DOI:10.11721/cqnuj20190514

具有风险厌恶型决策者的有限阶段马尔可夫决策过程*

楼振凯[1], 楼旭明[2], 侯福均[1]
(1. 北京理工大学 管理与经济学院, 北京 100081; 2. 西安邮电大学 经济与管理学院, 西安 710121)

摘要:【目的】在分析了期望最大化准则无法控制方差的局限性的基础上,考虑具有风险厌恶型决策人的有限阶段马尔可夫决策过程,为风险厌恶决策者提供决策方法。【方法】建立了悲观准则下有限阶段马尔可夫决策过程的数学模型,并基于动态规划原理和同向不等号相加的保号性给出了向后递推算法。【结果】得到了每个阶段所有可能状态的最优策略和到阶段结束至少可获得的报酬,并证明所得到的最优策略矩阵满足风险厌恶型决策者的要求。然后,针对连续性策略、成本最小化和风险偏好型决策者等情形下有限阶段马氏过程最优策略的求解进行了一些理论延伸。【结论】给出了一个三阶段马尔可夫过程的算例分析,验证了所提出的模型。

关键词:风险厌恶;悲观准则;动态规划;向后递推算法;最优策略矩阵
中图分类号:O221 **文献标志码**:A **文章编号**:1672-6693(2019)05-0086-06

马尔可夫决策过程是一种典型的随机动态规划,由于它具有无后效性、序贯决策的特点,因此被广泛地应用于各个领域[1-2]。根据决策时域范围的不同,可以将马尔可夫决策过程分为无限阶段过程和有限阶段过程。无限阶段马尔可夫过程的模型主要有折扣模型和平均模型,本文研究的对象是有限阶段马尔可夫决策过程。

近些年来,对有限阶段马尔可夫决策过程的研究文献较多。Nielsen 等人[3]研究了寻找第 k 个最优的马尔可夫策略的问题,并结合有向图建立了期望成本最小化的模型。Lovejoy[4]对部分可观测的有限阶段马尔可夫决策过程进行了研究,建立了期望报酬最大化的模型,并提出了求解模型的几种近似算法。另外,Monahan[5]和 Smallwood 等人[6]在期望报酬最大化准则下对部分可观测的有限阶段马尔可夫决策过程的控制问题进行了相关研究。王薇等人[7]将可变限速控制问题建模成有限阶段马氏过程,并通过仿真验证该模型对提高交通流量的有效性。李江波等人[8]基于风险理论建立效用函数,并用有限阶段马氏过程解决电网的实时电价问题。

上述文献都以期望准则作为决策的基础,另外一些学者注意到有限阶段马尔可夫决策中的方差问题。Henderson 等人[9]对马尔可夫过程进行模拟仿真,并提出了控制变量来减少方差的措施。Sobel[10]以最大化均值—方差系数为目标函数,对有限阶段马尔可夫决策过程进行研究。Filar 等人[11]提出了带有方差惩罚的马尔可夫决策问题,并与折扣模型和平均模型相结合,寻找决策过程的最优策略。

方差的存在伴随着决策过程中收益波动的发生,而人们在面对收益波动或损失风险时常表现出较高的敏感性[12]。因此对于风险厌恶的决策者而言,控制决策过程的风险显得十分重要。Shen 等人[13]分别对无限阶段的折扣模型和平均模型提出对应的风险敏感性标准,给出了相应的算法并证明最优解的存在性。Ruszczynski[14]和 Cavus 等人[15]都研究了具有风险厌恶决策者的马尔可夫风险控制问题,并给出了动态递归算法。

本文对决策人为风险厌恶型的有限阶段马尔可夫过程进行分析,基于动态规划原理给出了向后递归算法,根据同向不等号相加的保号性,得到每个阶段所有可能状态的最优策略和到阶段结束至少可获得的报酬。求解过程类似于决策者和自然的博弈,每一步向后递推都采取悲观准则。对于风险厌恶型决策者而言,知道每个阶段所有可能状态采取什么策略能保证至少获得的报酬最多,比知道获得更多期望报酬的策略更重要。

* **收稿日期**:2018-12-05 **修回日期**:2019-08-30 **网络出版时间**:2019-09-26 11:24
资助项目:国家自然科学基金面上项目(No. 71571019)
第一作者简介:楼振凯,男,博士研究生,研究方向为决策理论与应用,E-mail: louzk@bit.edu.cn;**通信作者**:楼旭明,男,教授,E-mail: louxuming@xupt.edu.cn
网络出版地址:http://kns.cnki.net/kcms/detail/50.1165.N.20190926.1123.010.html

1 有限阶段马氏决策描述

本文讨论的有限阶段马尔可夫决策过程为状态数量有限、策略数有限且齐次的。为了便于讨论，对这类决策过程做出一些描述[1,16]。

1.1 符号说明

1）选取策略的时间点称为决策时刻，用 T 表示所有决策时刻的点集，有限阶段马氏决策过程的决策时刻集记为 $T=\{0,1,\cdots,N\}$，N 为阶段数。

2）所有可能状态的集合称为状态空间，记有限状态空间 $\Omega=\{1,2,\cdots,k\}$，第 n 时刻的状态记为 $S_n=i,i\in\Omega$。

3）状态 i 对应的可用策略集记为 $A(i)$，也称为行动空间，令 $A=\bigcup_{i\in\Omega}A(i)$ 为所有可用策略集之并。一个 N 阶段马氏决策规则序列 $\pi=\{\pi_0,\pi_1,\cdots,\pi_N\}$ 称为随机马氏策略，其中 π_t 是决策时刻 t 的策略，它与 t 以前的行动及状态无关。全体随机马氏策略组成的集合记作 $\Delta,\pi\in\Delta$。

4）在任何一个决策时刻 n，决策过程处于状态 $i\in\Omega$，在采取行动 $a\in A(i)$ 后，下一时刻转移到状态 j 的概率记为 $p(j\mid i,a)$，记 a_n 为第 n 时刻采取的行动。

5）在任何一个决策时刻 n，决策过程处于状态 $i\in\Omega$，在采取行动 $a\in A(i)$ 后，下一时刻转移到状态 j 所能获得的报酬记为 $r(i,a,j)$。

上述符号构成的五元组 $\{T,\Omega,A(i),p(j|i,a),r(i,a,j)\}$ 称为一个有限阶段马尔可夫决策过程。

另外，相继的状态和行动组成马尔可夫决策的一条轨迹，从 0 时刻到 t 时刻的一条轨迹记为 $h_t:=(S_0,a_0,S_1,a_1,\cdots,S_{t-1},a_{t-1},S_t),t\geqslant 0$。

1.2 模型和迭代算法

对 $N\geqslant 0$，初始状态为 $i\in\Omega$，在随机策略 π 下 N 阶段期望总报酬定义为：

$$U_N(i,\pi)=\sum_{n=0}^{N-1}E_\pi^i[r(S_n,a_n)]+E_\pi^i[r(S_N)]。\tag{1}$$

这里 $r(S_N)$ 为过程终止时刻的剩余价值，某些情况下 $r(S_N)=0$。$r(S_n,a_n)$ 表示一步期望函数，当报酬函数既依赖于当前状态与当前策略，也与下一时刻的状态有关时，$r(i,a_n)=\sum_{j\in\Omega}r(i,a_n,j)p(j\mid i,a_n)$。

为了得到最优报酬的递推公式，在随机策略 π 下从时刻 t 到时刻 N 的期望报酬之和定义为：

$$u_t^\pi(h_t)=E_\pi\Big\{\sum_{n=t}^{N-1}r(S_n,a_n)+r(S_N)\mid h_t,S_t=i\Big\}。\tag{2}$$

下面给出有限阶段马氏决策过程的迭代算法。

步骤 1，令 $t=N$，对一切 $h_N=(h_{N-1},a_{N-1},S_N)$，$u_N{}^\pi(h_N)=r(S_N)$。

步骤 2，如果 $t=0$，停止；否则，令 $t-1=t$，进入下一步。

步骤 3，对 t 时刻的每个状态 $i\in\Omega$ 和每条轨迹 $h_t=(h_{t-1},a_{t-1},S_t)$，计算 $u_t^\pi(h_t)$：

$$u_t^\pi(h_t)=r(i,a(h_t))+\sum_{j\in\Omega}p(j\mid i,a(h_t))u_{t+1}^\pi(h_t,a(h_t),j),\tag{3}$$

$a(h_t)$ 表示 t 时刻行动的选择依赖轨迹 h_t。

步骤 4，返回步骤 2。

算法中 $u_t^\pi(i)$ 和 $U_N(i,\pi)$ 的区别在于，前者是从时刻 t 到决策过程终止的总报酬。因此，对 $\forall i\in\Omega$，有 $u_0^\pi(i)=U_N(i,\pi)$。

在本文设定的状态数量和策略数量离散有限的前提下，对 $i\in\Omega$，记最优策略 π^*，则：

$$\pi^*(i)\in\arg\max_{\pi\in\Delta}u_0^\pi(i)。\tag{4}$$

2 悲观准则下有限阶段马氏决策

2.1 有限阶段期望准则的局限性

有限阶段马尔可夫决策通常采用期望报酬最大化或期望成本最小化准则，然而对风险厌恶型决策者来说，该准则存在一定的局限性。

记 n 时刻期望最大的策略为 a_n^*,根据期望报酬最大化准则得到最优策略集 A^*:

$$A^*=\{a_n^* \mid a_n^* \in \arg\max_{a_n \in A_i} r(i,a_n)\}; \tag{5}$$

记 $v(i,a_n)$为采取策略 a_n 时的方差,则有:

$$v(i,a_n)=\sum_{j\in\Omega}(r(i,a_n,j)-r(i,a_n))^2 p(j \mid i,a_n); \tag{6}$$

记 n 时刻方差最小的策略为 a_n':

$$A'=\{a_n' \mid a_n' \in \arg\min_{a_n \in A_i} v(i,a_n)\}。 \tag{7}$$

一般来说,E-V 准则希望找到在最大化期望的同时最小化方差的解。然而多数情况下 E-V 准则很难满足,因此有些情况下会发生 $A^* \cap A=\varnothing$的情形。对于风险厌恶型决策者来说,提高了期望报酬,增加了报酬的方差,并不是最好的策略。

此外,风险厌恶型决策者在做决策的时候通常会采用悲观准则,即在决策过程中,每一步决策之前想知道最少能获得的报酬是多少。在这种情况下,期望报酬最大化准则给出的最优策略集并没有多大参考价值。

2.2 考虑悲观准则的向后迭代算法

对于价值或报酬的悲观准则又称极大化极小准则,是 Wald 在 1950 年提出的[17]。对一步期望报酬函数来说,考虑悲观准则指的是,在 n 时刻选择策略 a_n^*,则有:

$$p(j|i,a_n)>0,\ r(i,a_n^*)=\max_{a_n\in A_i}\min_{j\in\Omega} r(i,a_n,j),j\in\Omega。 \tag{8}$$

基于悲观准则,给出具有风险厌恶型决策者的有限阶段马尔可夫决策过程的迭代算法,形式上类似于期望报酬最大化的向后递归过程。

步骤 1,令 $t=N$,此时效用函数 $u_N(S_N)=r(S_N)$,$S_N\in\Omega$,为了方便后续递推,不妨记 $u_N(S_N)\geqslant r(S_N)$。

步骤 2,如果 $t=0$,停止;否则,令 $t-1=t$,进入下一步。

步骤 3,对 t 时刻的任意状态 i,在悲观准则下到决策过程结束至少能获得的报酬记为 $u_t(S_t)$,如果 $p(j|i,a_t)>0$,则最优策略 a_t^* 由

$$a_t^* \in \arg\max_{a_t\in A_i}\min_{j\in\Omega}\{r(i,a_t,j)+u_{t+1}(S_{t+1})\} \tag{9}$$

确定。采取策略 a_t^* 后可以得到报酬 $u_t(S_t)$的取值下限:

$$u_t(S_t)\geqslant r(i,a_t^*,S_{t+1}) + u_{t+1}(S_{t+1})。 \tag{10}$$

步骤 4,返回步骤 2。

通过上述算法,可以得到悲观准则下每个阶段所有可能状态的最优策略和到阶段结束至少可获得的报酬。要确定所得到的策略和报酬下限符合风险厌恶型决策者的要求,需要证明以下两个结论。

结论 1 每个阶段所有可能状态在最优策略下所获得的报酬下限均可达到,即无法提高任何一个报酬的下限值。

证明 不失一般性,设 m 时刻采取最优策略 a_m^* 所能获得报酬下限为 $u_m(S_m)$,$m\in\{0,1,\cdots,N-1\}$,则根据算法可以递推得到:

$$u_m(S_m)\geqslant r(S_m,a_m^*,S_{m+1})+u_{m+1}(S_{m+1})\geqslant\sum_{n=m}^{N-1} r(S_n,a_n^*,S_{n+1})+r(S_N)。 \tag{11}$$

根据算法的规则,每一步递推中报酬函数对应的转移概率 p 都大于 0,即:

$$\forall m\in\{0,1,\cdots,N-1\},p(S_{m+1}\mid S_m,a_m^*)>0\Rightarrow\prod_{n=m}^{N-1} p(S_{n+1}\mid S_n,a_n^*)>0。 \tag{12}$$

上述两式表明,在 m 时刻采取最优策略 a_m^* 时,获得报酬下限 $\sum_{n=m}^{N-1} r(S_n,a_n^*,S_{n+1})+r(S_N)$ 是可能达到的。 证毕

结论 2 在 m 时刻采取除了 a_m^* 以外的策略,所获得的报酬的下限都不可能高于采取 $u_m(S_m)$。

证明 根据悲观准则,这个结论的正确性是显然的。事实上,假设存在策略 a_m',使得过程在 m 时刻采取策略 a_m' 所获得报酬的下限大于 $u_m(S_m)$,则 a_m^* 不满足最优策略的定义,即:

$$a_m^* \neq \arg\max_{a_m\in A_i}\min_{j\in\Omega}\{r(i,a_m,j)+u_{m+1}(S_{m+1})\}。$$

至此可以确定，所得到的策略和报酬下限符合风险厌恶型决策者的要求。不同于期望报酬最大化准则得到的序贯策略集，基于悲观准则的向后迭代算法得到的是策略矩阵，矩阵中每一个元素为某阶段到达某状态时在悲观准则下应该采取的最佳策略。策略矩阵表示如下：

$$\begin{bmatrix} a^*_{0,1}(u_0(1)) & a^*_{0,2}(u_0(2)) & \cdots & a^*_{0,k}(u_0(k)) \\ a^*_{1,1}(u_1(1)) & a^*_{1,2}(u_1(2)) & \cdots & a^*_{1,k}(u_1(k)) \\ \vdots & \vdots & \vdots & \vdots \\ a^*_{N-1,1}(u_{N-1}(1)) & a^*_{N-1,2}(u_{N-1}(2)) & \cdots & a^*_{N-1,k}(u_{N-1}(k)) \end{bmatrix}。$$

其中任意元素 $a^*_{i,j}(u_i(j))$ 表示在 i 时刻当状态为 j 时，采取最佳策略 $a^*_{i,j}$ 的情况下，所能获得的报酬的下限为 $u_i(j)$。　　证毕

2.3 相关理论延伸

下面对悲观准则下向后迭代算法做一些理论延伸。

若策略为连续有界变量，报酬为策略的连续函数，则仍然可以运用悲观准则的向后迭代算法。不同的是，外层的 max 需要对连续函数在闭区间内求极大值。

悲观准则下的向后迭代算法除了应用于求解本文中的报酬问题，还可以用于求解成本问题。所不同的是，对成本问题需要采用 min max 准则，其余过程完全一样，且能得到类似的结果。

对于风险偏好型决策者而言，还可以对有限阶段马尔可夫决策过程采用乐观准则，即对每阶段所有可能的状态所获得的报酬采用 max max 准则，类似于上一节的算法步骤，得到乐观准则下的向后迭代算法，进而得到乐观准则下每个阶段所有可能状态的最优策略和到阶段结束最多可获得的报酬。

另外，本文给出的算法中并没有考虑当多个策略都满足(9)式的情况下如何比较哪个策略更好。结合风险厌恶型和风险偏好型决策结果，可以得到每个阶段所有可能状态所能获得报酬的取值范围。在策略下限相等的情况下，可以比较策略的上限值，取上限较高的那个策略。

3 算例分析

考虑一个具有 3 个状态 $\Omega=\{1,2,3\}$，3 个阶段的马尔可夫决策过程。假设状态之间的转移概率不随策略的不同而发生变化(当采取的策略影响转移概率时，需要已知不同策略下的转移概率矩阵，其他过程不变)，转移概率矩阵为 $\boldsymbol{P}=\begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{bmatrix}$。过程终止时不同状态的剩余价值为：$u_3(1)=r(S_3=1)=1$，$u_3(2)=r(S_3=2)=0$，$u_3(3)=r(S_3=3)=3$。

在采取不同的策略 a_1,a_2,a_3 情况下，不同状态之间转移产生报酬矩阵分别为 $\begin{bmatrix} 2 & 1 & 3 \\ 1 & 4 & 2 \\ 3 & 2 & 1 \end{bmatrix}$，$\begin{bmatrix} 3 & 0 & 2 \\ 2 & 1 & 2 \\ 4 & 2 & 1 \end{bmatrix}$，$\begin{bmatrix} 5 & 3 & 2 \\ 3 & 2 & 1 \\ 1 & 4 & 2 \end{bmatrix}$。

采用悲观准则下的向后迭代算法，第三阶段起始状态 $S_2=1$ 时，根据悲观准则，策略 a_1 下能获得的最低报酬为 3，策略 a_2 下能获得的最低报酬为 4，a_3 下能获得的最低报酬为 5，根据(9)式有：

$$\arg\max_{a_i\in A}\min_{j\in\Omega}\{r(1,a_i,j)+u_3(S_3)\}=a_3。$$

类似可得到 $S_2=2$ 时的最优策略为 a_1，对应的报酬下限为 2；$S_2=3$ 时的最优策略为 a_1，所能获得的报酬下限为 2。依次类推，得到策略-报酬矩阵如下：

$$\begin{bmatrix} a^*_{0,1}(u_0(1)) & a^*_{0,2}(u_0(2)) & \cdots & a^*_{0,k}(u_0(k)) \\ a^*_{1,1}(u_1(1)) & a^*_{1,2}(u_1(2)) & \cdots & a^*_{1,k}(u_1(k)) \\ a^*_{3,1}(u_3(1)) & a^*_{3,2}(u_3(2)) & \cdots & a^*_{3,k}(u_3(k)) \end{bmatrix} = \begin{bmatrix} a_1(7) & a_1(6) & a_3(6) \\ a_1(5) & a_1(4) & a_3(4) \\ a_3(5) & a_1(2) & a_1(2) \end{bmatrix}。$$

策略矩阵中,个别策略的下限值相等的情形下只取了下标号较小的策略。

4 结论

本文对具有风险厌恶决策人的有限阶段马尔可夫决策过程进行了研究,本文的贡献可以总结如下。

1) 分析了期望准则具有方差不可控以及无法确定地给出最低报酬等局限性,指出该准则不适用于具有风险厌恶型决策者的有限阶段马氏决策。

2) 给出悲观准则,并设计了基于悲观准则的向后迭代算法,得到每个阶段所有可能状态下的最优决策和至少能获得的报酬,并证明了该算法满足风险厌恶型决策者的需要。

3) 分别从策略连续、成本效用函数和风险偏好等3个方面对模型和算法做了一些延伸,并指出这些延伸本质上和本文给出的过程类同。

由于是建立在悲观准则的基础上,本文并没有考虑策略发生的概率,只考虑发生和不发生。在策略下限发生概率较低的情况下,决策人仍然考虑该事件的发生,可以认为决策人是极端悲观的。

另外,本文是在风险厌恶情况下对有限阶段的期望准则做出的改进,并没有讨论无限阶段的情形,且所有参数都是已知的。Suresh 等人[18]和 Wolfram 等人[19]将 max-min 准则应用于求解参数不确定的无限阶段马氏过程鲁棒性决策问题,关于这方面理论仍然有一些未完美解决的地方,后续的研究将对该问题进行深入分析。

参考文献:

[1] 刘克,曹平. 马尔可夫决策过程理论与应用[M]. 北京:科学出版社,2015.
LIU K,CAO P. Theory and application of Markov decision process[J]. Beijing:Science Press,2015.

[2] HAZEGHI K. Markov decision processes:discrete stochastic dynamic programming[J]. Journal of the American Statistical Association,1995,90:392-429.

[3] NIELSEN L R,KRISTENSEN A R. Find the k best policies in a finite-horizon Markov decision process[J]. European Journal of Operational Research,2006,175(2):1164-1179.

[4] LOVEJOY W S. A survey of algorithmic methods for partially observed Markov decision processes[J]. Annals of Operations Research,1991,28(1):47-66.

[5] MONAHAN G E. State of the art:a survey of partially observable Markov decision processes:theory,models,and algorithms[J]. Management Science,1982,28(1):1-16.

[6] SMALLWOOD R D,SONDIK E J. The optimal control of partially observable Markov processes over a finite horizon[J]. Operations Research,1973,21(5):1071-1088.

[7] 王薇,杨兆升,赵丁选. 有限阶段马尔可夫决策的可变限速控制模型[J]. 交通运输工程学报,2011,11(5):109-114.
WANG W,YANG Z S,ZHAO D X. Control model of variable speed limit based on finite horizon Markov decision making[J]. Journal of Traffic and Transportation Engineering,2011,11(5):109-114.

[8] 李江波,王波,高岩,等. 马尔可夫决策过程下的智能电网实时电价模型[J]. 系统仿真学报,2016,28(11):2756-2763.
LI J B,WANG B,GAO Y,et al. Optimal real-time pricing model of smart grid based on Markov decision process[J]. Journal of System Simulation,2016,28(11):2756-2763.

[9] HENDERSON S G,GLYNN P W. Approximating martingales for variance reduction in Markov process simulation[J]. Mathematics of Operations Research,2002,27(2):253-271.

[10] SOBEL M J. The variance of discounted Markov decision processes[J]. Journal of Applied Probability,1982,19(4):794-802.

[11] FILAR J A,KALLENBERG L C M,LEE H M. Variance-penalized Markov decision processes[J]. Mathematics of Operations Research,1989,14(1):147-161.

[12] TOM S M,FOX C R,TREPEL C. The neural basis of loss aversion in decision-making under risk[J]. Science,2007,315(5811):515-518.

[13] SHEN Y,STANNAT W,OBERMAYER K. Risk-sensitive Markov control processes[J]. SIAM Journal on Control and Optimization,2013,51(51):3652-3672.

[14] RUSZCZYNSKI A. Risk-averse dynamic programming for Markov decision processes[J]. Mathematical Programming,2010,125(2):235-261.

[15] CAVUS O,RUSZCZYNSKI A. Risk-averse control of undiscounted transient Markov models[J]. SIAM Journal on

Control and Optimization,2014,52(6):3935-3966.

[16] PUTERMAN M L. Stochastic models[M]. Netherlands: Elsevier,1990.

[17] BALLESTERO E. Strict uncertainty:a criterion for moderately pessimistic decision makers[J]. Decision Sciences, 2002,33(1):87-108.

[18] SURESH K,EDVIN K P C,NESS B S. Markov decision processes with uncertain transition rates: sensitivity and robust control[C]//Proceeding of the 41st IEEE Conference on Decision and Control. Las Vegas,US:IEEE,2002.

[19] WOLFRAM W,DANIEL K,RUSTEM B. Robust Markov decision processes[J]. Mathematics of Operations Research,2013,38(1):153-183.

Finite Horizon Markov Decision Processes for Risk-averse Decision Makers

LOU Zhenkai[1], LOU Xuming[2], HOU Fujun[1]

(1. School of Management and Economics, Beijing Institute of Technology, Beijing 100081;
2. School of Economics and Management, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

Abstract: In dual-channel supply chain, how to sell heterogeneous products in direct channel and traditional channel reasonably is an issue to manufacturer. It is of a great significance to design product layout in business practice. [**Methods**]Mathematical models and backward induction were used to research manufacturers' product layouts and pricing when they introduce online channels, and the influence of relevant parameters on layout was studied by numerical analysis. [**Findings**]Results show that, consumers' acceptance of direct channel, high-end products' quality and cost will influence the strategy of manufacturer's online product layout. When the consumers' acceptance of direct channel and the quality of high-end products are high, the cost is low, manufacturer will choose the layout of high-end products. Otherwise, it will choose another layout. Direct price of manufacturer will increase and retail price of two retailers will decrease with the increase of consumers' acceptance of direct channel. [**Conclusions**]There is no strict dominant layout strategy for manufacturers, all kinds of factors are supposed to be considered.

Keywords: risk averse; pessimism criterion; dynamic programming; backward recurrence algorithm; optimal policy matrix

(责任编辑 黄 颖)