赛区评阅编号（由赛区组委会填写）：

2020 高教社杯全国大学生数学建模竞赛

承　诺　书

我们仔细阅读了《全国大学生数学建模竞赛章程》和《全国大学生数学建模竞赛参赛规则》（2019 年修订稿，以下简称为“竞赛章程和参赛规则”，可从全国大学生数学建模竞赛网站下载）。

我们完全清楚，在竞赛开始后参赛队员不能以任何方式，包括电话、电子邮件、“贴吧”、QQ 群、微信群等，与队外的任何人（包括指导教师）交流、讨论与赛题有关的问题；无论主动参与讨论还是被动接收讨论信息都是严重违反竞赛纪律的行为。

我们完全清楚，抄袭别人的成果是违反竞赛章程和参赛规则的行为；如果引用别人的成果或资料（包括网上资料），必须按照规定的参考文献的表述方式列出，并在正文引用处予以标注。

我们以中国大学生名誉和诚信郑重承诺，严格遵守竞赛章程和参赛规则，以保证竞赛的公正、公平性。如有违反竞赛章程和参赛规则的行为，我们将受到严肃处理。

我们授权全国大学生数学建模竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

我们参赛选择的题号（从 A/B/C/D/E 中选择一项填写）：B

我们的报名参赛队号（12 位数字全国统一编号）：

参赛学校（完整的学校全称，不含院系名）：南京大学

参赛队员 (打印并签名) ：1. 曾嘉馨 171840063 数学系

2. 徐佳美 181860117 计算机科学与技术系

3. 朱倩 181860154 计算机科学与技术系

指导教师或指导教师组负责人 (打印并签名)：

（指导教师签名意味着对参赛队的行为和论文的真实性负责）

日期: 2020 年 08 月 02 日

（请勿改动此页内容和格式。此承诺书打印签名后作为纸质论文的封面，注意电子版论文中不得出现此页。以上内容请仔细核对，如填写错误，论文可能被取消评奖资格。）

赛区评阅编号（由赛区组委会填写）：

2020 高教社杯全国大学生数学建模竞赛

编 号 专 用 页

赛区评阅记录（可供赛区评阅时使用）：

评阅人						
备注						

送全国评阅统一编号（由赛区组委会填写）：

全国评阅随机编号（由全国组委会填写）：

（请勿改动此页内容和格式。此编号专用页仅供赛区和全国评阅使用，参赛队打印后装订到纸质论文的第二页上。注意电子版论文中不得出现此页。）

穿越沙漠游戏策略设计

摘要

在穿越沙漠游戏中，玩家要在规定时间内到达终点，并保留尽可能多的资金。本文在单个、多个玩家，已知所有天气、仅知当天天气几个条件两两组合的四种情境下分别探讨最佳或较好的游戏策略。

对于问题一：采用**动态规划**，初始状态为第 0 天在起点。通过状态转移方程，利用第 $i-1$ 天在 j 地及其邻居的最大剩余资金量求出第 i 天在 j 地的最大剩余资金量，进而求出截止日期当天在终点的最大剩余资金量，从而得到最优策略。将关卡一、二中的数据代入，求得关卡一最大剩余资金量为 12820 元，关卡二最大剩余资金量为元。

对于问题二：贪心 +Q-learning

对于问题三：多人不完备信息博弈

对于问题四：博弈论

关键字： 动态规划　多属性决策　多人博弈

1 问题重述

1.1 问题背景

在如下小游戏中：玩家凭借一张地图，从起点出发，在沙漠中行走，途中会遇到不同的天气。可在起点用初始资金购买一定数量的资源（水和食物），也可在村庄购买，并可在矿山补充资金。目标是在规定时间内到达终点，并保留尽可能多的资金。

游戏的基本规则如下：

（1）游戏目标：在截止日期或之前到达终点，并保留尽可能多的资金。未到达终点前穿越沙漠所需的资源（水或食物）不能耗尽，每天玩家拥有的水和食物质量之和不能超过负重上限。

（2）移动策略：每天玩家可移动到与目前区域相邻的另一区域（有公共边界），也可在原地停留。玩家在原地停留一天消耗的资源数量称为基础消耗量，行走一天消耗的资源数量为基础消耗量的 2 倍。

（3）天气影响：每天的天气为“晴朗”、“高温”或“沙暴”，沙漠中所有区域的天气相同。不同天气玩家的资源基础消耗量不同，且沙暴日玩家必须在原地停留。

（4）资源获取：玩家可在起点处用初始资金以基准价格购买水和食物（只有第 0 天可以购买），也可在村庄用剩余的资金购买水和食物，每箱价格为基准价格的 2 倍。

（5）资金获取：玩家在矿山停留时，可通过挖矿获得资金，挖矿一天获得的资金量称为基础收益, 消耗的资源数量为基础消耗量的 3 倍；不挖矿消耗的资源数量为基础消耗量。到达矿山当天不能挖矿，沙暴日也可挖矿。玩家到达终点后可退回剩余的水和食物，每箱退回价格为基准价格的一半。

（6）多人游戏：当有多名玩家时，若某天其中的任意 k 名玩家均从区域 A 行走到区域 $B(A \neq B)$，则任一玩家消耗的资源数量均为基础消耗量的 $2k$ 倍；若某天其中的任意 k 名玩家在同一矿山挖矿，则任一玩家消耗的资源数量均为基础消耗量的 3 倍，且一天通过挖矿获得的资金是基础收益的 $\frac{1}{k}$；若某天其中的任意 k 名玩家在同一村庄购买资源，每箱价格均为基准价格的 4 倍。其他情况下消耗资源数量与资源价格与单人游戏相同。

1.2 求解问题

根据游戏的不同设定，建立数学模型，解决以下问题:

1. 只有一名玩家，玩家已知每天天气状况，给出一般情况下玩家的最优策略。求解附件中的“第一关”和“第二关”。

2. 只有一名玩家，玩家仅知道当天的天气状况，据此决定当天的行动方案，给出一般情况下玩家的最佳策略，并对附件中的“第三关”和“第四关”进行具体讨论。

3. 有 n 名玩家，他们有相同的初始资金，且同时从起点出发。

（1）假设在整个游戏时段内每天天气状况事先全部已知，每名玩家的行动方案需在第 0 天确定且此后不能更改。给出一般情况下玩家应采取的策略，并对附件中的“第五关”进行具体讨论。

（2）假设所有玩家仅知道当天的天气状况，从第 1 天起，每名玩家在当天行动结束后均知道其余玩家当天的行动方案和剩余的资源数量，随后确定各自第二天的行动方案。给出一般情况下玩家应采取的策略，并对附件中的“第六关”进行具体讨论。

2 问题分析

2.1 问题一分析

由于每天天气状况已知，游戏中不存在任何未知因素，因此一定存在最优解。游戏目标是在规定时间到达并保留尽可能多的资金，该目标可转化为求出截止日期当天在终点的剩余最大资金量 m，而 m 的求解实际上可转化为前一天在终点或终点邻居的剩余最大资金量的求解。这样一直迭代到第一天，可以得到动态规划的雏形。考虑用二维数组 $c[i][j]$ 记录在第 i 天到达 j 地最大剩余资金数对应的资源消耗和资金赚取，通过前一天的邻居节点 k（考虑到原地停留，还包括自身）对应的 $c[i][k]$ 和第 i 天到达 j 地的资源消耗、资金赚取情况获得状态转移方程。

2.2 问题二分析

对于玩家来说，本身存在最优策略，但因为玩家仅知晓当天天气和后续天气的部分条件设定（如，难以得出最优解。首先计算起点到终点的最短路径，结合对天气的估计，能到吗？如果不能到或很危险，那么首先考虑去村庄。时间和物资宽裕吗？高温天走还是停留？（晴天停留无意义，高温天可以考虑停留）

物资和时间之前的平衡考量

是否去矿山？矿山收益如何，值得去吗？去的话村庄距离多远？

该去村庄吗？

贪心，机器学习

2.3 问题三 (1) 分析

所有玩家已知所有天气，存在最优解，但为避免和其他玩家同行，该做怎么样的调整？博弈论

多人静态不完全信息博弈，机器学习？ＭＣ抽样算法是现今应用于非完备信息博弈中的一个基本方法，它通过随机抽样将非完备信息博弈问题转换为完备信息博弈问题，同时通过大规模的抽样次数来逼近真实的情况．ＵＣＴ算法引入第 3 方即 3 个人类选手博弈树

2.4 问题三 (2) 分析

以避开人群 + 到终点 + 多赚钱为目的多人动态完全信息博弈

模糊集（军棋子粒猜测）蒙特卡洛抽样算法（非完备信息）学习算法建立对手模型动态博弈模型ＧＤＧＭ（扑克牌，多人动态完全博弈）

MU 算法（MC 抽样算法）（不完备 UCT 算法搜索最优策略 maxn 搜索（多人博弈基本搜索算法 Paranoid 算法（多人博弈

状态转移函数博弈树

3 模型准备

3.1 背景知识

3.1.1 多人非完备信息博弈

3.1.2 动态博弈

3.2 模型假设

3.3 符号说明

下表列出了我们在建模过程中使用的符号及其含义，部分符号我们会在描述模型时进一步给出说明。

表 1　符号说明

符号	符号说明	单位
$weightLimit$	负重上限	千克
$originMoney$	初始资金	元
$deadline$	截止日期	天
$basicGain$	基础收益	元
$money$	在矿山赚取的资金	元
$foodNum$	消耗的食物	箱
$waterNum$	消耗的水	箱
$foodSize$	食物每箱质量	千克
$waterSize$	水每箱质量	千克
$foodPrice$	食物基准价格	元/箱
$waterPrice$	水基准价格	元/箱

4 模型的建立与求解

4.1 问题一：动态规划

游戏目标是规定时间达到终点并保留最大剩余资金，剩余资金量可用初始资金 + 挖矿所得 + 终点变卖资源 - 购买资源花销表示，由于初始资金一定，购买资源花销可以通过食物和水的消耗数量计算出，而最优策略中一定会避免到终点还有剩余资源的情况（资源变卖会折损资金），因此仅记录消耗的水、食物和挖矿所得资金即可求得剩余最大资金量。

令 $c[i][j]$ 表示第 i 天在 j 地的累计资源（水、食物）消耗情况和资金赚取情况, 即 $c[i][j]$ 实际包括消耗的水 $c[i][j].waterNum$, 消耗的食物 $c[i][j].foodNum$, 赚取的资金 $c[i][j].money$ 三个数据。为使剩余资金最大,在对可能方案的衡量中应尽量使 $c[i][j].waterNum * waterPrice + c[i][j].foodNum * foodPrice - c[i][j].money$ 取最小。

考虑到天气对消耗资源量的影响,用 $resource[i]$ 表示第 i 天的天气,$resource[weather[i]]$ 表示第 i 天对应天气的资源消耗情况。

现在考虑第 i 天处于 j 地所有可能的决策：①原地停留不挖矿（基础消耗），②移

动（2 倍的基础消耗），③挖矿（3 倍的基础消耗，收获基础收益）。

下面计算所有决策产生的结果（以下的加法运算表示各部分的 $waterNum$,$foodNum$,$money$ 各自相加）：

决策一：原地停留

$$res1 = cost[i-1][j] + resource[weather[i]]$$

决策二：非沙暴天，可以移动

$$res2 = cost[i-1][k] + 2 * recource[weather]$$

决策三：挖矿

$$res3 = cost[i-1][j] + resource[weather[i]] * 3 + basicGain$$

分别计算 $res1$,$res2$,$res3$ 对应的资金消耗与资金赚取的差值，取最小者，用其对应的 res 值更新 $c[i][j]$。

4.2 问题二：贪心

为实现最佳策略，玩家在确定行进目标后理论上一定会选择最短路径，即与目标地点之间相隔区域最少的路径。因此考虑对地图进行简化，只将起点、终点、矿山和村庄看作图的顶点，将各点之间的最短路径看作图的边，最短路径上经过区域的个数为边的权值，这样可以得到如图 1 所示的压缩图（以一个村庄，一个矿山为例，多个村庄、矿山情况下同理）：

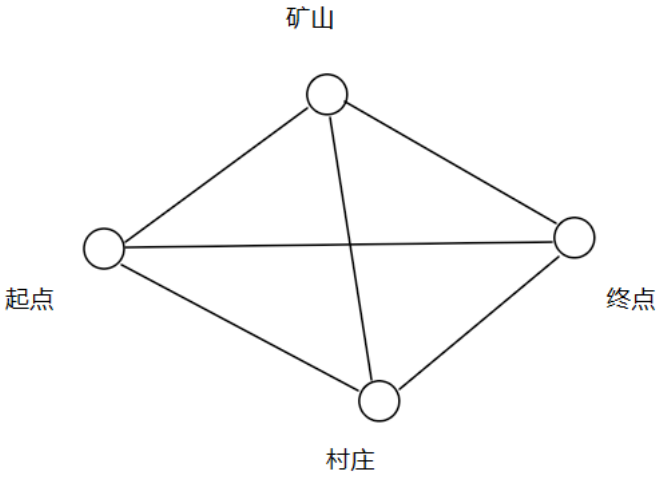

图 **1**　压缩图

由于后续天气状况未知（或仅知道一定天气出现的天数限制），到达压缩图中特定顶点的天数、消耗资源情况都是未知的，但有上限、下限和平均情况，这里考虑在题目

对天气的限制基础（若有）上对后续天气状况进行随机模拟（or 各取 $\frac{1}{3}$?），以此预估到达特定顶点的资源消耗情况和最短天数，据此决定行进目标。直接到达终点先去矿山挖矿，后到达终点村庄-矿山-村庄-终点平均情况背包容纳

p-greedy Q-learning 模型，在探索和经验中进行平衡。我们使用强化学习来解决仅知道当前天气的情况下，玩家应该采取的行动。Q 为动作效用函数（action-utility function），用于评价在特定状态下采取某个动作的优劣。它是智能体的记忆。在 p-greedy Q-learing 模型中，包括状态，动作，奖赏三个要素，智能体—玩家根据当前状态来选择下一步要采取的动作，并记录被反馈的奖赏，以便下次采取更优的策略。

状态的选择：我们把当前的天气 α，当前地点的类型 β 和当前已使用的物资占比作为状态 γ。γ 用背包剩余容量/最快到达终点所需物资来计算，数值越大，表示物资越充足。则状态的表达为（α，β，γ），共 40 种。

动作的选择：玩家有停留，挖矿，行走，购买物资，四种动作。奖赏的选择：若采取动作后，物资保证还能继续到达终点，给予 1 的奖赏，若其中有挖矿得到收益，给予 2 的奖赏；若物资不能保证到达终点，给予-2000 的奖赏。

在模型中，状态和动作的组合是有限的，若状态 m 中，动作 n 种，把 Q 当做一个表格，则共有 mn 行，每个动作有一个效用值。

训练

初始化 Q = ; while Q 未收敛：初始化玩家位置为起点，开始新一轮游戏 while S != 物资不足：使用策略 π，获得动作 x=π(S) 使用动作 x 进行游戏，获得玩家的新位置状态 S’ 与奖励 R(S,x) Q[S, X] ←(1 - p)*Q[S, X] + p * (R(S, X) + q * max Q[S’,X]) // 更新 Q S ← S’

其中，R 为眼前利益，max 为以前学到的利益。p 为学习速率，q 为折扣因子。根据公式可以看出，p 越大，保留之前训练的效果越少，越重视眼前利益，q 越大，max 所起到的作用就越大，越重视以往经验。

参数的设置：α0（晴朗）1（高温）2（沙尘暴）β 0（普通地点）2（矿山）3（村庄）γ 0，1,1.5,2，a (停留) b(挖矿) c（行走）d（购买物资）p:TODO q:TODO

将普通地点视为路径，起点、终点、村庄、矿山视为点，得到缩略图

4.3 问题三（1）

假设所有玩家绝对理性，他们趋向于达成纳什均衡，而不是破坏规则，使局面变得更混乱不可控。假设所有玩家都是风险厌恶型，他们趋向于携带充足的物资，离开起点时，他们会尽量将物资补充到最满，离开村庄前，他们会补充自己此前的消耗，直到资金耗尽。假设玩家判断自己物资耗尽前无法抵达下一个补充物资的地点，他们会停留在原地，不再对其他玩家产生影响。

（1）先用第一题的办法搜索出最佳路径，确定前往特殊地点（村庄、矿山）的顺序，

由于每个人都会选择对自己最有利的方案，所以他们都会按照最佳路径行走。如果最佳路径为直达终点的最短路，则玩家需要计算路上的最大消耗，并按消耗购买物资。如果路径经过村庄或矿山，那么玩家将会同时抵达第一个村庄或矿山。他们可以选择不同的路径抵达，但所耗时间和物资都是一样的。由于玩家事前不知道其他玩家的策略，因此他们会以相等的概率随机选择一条最短路径。在村庄，玩家可能会选择休息一天，让一部分玩家先离开，从而用较低价格购入村庄物品，但是玩家的智力水平是一致的，所以留下的玩家大概率会超过 2 个，那么这一天的停留就没有意义，因此，所有玩家的选择都是购入物资后离开。在矿山，玩家可能会选择挖矿、休息或者直接离开。某一玩家 i，如果不处于紧急状态，即不会因为停留在矿山而导致无法走到终点，就可以停留在矿山。玩家 i 可以预估出当前矿山的玩家数量，列出在矿山的 k 个玩家的所有可能策略，然后求得纳什均衡，判断这三类玩家分别占总数的多少，并以此为概率随机选择一种方案。

一般策略：随机选择一条最佳路径，按特殊地点分为若干段，每一段的起点和终点都是特殊地点，而且途中不经过特殊地点。预估一个特殊地点可能遇到的人数，列出可能策略并求出纳什均衡，按概率随机选择一种策略，然后继续对下一段路程实行同样的操作。在最后敲定行走路径和停留时间后，按最大消耗购买物资。

对于第五关，最佳路径只有两种，一种是走最短路径抵达矿山以后，按照每天的收益决定挖矿、休息或是离开，另一种路径是走最短路径直达终点。我们分别计算晴朗天气挖矿一天的收益和高温天气挖矿一天的收益：晴朗：35 高温：-205

于是，经过矿山的最佳路径是走最短路，第三天抵达矿山，4/5/6 进行挖矿，第 7 天离开矿山前往终点。这一路线的总收益 =-925

总消耗水食物负重资金 93 106 491 925

同样，我们可以计算出直达终点的总收益为-490

总消耗水食物负重资金 30 34 158 490

因此，最佳路径为直达终点。为了不让对方获得比自己更高的收益，两位玩家都会选择走直达终点的最佳路径，而直达终点的最佳路径有且仅有一条，因此他们的总消耗为

水食物负重资金 60 68 316 980

最后，他们各自剩下 9020 元。

（2）物资的三种状态：物资耗尽、物资告急、物资充足物资耗尽：剩下的水和食物不够，哪怕天天晴朗都无法抵达下一个物资补充地点物资告急：剩下的水和食物数量有限，假设接下来的天气随机变化，抵达下一个物资补充地点路上损耗的物资期望超过了现有物资，如果再不离开有可能因为极端天气而被淘汰物资充足：正常的状态

时间的两种状态：时间充足、时间告急时间告急：剩下的时间有限，假设接下来的天气随机变化，抵达终点消耗的时间期望超过了剩余时间，如果再不离开有可能因为极端天气而被淘汰时间充足：正常的状态

假定从一个特殊地点前往下一个特殊地点的路上，除非出现沙暴天气，否则所有人都不停顿。每天行动结束后，玩家会根据地图判断接下来的路是否有多种选择，如果有多种选择，他们将以同等的概率随机选择其中一条路线。

抵达村庄后，玩家有两种选择：在村庄停留一天再离开或直接离开。我们首先判断这里的所有人是否已经时间告急，如果是，所有人都将继续前进，如果不是，我们可以分析当前在村庄且不处于时间告急状态的 k 个玩家所采取的策略。这 k 个玩家中，有 s 个玩家是物资告急，他们下一步一定是离开村庄，而剩下的 k-s 个玩家中，一部分玩家选择购买物资后继续前进，另一部分玩家选择休息一天再走，对于这 k-s 个玩家，我们根据他们可能选择的策略求出达到纳什均衡时，选择留下和选择前进的人数分别是多少，并认为这两种人数的占比就是一个玩家选择留下或前进的概率。这两种策略的收益函数分别是：

抵达矿山后，玩家有三种选择：休息一天，挖矿或直接离开。我们首先判断这里的所有人是否已经时间告急，如果是，所有人都将继续前进，如果不是，还可以分析当前在村庄且通过了判断函数的检验未被淘汰的 k 个玩家所采取的策略。这 k 个玩家中，有 s 个玩家物资告急，他们的下一步一定是离开，剩下的 k-s 个玩家中，一部分玩家选择休息一天，一部分玩家选择挖矿，剩下的玩家直接离开，我们同样对他们的策略求得纳什均衡，计算出玩家采取各种策略的概率。这三种策略的收益函数分别是：

4.4 问题三（2）

5 模型评价

5.1 优点

5.2 缺点

5.3 改进方式

参考文献

[1] https://github.com/BlankerL/DXY-COVID-19-Data

[2] 曹盛力，冯沛华，时朋朋. 修正 SEIR 传染病动力学模型应用于湖北省 2019 冠状病毒病 (COVID-19) 疫情预测和评估 [J]. 浙江大学学报，2020，02：178-184