

## 2018 级数学建模新生赛

### 承 诺 书

我们仔细阅读了中国大学生数学建模竞赛的竞赛规则。我们完全明白，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、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们参赛选择的题号是（从 A/B 中选择一项填写）：   A  

我们的参赛报名号为（如果赛区设置报名号的话）：      F186     

所属学校（请填写完整的全名）：                     电子科技大学                    

参赛队员（打印并签名）：      朱中凯     

     孙夕嫒     

     郭志猛     

日期：   2018   年   4   月   23   日

# 电子科技大学 2018 级 数学建模新生赛论文

题目：对于数模参赛和选拔的优化模型

队员：朱中凯 2017190301009 英才实验学院

孙夕嫒 2017110901002 经济与管理学院

郭志猛 2017080201005 英才实验学院

2018 年 4 月 15 日

# 目 录

承诺书.....	1
封 面.....	2
目 录.....	3
一、摘要.....	4
二、问题的重述.....	5
三、问题的分析.....	5
四、模型假设.....	5
五、符号说明.....	6
六、模型建立与求解	
（一）数学建模最优参赛策略.....	6
第一步：最佳阵容.....	6
第二步：聚类生成.....	9
第三步：经验曲线模型的推广.....	11
（二）美赛校队选拔方案.....	16
第一步：现状分析.....	16
第二步：参数估计.....	17
第三步：分析与结论.....	17
七、模型推广与评价	
（一）模型的优点.....	18
（二）模型的缺点.....	18
（三）模型的推广.....	18
八、参考文献.....	18
九、附录	
附录 1：四级、雅思、托福与六级换算准则.....	19
附录 2：各个奖项的概率.....	19
附录 4：推荐信.....	20

## 2018 级数学新生赛摘要专用页

### 一、 摘要

本文主要研究电子科技大学学生参加数学建模比赛的最佳参赛策略,和协助数学建模校队老师挑选出最适合参加“美国大学生数学建模比赛”的校队选择方案,目标是通过对影响数学建模获奖的几大要素进行分析得出本队伍的最佳参赛策略,并得出校队选拔的最佳方案,进而大大提高获奖概率,节省老师、团队的工作量。

第一问通过对 matlab 模拟的所有参赛者数据进行有监督聚类,得出十四个凝聚点,即十四个获奖向量。分析本队成员的时间—经验曲线,来预估不同时间段本队成员的自身能力向量,再取其中一个时间点的能力向量,求出与十四个凝聚点间的直线距离,进而决定什么时间参加什么比赛的最佳参赛策略。

第二问根据以往通过校队选拔的队伍内的最高英语水平、最高在校成绩平均分、最高周末班考试成绩、最高以往获奖经历,以及队伍平均英语水平、平均在校成绩平均分、平均周末班考试成绩、平均以往获奖经历(自变量)以及选拔后参赛是否获奖(因变量)进行机器学习,即用 MATLAB 做多元回归(得出结论)

**关键字:** 有监督聚类、线性回归、经验曲线、数据统计, 机器学习

## 二、问题重述

所给问题是建立在建模比赛投入与产出不成正比，校赛选拔效率、准确率不高的背景下，让我组结合关于获奖的各种因素，如，英语水平、在校平均成绩等相关的各种因素，寻找两种模型，对本队成员参赛策略与美赛校队选拔方案进行讨论，从而达到团队、学校利益最大化。

按照题目要求，通过分析，我们需要解决以下问题：

- 1、如何合理挑选队友
- 2、如何科学制定问题中涉及的变量。
- 3、如何合理的将其中某些变量量化
- 4、如何根据往年的数据，定性或定量分析各个因素对获奖的影响权重和不同奖项所需获奖因素的不同
- 5、如何合理评估个人能力随时间的变化。
- 6、在以上问题的基础上，再把抽象成数学问题的模型还原成实际生活问题，从而进一步改进优化。

## 三、问题分析

针对以上问题，我们小组对第一问的分析是，首先通过数据统计确定队伍的最最佳阵容，再在考虑影响小组参赛策略的五大因素，即：英语水平、在校成绩平均分、周末班考试成绩、以往获奖经历、比赛消耗时间成本的基础上，运用有监督聚类模型，对所有参赛队伍数据进行聚类，从中选出十四个凝聚点，即：美赛 O、M、H；国赛国一、国二、国三、省一、省二、省三；校内赛一、二、三等奖；美赛模拟一、二、三等奖的凝聚点。最后通过数据分析产生的本队成员的时间—经验曲线，来预估不同时间段本队成员的自身能力向量与各大赛事凝聚点间的直线距离，进而决定什么时间参加什么比赛得奖概率最大。最后根据三个队员的发展规划，决定最优参赛策略。

我们小组对第二问的分析是，根据以往通过校队选拔的队伍内的最高英语水平、最高在校成绩平均分、最高周末班考试成绩、最高以往获奖经历，以及队伍平均英语水平、平均在校成绩平均分、平均周末班考试成绩、平均以往获奖经历（自变量）以及选拔后参赛是否获奖（因变量）用 MATLAB 做多元回归，得出所有因素对队伍获奖的权重，进而决定选拔方案。

## 四、模型假设

- 1、假设英语水平可以由一个英语考试成绩衡量，且不累加。
- 2、假设比赛经验可以由参赛前获奖的累计分数衡量。
- 3、排除不可控因素，假设平均成绩、周末班成绩可以完全反映数学、编程能力，同时假设参赛者的必修课中都有数学、编程相关课程。
- 4、假设周末班不统计出勤率，仅有周末班成绩一个指标。
- 5、排除团队成员遇到重大意外而导致的能力增长不符合增长曲线的情况以及因重大意外而不能参加某次比赛的情况。
- 6、假设团队成员只有出国、读研、创新创业这三个方向。
- 7、假设队友确定后将不再产生人员变更。
- 8、假设每个学生都正常并合格地完成了学习任务，掌握了课程要求的能力。

## 五、符号说明

符号说明

符号	含义说明
$M_i$ ( $i=o, m, h$ )	获得美赛 O/M/S/H 队伍的获奖向量
$G_i$ ( $i=1, 2, 3$ )	获得国赛 1/2/3 队伍的获奖向量
$S_i$ ( $i=1, 2, 3$ )	获得省 1/2/3 队伍的获奖向量
$X_i$ ( $i=1, 2, 3$ )	获得校建模赛 1/2/3 队伍的获奖向量
$pM_i$ ( $i=1, 2, 3$ )	获得美赛模拟赛 1/2/3 队伍的获奖向量
$S$	队伍能力向量
$e_i$ ( $i=1, 2$ )	队伍平均、最高英语水平

$c_i$ ( $i=1, 2$ )	队伍平均、最高在校成绩
$z_i$ ( $i=1, 2$ )	队伍平均、最高周末班成绩
$h_i$ ( $i=1, 2$ )	队伍平均、最高以往获奖成绩
$b_i$ ( $i=1, 2$ )	队伍平均、最高准备比赛的时间成本
$D$	能力向量与获奖向量间的距离
$A$	是否获奖情况

表 5-1：符号说明

## 六、模型建立与求解

### （一）数学建模最优参赛策略

#### 第一步：最佳阵容

第一题需要我们建立一个最优参赛策略并根据我队队员的实际情况给我自己队伍的最优参赛策略。我们将寻找最优参赛策略分为三个过程，分别是最佳阵容，然后建立一个聚类模型，最后引入一个经验公式对模型进行优化并且找出我队参赛的最优参赛策略。

最佳阵容的确定我们通过抽样调查得出新生在寻找队友时注重的方面，结果如图：

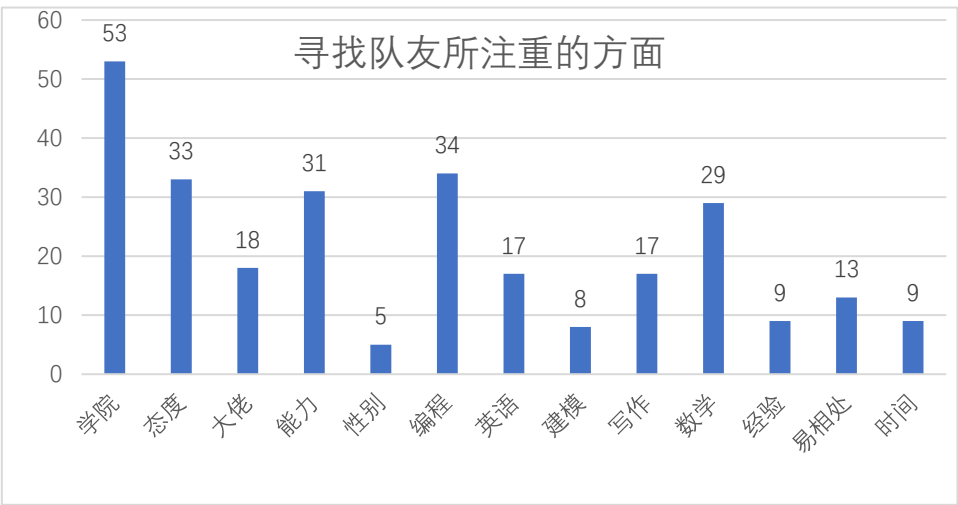


图 6-1：寻找队友所注重的方面

我们发现，出现频次最高的关键词是学院，其次是编程、数学能力和态度，再次是“大佬”、英语和写作能力，最后还有人提到容易相处、参赛经验、

时间充足、建模能力和性别。

对此，我们认为这与新生赛的参赛主体是新生有着密切的联系。一名新生在考量自己的队友时，由于大部分同学对建模的**基础有限**，与建模相关的获奖经历、参赛经验往往偏低，并且曾经得奖的大一新生多数情况下可能会延续之前的队伍配置，不会再次寻找队友。所以这造成寻找队友时大部分同学不会过多关注参赛经验，只会抱着试试看的心态寻找“大佬”队友。同时，由于“容易相处”等概念**难以量化**，也会导致同学的关注度偏低。

**在以上前提下，我们展开对其它关键词的分析。**

在难以量化和基础有限的前提下，考量一个人能力的基础坐标就变成了“学院”。在一般情况下我们可以认为，在相应领域，对应专业的同学的平均能力会高于非该专业同学的平均能力。所以，英语专业的同学更有机会承担队伍中英文论文的写作，计算机专业的同学更有机会承担队伍中的编程任务。当然，我们还应注意到大佬等人士可能会在掌握本专业基础能力的条件下还有更强的其它专业的能力，有鉴于此，我们将其视为异常值，因为大佬的标准难以量化衡量，一个学院的平均水平往往可以被作为参考值。

然后，同学的关注度对于各项能力也有所差异，如图：

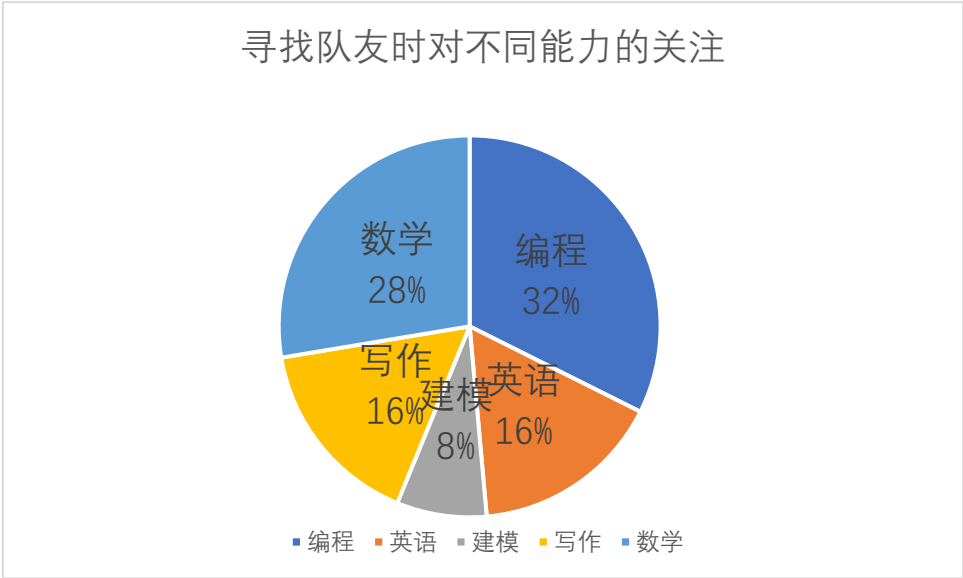


图 6-2：寻找队友时对不同能力的关注

可以看到参赛同学对队友能力的关注大致分为三个等级，第一等级是对“编程”和“数学”的关注。我们认为大一新生的数学基础可能更趋近于一致，不会有巨大差异，并且比赛对数学能力的要求并不是主要考察标准。但是不同同学的编程能力可能会有很大差异，有人大一已经加入了 ACM 校队，而有的可能大一才开始学 C 语言，甚至还没有开始学习语言。所以对编程的关注度高，而数学能力可以通过“微积分”、“线性代数”等课程进行量化，所以关注度偏高。而写作能力和英语能力在建模比赛中也不可或缺，所以紧随其后。而建模能力无法量化且同学们基础有限，所以对队友的建模能力也普遍要求较低。

通过以上分析，我们得出结论。考虑到参赛同学基础有限且很多能力难以量化，所以最佳阵容的确定应该主要通过学院和可量化的能力，如微积分成绩、C 语言成绩等进行选择。

对于学院来说，应当假设各位同学都是平均生，都已经达到了学院的基本

能力要求，并且同学在开课之后都具备了课程目标要求的能力。根据三个模块：数学建模、编程、写作对阵容进行确定。

在数学模型方面，开设相关课程的学院的同学可能更具优势。

课程序号	课程名称	课程类别	开课院系
F1152440.01	数学建模	学科拓展课程	数学科学学院
I1108760.01	数学建模方法	素质教育选修课 (自然科学类)	数学科学学院
I1108860.01	数学建模基础	素质教育选修课 (自然科学类)	数学科学学院
I1109820.01	数学建模实验	素质教育选修课 (跨专业课外选修实验类)	数学科学学院
L1104520.01	数学建模课程设计	实习实训	数学科学学院
课程序号	课程名称	课程类别	开课院系
L2703820.01	数学综合创新项目	实践创新与科研训练	英才实验学院

表 6-1：数学建模相关课程情况

由上述表格可以看到，数学科学学院、英才实验学院、以及其它学院选上数学建模相关课程的学生可以承担数学建模的任务分工。

在编程方面，我们通过在信息门户的全校开课平台查询，发现程序设计涵盖了大部分学院，所以我们认为这些学院的同学已经具有了承担编程的任务分工的能力。

在写作方面，考虑到日后参加的美赛和美赛模拟赛，英语能力必不可少。所以英语能力也是一项判断的依据。英语专业的学生有较大优势。考虑到对于大一新生，英语是一门必修课程，所以不同学院同学之间的差异并不是很大，差异更多集中在个体上。所以我们认为英语专业学生和其它专业英语成绩突出的学生可以承担英语写作任务。

对于个人素质来说，在相应的课程上取得更优秀的成绩的同学可以视为相应能力更为突出。

综上所述，我们认为要组成一个最佳阵容，如果之前有过参赛获奖经历，可以选择之前的队友再次组队。如果没有参赛获奖经历，那么上过相关课程的同学在相应方面更有优势。并且成绩优秀的同学的阵容更好。当然，这是平均来说，但是个体的差异实际上大于学院成绩的差异，所以最佳阵容的确定还是要在建模、编程、写作（英语）各有突出之处的同学之中选择，且一个团队中某项能力的最大值的重要性要高于平均值的重要性。

## 第二步：聚类生成

在阵容已经确定的前提下，我们需要制订什么时间参加什么比赛的参赛策略。通过数据统计，在考虑影响小组参赛策略的五大因素，即：英语水平、在校成绩平均分、周末班考试成绩、以往获奖经历、比赛消耗时间成本的基础上，运用聚类模型，将不同比赛获奖团队数据运用 K-means 计算方法反复循环得到此比赛某奖项的凝聚点，即：美赛 O、F、M、S、H；国赛国一、国二、国三、省一、省二、省三；校内赛一、二、三等奖；美赛模拟一、二、三等奖的凝聚点。



## 1. 数据的得来

首先，我们根据问卷调查和合理的预估，充分考虑到能参加比赛的选手的水平，采用 matlab 生成大量符合选手真实水平的数据。生成的数据有在校成绩平均分、英语成绩、以往获奖成绩、周末班成绩、准备比赛的时间成本五个维度。

下面是部分数据：

	c1	e1	h1	z1	b1
选手 1	73	525	5	77	113
选手 2	84	582	4	80	106
选手 3	79	621	8	86	195
选手 4	83	473	4	84	138
选手 5	97	637	10	88	169

表 6-2： 部分参赛选手的成绩和准备时间的数据

其中在校数据平均分和周末版成绩我们采取百分制，英语成绩取四六级的 710 分制，以往获奖成绩的最高取含金量（含金量是得奖概率的倒数）的总和，准备比赛的时间成本我们以小时为单位并通过问卷调查取得。

## 2. 数据的预处理

数据的指标单位不同，我们为了消除量纲的影响，采取极差正规化变换（Range 0-1）。首先所有样本表示为：

$$X = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nm} \end{pmatrix}$$

极差表示为：

$$R = \max_{1 \leq i \leq n} x_{ij} - \min_{1 \leq i \leq n} x_{ij}$$

那么每列数据可以得到一个极差，如下表：

	c1	e1	h1	z1	b1
极大值	100	684	11	100	229
极小值	66	401	0	71	36
极差	35	283	12	29	193

表 6-3： 全部参赛选手各项参数的极大值、极小值和极差

极差正规化变化的表达式为：

$$x_{ij}^* = \begin{cases} \frac{x_{ij} - \min_{1 \leq i \leq n} x_{ij}}{R_j}, R_j \neq 0 & \begin{pmatrix} i = 1, 2, \dots, n \\ j = 1, 2, \dots, n \end{pmatrix} \\ 0.5, R_j = 0 \end{cases}$$

这样变化后的数据最小为 0，最大为 1，其余在区间 [0, 1] 内，极差为 1，无量纲。

以下为极差正规化变化得到的部分数据：

	c1	e1	h1	z1	b1
选手 1	0	0.4381625	0.416666667	0.206896552	0.398963731
选手 2	0.5142857	0	0.333333333	0.310344828	0.362694301
选手 3	0.3714286	0.7773852	0	0.517241379	0.823834197
选手 4	0.4857143	0.254417	0.333333333	0	0.528497409
选手 5	0.8857143	0.8339223	0.833333333	0.586206897	0

表 6-4：部分选手成绩经极差正规化变化后所得数据

在这些部分的基础上，我们为了描述变量之间的相关性，通过求不同选手的成绩向量之间的距离，来判断他们成绩的亲密关系。为此，我们采用了欧式距离：

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

通过求不同选手的成绩向量所表示的坐标之间的距离，来确定两个选手成绩的相关性。

部分数据如图：

	x1	x2	x3	x4	x5
y1	0	0.397875	0.695549	0.444277	1.014154
y2	0.397875	0	0.637155	0.442357	0.779989
y3	0.695549	0.637155	0	0.699749	0.564241
y4	0.444277	0.442357	0.699749	0	0.889185
y5	1.014154	0.779989	0.564241	0.889185	0

表 6-5：部分选手成绩向量表示坐标之间的距离

这样，我们就得到了关系矩阵。

### 3. 数据的聚类及确定最佳分类

接下来，我们将对所有得到的向量距离采用 K-means 算法进行数据的集聚。考虑到这种算法集聚的随机性，我们采用读取初始聚类中心的方法，选择出可能得奖的选手的 14 个向量（分别对应包括美赛、国赛、校赛、美模模拟赛的各个小类共 14 个奖项）。然后不断进行迭代，得出新的迭代中心值。对于这一过程，我们采用软件 Spss 进行计算，下图为迭代过程中的历史记录：

迭代历史记录 <sup>a</sup>														
迭代	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	16.812	5.201	14.126	.000	13.481	11.153	16.893	7.524	5.794	9.534	19.174	8.109	4.619	10.696
2	5.377	4.604	5.603	.000	10.851	3.093	5.755	1.473	5.423	5.866	8.203	.000	2.867	3.354
3	1.590	3.574	3.448	.000	7.042	.000	.000	3.336	5.040	.000	4.545	.000	1.601	2.094
4	1.478	2.614	.000	.000	4.747	.000	.000	1.943	1.152	.000	.000	.000	.000	1.060
5	.000	.000	3.140	.000	.000	4.916	.000	1.976	1.157	4.581	.000	.000	.000	2.037
6	.000	.814	.000	.000	.000	8.297	3.957	1.402	.000	.000	.000	.000	.000	2.197
7	.000	.000	.000	.000	.000	4.143	.000	.000	.000	.000	.000	.000	.865	.000
8	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

a. 由于聚类中心中不存在变动或者仅有小幅变动，因此实现了收敛。任何中心的最大绝对坐标变动为 .000。当前迭代为 8。初始中心之间的最小距离为 39.281。

表 6-6：14 个聚类中心点的迭代历史记录

最终，我们得到了需要的 14 个聚类向量：

	c1	e1	h1	z1	b1
美赛 O、F 奖	92	640	17	90	210
美赛 M 奖	87	585	7	88	189
美赛 H 奖	82	569	7	82	176
国赛一等奖	94	685	15	93	213
国赛二等奖	90	609	14	85	186
省赛一等奖	86	566	9	82	165
省赛二等奖	84	546	8	82	141
省赛二等奖	82	521	8	82	132
校赛一等奖	88	541	2		53
校赛二等奖	81	514	2		55
校赛三等奖	74	481	2		48
美赛模拟一等奖	85	534	19		62
美赛模拟二等奖	80	527	0		39
美赛模拟三等奖	76	506	0		34

表 6-7：14 个聚类中心及其对应比赛情况（缺失数据是因为参加相关比赛时未参加周末班）

以上 14 个聚类向量即为最终的聚类中心，分别表示各个奖项获奖队伍的近似值。

### 第三步：经验曲线模型的推广

#### 经验曲线概念的引出：

学习曲线又称经验曲线，由 Wright 于 1936 年提出，用以刻画生产效率随时间变化的规律。Wright 发现在产品生产过程中，随着累计产量的增加，产品单位工时会逐渐下降，生产效率逐渐上升，但当累计产量达到一定数量后，产品的单位工时基本保持不变，生产效率趋于稳定，常用经验曲线的数学模型如下：

$$y = ax^{-b}$$

一些参数的说明：

a	生产第一件产品时所耗时数
x	生产的产品数
b	经验系数
y	生产 x 件产品时单位产品所耗费的平均时数

表 6-8：部分参数的补充说明

其图像如下：

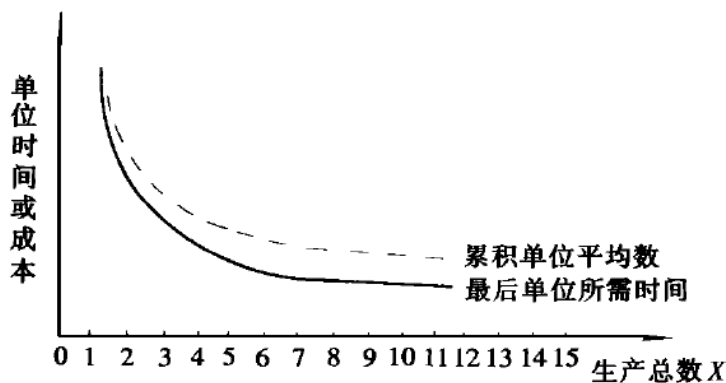


图 6-2：单位时间或成本与生产总数之间的关系

所以由此生产效率与时间之间的关系如图所示，其方程应该为  $y = \frac{1+x^{-b}}{ax^{-b}}$  【5】，其中 y 表示生产效率，其余各值与上述相同。

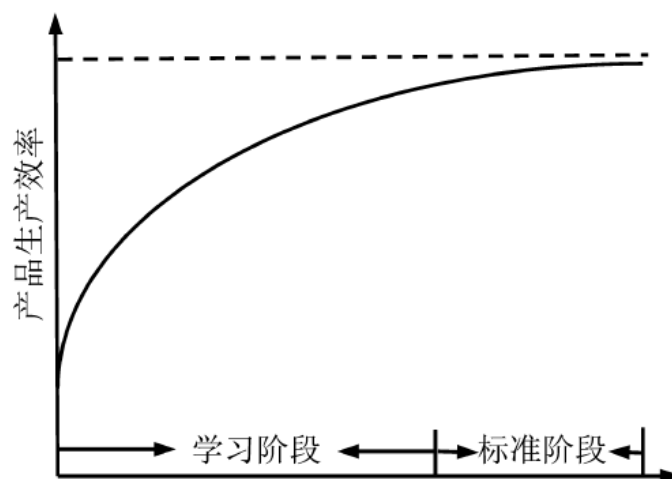


图 6-3：产品生产效率和时间的关系

图中的曲线包含学习阶段和标准阶段。学习阶段产品的生产效率随累计产品数量的增加而增加，并且生产效率的增长速度呈下降趋势(曲线的斜率逐渐减小)；在标准阶段，当累计产品生产数量增长到一定值后，生产效率基本保持不变。因此，学习曲线的特点是在开始阶段上升速度很快，之后上升速度逐渐递减，最终上升速度趋于零，曲线趋于平坦。显然，对于不同产品或不同企业，组织的学习效率存在着差异，影响曲线的变化速度。因此在实际研究中，需要根据企业的实际特点来测算相应的学习率。【4】

建模的过程：

所以我们将这个模型应用于数学建模中来，我们可以发现：产品生产效率可以转化为建模效率以及建模质量，学习阶段可以等效为学习时间长短，因此我们

可以表示出相应的曲线的方程为： $y = \frac{1+x^b}{a}$ ，下一步我们应该将 a, b 参量求出来。

模型的求解：

首先，我们统计了 130 位建模的同学，其数据如下：有些同学由于经验限制，并没有完成较多比赛，但也参加本次统计。

	A	B	C	D	E	F	G	H	I	J
	时间 频次 建模次数	两天	三天	四天	五天	六天	七天		平均用时/天	总人数
1	第一次建模	0	45	12	16	11	46		5.007692308	130
2	第三次建模	12	18	14	13	0	10		4.014925373	67
3	第五次建模	16	16	12	11	0	0		3.327272727	55
4	第七次建模	17	15	13	0	0	0		2.911111111	45

表 6-9：问卷调查得到建模同学对完成建模工作且得奖所用时间的估计

使用 matlab 拟合出的图像：

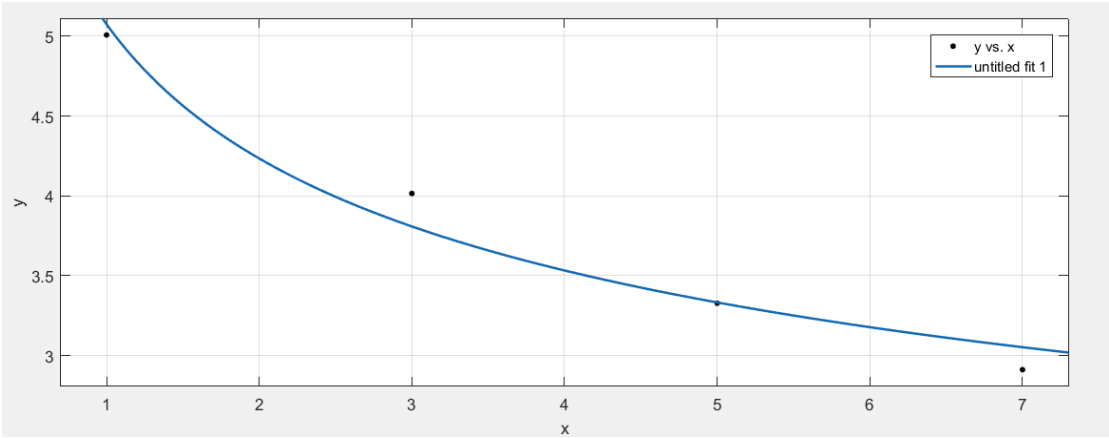


图 6-4：根据上表拟合出的图像

相应的数据为：

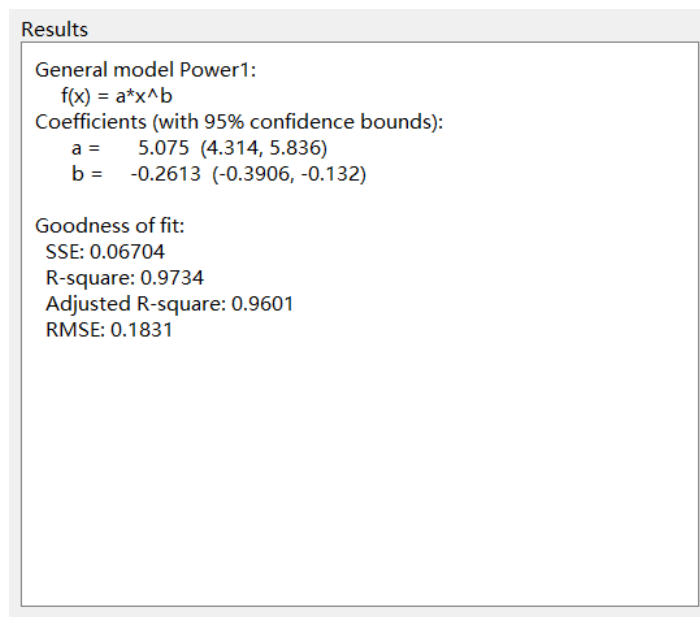


图 6-5：通过计算得到公式中参数的值

此便可以近似为数学模型的经验曲线，我们将其转变为能力曲线，坐标轴的 x 向量转化为学习的时间，y 轴可以体现出我们的建模能力，则曲线的方程

为：

$$y = \frac{1 + x^{0.2613}}{5.075}$$

图像为：

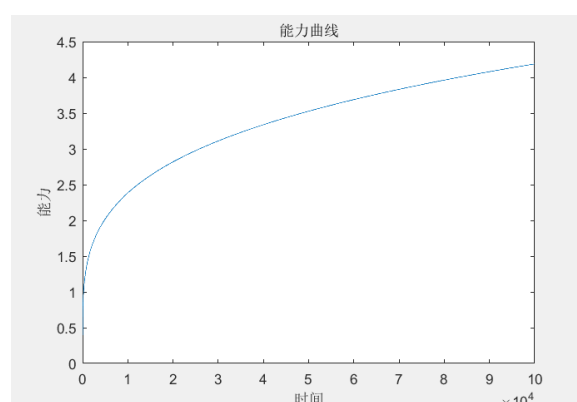


图 6-6：所得公式对应的曲线

我们由前文得到了 14 个向量，我们将平均分，周末班成绩的平均值（没有的取 0），获奖情况，英语水平乘以一个过渡矩阵：

$$\begin{pmatrix} \frac{3}{\sqrt{10}} & \frac{3}{\sqrt{11}} & \frac{1}{\sqrt{11}} \\ 0 & \frac{1}{7.1\sqrt{11}} & \frac{3}{7.1\sqrt{11}} \\ \frac{2}{\sqrt{10}} & \frac{2}{\sqrt{11}} & \frac{2}{\sqrt{11}} \end{pmatrix}$$

其结果可以表示出建模，编程，写作的能力，并将它们取平均后成为能力值，横坐标为准备的比赛的时间乘以 k 倍，相应的数据如下

	在校平均分	英语成绩	获奖成绩	周末班成绩	时间成本	建模能力	编程能力	写作能力	平均成绩
Mo	92	640	17	90	210	97.08192	119.7425	119.2244	112.0163
Mm	87	585	7	88	189	87.43698	108.2107	105.1319	100.2599
Mh	82	569	7	82	176	82.21922	102.5563	101.4352	95.40358
G1	94	685	15	93	213	98.18872	122.7087	124.5051	115.1342
G2	90	609	14	85	186	91.86417	113.4511	112.4106	105.9086
S1	86	566	9	82	165	85.3815	105.444	102.8621	97.89587
S2	84	546	8	82	141	83.80036	103.0872	99.40956	95.43236
S3	82	521	8	82	132	82.85167	101.121	95.92307	93.29857
X1	88	541	2	0	53	43.00698	63.97986	83.39549	63.46077
X1	81	514	2	0	55	39.68658	59.6674	78.90042	59.41813
X3	74	481	2	0	48	36.36619	55.10014	73.64096	55.03576
pM1	85	534	19	0	62	52.3357	72.57718	92.30281	72.40523
pM2	80	527	0	0	39	37.94733	58.56115	79.19981	58.56943
pM3	76	506	0	0	34	36.04997	55.86028	75.92141	55.94389

表 6-11：聚类得到的 14 个聚类中心点及其参数

但是由于参量的改变，对于函数，我们的参数也要相应的改变，那么我们根据数据的大小，我们在原来方程的基础上增加系数，最后经过拟合，我们取

$$y = 300 \frac{1+x^{0.2613}}{5.075}$$

k 取 0.00012

图像如下：

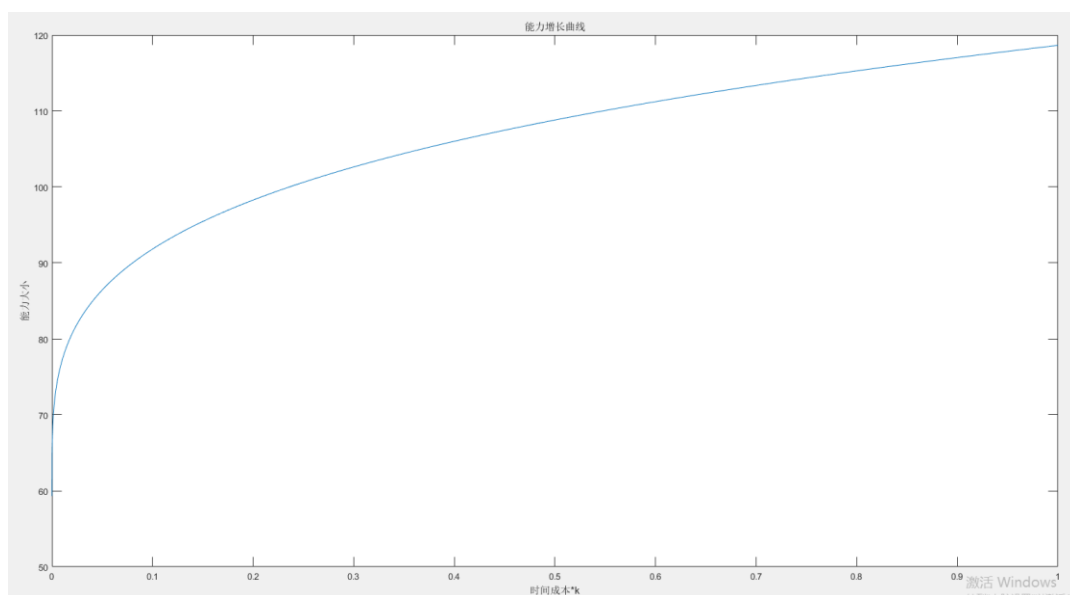


图 6-7：函数表达式对应的图像（x 轴为时间成本乘以 k，y 轴为能力（用平均成绩度量））

其中，时间成本即随着培训时间的向前推进，由图可得，我们的经验能力也在不断的增长。不仅如此，我们还可以在平时进行训练，增加自己的能力，并且我们平时的学习中，我们也经常会学到数模的内容，我们将这些知识并到平时的学习中，那么我们选取最好的排列方式便是最好的学习模式了。

即，在每次的比赛中，距离下次的比赛中，我们将自己的能力通过过渡矩阵进行量化，与下一次比赛需要的能力进行比较，如果相差较大，我们在保证自己学业不耽误的情况下，以拼得最低奖的姿态进行准备，之后做题，当是对自己的训练。如果感觉自己能力与目标差距不大，我们便向最高奖发起冲锋，大大增加获奖概率，直到最后得奖。

容易证明，这个方法是可行的，原因如下：最差是第二次国赛差距较大，之前一直以最低奖为目标，在美赛之前，那么时间成本为：604 小时，折合成能力为 113.8622，显然，这是可以冲击美赛的。

以我们组为例，我们的各项参数如图所示：

在校平均分	英语成绩	获奖成绩	周末班成绩	时间成本	建模能力	编程能力	写作能力	平均成绩
89	613	0	0		42.21640676	66.28365946	91.51293972	66.67100198

表 6-12：我组对应的各项参数

对照图 6-7，我们组原始能力为 66.671（y 轴），相当于从零基础消耗 1.6 小时（x 轴）时间成本获得的能力。我们可以向校赛一等奖发起冲击，所以我们花费 53 小时的时间成本，成功后我们的能力提升到 75.2746（分）。现在我们距离国赛还有一定的距离，所以我们放弃国赛，但是我们要进行日常培训，以追赶他人，那么，我们将会使用 132 小时的时间成本，能力提升至 81.3098（分），之后是美赛校内赛，我们将会以 62 小时的时间成本冲击一等奖，最终能力提升到 83.0207（分）。我们会竞选美赛校队，由于能力不足，便以 176 小时的时间成本自学向 h 奖发起冲击，来年，首先得校内赛，由于经验足够，我们将会以 53 小时的时间成本向一等奖发起冲击，能力提升到 91.9856（分）。此时我们会主动以 213 小时的时间成本向国奖发起冲击，能力提升到 93.9531（分），此能力有大概率获奖，之后便以最佳状态冲击美赛，最后的能力将会冲击美赛。

注：能力通过平均成绩量度，即消耗时间成本乘以系数 k（k=0.00012）为自变量，能力也就是平均成绩为因变量。

## （二）美赛校队选拔方案

### 第一步：现状分析

据调查统计美赛国赛都有赛前集训和周末培训班，投入大量时间却没有获奖的人不在少数。2016 年美国大学生数学建模竞赛和交叉学科建模竞赛，我校选派了 72 支队伍就有 19 支没获得 h 及以上奖，2015 年我校选派的 44 支参赛队伍，有 11 支没有获得 h 及以上的奖，为了提高校队选拔质量与效率必须建立合理的选拔方案。

根据上文模型假设，美赛校队始终以团队形式进行选拔。为了分析究竟团队中究竟是各项指标的最高值影响较大还是平均值影响较大，也就是判断一个团队整体实力强获奖概率大还是最高水平高获奖概率大，以及哪一项指标对于团队获奖影响较大，我们建立了以下模型： $A = \alpha c1 + \beta z1 + \gamma h1 + \delta b1 + \epsilon c2 + \epsilon z2 + \theta h2 + \vartheta b2 + K$ ，其中  $\alpha$ 、 $\beta$ 、 $\gamma$ 、 $\delta$ 、 $\epsilon$ 、 $\epsilon$ 、 $\theta$ 、 $\vartheta$  为权重系数 k 为常数项，通过编程对以往数据进行机器学习，得出合理的各个因素对获奖与否的权重系数。

其中 e 为雅思、托福、四级换算成的六级成绩，h 为根据公式

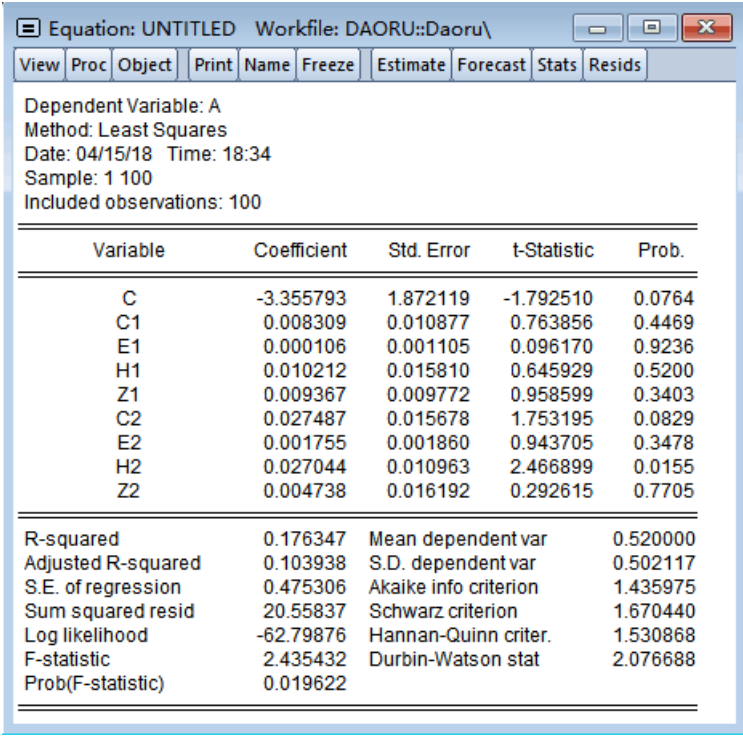
$$h(\text{奖项分数}) = 1 / \text{获奖概率}$$

得到的以往获奖经历量化后的数值。c、z 均为以百分制计算的分数，他们的单位都是分。



第二步：参数估计

运用软件 Eviews, 确定函数的自变量与因变量, 编写计算回归方程程序  
模型的估计结果为



The screenshot shows the EViews software interface with the 'Equation: UNTITLED' window. The window displays the following information:

- Dependent Variable: A
- Method: Least Squares
- Date: 04/15/18 Time: 18:34
- Sample: 1 100
- Included observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.355793	1.872119	-1.792510	0.0764
C1	0.008309	0.010877	0.763856	0.4469
E1	0.000106	0.001105	0.096170	0.9236
H1	0.010212	0.015810	0.645929	0.5200
Z1	0.009367	0.009772	0.958599	0.3403
C2	0.027487	0.015678	1.753195	0.0829
E2	0.001755	0.001860	0.943705	0.3478
H2	0.027044	0.010963	2.466899	0.0155
Z2	0.004738	0.016192	0.292615	0.7705

R-squared	0.176347	Mean dependent var	0.520000
Adjusted R-squared	0.103938	S.D. dependent var	0.502117
S.E. of regression	0.475306	Akaike info criterion	1.435975
Sum squared resid	20.55837	Schwarz criterion	1.670440
Log likelihood	-62.79876	Hannan-Quinn criter.	1.530868
F-statistic	2.435432	Durbin-Watson stat	2.076688
Prob(F-statistic)	0.019622		

图 6-8：Eviews 计算得到的数据

表中 coefficient 为系数, 而 prob 为显著程度, 得到关于 A 的八元一次方程  $A=0.008309c1 + 0.009367z1 + 0.010212h1 + 0.000106e1 + 0.027487c2 + 0.004738z2 + 0.027044h2 + 0.001755e2 - 3.355793$

第三步：分析和结论

(1) 拟合优度和统计检验:

由模拟结果可知, 残差较小, 拟合度良好。

(2) 结论:

从结果可以看出团队中英语最高分对美赛结果影响最为显著, 同时在校平均成绩最高分对影响也较为显著, 而除了周末班成绩以外, 其它因素的最高值都比团队平均值对于美赛能否获奖影响显著, 最后将所有参加选拔的团队的成绩带入公式按照排序录取。同时应该特别考虑一人能力突出的团队, 和英语水平、在校成绩较高的个人。

## 七、模型的推广与评价

### 1、 模型的优点：

1、思考过程全面，对于解决第一问的模型，充分的考虑选择队友的最佳阵容，反映队员能力变化的经验曲线，以及样本的多样性等问题，力求计算出科学、灵活的解决方案。对于第二问的模型，不但考虑了队伍的平均水平更考虑了队伍的最高水平，解决了例如团队中一人达标能否获奖的问题，为校队选拔增加了更多的选择

2、采用 matlab 预估数据并结合已有的数据，使得数据产生尽可能合理，同时有监督聚类模型可以防止数据偏差较大的情况发生。

### 2、 模型的缺点：

1、对于影响比赛能否获奖的因素，本文在第一题的模型中采取了五个因素，在第二个模型中采取了四个因素，并忽略其他因素。这必然会造成一定的误差和不客观

2、由于数据的缺少，通过问卷调查和 matlab 合理预估产生的数据必然和真实数据存在着一定的偏差。

3、经验曲线通过抽样调查数据计算而成，存在较大的主观判断，不具有普遍性。

### 3、 模型的推广：

本模型虽然存在一些不合理的地方，不过主要是因为无法获取真实数据造成的，如果获取真实数据模型会较为准确。因此，本模型只须更换因变量，并带入各个比赛真实的历史数据，即可以适用于学校大部分比赛的选拔，同时也可以帮助同学们为自己是否参加某些比赛做出最优决策。

## 八、参考文献

【1】姜启源，谢金星，叶俊.《数学模型》.北京：高等教育出版社.2013。

【2】彭放，罗文强，肖海军，何水明 .《数学建模方法》.北京：科学出版社，2007。

【3】excel 文档.《全国赛集训选拔录取名单》。

【4】雷田礼，杨全.《关于经验曲线数学模型的证明及经济意义》. 2003。

【5】邵真，冯玉强，刘洋.《基于学习曲线的 ERP 实施培训时间决策模型》2010.8。

九、附录

附录 1、四级、雅思、托福与六级换算准则

根据大量搜集数据取平均值计算后，用六级成绩换算雅思托福四级成绩。分级标准见下表：

六级成绩(分)	托福成绩（分）
425	60±5
500	80±5
600	90±5
650	105±5

表 9-1：六级与托福成绩的换算

六级成绩（分）	雅思成绩（分）
500	5.5-6
550	6.5-7
600-650	7-7.5

表 9-2：六级与雅思成绩的换算

四级成绩（分）	六级成绩（分）
500	425
550	480
600	540

表 9-3：四级与六级成绩的换算

附录 2、各个奖项的概率

（1）美国大学生数学建模大赛获奖概率

特等奖	0.16%
特等奖提名	0.27%
优异奖	8.88%
荣誉奖	37.97%
成功参与奖	51.55%

（2）中国大学生数学建模大赛获奖概率

一等奖	1.40%
二等奖	6.50%

（3）电子科技大学美模赛模拟赛获奖概率

一等奖	1.74%
二等奖	2.78%
三等奖	12.15%

数据来源：学校网站、百度百科、网页

注：综合以上数据信息，我们将获奖概率整合，并且将美赛的特等奖和特等奖提名视为特等奖，优异奖视为一等奖，荣誉奖视为二等奖，成功参与奖视为三等奖，并用概率的倒数表示该项奖项的含金量

附录 3、代码（已省略）

#### 附录 4、推荐信

尊敬的各位校队老师：

您好！

首先，真挚地感谢您抽出时间查看我们的材料，您目光的每一秒停留都让三个渴望成功的心激动不已。

数学建模大赛是有益于我们人生发展的比赛，为了实现我们在大学期间参赛利益的最大化和参赛成本的最小化，我们对参赛策略进行了定量分析和数学建模。建模的过程体现了我们的诸多优势。

首先，我们掌握了资料的检索能力。在数学建模比赛开始之前，我们通过图书、历年优秀论文、咨询学长老师、网上学习等手段快速地对一些基础模型进行了熟悉，了解清楚了其适用条件和使用方法，在短期内积累了大量的前人经验。

其次，我们有着优秀的合作能力。在试题发下来的第一时间我们就充分讨论并确定了题目的选择，之后大家各抒己见，提出了自己对题目的见解并最终统一了结论和做题的顺序、方法和进度安排。

并且，我们的基础知识牢固，模型掌握充分。我们初期提出了大量的模型，如聚点模型、灰色地带模型、Leslie 模型等，然后经过讨论交流，大家对各种模型提出质疑和修改方案，最终采纳了聚点模型，并对我们今后如何参加数学建模比赛最为合适进行了科学预测。

而且，我们决策果断，效率较高。新生赛只有不到三天的时间，在此期间，我们没有过多地深究能力范围之外的优化问题，而是紧扣重点，对能力范围之内的模型建设倾注了大量心血。在旧的模型不适合我们要建的模型的时候，我们及时调整思路，采取更为有效的方案，并最终得到了符合实际情况的解的方案。

还有，我们全心投入，态度端正。我们每个人一起努力，不放弃，不滑水，在互相帮助，互相鼓励的氛围中一齐思考问题的解决方案并努力实施。即便个人有事情也为了团队的利益及时放弃，保证了团队工作的正常运行。

我们同舟共济，一齐努力。在提交论文的最后关头，一位同学的电脑突然出现问题，大量资料丢失，我们妥善地给出解决方案，及时更换电脑，防止沉没成本的二次危害。最终化险为夷。

我们坚定信念，绝不动摇。在选定 A 题之后，我们发现身边的小伙伴大多选择了 B 题，并且他们在对 A 题的评价中不乏“太难了”、“没思路”等话语，但我们没有动摇，努力思考解决方案，一步一步逼近问题的最佳策略，并最终得到了成功，能够用自己的模型验证了很多真实数据，保证了模型的准确率。

.....

这些优势或许在其他队伍队员身上或许并不鲜见，但是我们有足够的信心去提升自己。正如我们在模型中对自己的预期，我们将不断参加比赛积累经验，提高自己的同时积极去争取更高的奖励，为自己未来的选择增加一个砝码。

所以，我们希望得到老师的第一次认可，让我们能够坚定住在数模道路上行走的信心，在这条道路上披荆前行，斩获更多的荣誉，过一个精彩的、充实的大学生活。

此致

敬礼！

全村人的希望队：朱中凯 孙夕嫒 郭志猛

朱中凯手机：13086678001

2018.4.15