日本オペレーションズ・リサーチ学会
2004年春季研究発表会

A Dynamic Decision Making Model with an Objective Function based on Fuzzy Preferences

01702986 北九州市立大学 *吉田祐治 YOSHIDA Yuji
01701690 千葉大学 安田正實 YASUDA Masami
01401530 千葉大学 中神潤一 NAKAGAMI Jun-ichi
01101550 千葉大学 蔵野正美 KURANO Masami

This talk presents a mathematical model for dynamic decision making with an objective function induced from fuzzy preferences. This model is related to decision making in artificial intelligence.

Let a *state space* $\mathbb{S}$ be a σ-compact convex subset of some Banach space, and the *states* are represented by elements of $\mathbb{S}$. The attributes of the states/objects can be represented as the d-dimensional coordinates when the Banach space is taken by d-dimensional Euclidean space $\mathbb{R}^d$. Let $\mathcal{S}$ be a subset of $\mathbb{S}$ such that $\mathcal{S}$ has finite elements. A map $\mu : \mathcal{S} \times \mathcal{S} \mapsto [0,1]$ is called a fuzzy relation on $\mathcal{S}$. Fuzzy preferences are introduced by fuzzy relations on $\mathcal{S}$.

Definition. A fuzzy relation μ on $\mathcal{S}$ is called a *fuzzy preference relation* if it satisfies the following conditions (a) - (b):

(a) $\mu(a,a) = 1$ for all $a \in \mathcal{S}$. (reflexive)

(b) $\mu(a,c) \geq \min\{\mu(a,b), \mu(b,c)\}$ for all $a,b,c \in \mathcal{S}$. (transitive)

(c) $\mu(a,b) + \mu(b,a) \geq 1$ for all $a,b \in \mathcal{S}$. (connected)

Here, $\mu(a,b)$ means the degree that the decision maker likes a than b. We introduce a ranking method of states, which is called a *score ranking function.*

Definition. For a fuzzy preference relation μ on $\mathcal{S}$, the following map r on S is called a score ranking function of states induced by the fuzzy preference relation μ:

$$r(a) = \sum_{b \in \mathcal{S}: b \neq a} \{\mu(a,b) - \mu(b,a)\}, \quad a \in \mathcal{S}. \tag{1}$$

We discuss a dynamic decision making model with fuzzy references and a time space $\{0,1,\cdots,T\}$. Let $\mathcal{S}_0$ be a subset of $\mathbb{S}$ such that $\mathcal{S}_0 := \{c^i | i = 1,2,\cdots,n\}$ has n elements and a partial order $\succsim$. $\mathcal{S}_0$ is called an *initial state space* and it is given as a *training set* in a learning model. Let μ_0 be a fuzzy preference relation on $\mathcal{S}_0$. Let $t(=0,1,2,\cdots,T)$ be a current time. An action space A_t at time $t(< T)$ is given by a compact set of some Banach space. At time t, a current *state* is denoted by s_t, and an initial state s_0 is given by an element in $\mathcal{S}_0$. Define a family of states until time t by $\mathcal{S}_t := \{c^1, c^2, \cdots, c^n, s_1, s_2, \cdots, s_t\}$. $u_t(\in A_t)$ means an *action* at time t, and $h_t = (s_0, u_0, s_1, u_1, \cdots, s_{t-1}, u_{t-1}, s_t)$ means a *history* with states $s_0, s_1, \cdots, s_t$ and actions $u_0, u_1, \cdots, u_{t-1}$. Then, a *strategy* is a map $\pi_t : \{h_t\} \mapsto A_t$ which is represented as $\pi_t(h_t) = u_t$ for some $u_t \in A_t$. A sequence $\pi = \{\pi_t\}_{t=1}^{T-1}$ of strategies is called a *policy.*

Let $\{\bar{\rho}_t\}_{t=1}^{T}$ be a sequence of nonnegative numbers. We deal with the case where a current state s_t is represented by a linear combination of the initial states $c^1, c^2, \cdots, c^n$ and the past states $s_1, s_2, \cdots, s_{t-1}$:

$$s_t = \sum_{i=1}^{n} \bar{w}_t^i c^i + \sum_{j=1}^{t-1} \bar{w}_t^{n+j} s_j, \tag{2}$$

for some weight vector $(\bar{w}_t^1, \bar{w}_t^2, \cdots, \bar{w}_t^{n+t-1}) \in \mathbb{R}^{n+t-1}$ satisfying $-\bar{\rho}_t \leq \bar{w}_t^i \leq 1 + \bar{\rho}_t$ $(i = 1,2,\cdots,n+t-1)$ and $\sum_{i=1}^{n+t} \bar{w}_t^i = 1$, where $\sum_{j=1}^{0} := 0$ and

$$\bar{w}_0^i := \begin{cases} 1 & \text{if } s_0 = c^i \\ 0 & \text{if } s_0 \neq c^i \end{cases} \quad \text{for } i = 1,2,\cdots,n. \tag{3}$$

The equation (2) means that the current state s_t is cognizable from the knowledge of the past states $\mathcal{S}_{t-1} = \{c^1, c^2, \cdots, c^n, s_1, s_2, \cdots, s_{t-1}\}$, which we call an *experience set.* Then, $\bar{\rho}_t$ is called a *capacity factor* regarding the range of cognizable states. The range becomes bigger as the positive constant $\bar{\rho}_t$ is taken greater in this model. If $\bar{\rho}_t = 0$ for all $t = 1, 2, \cdots, T$, the decision maker is conservative and the range of cognizable states at any time t is the same as the initial cognizable scope, which is the convex full of $\mathcal{S}_0 = \{c^1, c^2, \cdots, c^n\}$. For $i = 1, 2, \cdots, n$, we define a sequence of weights $\{w_t^i\}_{t=0}^T$ inductively by

$$w_0^i := \bar{w}_0^i \quad \text{and} \quad w_t^i := \bar{w}_t^i + \sum_{j=1}^{t-1} \bar{w}_t^{n+j} w_j^i \quad (t = 1, 2, \cdots, T). \tag{4}$$

Then it holds that $\sum_{i=1}^n w_t^i = 1$ and $s_t = \sum_{i=1}^n w_t^i c^i$. Let $t(= 1, 2, \cdots, T)$ be a current time. We define a fuzzy relation μ_t on $\mathcal{S}_t$ by induction on t as follows: $\mu_t := \mu_{t-1}$ on $\mathcal{S}_{t-1} \times \mathcal{S}_{t-1}$, $\mu_t(s_t, s_t) := 1$,

$$\mu_t(s_t, a) := \sum_{i=1}^n \bar{w}_t^i \mu_t(c^i, a) + \sum_{j=1}^{t-1} \bar{w}_t^{n+j} \mu_t(s_j, a) \text{ and } \mu_t(a, s_t) := \sum_{i=1}^n \bar{w}_t^i \mu_t(a, c^i) + \sum_{j=1}^{t-1} \bar{w}_t^{n+j} \mu_t(a, s_j)$$

for $a \in \mathcal{S}_{t-1}$.

Lemma. *Define a sequence of capacities $\{\rho_t\}_{t=1}^T$ by $\rho_1 := \bar{\rho}_1$ and $\rho_{t+1} := \rho_t + \bar{\rho}_{t+1}(1 + t + t\rho_t)$ for $t = 1, 2, \cdots, T-1$. Then, it holds that $-\rho_t \le w_t^i \le 1 + \rho_t$ for $i = 1, 2, \cdots, n; t = 1, 2, \cdots, T$.*

Let (Ω, P) be a probability space. Let π be a policy and let $t(= 0, 1, 2, \cdots, T)$ be a current time. Then, maps $X_t^\pi : \Omega \mapsto \mathbb{S}$ denote random variables taking values in states. We put the transition probability from a current state s_t to a next state s_{t+1} by $P_{h_t}(X_{t+1}^\pi = s_{t+1})$ when a history $h_t = (s_0, u_0, s_1, u_1, \cdots, s_{t-1}, u_{t-1}, s_t)$ is given. For $t = 1, 2, \cdots, T$, we define a scaling function

$$\varphi_t(x) := \frac{x}{2K(n,t)} + \frac{1}{2}, \tag{5}$$

where $K(n,t) := (n+1)(n + t - 2 + (n+1)\sum_{m=1}^{t-1} \rho_m)$. Then, the scaling function φ_t is a map $\varphi_t : [-K(n,t), K(n,t)] \mapsto [0,1]$. Here, we deal with only strategies such that the random variable X_t^π is represented by

$$X_t^\pi = \sum_{i=1}^n \bar{W}_t^i c^i + \sum_{j=1}^{t-1} \bar{W}_t^{n+j} s_j, \tag{6}$$

for some sequence of real random variables $\{\bar{W}_t^i\}_{i=1}^{n+t-1}$ satisfying $-\bar{\rho}_t \le \bar{W}_t^i \le 1 + \bar{\rho}_t$ $(i = 1, 2, \cdots, n + t - 1)$ and $\sum_{i=1}^{n+t-1} \bar{W}_t^i = 1$, where $\bar{W}_0^i := 1_{\{X_0^\pi = c^i\}}$ for $i = 1, 2, \cdots, n$. Let $t(= 0, 1, 2, \cdots, T)$ be a current time. We introduce total values $V_t^\pi(h_t)$ at time t by

$$V_t^\pi(h_t) := E_{h_t}\left[\sum_{m=t}^T \varphi_m(r_m(X_m^\pi))\right], \tag{7}$$

where $E_{h_t}[\cdot]$ denotes the expectation with respect to paths with a history h_t and

$$r_t(X_t^\pi) := \sum_{a \in \mathcal{S}_t} \{\mu_t(X_t^\pi, a) - \mu_t(a, X_t^\pi)\}. \tag{8}$$

Qwing to the scaling function (5), we can take a balance among the scores $\varphi_t(r_t(X_t^\pi))$ $(t = 0, 1, \cdots, T)$. The optimal total values $V_t(h_t)$ is defined by $V_t(h_t) := \sup_\pi V_t^\pi(h_t)$.

Theorem. (The optimality equation). *Let a history $h_t = (s_0, u_0, s_1, u_1, \cdots, s_{t-1}, u_{t-1}, s_t)$ for $t = 0, 1, 2, \cdots, T-1$. Then, it holds that*

$$V_t(h_t) = \sup_\pi E_{h_t}[\varphi_t(r_t(s_t)) + V_{t+1}((h_t, u_t, X_{t+1}^\pi))] \tag{9}$$

for $t = 0, 1, 2, \cdots, T-1$, and $V_T(h_T) = \varphi_T(r_T(s_T))$ at terminal time T.