words, an unequal distribution can be just when it maximizes the minimum benefit to those with the lowest allocation of welfare-conferring resources.

Motivated by different application domains, we propose four maximin fairness criteria and develop corresponding algorithms for computing the optimal policy: maximizing the minimum expected utility of agents (MMEU), maximizing the expected minimum utility of agents (MEMU), greedy MMEU, and greedy MEMU.

For brevity, we use $\psi(i, \pi)$ to denote a random variable that represents the sum of rewards gained by agent i over the horizon H under policy π:

$$\psi(i, \pi) = \sum_{t=1}^{H} R_i(s_t, a_t, s_{t+1}) \tag{1}$$

where s_t and a_t are the state and the action chosen under policy π at time t, resulting in the next state s_{t+1}.

Maximizing the Minimum Expected Utility (MMEU)

The first maximin fairness criterion we propose is called MMEU fairness (Maximizing the Minimum Expected Utility), where the central decision maker maximizes the expected utility of the agent with the least. This criterion assumes agents' individual objectives are relatively independent and agents are concerned with fairness based on the expected long-term utility over the whole horizon instead of fairness at every decision step.

For example, in cloud computing (Perez et al. 2009), when allocating sharable resources (CPU cycles or network bandwidth), the system needs to consider the fairness among different classes of customers to avoid "unhappy" customers. As long as it is fair, a customer is usually not concerned with the resource utilization of other customers. In addition, a customer only cares about the expected quality of the service it received (e.g., how fast its jobs are finished or the response time of its web services) over a time period, and does not care how computing resources are actually allocated to its jobs at each time step during the execution. Therefore, based on the maximin principle, the system is considered to be fair by customers if it maximizes the minimum expected service quality received by customers.

Formally, under the MMEU fairness criterion, the goal of solving multi-agent MDPs is to find the optimal policy that maximizes the following stage-to-go value function V^{π}_{mmeu}:

$$V^{\pi}_{mmeu} = \min_{i \in I} \mathbf{E}[\psi(i, \pi)|\pi, s_1] \tag{2}$$

where the expectation operator $\mathbf{E}(\cdot)$ averages over stochastic transitions. The optimal policy $\pi^*_{mmeu} = \text{argmax}_{\pi} V^{\pi}_{mmeu}$.

We develop a linear programming (LP) approach to computing the optimal MMEU fairness policy. Similar to the linear programming formulation of single-agent MDPs (Puterman 2005), our approach uses frequencies of state-action visitations to reformulate the maximin objective function as defined in (2). For a multi-agent MDP, given a policy and the initial state, frequencies of visiting state-action pairs at each time step are uniquely determined. We use $x^t_{\pi}(s, a)$ to denote the probability, under policy π and initial state s_1, that the system occupies state s and chooses action a at time t. Using this frequency function, we rewrite the MMEU objective function as follows:

$$V^{\pi}_{mmeu} = \min_i \sum_{t=1}^{H} \sum_{s,s' \in S} \sum_{a \in A} x^t_{\pi}(s, a) T(s'|s, a) R_i(s, a, s')$$

The MMEU optimization is formulated as a LP problem:

$$\begin{aligned}
&\text{Maximize}_{x,z} \quad z \\
&\text{Subject to} \\
&z \le \sum_{t=1}^{H} \sum_{s \in S} \sum_{a \in A} \sum_{s' \in S} x^t_{\pi}(s, a) T(s'|s, a) R_i(s, a, s'), \forall i \in I \\
&\sum_{a \in A} x^1(s, a) = b(s), \\
&\sum_{a \in A} x^t(s, a) = \sum_{s' \in S} \sum_{a \in A} T(s|s', a) x^{t-1}(s', a), t = 2, \ldots, H \\
&\text{and } x^t(s, a) \ge 0, \forall a \in A, \forall s \in S, \text{ and } t = 1, \ldots, H
\end{aligned} \tag{3}$$

where b is the vector with all zeros except for the position for initial state s_1 with $b(s_1) = 1$. The first set of constraints is used to linearize the objective value function by introducing another variable z, which represents the minimum expected total reward among all agents. The remaining constraints are included to ensure that $x^t(s, a)$ is well-defined. The second constraint set is to ensure the probability of visiting state s at the initial time step is equal to the initial probability of state s. The third set of constraints requires that the probability of visiting state s' is equal to the sum of all probabilities of entering into state s'.

We can employ existing LP solvers (e.g., interior point methods) to compute an optimal solution x^* for problem (3) and derive a state-to-go policy π^* from x^* by normalization:

$$\pi^t(s, a) = \frac{x^t(s, a)}{\sum_{a \in A} x^t(s, a)} \tag{4}$$

Note that the optimal MMEU fairness policy can be stochastic. This is because this LP problem has $|H||S| + |I|$ rows with only $|H||S|$ variables, and, as a result, for some state s at time t, it is possible that $x^t(s, a_s) > 0$ for more than one action $a_s \in A$.

Although we define the MMEU fairness criterion for the finite horizon case, it can be readily adapted to the infinite horizon case by multiplying the reward at time t by a discount factor λ^t. For the infinite horizon, the optimal fairness policy can be stationary, which does not depend on the time. Hence we can use the stationary frequency function $x(s, a)$ that does not depend on the time. The second and third constraint sets in the LP problem need to be substituted by the following constraint set:

$$\sum_{a \in A} x(s', a) = b(s') + \sum_{s \in S} \sum_{a \in A} \lambda T(s'|s, a)x(s, a), \forall s' \in S$$

which ensures that the probability of visiting state s' is equal to the initial probability of state s' plus the sum of all probabilities of entering into state s'. The optimal MMEU policy for the infinite horizon case can be derived through normalization from the solution of the LP problem.

Maximizing the Expected Minimum Utility (MEMU)

Although the MMEU criterion provides a fairness solution for many decision-making problems, it may not be appropriate for applications where agents' objectives interact. For example, in manufacturing, resources need to be dynamically and fairly allocated to different work cells in an assembly line in order to optimize its throughput. The objective of each work cell is to maximize its own throughput. Work cells' objectives interact because the throughput of the assembly line is determined by the minimum throughput among work cells. In order to maximize the throughput for a given horizon, it is more sensible for the system to maximize the expected minimum total throughput of work cells.

In order to address such similar decision-making problems, we propose a MEMU (Maximizing the Expected Minimum Utility) fairness criterion, where the central decision maker aims to maximize the expected minimum total rewards gained by agents over the time horizon. Under the MEMU fairness criterion, the stage-to-go value function V^{π}_{memu} for policy π is defined as follows:

$$V^{\pi}_{memu} = \mathbf{E}[\min_{i \in I} \psi(i, \pi)|\pi, s_1] \tag{5}$$

where the expectation operator $\mathbf{E}(\cdot)$ averages over stochastic transitions. The optimal policy $\pi^*_{memu} = \arg\max_{\pi} V^{\pi}_{memu}$.

The MEMU objective function looks similar to that of MMEU fairness except for the order of the expectation and minimization operators. However, this difference leads to a more fine-grained fairness than that of MMEU. MMEU fairness focuses on individual agents' objectives and attempts to maximize the total reward for the agent with the least, while the MEMU fairness solution focuses more on the system objective based on execution traces with finite horizon H, maximizing the expected minimal total rewards gained by agents over an execution trace. (Note that different execution traces of a multi-agent MDP may have different agents with the least total reward.)

The MEMU optimization problem may be difficult to solve directly, since evaluating the expectation of the minimum can be hard. However, as shown by the following proposition, we can use the MMEU fairness policy to provide an approximate MEMU solution with both lower and upper bounds.

Proposition 1. *Let V^*_{memu} be the value of the optimal MEMU fairness policy, V^{ub} be the value of the optimal MMEU policy π^*_{mmeu} evaluated with its own objective function defined by (2), and V^{lb} be the value of the optimal MMEU policy π^*_{mmeu} evaluated with the MEMU value function defined by (5). We then have:*

$$V^{lb} \leq V^*_{memu} \leq V^{ub}$$

Proof. First, by definition, the value of any policy evaluated with the MEMU objective function will be less than or equal to V^*_{memu}. Therefore, $V^{lb} \leq V^*_{memu}$. Furthermore,

$$\begin{aligned} V^*_{memu} &= \max_{\pi} \mathbf{E}[\min_{i \in I} \psi(i, \pi)|\pi, s_1] \\ &\leq \max_{\pi} \min_{i \in I} \mathbf{E}[\psi(i, \pi)|\pi, s_1] \\ &= V^{ub} \end{aligned} \tag{6}$$

Note that $\psi(i, \pi)$ is a linear function and convex and the minimum operator is a concave function. The inequality in the proof holds due to Jensen's Inequality (Jensen 1906), which states that the concave transformation of a mean is greater than or equal to the mean after concave transformation. □

The lower bound on the value of the MEMU policy is readily evaluated through Monte Carlo simulation. When V^{lb} and V^{ub} are close, we can conclude that the policy π^*_{mmeu} is almost optimal under the MEMU criterion.

Similar to the MMEU criterion, the MEMU criterion can also be adapted to the infinite horizon by introducing a discount factor. Since Jensen's inequality holds for an infinite discrete form, Proposition 1 still holds and the approximate computation approach is valid for the infinite horizon case.

Greedy MMEU

Both MMEU and MEMU fairness criteria assume agents are interested in fairness based on long-term utilities (e.g., total rewards gained over the whole horizon). However, there are applications where the system needs to take account of both extended-time and immediate fairness. For example, in traffic light control (Houli, Zhiheng, and Yi 2010), the system not only balances the extended-time traffic flow, but also considers the waiting time of current drivers at different directions. When using MMEU or MEMU fairness policy, it is likely that the interval of changing traffic lights will be relatively long, because frequent changes will slow down the traffic flow. As a result, some drivers will wait for quite a long time to pass an intersection.

To address this type of optimization problems, we propose a variant of the MMEU fairness criterion, called greedy MMEU. The greedy MMEU criterion aims to maximize the minimum "expected utility" at every time step. The "expected utility" under the greedy MMEU criterion is defined

differently from MMEU. As shown below, its stage-to-go value function V^{π}_{gmmeu} for policy π is defined iteratively:

$$V^{\pi^t}_{gmmeu}(s_t) = \min_i \mathbf{E}[R_i(s_t, \pi^t(s_t), s_{t+1}) + V^{\pi^{t+1}}_{gmmeu}(s_{t+1})]$$
$$V^{\pi^H}_{gmmeu}(s_H) = \min_i \mathbf{E}[R_i(s_H, \pi^H(s_H), s_{H+1})] \quad (7)$$

where $t = 1, \ldots, H-1$ and the expectation operator $\mathbf{E}(\cdot)$ averages the stochastic transition function. From the definition, we can see that the greedy MMEU policy attempts to achieve fairness at each decision step with a consideration of long-term expected utilities starting from that step. Therefore, greedy MMEU provides a more fine-grained fairness than MMEU, which focuses on fairness for the starting step based on expected utilities gained over the whole horizon.

Exploiting the iterative definition of the greedy MMEU criterion, we design a backward induction algorithm to compute the optimal stage-to-go value function $V^{t^*}_{gmmeu}$ and the optimal policy $\pi^{t^*}_{gmmeu}$, which is described as following:

1. Set time $t = H$ and compute $V^{H^*}_{gmmeu}(s_H)$ and $\pi^{H^*}_{gmmeu}(s_H)$ for all state $s_H \in S$ by

$$V^{H^*}_{gmmeu}(s_H) = \max_{a \in A} \min_i \sum_{s \in S} T(s|s_H, a) R_i(s_H, a, s)$$

$$\pi^{H^*}_{gmmeu}(s_H) = \operatorname*{argmax}_{a \in A} \min_i \sum_{s \in S} T(s|s_H, a) R_i(s_H, a, s)$$

2. Substitute $t-1$ for t and compute $V^{t^*}_{gmmeu}(s_t)$ and $\pi^{t^*}_{gmmeu}(s_t)$ for each $s_t \in S$ by

$$V^{t^*}_{gmmeu}(s_t) = \max_{a \in A} \min_i \sum_{s \in S} T(s|s_t, a)(R_i(s_t, a, s) + V^{t+1^*}_{gmmeu}(s))$$

$$\pi^{t^*}_{gmmeu}(s_t) = \operatorname*{argmax}_{a \in A} \min_i \sum_{s \in S} T(s|s_t, a)(R_i(s_t, a, s) + V^{t+1^*}_{gmmeu}(s))$$

3. If t = 1 stop; otherwise, go to step 2.

It is quite straightforward to prove the optimality of this backward induction algorithm by using induction. The complexity of this algorithm is linear with the horizon. Unlike the MMEU criterion, the greedy MMEU criterion has a deterministic optimal policy.

Greedy MEMU

Similarly, we also propose a variant of the MEMU fairness criterion, called greedy MEMU, which aims to maximize the expected total minimum rewards at every time step. Under the greedy MEMU fairness criterion, the stage-to-go value function V^{π}_{gmmeu} for policy π is defined as follows:

$$V^{\pi}_{gmemu} = \mathbf{E}[\sum_{t=1}^{H} \min_i R_i(s_t, \pi^t(s_t), s_{t+1}) | \pi, s_1] \quad (8)$$

where the expectation operator $\mathbf{E}(\cdot)$ averages the stochastic transition function.

The greedy MEMU criterion intends to achieve fairness at each decision by maximizing the expected the sum of minimum rewards received by agents at each step, while the MEMU criterion focuses on fairness on the starting step by maximizing the expected minimum of total rewards gained by agents over the whole horizon. Therefore, greedy MEMU provides a more fine-grained fairness than MEMU.

We can derive the Bellman equation for the stage-to-go value function V^{π}_{gmemu} as follows:

$$V^{\pi^t}_{gmemu}(s_t) = \mathbf{E}[\min_i R_i(s_t, \pi^t(s_t), s_{t+1}) + V^{\pi^{t+1}}_{gmemu}(s_{t+1})]$$
$$V^{\pi^H}_{gmemu}(s_H) = \mathbf{E}[\min_i R_i(s_H, \pi^H(s_H), s_{H+1})] \quad (9)$$

where $t = 1, \ldots, H-1$. This iterative form looks similar to the definition of the greedy MMEU value function except for the order of the expectation and minimization operators. This reverse order results in a more fine-grained fairness of greedy MEMU than that of greedy MMEU. This is because greedy MEMU assumes agents are interested in actual rewards received at each step while greedy MMEU assume agents are interested in expected rewards at each step.

The greedy MEMU criterion can find applications in domains that require very strict fairness, such as telecommunication wireless systems (Eryilmaz and Srikant 2005) and sewage flow control systems (Aoki, Kimura, and Kobayashi 2004). For example, in telecommunication systems, the greedy MEMU policy for allocating wireless bandwidth is more desirable than that of greedy MMEU, because users care more about actual bandwidth allocated to them at each time step than expected bandwidth (which has uncertainty).

By exploiting the Bellman equation of the stage-to-go value function, we can revise the backward induction algorithm described in the previous section by changing the order of the expectation and minimization operators, and use this revised algorithms to compute the optimal greedy MEMU fairness policy.

The greedy MEMU policy for a multi-agent MDP is equivalent to the utilitarian policy of the single-agent MDP derived from the multi-agent MDP by using a reward function $R(s_t, \pi^t(s_t), s_{t+1}) = \min_i R_i(s_t, \pi^t(s_t), s_{t+1})$. Therefore, we can adapt the definition of the greedy MEMU criterion in (8) to the infinite horizon case by introducing a discount factor and use existing techniques (e.g., value iteration, linear programming) to solve the derived single-agent MDP to find an optimal greedy MEMU fairness policy, which is deterministic and stationary.

Discussions

To better understand fairness criteria proposed in the previous section, in this section, we will analyze their connections and discuss and compare their characteristics. This analysis and comparison are intended to help choose proper fairness criteria for different applications.

Connections among Fairness Criteria

We first formally analyze the ordering relationships of the objective function among four fairness criteria, which

are shown by the following two propositions. Using these propositions, we then discuss how different granularities of fairness affects both system and individual performance.

Proposition 2. *For any policy π, $V^{\pi}_{gmemu} \leq V^{\pi}_{memu} \leq V^{\pi}_{mmeu}$.*

Proof. The first inequality holds because $\sum_{t=1}^{H} \min_i R_i \leq \min_i \sum_{t=1}^{H} R_i$ for any sequence number.

$$\begin{aligned} V^{\pi}_{gmemu} &= \mathbf{E}[\sum_{t=1}^{H} \min_i R_i(s_t, \pi^t(s_t), s_{t+1}) | \pi, s_1] \\ &\leq \mathbf{E}[\min_i \sum_{t=1}^{H} R_i(s_t, \pi^t(s_t), s_{t+1}) | \pi, s_1] \\ &= V^{\pi}_{memu} \end{aligned}$$

Because the minimization operator is a concave function, using Jensen's inequality, we have :

$$V^{\pi}_{memu} = \mathbf{E}[\min_{i \in I} \psi(i, \pi) | \pi] \leq \min_{i \in I} \mathbf{E}[\psi(i, \pi) | \pi] = V^{\pi}_{mmeu}$$

□

Although Proposition 2 is proved for the finite horizon case, it also holds for the infinite horizon case.

Proposition 3. *For any stage-to-go policy π, for all state $s \in S$, $V^{\pi}_{gmemu}(s) \leq V^{\pi}_{gmmeu}(s) \leq V^{\pi}_{mmeu}(s)$.*

Proof. The first inequality is proved by induction. For time $t = H$, using Jensen's inequality, we directly have

$$\begin{aligned} V^{\pi^H}_{gmemu}(s_H) &= \mathbf{E}[\min_i R_i(s_H, \pi^H(s_H), s_{H+1})] \\ &\leq \min_i \mathbf{E}[R_i(s_H, \pi^H(s_H), s_{H+1})] \\ &= V^{\pi^H}_{gmmeu}(s_H) \end{aligned}$$

Assume the first inequality holds for time $t+1$. Now let us show it holds for time t.

$$\begin{aligned} V^{\pi^t}_{gmemu}(s_t) &= \mathbf{E}[\min_i R_i(s_t, \pi^t(s_t), s_{t+1}) + V^{\pi^{t+1}}_{gmemu}(s_{t+1})] \\ &\leq \mathbf{E}[\min_i R_i(s_t, \pi^t(s_t), s_{t+1}) + V^{\pi^{t+1}}_{gmmeu}(s_{t+1})] \\ &\leq \min_i \mathbf{E}[R_i(s_t, \pi^t(s_t), s_{t+1}) + V^{\pi^{t+1}}_{gmmeu}(s_{t+1})] \\ &= V^{\pi^t}_{gmmeu}(s_t) \end{aligned}$$

We can also prove the second inequality by induction. For time $t = H$, by definition,

$$V^{\pi^H}_{gmmeu}(s_H) = \min_i \mathbf{E}[R_i(s_H, \pi^t(s_H), s_{H+1})] = V^{\pi^H}_{mmeu}(s_H)$$

Assume the second inequality holds for time $t+1$. Now let us show it holds for time t.

$$\begin{aligned} V^{\pi^t}_{gmmeu}(s_t) &= \min_i \mathbf{E}[R_i(s_t, \pi^t(s_t), s_{t+1}) + V^{\pi^{t+1}}_{gmmeu}(s_{t+1})] \\ &\leq \min_i \mathbf{E}[R_i(s_t, \pi^t(s_t), s_{t+1}) + V^{\pi^{t+1}}_{gmmeu}(s_{t+1})] \\ &= \min_i \mathbf{E}[R_i(s_t, \pi^t(s_t), s_{t+1}) \\ &\quad + \min_i \mathbf{E}[\sum_{k=t+1}^{H} R_i(s_k, \pi^t(s_k), s_{k+1})]] \\ &\leq \min_i \mathbf{E}[R_i(s_t, \pi^t(s_t), s_{t+1}) \\ &\quad + \mathbf{E}[\sum_{k=t+1}^{H} R_i(s_k, \pi^t(s_k), s_{k+1})]] \\ &= V^{\pi^t}_{mmeu}(s_t) \end{aligned}$$

□

As discussed in the previous section, among four fairness criteria, MMEU is the most coarse-grained and greedy MEMU is the most fine-grained fairness. Proposition 2 and 3 indicate that, for any policy, the more fine-grained the fairness criterion, the lower the objective value of a given policy. Immediately following Proposition 2 and 3, Corollary 1 shows the ordering relationships of the optimal objective value among four fairness criteria.

Corollary 1. *For all state $s \in S$, $V^*_{gmemu}(s) \leq V^*_{gmmeu}(s) \leq V^*_{mmeu}(s)$ and $V^*_{gmemu}(s) \leq V^*_{memu}(s) \leq V^*_{mmeu}(s)$.*

The objective value of a policy indicates the long-term expected utility gained by the system. Therefore, from the system perspective, the optimal objective value reflects its best performance under a particular criterion, even though the value function is defined differently with different criteria. Corollary 1 implies that the more fine-grained the fairness criterion, the lower the system optimal performance. In contrast, from the individual agent's perspective, the expected total rewards (i.e., $\mathbf{E}[\sum_{t=1}^{H} R_i^t]$) reflects its performance. By definition, the optimal policy under the MMEU fairness criterion maximizes the expected total rewards of the agent with the least performance. Therefore, for this agent, the optimal policy under this most coarse-grained fairness criterion (i.e., MMEU) yields its highest individual performance among four criteria. In other words, using a more fine-grained fairness potentially results in lower system and individual performance.

Comparison of Fairness Criteria

Table 1 summarizes some characteristics of four fairness criteria we proposed. All fairness criteria except Greedy MMEU can be defined for both finite and infinite horizon. In this paper, we developed exact algorithms for computing the optimal policy under fairness criteria of MMEU, greedy MMEU, and greedy MEMU. It is not clear whether there exists a polynomial exact algorithm for computing the optimal MEMU policy. The complexity of the backward induction algorithm for greedy MMEU and greedy MEMU is

Table 1: Comparison of different fairness criteria

Criterion	Horizon	Algorithm	Complexity	Optimal policy	Granularity
MMEU	finite or infinite	exact	polynomial	stochastic	coarse-grained
MEMU	finite or infinite	approximate	polynomial	stochastic	coarse-grained
Greedy MMEU	finite	exact	linear	deterministic	fine-grained
Greedy MEMU	finite or infinite	exact	linear or polynomial	deterministic	fine-grained

$O(H|I||A||S|^2)$, which is linear in the horizon H. Therefore, solving the greedy MMEU or greedy MEMU fairness policy is much more efficient than solving the MMEU problem, which uses LP and has complexity $O(n^{3.5})$ (where $n = H|A||S|$ is the number of variables).

The LP approach for solving MMEU and MEMU fairness problems computes a stochastic policy, while the backward induction algorithm computes a deterministic optimal policy for both greedy MMEU and greedy MEMU criteria. Under the MMEU criterion, the optimal deterministic policy may exist for some problems, but not for other problems. While in most applications there is no reason to exclude stochastic policies a priori, there can be cases when stochastic policies are clearly undesirable or even unethical. For example, if the policy determines the critical medical equipment usage among a group of patients, then flipping a coin to determine the course of action may be inappropriate.

As discussed in the previous section, MMEU provides the most coarse-grained fairness among four proposed criteria, greedy MEMU provides the most fine-grained fairness, and MEMU and greedy MMEU provides in-between fairness. Since the MMEU criterion optimizes the fairness based on the long-term objectives of individual agents, one of its strengths is that its optimal fairness provides the greatest minimum expected long-term utility among all four criteria, i.e., the best expected performance from the system perspective. However, with the MMEU policy, for a particular execution trace, the minimum total rewards of agents might be low for applications with a high variance of individual rewards from one run to another. Therefore, the MMEU criterion works best for problems where decision process is repeated, the decision maker is concerned with individual agents' objectives but not with their interactions, and agents are not myopic and are interested in expected long-term utilities.

In contrast to the MMEU criterion that considers the minimum expected total reward averaged over repeated runs, the MEMU criterion more specially considers more detailed information, the minimum total rewards for individual runs. The MMEU criterion works best for decision-making problems where the decision maker needs to consider interactions between individual agents' objectives and focuses on the long-term performance of the system instead of individual agents. Greedy MMEU and greedy MEMU criteria attempt to achieve fairness at every decision, which potentially results in lower optimal system and individual performance. Therefore, it is better to only apply them to problems where agents are myopic and fairness is necessary at every decision, unless the computational complexity is of concern.

Related Work

The notion of maximin fairness is widely used in various areas of networking, such as bandwidth sharing, congestion control, routing, load-balancing and network design (Bonald and Massoulié 2001; Nace and Pióro 2008). Maximin fairness has been applied in combinatorial optimization in multi-agent settings, where agents have different evaluation functions about combinatorial problems (Escoffier, Gourvès, and Monnot 2013). Unlike our work, these works are dedicated to one-shot deterministic decision making problems and does not consider the dynamics and uncertainty of users and the environment (e.g., dynamic changes to users' demands and resource availability).

Our multi-agent MDPs can be viewed as multi-objective MDPs (Roijers et al. 2013). Existing work on multi-objective MDPs focuses on linear scalarization function (similar to the utilitarian criterion) and strictly monotonically increasing scalarization functions (similar to our MMEU criterion) in the infinite horizon case. In contrast, this paper studied fairness in infinite and finite horizon with three additional solution criteria.

Fairness has also been studied in goods division (Chen et al. 2010; Chevaleyre et al. 2007; Procaccia 2009). Fair division theory focuses on proportional fairness and envy-freeness. Most existing work in fair division involves a static setting, where all relevant information is known upfront and fixed. Only a few approaches deal with dynamics of agent arrival and departures (Kash, Procaccia, and Shah 2013; Walsh 2011). However, unlike our work, these approaches in fair division do not address uncertainty or other dynamics, such as changes of resource availability and demands.

Conclusion

Fairness is an important criterion for decision making in many practical domains. In this paper, we proposed four different fairness solution criteria for sequential decision making under uncertainty and developed optimal computational algorithms for three of them and a bounded approach for the remaining one. In order to help choose proper fairness criteria for different applications, we formally analyzed on the relationships among these solution criteria and discussed and compared their characteristics. By introducing solution concepts and baseline algorithms, this paper provides an initial effort for studying fairness in sequential decision making under uncertainty. In future work, we are interested in developing methods for computing decentralized fairness policies, algorithms for learning fairness policies, and more scalable approaches for computing fairness policies by exploiting problem structures.

References

Andrews, M.; Kumaran, K.; Ramanan, K.; Stolyar, A.; Whiting, P.; and Vijayakumar, R. 2001. Providing quality of service over a shared wireless link. *Communications Magazine, IEEE* 39(2):150–154.

Aoki, K.; Kimura, H.; and Kobayashi, S. 2004. Distributed reinforcement learning using bi-directional decision making for multi-criteria control of multi-stage flow systems. In *The 8th Conference on Intelligent Autonomous Systems*, 281–290.

Bonald, T., and Massoulié, L. 2001. Impact of fairness on internet performance. In *Proceedings of the 2001 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, 82–91.

Brams, S. J. 1995. On envy-free cake division. *Journal of Combinatorial Theory, Series A* 70(1):170–173.

Chen, Y.; Lai, J.; Parkes, D. C.; and Procaccia, A. D. 2010. Truth, justice, and cake cutting. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*.

Chevaleyre, Y.; Endriss, U.; Estivie, S.; and Maudet, N. 2007. Reaching envy-free states in distributed negotiation settings. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 1239–1244.

Eryilmaz, A., and Srikant, R. 2005. Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, volume 3, 1794–1803. IEEE.

Escoffier, B.; Gourvès, L.; and Monnot, J. 2013. Fair solutions for some multiagent optimization problems. *Autonomous Agents and Multi-Agent Systems* 26(2):184–201.

Guestrin, C. E. 2003. *Planning under uncertainty in complex structured environments*. Ph.D. Dissertation, Stanford University, Stanford, CA, USA.

Houli, D.; Zhiheng, L.; and Yi, Z. 2010. Multiobjective reinforcement learning for traffic signal control using vehicular ad hoc network. *EURASIP journal on advances in signal processing* 2010:7.

Jensen, J. L. W. V. 1906. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica* 30(1):175–193.

Kash, I. A.; Procaccia, A. D.; and Shah, N. 2013. No agent left behind: dynamic fair division of multiple resources. In *International conference on Autonomous Agents and Multi-Agent Systems*, 351–358.

Kok, J. R., and Vlassis, N. 2006. Collaborative multiagent reinforcement learning by payoff propagation. *Journal of Machine Learning Research* 7:1789–1828.

Nace, D., and Pióro, M. 2008. Max-min fairness and its applications to routing and load-balancing in communication networks: A tutorial. *IEEE Communications Surveys and Tutorials* 10(1-4):5–17.

Perez, J.; Germain-Renaud, C.; Kégl, B.; and Loomis, C. 2009. Responsive elastic computing. In *Proceedings of the 6th international conference industry session on Grids meets autonomic computing*, 55–64. ACM.

Procaccia, A. D. 2009. Thou shalt covet thy neighbor's cake. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence, 2009*, 239–244.

Puterman, M. L. 2005. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Willey Interscience.

Rawls, J. 1971. *The theory of justice*. Cambridge, MA: Harvard University Press.

Roijers, D. M.; Vamplew, P.; Whiteson, S.; and Dazeley, R. 2013. A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research* 48(1):67–113.

Walsh, T. 2011. Online cake cutting. In *Algorithmic Decision Theory - Second International Conference*, volume 6992 of *Lecture Notes in Computer Science*, 292–305.

Zhang, C., and Lesser, V. R. 2011. Coordinated multi-agent reinforcement learning in networked distributed POMDPs. In Burgard, W., and Roth, D., eds., *AAAI*. AAAI Press.