**Summary**

当我们选择一份暑假时,我们会多角度考虑各种因素. 问题A的主要任务是分析本科生在选择工作时考虑的因素,

**Contents**

[1 Introduction 3](#_Toc56969913)

[1.1 Background 3](#_Toc56969914)

[1.2 Restatement of the problem 3](#_Toc56969915)

[2 Overall Analysis 3](#_Toc56969916)

[3 Assumptions 4](#_Toc56969917)

[3 Notation 5](#_Toc56969918)

[4 Analysis and Modeling 6](#_Toc56969919)

[4.1 Posts Distribution Model 6](#_Toc56969920)

[4.2 Preference. describing 8](#_Toc56969921)

[4.3 Job choosing 10](#_Toc56969922)

1. Introduction
   1. Background

对于暑假，每个本科生都有独特的计划。尽管距离暑假还有一段时间，但我们已经开始为summer job做准备了。打开招聘网站，各种各样的职位让我们眼花缭乱。在如此繁多的选择中，我们希望找到对于我们而言“最好的”选择。

这些工作里，有的允许我们进行Remote working,有的距离我们的家有一定的距离。有的工作并不要求我们的能力，比如一些体力活动。有的工作会要求我们具有一定的能力，比如销售员。有的工作则要求我们具有一定的专业知识或者技能，比如家庭教师。

我们对暑假的规划各不相同，赚些钱当然可以，但是我们中的一些人也可能并不想占用太多时间，有些人也可能想做些有挑战性的工作。当然，也有人想赚到尽可能多的钱。

* 1. Restatement of the problem

我们有各种各样的工作可以选择，我们对于工作的个人偏好也有所不同。

基于以上背景，我们需要建立数学模型来解决下列问题：

1. 分析找summer job时应该考虑哪些因素。
2. 如何从每个人对工作的偏好出发，找到他们最满意的工作？
3. 展示我们的工作成果。
4. Overall Analysis

为了解决上述问题,首先,我们需要考虑两大部分的内容.一个是如何表示这些展示在本科生面前的工作,另一个是如何表示和使用每个人的偏好.再者,我们需要选择一个有效的算法,能够基于每个人偏好,评价出让此人最满意的工作.

我们的工作将会被划分成如下几个步骤：

1. 分析本科生找summer job时会考虑的因素

对于一份工作,我们会关心工作的地点,时薪(dollars per hour),要求我们工作的时间长短(hours per week),以及这份工作的门槛.

1. 确定工作所具有的属性.

通常在招聘网站上,我们能够看到每个工作的时薪,工作时长以及工作门槛.

考虑到大学生可能会希望进行远程工作,对于线下工作,则会考虑通勤距离.还需要考虑的是,实际情况下每个地区的工作机会是不相同的.因此,我们将引入一个“地图”来表现全体工作的分布.这使得每个岗位除了具有时薪,工作时长和工作难度的属性外,还将拥有一个坐标.

1. 针对大学生们输入的偏好,寻找最好的工作

为了能够充分地从每位本科生的偏好出发, 我们将采用层次分析法来寻找定量化的偏好描述. 基于该描述寻找到最适合用户的工作

1. 展示我们的工作成果.

我们将用……方法来呈现我们的工作成果.

3 Assumptions

To simplify and clarify our problems, we make the following basic assumptions.

1. 假设寻找summer job的大学生的家在一个区域的任意一点.
2. 假设每个大学生都能胜任他所选择的工作.
3. 假设所有工作机会都是公开的,每个大学生都能够浏览到所有工作.
4. 假设每个工作按周结算.
5. 假设不同的工作按难度分为五个种类.
6. 假设不同地区能够提供的工作岗位种类相同,但其所占的比例不同.
7. 假设每个地区能够提供的工作岗位随机分布.
8. Notation

|  |  |  |
| --- | --- | --- |
| **Abbreviation** | **Description** | **Unit** |
|  | Location of a post. |  |
|  | Location of an undergraduate. |  |
|  | Distance between and | km |
|  | Whether to accept offline job. |  |
|  | Wage of a post. | dollars/h |
|  | Required working time. | h/week |
|  | Difficulty of a post. |  |
|  | Difficulty of the post the student wants to do. |  |
|  | Attributes of jobs may be compared with each other. |  |
|  | Importance scale of M compare to N. |  |
|  | Consistency index. |  |
|  | Random consistency index. |  |
|  | Consistency ratio. |  |
|  | Judgement matrix of user’s preference. |  |
|  | Judgement matrix of attributes of posts. |  |
|  | User preference weight |  |
|  | Score of posts. |  |
|  |  |  |

1. Analysis and Modeling
   1. Posts Distribution Model
      1. Posts attribute setting

基于大学生选择工作时所考虑的因素(距离,时薪,工作时长,工作难度),我们也将工作设定为五个属性：工作的种类,所在的地点,通勤距离,时薪,每周工作时长.

具体的设定如下:

1. 工作种类:

我们将工作种类设为一个在1到5之间的一个常数.即按工作难度和所需求的技能程度,分为:

低难度工作.如快递送货员,工地施工者.

中低难度工作.如餐厅洗碗工,传单发放.

中难度工作.如推销员,接待员.

中高难度工作.如普通职员,小型企业的实习生.

高难度工作.如家庭教师,助教,大型企业的实习生.

1. 工作地点

对于线下的工作,我们将工作地点设为区域内的平面坐标系中的一点

而对于远程工作,我们认为工作地点即为求职者的家中,则不需要设定工作地点的坐标.

1. 通勤距离

我们将全体本科生的家设定为在一个区域中的任意一点,通勤距离设定为家坐标与工作地点坐标的直线距离.

如果这份工作是远程的,那么我们认为通勤距离为0.

1. 时薪

对于每一份工作,我们认为时薪是一个不变的量,其单位是dollars per hour.

在不同地区,可能相同种类工作所提供的时薪却并不相同,这将在之后的Posts generation and distribution步骤中体现.

1. 工作时长

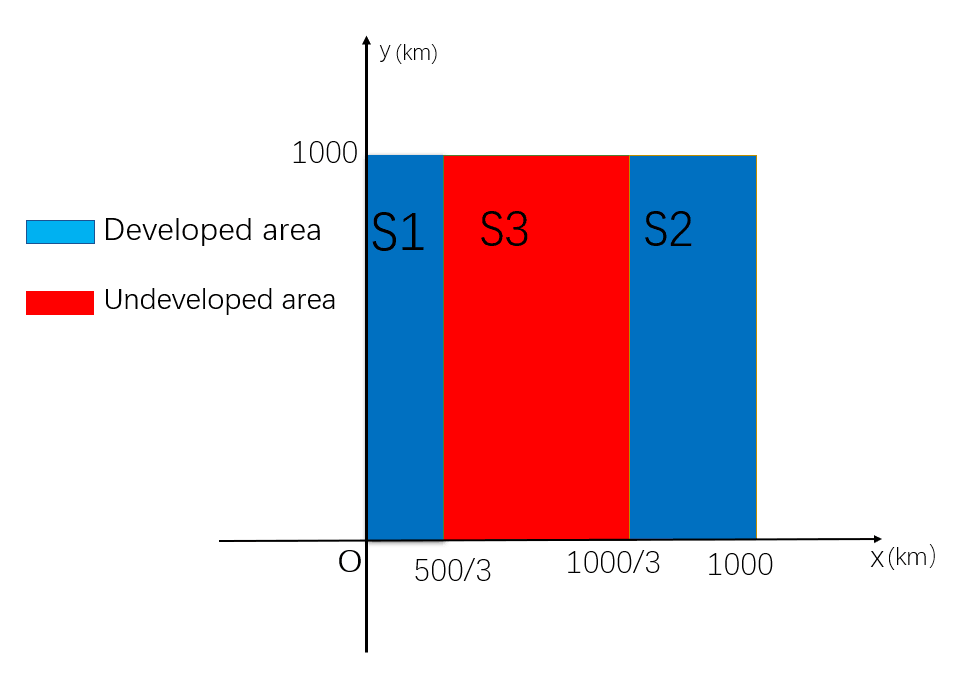
对于每一份工作,我们认为工作时长是一个不变的量,其单位是hours per week.

* + 1. Posts generation and distribution

为了能够合理地生成工作,我们首先建立了一个模型,更确切的说,一个地图,用于表示全体找工作的本科生的家所在地,以及线下工作的工作地点.

我们将这份地图设置为一个国家,包括developed area 和less developed area.这个地图将影响线下工作机会的分布.

为了合理简化模型,我们参考了美国本土的经济发展情况,如下图,我们将中部的1/2划分为less developed area,东部的1/3和西部的1/6划分为developed area.南北跨度为1000km, 东西跨度为1000km.



**Figure 1** Division of developed and less developed area.

为了将工作生成并投放入地图中,我们查询了著名的求职网站,并整理了网站里有关summer job的信息.

|  |  |  |  |
| --- | --- | --- | --- |
| Requirement | Wage | Hours per week | Proportion |
| Low | $3-$5 /h | 20-40 h/w | 0.0400 |
| Mid-Low | $3-$10 /h | 20-30 h/w | 0.1000 |
| Medium | $10-$15 /h | 20-50 h/w | 0.2600 |
| Mid-High | $20-$30 /h | 30-50 h/w | 0.3600 |
| High | $30-$50 /h | 40-50 h/w | 0.2400 |

**Table 1** Distribution of posts in developed area [1]

|  |  |  |  |
| --- | --- | --- | --- |
| Requirement | Wage | Hours per week | Proportion |
| Low | $1-$3 /h | 20-40 h/w | 0.1535 |
| Mid-Low | $3-$8 /h | 30-40 h/w | 0.4386 |
| Medium | $8-$10 /h | 30-50 h/w | 0.3289 |
| Mid-High | $10-$20 /h | 30-50 h/w | 0.0768 |
| High | $20-$25 /h | 40-50 h/w | 0.0022 |

**Table 2** Distribution of posts in less developed area [2]

我们假设developed area 和 less developed area提供的线下工作岗位一样多.全国共有100000个线下职位供我们的本科生选择.

我们将按比例随机生成50000个在developed area里的不同种职位,和50000个less developed area里的职位. 接着随机投放入地图中.

其中, 对于Table 1 和 Table 2 中的每个范围值, 我们使用随机数生成函数进行随机取值.例如, 对于Table 1 中的一个High requirement job (that is ) , 我们可能会生成一个, 的工作岗位.

接着,将两组50000个职位按比例均匀投放到developed area和less developed area中.

这样,我们就得到了可供本科生选择的全部summer job.

* 1. Preference describing

4.2.1 Judgement matrix constructing

我们认为每个学生的偏好是指, 该学生评判一个工作所具有的属性对于他或她而言的重要程度.

为了得到更精确的分析结果, 我们首先会请用户确定他希望做线上还是线下工作, 再选择工作的难度.

我们选择了1到9之间的常数作为标度.如果工作属性M和工作属性N比较有如下情况:

|  |  |
| --- | --- |
| Scale | Description |
| 1 | M and N are of equal importance. |
| 3 | M is a bit more important than N. |
| 5 | M is more important than N. |
| 7 | M is mightily more important than N. |
| 9 | M is extremely more important than N. |
| 2,4,6,8 | The median of the two adjacent judgments above |
|  | Measurement of N compare to M. |

For example, if represents Important scale of wage compare to working time, then represents important scale of working time compare to wage and .那么, 针对远程工作和线下工作, 我们可以设置本科生对工作属性的判断矩阵.

远程工作的判断矩阵如下:

线下工作的判断矩阵如下:

判断矩阵是完全基于用户所给出的偏好来构造的.

4.2.2 Quantitative description for preference

对于线上工作而言, 需要比较的只有时薪和工作时长. 用户对两个属性的评价是确切的. 因此这个判断矩阵我们认为是合理的.

然而, 对于线下工作而言, 这个判断矩阵是不够严谨的. 因为用户对于三个属性的评价可能是不明确的. 例如, 用户可能会认为时薪比工作时长更重要, 工作时长比距离重要, 距离比时间重要, 这样就构成了一个自相矛盾的状况. 因此, 我们需要对判断矩阵进行一致性检验.

对判断矩阵(2) 求解特征根, 找到其最大特征根 , 对应的特征向量, 则一致性指标(为矩阵的秩):

通过查表得知, 当矩阵秩为三时, .

一般认为一致性比率时，认为判断矩阵的不一致程度在容许范围之内，有满意的一致性. 若所求出的, 我们则会请用户重新思考自己的偏好.

为了进一步明确用户对各个属性的偏好,我们还需计算用户对各个属性的偏好权重.

若,则

若,则

由以上分析, 我们就得出了用户偏好的定量化描述.

* 1. Job choosing

4.3.1 Judgement matrix constructing

我们先对10个工作进行评判并寻找最优.

当, 即用户只希望做远程工作时, 构造判断矩阵. 当, 即用户希望做线下工作时, 构造判断矩阵.

以的情况为例进行说明.在用户输入之后, 开始对全部满足要求的工作进行分析.

对于工作所具有的属性而言, 它们均不一定为1-9的常数, 因此不能够作为判断的标度. 首先我们将对这些工作去量纲化并化为能够形成判断矩阵的标度.

选择个工作的薪水构成如下的矩阵

其中, 若(且)

选择个工作的工作时长构成如下矩阵

其中,若且)

请注意与的构造方法与的构造方法不同.

参考4.2.2中的一致性检查方法

进行一致性检验时, 值查询如下表:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|  | 0 | 0 | 0.58 | 0.90 | 1.12 | 1.24 | 1.32 | 1.41 | 1.45 | 1.49 |

Table 3 Random consistency index

4.3.2 Best job choosing base on user’s preference

构造出三个判断矩阵后, 可以求出每个工作从这三个维度出发的评分.

每个工作在时薪方面的得分为

同理, 每个工作在工作时长方面和通勤距离方面的得分分别为

为了找到最满足用户需求的工作, 我们求出对于该用户而言, 这个工作里的每一个工作的总得分

总得分最高的工作, 即为这个工作中最符合用户要求的工作.

在应用层次分析法时, 通常不会大于10. 因此可供选择的工作数量远远超出我们一次性能够分析的工作数. 为了分析全部可供选择的工作, 我们将每次总得分最高的工作保留, 并与下面还未参与分析的个工作共同作为下一次分析的输入参数. 以此得出在全部工作中最符合用户要求的工作.

5. Model verification