**Summary**

**Contents**

[**1 Introduction 3**](#_Toc56947337)

[**1.1 Background 3**](#_Toc56947338)

[**1.2 Restatement of the problem 3**](#_Toc56947339)

[**2 Overall Analysis 3**](#_Toc56947340)

[**3 Assumptions 4**](#_Toc56947341)

[**3 Notation 5**](#_Toc56947342)

[**4 Analysis and Modeling 6**](#_Toc56947343)

[**4.1 Posts Distribution Model 6**](#_Toc56947344)

[**4.2 Analytic Hierarchy Process 8**](#_Toc56947345)

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. 确定求职者们选择工作的流程。

这将帮助我们简化在评价工作时所使用的算法.通过让求职者首先输入一些条件,将使得工作机会经过筛选后,再呈现于用户的面前.

1. 针对大学生们输入的偏好,评价工作是否适合他们.

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评价工作对于每个大学生的合适程度.

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 |
|  | Maximum distance acceptable for student. | km |
|  | 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. |  |
|  |  |  |

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

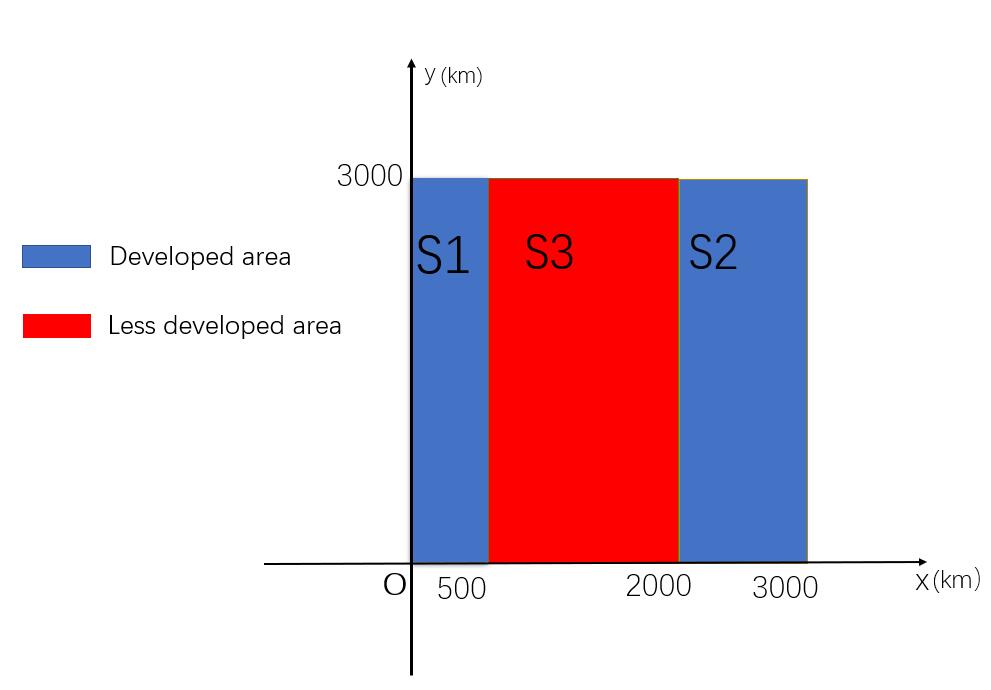
对于每一份工作,我们认为工作时长是一个不变的量,其单位是hours per week.

* + 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.

4.2 Analytic Hierarchy Process

4.2.1 Description of preference.

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.为了将重要程度量化,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成对比较矩阵.

远程工作的成对比较矩阵如下:

线下工作的成对比较矩阵如下:

成对比较矩阵是完全基于用户所给出的偏好构造的.

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.

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, 我们需要对成对比较矩阵进行一致性检验.