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 安田正宣 YASUDA Masami
01401530 千葉大学 中神潤一 NAKAGAMI Jun-ichi
01101550 千葉大学 蕭野正美 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 $S$ be a $\sigma$-compact convex subset of some Banach space, and the states are represented by elements of $S$. The attributes of the states/objects can be represented as d-dimensional coordinates when the Banach space is taken by d-dimensional Euclidean space $\mathbb{R}^d$. Let $S$ be a subset of $S$ such that $S$ has finite elements. A map $\mu : S \times S \mapsto [0,1]$ is called a fuzzy relation on $S$. Fuzzy preferences are introduced by fuzzy relations on $S$.

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

(a) $\mu(a,a) = 1$ for all $a \in S$. (reflexive)
(b) $\mu(a,c) \geq \min\{\mu(a,b),\mu(b,c)\}$ for all $a,b,c \in S$. (transitive)
(c) $\mu(a,b) + \mu(b,a) \geq 1$ for all $a,b \in S$. (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 $\mu$ on $S$, the following map $r$ on $S$ is called a score ranking function of states induced by the fuzzy preference relation $\mu$:

\[ r(a) = \sum_{b \in S : b \neq a} \{\mu(a,b) - \mu(b,a)\}, \quad a \in S. \]  \hspace{1cm} (1)

We discuss a dynamic decision making model with fuzzy references and a time space $\{0, 1, \cdots, T\}$. Let $S_0$ be a subset of $S$ such that $S_0 := \{c^i | i = 1, 2, \cdots, n\}$ has $n$ elements and a partial order $\succ$. $S_0$ is called an initial state space and it is given as a training set in a learning model. Let $\mu_0$ be a fuzzy preference relation on $S_0$. 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 $S_0$. Define a family of states until time $t$ by $S_t := \{c^1, c^2, \cdots, c^n, s_1, s_2, \cdots, s_t\}$. $u_t (\in A_t)$ 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 $\pi : \{h_t\} \mapsto A_t$ which is represented as $\pi_t(h_t) = u_t$ for some $u_t \in A_t$. A sequence $\tau = \{\pi_t\}_{t=1}^T$ of strategies is called a policy.

Let $\{\tilde{\mu}_t\}_{t=1}^T$ 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 $s_1, s_2, \cdots, s_{t-1}$:

\[ s_t = \sum_{i=1}^n \tilde{\mu}_t^i c^i + \sum_{j=1}^{t-1} \tilde{\mu}_t^{n+j} s_j, \]  \hspace{1cm} (2)

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

\[ \tilde{\mu}_0^i := \begin{cases} 1 & \text{if } s_0 = c^i, \\ 0 & \text{if } s_0 \neq c^i \end{cases} \quad \text{for } i = 1, 2, \cdots, n. \]  \hspace{1cm} (3)
The equation (2) means that the current state $s_t$ is cognizable from the knowledge of the past states $S_{t-1} = \{c^1, c^2, \cdots, c^n, s_1, s_2, \cdots, s_{t-1}\}$, which we call an experience set. Then, $\bar{p}_t$ is called a capacity factor regarding the range of cognizable states. The range becomes bigger as the positive constant $\bar{p}_t$ is taken greater in this model. If $\bar{p}_t = 0$ 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 $S_0 = \{c^1, c^2, \cdots, c^n\}$. For $i = 1, 2, \cdots, n$, we define a sequence of weights $\{w^i_t\}_{t=0}^T$ inductively by

$$w^i_0 := w^i_0$$ and $$w^i_t := w^i_t + \sum_{j=1}^{t-1} w^{n+j}_t w^i_j \quad (t = 1, 2, \cdots, T).$$

(4)

Then it holds that $\sum_{i=1}^{n} w^i_t = 1$ and $s_t = \sum_{i=1}^{n} w^i_t c_i$. Let $t(=1, 2, \cdots, T)$ be a current time. We define a fuzzy relation $\mu_t$ on $S_t$ by induction on $t$ as follows: $\mu_t := \mu_{t-1}$ on $S_{t-1} \times S_{t-1}$, $\mu_t(s_t, s_t) := 1, \mu_t(s_t, a) := \sum_{i=1}^{n} w^i_t \mu_t(c^i, a) + \sum_{j=1}^{t-1} w^{n+j}_t \mu_t(s_j, a)$ and $\mu_t(a, s_t) := \sum_{i=1}^{n} w^i_t \mu_t(a, c^i) + \sum_{j=1}^{t-1} w^{n+j}_t \mu_t(a, s_j)$ for $a \in S_{t-1}$.

**Lemma.** Define a sequence of capacities $\{p_t\}_{t=1}^T$ by $p_1 := \bar{p}_1$ and $p_{t+1} := \bar{p}_t + \bar{p}_{t+1}(1 + t + t p_t)$ for $t = 1, 2, \cdots, T - 1$. Then, it holds that $-\bar{p}_t \leq w^i_t \leq 1 + \bar{p}_t$ for $i = 1, 2, \cdots, n, t = 1, 2, \cdots, T$.

Let $(\Omega, P)$ be a probability space. Let $\pi$ be a policy and let $t(=0, 1, 2, \cdots, T)$ be a current time. Then, maps $X^\pi_t : \Omega \rightarrow S$ 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 $P_{ht}(X^\pi_{t+1} = s_{t+1})$ 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

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

(5)

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

$$X^\pi_t = \sum_{i=1}^{n} \tilde{W}^i_t c^i + \sum_{j=1}^{t-1} \tilde{W}^{n+j}_t s_j,$$

(6)

for some sequence of real random variables $\{\tilde{W}^i_t\}_{i=1}^{n+t-1}$ satisfying $-\bar{p}_t \leq \tilde{W}^i_t \leq 1 + \bar{p}_t$ for $i = 1, 2, \cdots, n + t - 1$ and $\sum_{i=1}^{n+t-1} \tilde{W}^i_t = 1$, where $\tilde{W}^i_0 := 1(X^\pi_{t-1} = c^i)$ for $i = 1, 2, \cdots, n$. Let $t(=0, 1, 2, \cdots, T)$ be a current time. We introduce total values $V^\pi_t(h_t)$ at time $t$ by

$$V^\pi_t(h_t) := E_{ht} \left[ \sum_{m=1}^{T} \varphi_m(r_m(X^\pi_m)) \right],$$

(7)

where $E_{ht}[\cdot]$ denotes the expectation with respect to paths with a history $h_t$ and $r_t(X^\pi_t) := \sum_{a \in S_t} \{\mu_t(X^\pi_t, a) - \mu_t(a, X^\pi_t)\}$.

(8)

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

**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

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

(9)

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

---

日本オペレーションズ・リサーチ学会 ©無断複写・複製・転載を禁ず