Markov决策过程不确定策略特征模式

黄镇谨[1,2] 陆 阳[1,3] 杨 娟[1] 方 欢[1]
(合肥工业大学计算机与信息学院 合肥 230009)[1] (广西工学院计算机工程系 柳州 545006)[2]
(安徽省矿山物联网与安全监控技术重点实验室 合肥 230088)[3]

摘 要 马尔科夫决策过程可以建模具有不确定性特征的复杂系统,而在进行模型分析时需要采用策略对不确定性进行处理。首先,研究不同策略下时空有界可达概率问题,给出不确定性解决策略的定义及分类方法。其次,在时间无关策略下,证明基于确定性选取动作和随机选取动作的时空有界可达概率的一致性,并且论证了时间依赖策略相对于时间无关策略具有更好的时空有界可达概率。最后结合实例简要阐述了结论的正确性。

关键词 马尔科夫决策过程,不确定性策略,时空有界可达概率
中图法分类号 TP301 **文献标识码** A

Property Patterns of Markov Decision Process Nondeterministic Choice Scheduler

HUANG Zhen-jin[1,2] LU Yang[1,3] YANG Juan[1] FANG Huan[1]
(School of Computer & Information, Heifei Univerisity of Technology, Hefei 230009, China)[1]
(Department of Computer Engineering, Guangxi University of Technology, Liuzhou 545006, China)[2]
(The Anhui Provincial Key Laboratory of Mine IoT and Mine Safety Supervisory Control, Hefei 230088, China)[3]

Abstract Markov decision process can model complex system with nondeterminism. Schedulers are required to resolve the nonderministic choices during model analysis. This paper introduced the time-and space-bounded reachability probabilities of markov decision process under different schedulers. Firstly, the formal definition and classification method of schedulers for nonderminism were proposed and then we proved that the reachability probabilities coincide for deterministic and randomized schedulers under time-abstract. Also, it was proved that time-dependent scheduler generally induces probability bounds that exceed those of the corresponding time-abstract. At the end of paper, two cases were illustrated for describing the correctness of the conclusion.

Keywords Markov decision process, Nonderministic scheduler, Time-and space-bounded reachability probability

1 引言

大量的复杂系统都具有不确定性特征,这种固有的属性反映了系统的不同演变行为,可以用于刻画建模过程中其他方式不能表达的概念,比如调度自由、实现自由、外部环境未知、不完全信息[1]等。马尔科夫决策过程[2]用于刻画具有不确定性和概率选择的系统,并广泛应用于各个领域。对于此类系统,在评估的过程中,常把对特定属性的验证转换为对应的可达概率值的求解。由于系统的演变行为不但受模型结构本身的影响,而且与不确定性解决策略[3]息息相关,因此,对马尔科夫决策过程中不确定性解决策略进行归类并分析不同类型策略下系统可达概率值之间的关系,将有助于在实际的系统设计中采取更优化的方案。

系统性能的可满足性与从给定的状态出发到达特定目标的可达概率值有关。为了验证系统的性能指标,很多学者研究了如何求解系统的可达概率值的问题。传统的连续时间Markov链需要求解其稳态和瞬态下的时间可达概率值[4]。采用具有回报信息的Markov回报链建模系统则需要求解其时间空间可达概率值[5],但这两种建模方式都不能表达系统的不确定性。马尔科夫决策过程、交互式马尔科夫链[6]、马尔科夫博弈[8]均可用于具有不确定性属性系统的建模,对于这类系统来说,系统的可达概率最值与采取的不确定性解决策略相关。近年来,对具有不确定性特征系统的时间有界可达概率问题的研究受到了关注。文献[7]对时间无关策略下时间可达概率最值进行了研究,指出对后续动作的确定性选取方式和随机选取方式具有相同的时间可达概率最值。文献[8]分析了连续时间随机博弈下的时间有界可达概率问题。文献[9]通过求解交互式马尔科夫链下的时间有界可达概率,验证系统的性能属性。在文献[10]中,Neuhäußer采用离散化的方法求解时间相关策略下马尔科夫决策过程的时间可达

到稿日期:2012-06-20 返修日期:2012-10-05 本文受国家自然科学基金资助项目(60873195,61070220),高等学校博士点基金资助项目(20090111110002)资助。

黄镇谨(1975—),男,博士生,主要研究方向为计算机控制、形式化技术,E-mail:schzj@163.com;陆 阳(1967—),教授,博士生导师,主要研究方向为人工智能、计算机控制、传感器网络;杨 娟(1983—),女,博士生,主要研究方向为人工智能、神经网络;方 欢(1982—),女,博士生,主要研究方向为Petri网、形式化方法。

概率最值。文献[11]针对马尔科夫决策过程和连续时间马尔科夫博弈，分析了时间有界可达概率最值与最优化策略的关系并对策略的结构进行了研究。文献[12]采用基于均匀化的数值方法计算马尔科夫决策过程中时间相关策略下的时间可达概率值。在这些文献中，忽略了评估系统的另一个重要属性参数——空间信息与不确定性策略的关系。然而，实际的性能分析过程中，除了要考虑时间因素是否满足，还要考虑系统的空间约束，比如成本、带宽、能耗等。文献[13]虽然分析了 Markov 过程系统的空间性能，但并没有讨论不确定性策略类型与时空可达概率值的关系。扩展空间因素的连续时间马尔科夫回报过程能够建模具有空间信息和不确定性信息的系统。因此，分析马尔科夫回报过程中不同不确定解决策略类型下的时间空间可达概率值之间的关系具有一定的意义。在文献[7]的基础上，本文给出连续时间 Markov 回报过程中时空可达概率最值与策略之间的关系并给予证明。

本文第 2 节介绍连续时间 Markov 决策过程中不确定性解决策略的定义及分类；第 3 节详细讨论在时间无关策略下采用后续动作的确定性选取和随机选取方式的时空可达概率之间的关系，随后论证相对于时间无关策略，采用时间相关策略具有更高的时空可达概率值；第 4 节结合实例简要阐述定理的正确性；最后是结论。

2 不确定性策略定义及分类

定义 1 连续时间马尔科夫决策过程(Continue time markov decision process, CTMDP)[10]是一个四元组(S, Act, R, v)，其中 S 和 Act 为非空有限状态集和动作集，$R: S\times Act\times S\to \mathbb{R}_{\geqslant 0}$是三维转移率矩阵，$v\in Dist(S)$为初始分布函数。

CTMDP $C(S, Act, R, v)$中的有限路径为序列 $\pi=s_0 \xrightarrow{t_0,\alpha_0} s_1 \xrightarrow{t_1,\alpha_1}\cdots\xrightarrow{t_{n-1},\alpha_{n-1}} s_n$，对于任意的 $i\leqslant n$，有 $s_i\in S, \alpha_i\in Act, t_i\in \mathbb{R}_{\geqslant 0}$。$n$ 为路径 π 的长度，表示为$|\pi|$。$\pi[k]=s_k$ 和 $\delta(\pi,k)=t_k$ 分别表示路径 π 上的第 k 个状态和在 k 状态上的停留时间。令 $\Delta\pi=\sum_{k=0}^{n-1} t_k$ 为路径 π 上的累计时间消耗。记 $\pi\downarrow=s_n$ 为路径 π 上的最后一个状态。$Distr(S)$，$Distr(Act)$为基于状态和动作上的概率分布。

定义 2 设 $M(S, Act, R, v)$为 CTMDP，基于 CTMDP 的策略是映射函数 $D: Path\to Distr(Act)$，对于任意的 $\pi\in Path$，$\alpha\in Act(\pi\downarrow)$，有 $D(\pi)(\alpha)>0$。

直觉上，一个策略表示当前状态下，系统选择下一个动作所采取的准则，选择的准则可依赖于从初始状态到当前状态的路径片段上的历史信息和动作选取方式。历史信息包括状态信息、动作信息和时间信息。选取方式包括确定性选择和随机选取。下面根据策略的时间相关性和策略的确定性与否对 CTMDP 中的不确定性解决策略进行分类，给出每种类别的形式化定义其含义。

对于时间无关性策略来说，系统选择下一个动作的准则与时间信息无关，主要包括以下几种类别：

静态位置策略(Time abstract stationary position scheduler, TASP)：对于确定性策略，有 $D: S\to Act$。且 $D(s)\in Act(s)$，随机策略则为 $D: S\to Distr(Act)$，且 $D(s)(\alpha)>0, \alpha\in Act(s)$。

位置策略(Time abstract deterministic position scheduler, TAP)：对于确定性策略，$D: S\times \mathbf{N}\to Act$ 且 $D(s,n)\in Act(s)$，随机策略则为 $D: S\times \mathbf{N}\to Distr(Act)$，且 $D(s)(\alpha)>0$，$\alpha\in Act(s)$。

历史依赖策略(Time abstract history dependent deterministic scheduler, TAH)：对于确定性策略，$D: (S\times Act)^*\times S\to Act$，且 $D(s_0 \xrightarrow{\alpha_1} s_1 \xrightarrow{\alpha_2}\cdots\xrightarrow{\alpha_{n-1}} s_n)\in Act(s)$，随机策略则为 $D(s_0 \xrightarrow{\alpha_1} s_1 \xrightarrow{\alpha_2}\cdots\xrightarrow{\alpha_{n-1}} s_n)(\alpha)>0, \alpha\in Act(s)$。

定义 3 设 $M=(S, Act, R, v)$为 CTMDP。$Path^n=S\times(Act\times R_{\geqslant 0}\times S)^n$ 是M 中长度为 n 的路径集；则 $Path^*(M)=\bigcup_{n\in N} Path^n$ 为 M 中的有限路径集，$Path^\omega(M)=(S\times Act\times R_{\geqslant 0})^\omega$ 为M 中无限路径集，$Path(M)=Path^*(M)\cup Path^\omega(M)$表示 M 中的所有路径。$Path^n_{abs}=S\times(Act\times S)^n$，同理可定义 $Path^*_{abs}$、$Path^\omega_{abs}$、$Path_{abs}$。

对于时间依赖策略来说，动作的选取除了要考虑经过路径中的状态信息、动作信息外，还要考虑路径过程中所花费的时间信息。

累计时间历史依赖策略(Total time history dependent scheduler, TTH)：$D: Path^*_{abs}\times\mathbb{R}_{\geqslant 0}\to Distr(Act)$。

累计时间位置策略(Total time positional scheduler, TTP)：$D: S\times\mathbb{R}_{\geqslant 0}\to Distr(Act)$

时间位置策略(Time positional scheduler, TP)：$D: S\times\mathbb{R}_{\geqslant 0}\to Distr(Act)$，其中$\mathbb{R}$ 为最后一个迁移的延迟时间。

根据以上的定义，下面给出两个策略相等的语义：

定义 4 设 M 为 CTMDP，D 为 M 上的策略，若 $\pi,\pi'\in Path^*(M)$，则有：

$D\in TASP\Leftrightarrow\pi\downarrow=\pi'\downarrow\Rightarrow D(\pi)=D(\pi')$

$D\in TAP\Leftrightarrow(\pi\downarrow=\pi'\downarrow\wedge|\pi|=|\pi'|)\Rightarrow D(\pi)=D(\pi')$

$D\in TAH\Leftrightarrow abs(\pi)=abs(\pi')\Rightarrow D(\pi)=D(\pi')$

$D\in TTHP\Leftrightarrow(abs(\pi)=abs(\pi')\wedge\Delta(\pi)=\Delta(\pi'))\Rightarrow D(\pi)=D(\pi')$

$D\in TTP\Leftrightarrow(\pi\downarrow=\pi'\downarrow\wedge\Delta(\pi)=\Delta(\pi'))\Rightarrow D(\pi)=D(\pi')$

$D\in TP\Leftrightarrow(\pi\downarrow=\pi'\downarrow\wedge\delta(\pi,|\pi-1|)=\delta(\pi',|\pi'-1|))\Rightarrow D(\pi)=D(\pi')$

策略类型具有层级结构，不考虑参数 n，可以把一个位置策略视为静态位置策略，同样，除了长度信息，忽略路径上的其他信息，一个历史依赖策略就衍化成位置策略。确定性策略可以看成在每一状态中选择后续动作时，选择某个动作概率为 1 而其他动作的概率都为 0 的随机策略。

3 时空可达概率性质

定义 5 连续时间马尔科夫回报决策过程(Continuous-time Markov reward decision process, CMRDP)是一个五元组(S, Act, R, v, ρ)，其中(S, Act, R, v)为 CTMDP，$\rho: S\times Act\to\mathbb{R}_{\geqslant 0}$，为回报函数。

定义 6 令 $C=(S, Act, R, v)$为 CMRDP，$G\subseteq S$，$s\in S$ 且 $z, r\in\mathbb{R}_{\geqslant 0}$，则：$p^{C,G}_{\max}: S\times\mathbb{R}_{\geqslant 0}\times\mathbb{R}_{\geqslant 0}\to[0,1]: (s,z,r)\to\sup \Pr^{\omega}_{v,D}(\Diamond^{[0,z]}_{[0,r]} G)$为策略 D 下时间界 z 和空间界 r 内到达目标

状态集 G 的最大时空有界可达概率。其中，$\Pr_{v,D}$ 为测度空间 $(Paths, \mathcal{F}_{paths})$ 上的概率测度。有关 CTMDP 概率测度的概念可以参考文献[14]。

定理 1 对 CMRDP，$G\subseteq S, s\in S$，且 $z,r\in\mathbb{R}_{\geqslant 0\cup\{\infty\}}$，若 $D\in TAHD, D'\in TAHR$，则有：

$$\sup \Pr^{\omega}_{v,D}(\Diamond^{[0,z]}_{[0,r]}G)=\sup \Pr^{\omega}_{v,D'}(\Diamond^{[0,z]}_{[0,r]}G)$$

证明：设有 CMRDP M，对于任意的 TAHD 策略 D，有：

$$\Pr^{\omega}_{v,D}(\Diamond^{[0,z]}_{[0,r]}G)=\lim_{n\to\infty}\Pr^{\omega}_{v,D}(\Diamond^{[0,z]}_{[0,r]\leqslant n}G)$$

其中标注 $\leqslant n$ 表示至少在 n 步迁移之内到达目标状态集 G。根据凸组合理论，则需证明对于给定的 $n\in\mathbb{N}$，存在有限的 TAHD 策略族 $(D_i)_{i\in J_n}$（J_n 为指标集），使得在 TAHR 策略 D' 下 n 步迁移路径柱集上的测度 $\Pr^{\omega}_{v,D'}$ 为 $\Pr^{\omega}_{v,D_i}(i\in J_n)$ 的凸组合。下面用数学归纳法证明：

令 $cyl(s_0\xrightarrow{t_0}s_1\xrightarrow{t_1}\cdots\xrightarrow{t_{n-1}}s_n)$ 表示起点为 s_0 的时间路径上的柱集，则当 $n=0$ 时，$\Pr^{\omega}_{v}$ 表示为起点为 v，包含 0 个变迁的路径片段上的测度，根据路径测度的定义，对于所有的策略，有 $\Pr^{\omega}_{v}=1$。

假设当 $n=k$ 时，存在有限的 TAHD 策略族 $(D_i)_{i\in J_n}$ 和 $p_i\in[0,1]$ 且 $\sum_{i\in J_n}p_i=1$，对于所有的路径上的柱集 $C=cyl(s_0\xrightarrow{t_0,\alpha_0}s_1\xrightarrow{t_1,\alpha_1}\cdots\xrightarrow{t_{k-1},\alpha_{k-1}}s_k)$，有：

$$\Pr^{\omega}_{v,D'}(C)=\sum p_i\cdot\Pr^{\omega}_{v,D}(C)\tag{1}$$

下面证明 $n=k+1$，上式成立。

首先求解 TAHD 和 TAHR 之间的关系。令 F 为经过 $n+1$ 变迁后的有限 TAHD 策略集：

$f:(S\times Act)^{n+1}\times s\to Act$，则有 $f(\pi\to s)\in Act(s)$。对于 $f\in F$，令 $u_f:(S\times Act)^{n+1}\times s\to Dist(Act)$ 为：

$$u_f(\pi)(\alpha)=\begin{cases}1, & f(\pi)=\alpha\\ 0, & \text{otherwise}\end{cases}$$

接下来利用 u_f 构造 $n+1$ 步变迁后的 TAHR 策略 η，由 η 的定义 $\eta:(S\times Act)^{n+1}\times s\to Dist(Act)$，将 η 构造为 f 函数的有限凸组合：

$$\eta=\sum_{f\in F}q_f\cdot u_f,\ \sum q_f=1\ \text{且}\ 0\leqslant q_f\leqslant 1$$

q_f 的求解过程如下：

选择 $f_1\in F$ 且对于所有的 π 及 α，有 $f_1(\pi)=\alpha$ 和 $\eta(\pi)(\alpha)>0$，令

$$q_{f_1}=\min\{\eta(\pi)(f_1(\pi))\mid\pi\in(S\times Act)^{n+1}\times S\}$$

$$\eta_1(\pi)(\alpha)=\begin{cases}\eta(\pi)(\alpha)-q_{f_1}, & \text{当 } f_1(\pi)=\alpha\\ \eta(\pi)(\alpha), & \text{其他}\end{cases}$$

接下来选择 $f_2\in F$ 且对于所有的 π 及 α，有 $f_2(\pi)=\alpha$ 和 $\eta_2(\pi)(\alpha)>0$，令 $q_{f_2}=\min\{\eta_1(\pi)(f_2(\pi))\mid\pi\in(S\times Act)^{n+1}\times S\}$

$$\eta_2(\pi)(\alpha)=\begin{cases}\eta(\pi)(\alpha)-q_{f_2}, & \text{当 } f_2(\pi)=\alpha\\ \eta_1(\pi)(\alpha), & \text{其他}\end{cases}$$

重复上述过程直至对于所有的 π 和 α，都有 $\eta_j(\pi)(\alpha)=0$。考虑函数 η，即 $\eta(\pi)(\alpha)=D'(\pi)(\alpha)$，令 q_f 为如上所求，则对于所有 $\pi\in(S\times Act)^{n+1}\times S$，有：

$$D'(\pi)=\sum_{f\in F}q_f\cdot u_f(\pi,\cdot)$$

令柱集 $C=cyl(s_0\xrightarrow{I_0,\alpha_0}\cdots\xrightarrow{I_{n-1},\alpha_{n-1}}s_n\xrightarrow{I_n,\alpha_n}s_{n+1})$，其中 I 为任意的时间区间，同样令 $C'=cyl(s_0\xrightarrow{I_0,\alpha_0}s_1\xrightarrow{I_1,\alpha_1}\cdots\xrightarrow{I_{n-1},\alpha_{n-1}}s_n)$，则：

$$\begin{aligned}\Pr^{\omega}_{v,D'}(C)&=\Pr^{\omega}_{v,D'}(C')\cdot D'(\pi)(\alpha_n)\cdot P(s_n,\alpha_n,I_n,s_{n+1})\\&=\sum p_i\cdot\Pr^{\omega}_{v,D}(C')\cdot\sum_{f\in F}q_f\cdot u_f(\pi,\alpha_n)\cdot P(s_n,\alpha_n,I_n,s_{n+1})\\&=\sum_{(i,f)\in J_n\times F}p_i\cdot q_f\Pr^{\omega}_{v,D}(C')\cdot u_f(\pi,\alpha_n)\cdot P(s_n,\alpha_n,I_n,s_{n+1})\\&=\sum_{(i,f)\in J_n\times F}p_{i,f}\Pr^{\omega}_{v,D}(C)\end{aligned}$$

因此当 $n=k+1$ 时，有式(1)成立，根据凸组合理论，命题得证。

定理 2 对 CMRDP，$G\subseteq S, s\in S$，且 $z,r\in\mathbb{R}_{\geqslant 0\cup\{\infty\}}$，令 THD 为添加了时间因素的 TAHD 策略，则有：

$$\sup_{D\in THD}\Pr^{\omega}_{v,D}(\Diamond^{[0,z]}_{[0,r]}G)\geqslant\sup_{D\in TAHD}\Pr^{\omega}_{v,D}(\Diamond^{[0,z]}_{[0,r]}G)$$

证明：令 $L=\{x\in I\mid\rho(s)\cdot x\in K\}$，$I=[0,z]$，$K=[0,r]$，则

$$\Pr^{\omega}_{v}(\Diamond^{I}_{K}G)=\int_0^{\sup L}\sum_{s'\in S}R(s,s')\cdot e^{-E(s,s')\cdot x}\cdot\Pr^{\omega}_{s'}(\Diamond^{I-x}_{K-\rho(s)\cdot x}G)\,dx$$

不失一般性，任意选取结点 s'，s' 为从始点 s 到目标结点集 G 路径上的一点，且 $Act(s')>1$，假设从 s 到 s' 所花的时间为 $t'=\sum_{i=0}^{k-1}t_i$，$y(t')=\sum_{i=0}^{k-1}\rho(s_i)\cdot t_i$，下面证明在点 s'，采用 THD 策略时到达 G 的概率大于 TAHD 策略的到达概率。设由 s 到 s' 的概率为 $f(t')$，$Act(s')=\{\alpha_1,\alpha_2,\cdots\alpha_n\}$，对于 TAHD 策略，有 $D_i(s',t')=\alpha_i$，$i=1\cdots n$。

$$\Pr^{\omega}_{v,D_i}(\Diamond^{I}_{K}G)=f(t')\int_0^{\sup L'}\sum_{s\in S}R(s',s'',D_i)\cdot e^{-E(s',s'',D_i)}\cdot\Pr^{\omega}_{s''}(\Diamond^{I-t'-x}_{K-y(t')-\rho(s')\cdot x}G)\,dx$$

令 $G_i(t')=\Pr^{\omega}_{v,D_i}(\Diamond^{I}_{K}G)$，因此有

$$\sup_{D\in TAHD}\Pr^{\omega}_{v,D}(\Diamond^{I}_{K}G)=\sup\Pr^{\omega}_{v,D_i}(\Diamond^{I}_{K}G)$$

在时间区间 $[0,I-t']$ 内，设函数 $G_i(t')$ 之间存在交点 t_1，t_2，$\cdots$，t_n，当 $D\in THD$ 时，对于每个时间段 $t_{i-1}\leqslant I-t'<t_i$，取

$$\Pr^{\omega}_{v,D}(\Diamond^{[0,z]}_{[0,r]}G)=\sup\Pr^{\omega}_{v,D_i}(\Diamond^{I}_{K}G)$$

因此，在整个时间区间 $[0,I-t']$，均有：

$$\sup_{D\in THD}\Pr^{\omega}_{v,D}(\Diamond^{[0,z]}_{[0,r]}G)\geqslant\sup_{D\in HD}\Pr^{\omega}_{v,D}(\Diamond^{[0,z]}_{[0,r]}G)$$

4 实例

下面分别就定理 1 和定理 2 给出一个简单的例子。针对定理 1，简要描述如何确定 TAHD 策略族凸组合的权系数 q_f。

例 1 设有路径 σ,σ'，$|\sigma|=|\sigma'|=n+1$，$Act(\sigma\downarrow)=\{\alpha,\beta,\delta\}$，$Act(\sigma'\downarrow)=\{\alpha,\gamma\}$，若有 HR 策略 $D'=\{\mu(\sigma)(\alpha)=\frac{3}{6},\mu(\sigma)(\beta)=\frac{1}{6},\mu(\sigma)(\delta)=\frac{2}{6},\mu(\sigma')(\alpha)=\frac{3}{5},\mu(\sigma')(\gamma)=\frac{2}{5}\}$，则构造系数 q_f：

(1)选取 $f_1(\sigma)=f_1(\sigma')=\alpha$，令 $q_{f_1}=\min(\eta(\pi)(\alpha))=\min(\frac{3}{6},\frac{3}{5})=\frac{1}{2}$，因此 $\eta_1(\sigma)(\alpha)=0$，$\eta_1(\sigma')(\alpha)=\frac{1}{10}$，其余 $\eta_1(\cdot)(\cdot)=\eta(\cdot)(\cdot)$。

(2)选取 $f_2(\sigma)=\beta, f_2(\sigma')=\alpha, q_{f_2}=\min(\eta_1(\pi)(\alpha))=\min(\frac{1}{6},\frac{1}{10})=\frac{1}{10}$，因此 $\eta_2(\sigma)(\beta)=\frac{1}{15}, \eta_2(\sigma')(\alpha)=0, \eta_2(\cdot)(\cdot)=\eta_1(\cdot)(\cdot)$。

依次类推，最后得 $q_{f_3}=\{\frac{1}{15}\mid f_3(\sigma)=\beta, f_3(\sigma')=\gamma\}$，$q_{f4}=\{\frac{1}{3}\mid f_4(\sigma)=\beta, f_4(\sigma')=\gamma\}$。由定理 1，任选 D' 下的任一策略 $\mu(\sigma)(\beta)$，有 $D'(\sigma)(\beta)=\sum_{f\in F}q_f\cdot u_f(\sigma,\beta)$，以上计算结果：$\sum_{f\in F}q_f\cdot u_f(\sigma,\beta)=q_{f_1}\cdot u_{f_1}(\sigma,\beta)+q_{f_2}\cdot u_{f_2}(\sigma,\beta)+q_{f_3}\cdot u_{f_3}(\sigma,\beta)+q_{f_4}\cdot u_{f_4}(\sigma,\beta)=\frac{1}{2}\cdot 0+\frac{1}{10}\cdot 1+\frac{1}{15}\cdot 1+\frac{1}{3}\cdot 1=\frac{3}{6}$，与 HR 策略中的假设值相等。

例 2　设有 CMRDP，如图 1 所示，初始点 $v=[s_0]$，目标结点 $G=[s_3]$，$Act=\{\alpha,\beta,\delta,\gamma\}$。不失一般性，假设 $D(s_0,0)=\{\alpha\mapsto 1\}$，选取 s_1 分析，设在 s_0 停留的时间为 t_0，令 $z_1=z-t_0, r_1=r-\rho(s_0)\cdot t_0, L_1=\{x_1\in I_1\mid \rho(s_1)\cdot x_1\in K_1\}$，其中 $I_1=[0,z_1], K_1=[0,r_1]$，则在 L_1 内通过不同的 $Act=\{\alpha,\delta,\gamma\}$ 由 s_1 到 s_3 的概率分别为：$P_\alpha(x)=\int_0^{\sup L_1}e^{-x}\,dx$，$P_\beta(x)=\int_0^{\sup L_1}(3e^{-3t_1}\cdot\int_0^{\sup L_2}e^{-t_2}\,dt_2)\,dt_1$，$P_\delta(x)=\int_0^{\sup L_1}(8e^{-8t_1}\cdot\frac{1}{2}\int_0^{\sup L_2}2e^{-2t_2}\,dt_2)\,dt_1$，其中 $L_2=\{x_2\in I_2\mid \rho(s')\cdot x_2\in K_2\}$，$I_2=[0,z_2]$，$K_2=[0,r_2]$，$z_2=z-t_0-t_1$，$r_2=r-\rho(s_0)\cdot t_0-\rho(s')\cdot t_1$，$s'$ 分别为 s_3,s_4。设 L_1 和 L_2 都是非空集合，则 $P_\alpha(x)=1-e^{-x}$，$P_\delta(x)=\frac{1}{2}-\frac{2}{3}e^{-2x}+\frac{1}{6}e^{-8x}$，$P_\beta(x)=1+\frac{1}{2}e^{-3x}-\frac{3}{2}e^{-x}$。因此，当 $x=z-t_0$ 取不同值时，到达 s_3 的时空可达概率如图 2 所示。由图可知，当 $x=z-t_0<t_1$ 时，取 $D(s_1,t_0)=\{\alpha\mapsto 1\}$，$t_1\leqslant z-t_0<t_2$ 时，取 $D(s_1,t_0)=\{\delta\mapsto 1\}$，$t_2\leqslant z-t_0$ 时，取 $D(s_1,t_0)=\{\gamma\mapsto 1\}$，则 $\Pr_v^{\sigma}(\Diamond^{\leqslant x}G)$ 的概率最大。

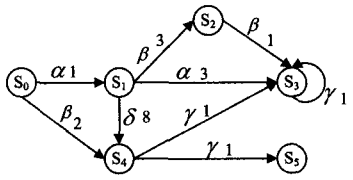

图 1　CMRDP 模型

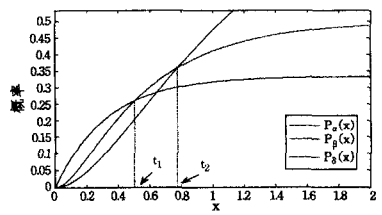

图 2　在 $z-t_0$ 时间内时空可达概率

结束语　马尔科夫决策过程可用于复杂系统的建模。系统的演变行为不但与模型结构有关，而且受到不确定性解决策略的影响。本文给出了马氏决策过程不确定性解决策略的形式化定义，给出了判别相等策略的方法。由前述可知，在时间无关条件下，从初始结点到目标结点集，对后续动作的随机选取和确定性选取具有相同的时空有界可达概率，考虑了时间因素的时间相关策略比时间无关策略更加灵活，并具有更大的时空有界可达概率。由于时间相关策略的优越性，进一步的工作主要研究在具体的时间策略下如何求解时空有界可达概率的值，同时对于给定的系统属性，给出模型验证的方法。

参 考 文 献

[1] 钮俊，曾国荪，吕新荣. 随机模型检测连续时间 Markov 过程[J]. 计算机科学，2011，38(9)：112-115

[2] Puterman M L. Markov Decision Processes：Discrete Stochastic Dynamic Programming[M]. NewYork：Wiley，1994

[3] Markus B. Optimal Schedulers for Time-Bounded Reachability in CTMDPs[D]. Homburg：Saarland University，2009

[4] Baier C，Cloth L，Haverkort B，et al. Model checking Markov chains with aciton and state labels[J]. IEEE Transactions on Software Engineering，2007，33(4)：209-224

[5] Cloth L. Model checking algorithms for Markov reward models [D]. Enschede，University of Twente，2006

[6] Hermanns H. Interactive Markov Chains [M]. Heidelberg：Springer，2002

[7] Baier C，Hermanns H，Katoen J-P，et al. Efficient computation of time-bounded reachability probabilities in uniform continuous-time Markov decision processes[J]. Theoretical Computer Science，2005，345(1)：2-26

[8] Brazdil T，Forejt V，Krcal J，et al. Continuous-time stochastic games with time bounded reachability[A] // Conference on Foundations of Software Technology and Theoretical Computer Science，2009[C]. 2009：61-72

[9] Lijun Z，Neuhäußer M R. Model checking interactive Markov chains[A]//Proceedings of TACAS，2010[C]. 2010：53-68

[10] Neuhäußer M-R，Zhang L. Time-bounded reachability probabilities in continuous-time Markov decision processes[A]//7th International Conference on Quantitative Evaluation of Systems，2010[C]. 2010：209-218

[11] Rabe M，Schewe S. Finite optimal control for time-bounded reachability in CTMDPs and continuous-time Markov games[A]// CoRR，2010[C]. 2010：1004-4005

[12] Buchholz P，Hahn E M，Hermanns H，et al. Model Checking Algorithms for CTMDPs[A] // Proceedings of the 23rd international conference on Computer aided verification，CAV'11，2011[C]. Berlin，Heidelberg：Springer-Verlag，2011：225-242

[13] 钮俊，曾国荪，王伟. 基于模型检测的时间空间性能验证方法[J]. 计算机学报，2010，33(9)：1624-1633

[14] Neuhäußer M-R. Model Checking Nondeterministic and Randomly Timed Systems[D]. Aachen，RWTH Aachen University，2010