基于马尔可夫决策的穿越沙漠游戏策略研究

邓晨晨　黄宇轩　齐泽坤　杨　松

摘要：“穿越沙漠”游戏是一款综合考虑资金、资源、天气、时间、博弈等多种因素在内的复杂策略游戏。文章将基于图论与马尔可夫决策有关模型，分析讨论玩家在未来信息已知与未来信息未知两种情形下的最优策略。该模型综合考虑了风险评估与多阶段决策理论，可为优化算法与企业决策提供一定借鉴意义。

关键词：沙漠掘金；图论；动态规划；马尔可夫决策；最优化理论

一、引言

“穿越沙漠”游戏是一款综合考虑资金、资源、天气、时间、博弈等多种因素在内的多阶段策略游戏。游戏要求玩家在沙暴天气原地停留、到达矿山当天不许挖矿并且保证在路途中不得耗尽资源。游戏允许玩家挖矿获得收益，并利用初始资金及收益在村庄随时补给资源。玩家必须在截止日期之前抵达终点，并保留尽可能多的留存收益。该情景策略游戏将野外求生中多变的天气与不定的决策通过情景模拟的方式真实呈现，对于玩家的数据意识、信息搜集与灵活决策能力以及风险防控都提出了很高要求。本文将基于图论与马尔可夫决策有关模型，综合考虑玩家在两种情形下所面临的现实困境，并对该最优策略展开具体讨论。

二、问题分析与求解

（一）未来信息已知：基于多阶段决策的动态规划模型

经济学中，期望收益为根据已知信息对未来收益的预判。在游戏中，玩家期望在规定的时间内获得尽可能多的资金。由于天气数据与地图完全已知，本文首先根据地图信息建立图论模型，接着使用动态规划将沙漠掘金问题划分为多阶段决策模型，从基本逻辑出发，首先规划出掘金路线，进而分析资源购置策略，在此基础上依据天气状况与资源情况求解挖矿策略，最终通过筛选期望收益的最大值来获取玩家的最优策略。

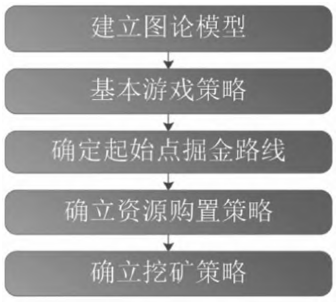

图 1　问题流程图

1. 图论模型

设地图共有 n 个区域，其中含有 k 个村庄，记为集合 $A=\{a_1、a_2\cdots a_k\}$；含有 m 座矿山，记为集合 $B=\{b_1、b_2\cdots b_m\}$。沙漠起始点记为 s_0，沙漠终点记为 s_n。$w_1(t)$ 为第 t 日水资源基础消耗量，$w_2(t)$ 为第 t 日食物资源基础消耗量。矿山的单日收益为 r，每箱水资源的质量为 m_1，基准价格为 p_1，每箱食物资源的质量为 m_2，基准价格为 p_2。玩家在第 t 天的剩余水资源质量为 $M_1(t)$、剩余食物资源质量为 $M_2(t)$、剩余资金为 $C(t)$。游戏时限为 t_0 天，承重上限为 W_{max}。

玩家在任一区域可选择“停留”状态，其时间递归式如下：

$$\begin{cases} M_j(t+1)=M_j(t)-w_j(t) & j=1,2 \\ C(t+1)=C(t) \end{cases}$$

在非沙暴天气时，玩家在任一区域可选择“移动”状态，其时间递归式如下：

$$\begin{cases} M_j(t+1)=M_j(t)-2\times w_j(t) & j=1,2 \\ C(t+1)=C(t) \end{cases}$$

玩家在矿山区域时，可以选择“挖矿”状态，其时间递归式如下：

$$\begin{cases} M_j(t+1)=M_j(t)-3\times w_j(t) & j=1,2 \\ C(t+1)=C(t)+r \end{cases}$$

玩家经过或停留村庄区域时，可以购置资源，购置资源产生的递归式如下：

$$\begin{cases} M_1(t)=M_1(t)+x\times m_1 \\ M_2(t)=M_2(t)+y\times m_2 \\ C(t)=C(t)-x\times 2p_1-y\times 2p_2 \end{cases}$$

其中 x 为购置水资源的箱数，y 为购置食物资源的箱数。由于村庄和矿山在游戏中的特殊性，将地图转化为如图 2 所示的图论模型。

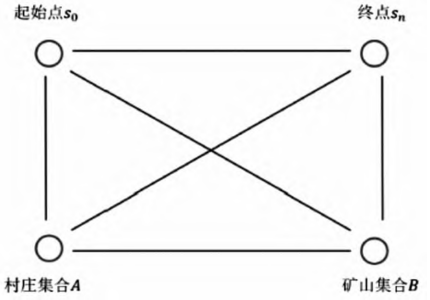

图 2　图论模型

该图 G 中点集 V 与边集 E 的表达式如下所示：

$$\begin{cases} G=\{V,E\} \\ V=\{s_0,s_n,A,B\} \\ E=\{l_{s_0s_n},l_{s_0A},l_{s_0B},l_{s_nA},l_{s_nB},l_{BA}\} \end{cases}$$

其中，s_0 为起始点区域，s_n 为终点区域，$A=\{a_1、a_2\cdots a_k\}$ 为村庄集合，$B=\{b_1、b_2\cdots b_m\}$ 为矿山集合。该模型将区域划分为四个区域集，图中的边集表示区域间的距离。由于玩家在游戏过程中不断经过以上四个区域集，将四个区域集的点作为掘金路线的基本点，则一条掘金路线 L 可写为：

$$L=s_0-s_1-s_2-\cdots-s_n$$

式中，$s_i\in V$。设第 t 天的玩家区域位置为 S_t，可将原问题分解为不同区域集下的多阶段决策问题，通过求解每一阶段下的最优策略建立动态规划模型。

2. 基本游戏策略

为获取更多的资金，有两种基本途径：收益最大化与支出最小化。我们从这两个基本途径延伸出四条基本游戏策略：

最短路原则：从区域 s_1 向区域 s_2 出发时，由于全部区域的天气状况相同，耗时相同的路径具有相同的资源消耗量，因此耗时最短的路径为资源消耗量最少的路径。故我们以区域之间的最短路设为区域 s_1、s_2 之间的距离：

$$l_{s_1s_2}=\min\{l_1,l_2,\cdots,l_n\}$$

式中，$l_{s_1s_2}$ 指区域 s_1、s_2 之间的距离，$l_1\sim l_n$ 为区域 s_1、s_2 之间所有可能的路径。

满载原则：资源不足时玩家需要前往村庄补充必备资源，多次重复前往村庄将增加路途资源的消耗，为减少前往村庄次数，除最后一次购置资源外，其余应使负重满载。设第 i 次经过(或停留)村庄集合时购买的水资源为 x_i 箱，食物资源为 y_i 箱，应有：

$$\{x_i,y_i\}=\{x,y\,|\,M_1(t)+xm_1+M_2(t)+ym_2\leqslant M_{max},\max xm_1+ym_2\}$$

式中，$M_1(t)$ 为玩家在第 t 天的剩余水资源质量，$M_2(t)$ 为玩家在第 t 天的剩余食物资源质量。

顺路原则：减少路途的时间支出可以获得更多的资金收入和更少的资源消耗。设起始点 s_0 至终点 s_n 的距离为 $l_{s_0s_n}$，沙漠掘金路线为 $s_0-s_1-s_2-\cdots-s_n$，其中 $s_i\in V$，则应有：

$$\{s_1,s_2,\cdots,s_{n-1}\}=\{s_1,s_2,\cdots,s_{n-1}\,|\min\sum_{i=1}^{i=n}l_{s_{i-1}s_i}-l_{s_0s_n}\}$$

其中，$s_1,s_2,\cdots,s_{n-1}$ 为符合顺路原则的最优村庄与最优矿山组合。

不剩余原则：在终点剩余资源将以基准价格的一半退回，造成资金浪费，结合满载原则，最后一次购置资源的数量 x_n、y_n 应有如下关系：

$$\{x_n,y_n\}=\{x_i-M_1(t)\,|_{s(t)=s_n},yi-M_2(t)\,|_{s(t)=s_n}\}\,|_{i=n}$$

3. 掘金路线策略

玩家在起始点面临三种选择：向村庄集合 A 出发购买必备资源；向矿山集合 B 出发来获取未来的资金收益；向终点 s_n 出发以结束游戏。由此引申出三种掘金路线：

(1)先前往村庄后前往矿山路线 $s_0-a_i-b_i-\cdots-s_n$

(2)先前往矿山后前往村庄路线 $s_0-b_i-a_i-\cdots-s_n$

(3)直接前往终点路线 s_0-s_n

其中，a_i 为村庄集合 $A=\{a_1、a_2\cdots a_k\}$、中的某一村庄，b_i 为矿山集合 $B=\{b_1、b_2\cdots b_m\}$ 中的某一矿山，s_0 为起始点，s_n 为沙漠终点。以 $plan=1,2,3$ 表示三种不同的路线方式：

$$\begin{cases} s_1\in A,plan=1 \\ s_1\in B,plan=2 \\ s_1=s_0,plan=3 \end{cases}$$

玩家在村庄补充资源后，若时间充裕将前往矿山采矿；玩家在矿山消耗大量资源后，若时间充裕将前往村庄补充资源。因此路线中村庄和矿山应交替出现，直至接近时限玩家向终点移动。则有：

$$s_i\in A\wedge s_{i+1}\in A=\varnothing,0<i<n$$

由于游戏目标为在规定时限内获取更多的资金，从期望收益角度分析三种路线，期望收益高的路线为最优掘金路线，并通过顺路原则求解相应村庄和矿山位置。

4. 资源购置策略

在沙漠起始点可以低廉价格购买一次资源，在沙漠途中经过或停留村庄时均可购置资源，资源购置量应匹配于资源消耗量。此外，由于水资源和食物资源的价格不等，在村庄购买将进一步拉大两种资源的差价，在不改变掘金路线的情况下，在初始点应以低廉价格购买尽可能多的贵重资源。

(1)初始点资源购置策略。对于一条已知的掘金路线(如 $s_0-s_1-s_2-\cdots-s_n$)，设初始的资源购置量为水资源 x_0 箱、食物资源 y_0 箱，考虑两种资源价格不等的情况，设 $p_1\leqslant p_2$，则应在初始点购买尽量多的食物资源。设 f_i 为第 i 次前往村庄前的路途、挖矿水资源消耗量。以 $state=1,2,3,4$ 代表玩家四种不同的状态："停留"、"移动"、"挖矿"、"购置资源"，第 t 天的水资源消耗量 $f(t)$ 为：

$$f(t)=\sum_{i=1}^{t}state\cdot w_1(t)$$

根据路线进行迭代，状态量 $state$ 的变化由下式决定：

$$\begin{cases} state=1,weather=sand \\ state=2,s(t)\in E \\ state=3,s(t)\in B \\ state=4,s(t)\in A \end{cases}$$

式中，$s(t)$ 为玩家在第天所处的位置，E、B、A 由图论模型给出。

为保证资源的配置不影响掘金路线，应有：

$$y_0=\{y_0\,|\max y_0,y_0\leqslant\left[\frac{W_{max}-\max(f_{i+1}-f_i)}{m_1}\right],i>0\}$$

对应的 x_0 由满载准则给出，关于资源价格 $p_1\geqslant p_2$ 的情况类似，本文不再赘述。

(2)村庄资源购置策略。在上节已经求得三种掘金路线，设第 i 次经过(或停留)村庄集合 B 时购买的水资源为 x_i 箱，食物资源为 y_i 箱，据满载原则，x_i 与 y_i 满足：

$$\{x_i,y_i\}=\{x,y|M_1(t)+xm_1+M_2(t)+ym_2\leqslant W_{max}\max xm_1+ym_2\},1\leqslant i<n$$

设最后一次经过村庄经过(或停留)村庄集合 B 时购买的水资源为 x_n 箱，食物资源为 y_n 箱，据不剩余原则有：

$$\{x_n,y_n\}=\{x_i-M_1(t)|_{s(t)=s_n},y_i-M_2(t)|_{s(t)=s_n}\}|_{i=n}$$

5. 挖矿策略

(1)前往终点决策。当剩余时间较少且资源不足时，玩家面临的选择为：前往村庄补给后返回矿山挖矿；直接前往终点，由此生成两种不同的决策方案。我们从期望收益角度分析两种方案，期望收益高的方案为最优决策。

(2)前往村庄时机。由于不同天气对路途的资源消耗不同，在给定所有天气数据的情况下可以选择合适的天气前往村庄购置资源。设 f_j 和 g_j 分别表示第次前往矿山挖矿，则前往村庄的最优决策为：

$$\min f_{j+1}+g_{j+1}-f_j-g_j$$

$$s.t.\begin{cases} f(t)=\sum_{i=1}^{t}state\cdot w_1(t) \\ g(t)=\sum_{i=1}^{t}state\cdot w_2(t) \end{cases}$$

(3)暂停挖矿。由于在村庄购置资源价格为基础价格的二倍，在沙暴或高温天

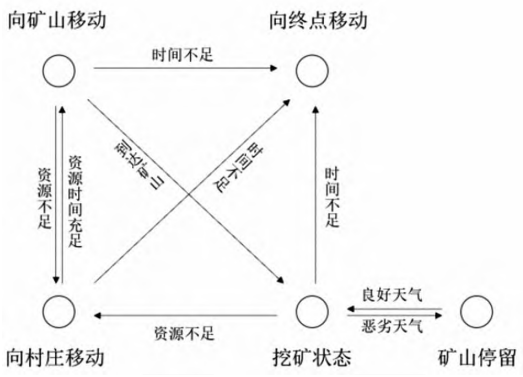

图 3　玩家策略示意图

气挖矿将承担高额的资源费用，在时间期限允许的条件下，可以尝试选择在沙暴或高温天气暂停挖矿休息，本策略的触发条件为：

$$\exists t_1、t_2, r-2(w_1(t_1)p_1+w_2(t_1)p_2<r-3(w_1(t_2)p_1+w_2(t_2)p_2)$$

综上，从玩家状态分析，策略示意图如图 3 所示。

（二）未来信息未知：基于风险预测的多阶段马尔可夫决策模型

经济学中，风险预测或风险评估指对未来不确定性的量化计算和预测。玩家在游戏中通过目前的状态与当天的天气情况(环境状态)，产生移动、停留挖矿等行为(对环境做出动作)，并通过环境的反馈调整决策。由于玩家的目标仍为到终点时资金的最大化，本文利用概率论相关理论的对期望收益进行定量评估预测，进而做出最优决策。由于情景的相似性，我们沿用前问题的变量符号与游戏基本策略。

1. 马尔可夫决策模型

马尔可夫决策(MDP)的流程图如图 4 所示。

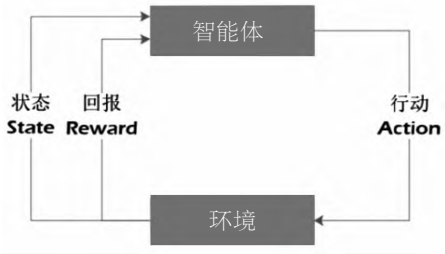

图 4　马尔可夫决策流程图

马尔可夫决策包含一组交互对象：智能体：MDP 中进行机器学习的代理，可以感知外界环境的状态进行决策、对环境做出动作并通过环境的反馈调整决策。环境：mdp 模型中智能体外部所有事物的集合，其状态会受智能体动作的影响而改变，且上述改变可以完全或部分地被智能体感知。

马尔可夫决策过程由五部分组成：$\{State, Action, P_{sa}, \gamma, Reward\}$，其中 $State$ 表示状态集合，$Action$ 表示行动集合，P_{sa} 表示状态转移概率，γ 表示阻尼系数，$Reward$ 表示该行动的回报。针对该游戏，$State$ 包含位置变量 $s(t)$、所剩资金变量 $C(t)$、所剩水资源变量 $M_1(t)$、所剩食物资源变量 $M_2(t)$；$Action$ 有四种情况："停留"、"移动"、"挖矿"、"购置资源"；阻尼系数 γ 在该问题中为 1；$Reward$ 由资金时间递推式决定。最终的路线与资源购置方案仍由期望收益最大给出：

$$\max C(t)|_{s(t)=s_n}$$

2. 风险预测模型

设晴朗、高温、沙暴天气的概率为 h_1~h_3，其对应的水资源基础消耗量为 w_{11}~w_{13}，食物资源基础消耗量为 w_{21}~w_{23}。利用概率论，对不同行为的资源消耗与资金变化进行预测。

(1)停留。玩家在任意天气均可以选择"停留"，两种资源的变化期望为：

$$M_j(t+1)=M_j(t)-\sum_{i=1}^{i=3}h_i w_{ji} \quad j=1,2$$

(2)移动。玩家在非沙暴天气可以选择"移动"。由于每次移动成功的概率为 h_1+h_2，需要求解移动距离为 l 时的时间消耗。这是一个帕斯卡过程，据概率论，帕斯卡分布的分布函数如下：

$$P(x=k)=C_{k-1}^{t-1}p^{r-1}\cdot q^{k-r}\cdot p, k=r, r+1, \cdots$$

因此，该过程的耗时与资源消耗期望如下：

帕斯卡过程算走r步的消耗期望

$$\begin{cases} T=\left[\dfrac{1}{h_1+h_2}\right]+1 \\ M_j(t+T)=M_j(t)-\left(\left[\dfrac{1}{h_1+h_2}\right]+1\right)\cdot\sum_{i=1}^{i=3}h_i w_{ji} \quad j=1,2 \end{cases}$$

考虑到天气的随机因素，需要多准备一些资源以应对极端情况(多次出现沙暴天气无法移动)。由于不同玩家有不同的游戏偏好，我们引入风险偏好系数 k，多准备 $k\cdot\sigma$ 的资源(σ 为标准差)，两种资源消耗的期望更正为：

$$M_j(t+T)=M_j(t)-\left(\left[\frac{1}{h_1+h_2}+k\cdot\frac{\sqrt{h_3}}{(h_1+h_2)^2}\right]+1\right)\cdot\sum_{i=1}^{i=3}2\cdot h_i w_{ji} \quad j=1,2$$

(3)挖矿。玩家在矿山区域可以选择"挖矿"，类比"停留"时的资源消耗，挖矿天的资源消耗与资金变化的期望为：

$$\begin{cases} M_j(t+T)=M_j(t)-T\cdot\sum_{i=1}^{i=3}3\cdot h_i w_{ji} \quad j=1,2 \\ C(t+T)=r\cdot T+C(t+T) \end{cases}$$

(4)资源购置。玩家经过或停留村庄区域时，可以购置资源，购置资源产生的递归式如下：

$$\begin{cases} M_1(t)=M_1(t)+x\times m_1 \\ M_2(t)=M_2(t)+y\times m_2 \\ C(t)=C(t)-x\times p_1-y\times p_2 \end{cases}$$

其中 x 为购置水资源的箱数，y 为购置食物资源的箱数。

3. 基本游戏策略

与前问题相比，游戏规则除天气情况未知外完全一致，因此最短路原则、满载原则、顺路原则与不剩余原则仍然适用。针对未来天气的未知性，根据风险预测原理，我们期望决策具有较小的风险性，由此引申出第五条基本策略：同路原则。

同路原则：由于区域间的最短路径可能存在多种行进方式，对于不同方案下的掘金路线，如 $L_1=s_0-s_1-s_2-\cdots-s_n$、$L_2=s_0-s_1'-s_2'-\cdots-s_n$，应尽量保证初始路线的重合，以便玩家进行更充分的决策。

4. 玩家策略

由于情景的相似性，我们仍沿用前问题模型中的掘金路线策略、资源购置策略与挖矿策略，并针对未来天气情况未知的条件进行改进。

(1)掘金路线策略。玩家在起始点仍具有三种掘金路线方式：①：先前往村庄后前往矿山路线 $s_0-a_i-b_i-\cdots-s_n$；②：先前往矿山后前往村庄路线；③：$s_0-b_i-a_i-\cdots-s_n$ 直接前往终点路线 s_0-s_n。分别计算三种策略的期望收益以判断最优路线。

(2)资源购置策略。资源购置策略与前问题基本一致，考虑到两种资源的价格不等，应在起始点购买尽量多的贵重资源。对于 $p_1\leqslant p_2$ 的情况，初始点的资源购买量由下式给出：

$$y_0=\{y_0 \mid \max y_0, y_0\leqslant\left[\frac{W_{max}-\max(f_{i+1}-f_i)}{m_1}\right], i>0\}$$

对于村庄资源的购置策略，由满载原则与不剩余原则给出：

$$\{x_i, y_i\}=\{x, y \mid M_1(t)+xm_1+M_2(t)+ym_2\leqslant W_{max}\max xm_1+ym_2\}, 1\leqslant i<n$$

$$\{x_n, y_n\}=\{x_i-M_1(t)|_{s(t)=s_n}, y_i-M_2(t)|_{s(t)=s_n}|_{i=n}$$

(3)挖矿策略。挖矿策略依然包含三部分内容：前往终点决策；前往村庄时机；暂停挖矿，与前问题保持一致。特别是当

VaR理论下房地产上市公司财务风险影响的问题研究

钱 桢 彭焱鑫

摘要：房地产作为我国经济发展不可或缺的一部分，房地产企业的发展往往促进了许多行业的繁荣和创新。但房地产业由于其工程周期长、资产负债率高，往往面临比其他企业更高的风险。文章以绿地控股上市公司为例，通过公司历史股价、财务报表等相关数据对房地产企业的财务风险进行深入探讨，并运用VaR财务风险评估体系模型结合因子分析法计算了当前条件下企业的财务风险。同时，根据理论分析和相关描述性统计，对如何提高我国房地产企业的财务安全提出了一些对策和建议。

关键词：房地产企业；VaR；财务风险；因子分析法

一、引言

资金安全是资本密集型行业房地产项目顺利发展的重要保障。长期以来，房地产业在我国经济发展中占有重要地位。防范房地产业系统性金融风险已经上升到国家安全的新高度，尤其是在我国房地产业总市值中占较大比重的房地产上市公司。如果公司突然出现较大的财务问题，必定会导致国家经济无法弥补的损失。

本文将参考一些研究模型，比较VaR值计算中不同方法的特点，最后选择使用蒙特卡罗模拟方法。针对样本的选取，本文将上海证券交易所绿地控股上市公司作为研究对象。以2000~2019年的年报数据为基础，分别计算样本公司12项财务指标和蒙特卡罗模拟法计算的VaR，最后完成对房地产上市公司财务风险评估体系的构建。在模型构建方面，借助因子分析法计算了样本财务风险评估指标体系的指标值，并结合研究结果，为房地产上市公司应对风险提供了对策和建议。

二、相关理论

风险是指由于经营决策、投融资方式或财务结构不合理而造成损失的可能性。目前我国在财务风险预警模型的构建和企业财务预警指标的选取方面已经取得了一定的成果。学者杨华通过对引入的非财务指标进行研究，借助因子分析法，建立预警模型。结果显示，在已知预警模型

满足下式时，启用暂停挖矿策略。

$$\exists j,-(w_{1j}+w_{2j})>r-3\cdot(w_{1j}+w_{2j})$$

三、结论与推广

本文根据图论、博弈论的有关原理建立数学模型，对游戏玩家多阶段、多目标的最优策略展开探究。基于马尔科夫的决策模型是本文的一大特色，综合考虑到模型拥有的优化功能，可以考虑将其推广至现实生活中个人面对复杂情境时的决策与权衡，以及企业面临变幻莫测的市场环境时合理的应对之策等相关议题的探讨之中。（注：本文数据与资料来源：2020年全国大学生数学建模比赛B题）

参考文献：

[1]Brihaye T,Geeraerts G,Hallet M,et al. On the termination of dynamics in sequential games[J].Information and Computation,2019,272:104505.

[2]Dhiman A,Uttam T,Balakrishna n S. Implementation of sequential game on quantum circuits [J].Quantum Information Processing,2020,19(04):1-16.

[3]陈如峰.自学习策略价值风险模型研究与应用[D].成都：电子科技大学，2020.

[4]王中玉，曾国辉，黄勃，方志军.改进A* 算法的机器人全局最优路径规划[J].计算机应用，2019,39(09):2517-2522.

（作者单位：西安交通大学）