

中南财经政法大学“博文杯”

实证创新基金项目

中南财经政法大学

“博文杯”大学生百项实证创新基金项目



项目主题： 基于决策树的个性化专车服务优化模型

---以 Uber 专车为例

指导老师： 熊平

主 持 人： 邓赓

团队成员： 苏以鹏 张玉潇 王胜

日 期： 2016 年 2 月

基于决策树的个性化专车优化模型

——以 Uber 专车为例

作者： 邓赓 苏以鹏 张玉潇 王胜

摘要 UBER TECHNOLOGIES INC（简称 Uber），是一家风险投资的创业公司和交通网络公司，其旗下 Uber(优步)是一款全球即时用车软件。Uber 打车是一个按需服务的 O2O（Online to Office）网站，网站以最简单的方式最快速的匹配时间，使得空闲的轿车资源网络化而得到最大化利用。每一个有需求的用户通过 iPhone、SMS、Android 向 Uber 发送请求，从而找到自己的私人司机进而再购买私家车搭乘服务。其特色之处便是“通过 GPS 追踪定位私家车，用户可以使用 Uber 发出打车请求；几分钟内一辆私家车就会开到你面前，支付小费或通过信用卡完成交易。”我们结合决策树模型的分析功能，以武汉的 Uber 市场为例，通过问卷调查，走访询问，查询资料等方式来分析探讨 Uber 推出个性化私人服务的可能性，以此来开发 Uber 专车服务的更深层潜力。

关键词 专车；优化模型；决策树；

一、 导言

如今，在“互联网+”的时代背景下，打车软件十分盛行，这种“出租车+互联网”模式成功地解决了出租车司机与打车者之间信息不对称的问题，极大的方便了用户的出行。而 UBER 便是这样的一个打车软件，但是与众不同的是，UBER 专注于成为平台，连接司机和乘客，让司机提供搭乘服务。更值得关注的是，UBER 司机多数为私家车，从而引导私家车参与到城市交通当中，这样在某种程度上面缓解了交通。

UBER 的这种特点，使得 UBER 推出个性化专车服务成为可能，从而使得出租车不再仅仅局限于提供搭乘服务，能针对用户的不同需求提供不同的专车服务。因此，探究 UBER 提供个性化专车服务的可行性具有重要意义。

二、 调研背景

(一)选题背景

以 Uber 为代表的一系列新兴科技企业出现爆发式增长，标志着共享经济时代的到来。随着中国城市化进程和生活水平的不断提高，在衣、食、住、行的基本生活要求之上，共享经济向更多的细分领域扩张。体现社会尊重及自我价值实现的个性化服务需求变得越来越强烈。例如婚礼咨询、家具设计、活动策划、插图设计等，人们已经逐渐习惯选择可定制的个性化服务。

正如《2015 年消费趋势：个性化、私人订制成大趋势》中说道“目前我国经济发展处于三期叠加的“新常态”阶段，即增长速度进入换挡期、结构调整面临阵痛期、前期刺激政策面临消化期，随着经济增速放缓，零售商营收增幅回落。消费结构受到商品消费分流、大众消费流转流、公务消费节流、高端消费外流、外部因素扰流的综合影响，个性化、定制化消费取代排浪式消费成为 2015 年新兴消费趋势”。在网络上，许多个性化需求都可以

得到满足，你甚至可以“在办公室点一杯星巴克，Uber 司机会为你送货上门”。而这，便是私人化定制服务的迷人之处。

个性化私人服务定制时代依然到来，而此前也有许多尝试试水案例，在《极致 Uber》中提到“Uber 在曼哈顿退出一项快递服务——由骑着自行车的快递员为用户提供按需快递服务，通过手机 app 上发送请求，快递员在很短时间内过来取件，然后配送到指定地点；之前 Uber 还尝试用冰激凌车给预定的人送冰激凌、在情人节递送玫瑰花、提供租赁直升飞机服务等多项打车以外的服务”。无疑，这便是一些个性化私人服务经典案例，Uber 跳出了传统的打车模式，相反推出了更贴近用户私人化的个性服务，用户可以有各种各样的要求，只要是关乎运输交通，Uber 及 Uber 司机都会想方设法满足用户的要求。

(二)选题现实意义

1. 有利于城市交通的缓解

Uber 以其强大的社会闲置汽车资源整合能力，将平时空闲的私家车整合起来提供服务，这使得汽车没有那么必要，只要 Uber 在手，任何需要运输的个性化私人服务都可以得到满足。因而，Uber 实行个性化私人接送和派送服务将会大大降低汽车的需求量从而缓解拥挤的城市交通。

2. 使用户更加方便使用

无疑，Uber 推出这样个性化的私人用户将会大大方便用户的使用，带来极其舒适的用户体验，无论身在何方无论有没有车，只要手机里面有一个 Uber App,所有的需求将会得到满足。

3. 有利于空闲资源的整合

在这样一个共享经济的时代下，追求的是资源的充分利用。而 Uber 恰恰符合这一特性，它整合了空闲的私家车资源，将它们加以利用，无论是对于私家车司机还是对于用户来说都是一种好的选择。

4. 有利于推动出租车行业发展

传统的出租车行业将不再适合现在的时代潮流，如今只有结合互联网技术进行改革才能继续发展，这是市场的发展趋势，专车的个性化私人服务尝试将会带来一场顺应时代潮流的大改革，市场即将迎来个性化私人服务时代。

(三) 研究目的及意义

由以上的叙述可知随着时代的发展技术的突飞猛进，市场的改革也在逐渐进行，可以预计的是在未来的市场中个性化的私人服务将会得到广泛的支持与应用。目前已经在各各业已经开始进行私人化服务的尝试，效果良好。而作为始终站在舆论关注焦点的 Uber 公司，我们从它入手探寻在专车领域个性化私人服务中需要注意的要素，从而得到在实行专车个性化私人服务中的关键环节，这对于我们在改革并构建新型的交通服务领域有很大帮助，同时也有利于我们理解个性化私人服务的内涵，以及对于这方面的大学生创业有着指导性的作用。因此，研究该选题不单单对于用户有着巨大的影响，同时对于个性化私人服务的市场改革有着重要的意义。

三、实证分析

(一) 调研方式

调研过程中，我们除了查阅网上的 Uber 相关数据还采取问卷调查的方式，用分别针对于 Uber 司机和用户的两份问卷，了解 Uber 乘客以及司机的相关信息及对个性化专车模型构建的一些看法与感受。为了全面了解各年龄段乘客对个性化专车的看法，我们采取网上发放问卷形式，对乘客进行访问。而为了了解各 Uber 司机对个性化专车的看法，我们分别找到 Uber 华中地区总部武汉及 Uber 中国西南地区总部成都，对分属不同两地的 Uber 司机进行问卷调查，将他们的回答进行汇总，进行详细分析，然后筛选出影响 Uber 个性化专车模型构建的属性分类，建立相应的 Uber 司机决策树和 Uber 用户决策树，从而提出具体的优化建议，更好的推进个性化专车在武汉，成都甚至全中国的推行。

(二) 调研内容

我们小组在调研前期做好了充足的准备，包括查阅资料，访谈 Uber 司机，Uber 用户，对本次调研所需的决策树模型以及注意事项进行了充分的学习，筛选出影响成都可能较高的八个属性，从而设置问题制作成问卷发放，与此同时，我们在调研地点与调研时间上做了精心的准备与选择。本次调研选取 Uber 华中地区总部武汉及 Uber 中国西南地区总部成都，对分属不同两地的 Uber 司机进行问卷调查，同时通过多名同学近期乘坐 Uber 的乘车记录联系到司机，从而分发问卷，保证问卷的分量与质量。

四、调查问卷结果及分析

(一) 问卷分析确定探究 Uber 个性化专车服务的影响因素

①Uber 司机调查问卷及个性化专车服务司机方面影响因素分析

11: 您属于哪种年龄段的人群?

通过这一个问题，可以了解到司机大概的年龄，因为个性化专车项目，与时代步伐一致甚至迅于时代步伐，年轻人对新鲜事物接受较快，而中年或老年人可能对此不太感兴趣，从而造成提供个性化专车服务的质量差异，进而影响用户体验，所以我们可以确定影响属性：司机年龄，记做 R1

4: 您是 Uber 的全职司机还是专职司机?

通过这个单选题，我们可以了解到司机是全职整天跑 Uber 还是在工作下班之余，开 Uber 赚赚小钱，与上题相同，是否全职很大程度上影响了 Uber 司机一天的工作量与投入的精力，从而对个性化专车服务造成影响，所以可以确定影响属性司机是否专职，记做 R2

2: 您以前是出租车司机吗? , 5: 作为 Uber 司机, 您有同时是其他搭车软件的专车司机么?

通过这两个问题，我们可以了解到司机是否只为一家搭车软件公司工作，而忠诚度决定了司机对 Uber 所投入的精力与感情，很大程度上会影响司机提供服务的质量，对乘客的个性化接送服务很可能存在不良的影响，所以我们可以确定影响属性：司机是否只用 Uber 一款打车软件，记做 R3

1: 您为什么当上 Uber 司机? , 3: 您是怎么了解到 Uber 的? 6: 支持您继续当 Uber 司机的理由是?,8:您对 Uber 的每周奖励政策满意么?

通过这四个问题, 我们可以了解到几乎所有司机都是慕名 Uber 良好的经济收益而来, 而这肯定直接影响了司机是否愿意提供个性化专车服务, 所以可以确定影响属性: 对 Uber 奖励政策的满意程度, 记做 R4

9: 您每天的平均接单量一般为多少?

通过这个问题, 我们能了解到司机每天的接单量, 根据司机反映的接单量, 我们可以确定司机每天提供接送乘客的服务次数, 而次数的多少很大程度上决定了司机提供个性化专车服务的能力大小。所以我们可以确定影响属性: 每天接单量, 记做 R5

12: 如果 Uber 推出定点接送儿童和老人的个性化接送服务, 您是否有兴趣? , 13: 如果 Uber 和快递点, 餐饮公司, 商家建立合作关系, 推出派送快递, 送餐, 送礼等个性化接送服务, 您是否有兴趣? 14: 您认为如果 Uber 推出个性化专车服务, Uber 公司还存在哪些不足?

通过这 3 个问题, 我们能了解到各个 Uber 司机对个性化接送, 派送服务的感兴趣程度, 而兴趣是一个人最好的老师, 对项目感兴趣与否很大程度上会影响个性化接送服务的质量。所以我们可以确定影响属性: 对提供个性化专车服务的感兴趣程度, 记做 R6

②Uber 司机调查问卷及个性化专车服务用户方面影响因素分析

1: 你的年龄是在?

通过这一个问题, 可以了解到用户各年龄段的接受程度, 因为个性化专车项目, 与时代步伐一致甚至迅于时代步伐, 年轻人对新鲜事物接受较快, 而中年或老年人可能对此不太感兴趣, 从而造成个性化专车服务被接纳程度不同, 这是需要考虑的重要因素, 所以我们可以确定影响属性: 用户年龄, 记做 V1

2: 请问你的职业领域是? 3: 你对 Uber 的了解程度? 4: 你更愿意选择哪种出行方式? 5. 你使用打车 APP 的机会多?

不同身份和社会背景的人对打车软件的接纳程度和使用情况, 存在很大差异。十分有必要弄清楚哪一类人群更能接受 Uber 的个性化专车服务, 所以把用户的职业领域作为影响属性, 记为 V2。

6: 你选择 Uber 有哪些原因? 11: 你通过 Uber 帮助你解决紧急派送任务, 你认为有哪些优点? 15: 你觉得 Uber 提供私人化接送服务有哪些优点?

用户选择 Uber 作为自己的出行选择, 一定有对应的原因。只有调查清楚最主要的选择原因, 才能最大化 Uber 的优势, 进一步进行优化, 最大限度的优化 Uber。因此, 把选择原因也作为影响属性, 记为 V3

7: 你一般在什么情况下使用 Uber? 8: 你是否在生活中某个时刻需要紧急用车? 9: 你在生活中是否遇到紧急派送服务? 13: 你在生活中是否遇到脱不开身, 又需要定点准时介绍人的情况?

Uber 推出个性化专车服务, 就需要了解, 在哪些情况下 Uber 被使用最多。只有清楚

了解哪些情况的使用频率最高，才能更好地，更具针对性推出个性化服务。使用情况也必须作为重要的考虑要素，因此把使用情况列为影响因素，记为 V4

12: 你对于 Uber 个性化服务有哪些顾虑 16: 你觉得 Uber 提供私人化的接送服务，有哪些缺点？

任何技术和思想都有其不足的一面，Uber 同样也是。分析清楚自身的不足，才能更长远的发展。因此用户对于 Uber 提供个性化专车服务的担心顾虑，必须放在我们的考虑范围之内。因此把用户的顾虑作为影响因素，记为 V5

14: 您对于将来 Uber 提供私人化接送服务，来解决你接送难的问题怎么看？

Uber 推出个性化专车服务，最终还是要服务于用户。因此用户的态度和看法就显得尤为重要。调查结果很大程度上，反映 Uber 推出个性化专车服务的可行性。因此用户的态度也要作为重要的影响要素，记为 V6

(二) 问卷结果汇总，建立司机数据集，用户数据集

表 1: 司机数据集 D1

B	C	D	E	F	G	H
用户年龄	职业领域	选择Uber的原因	使用情况	紧急派送需求	用户顾虑	用户态度
A	A	ABCD	B	B	BC	B
B	A	A	B	B	BCD	C
B	A	D	B	B	B	B
B	A	A	B	B	BC	B
B	A	A	B	A	BC	B
B	A	BCD	C	A	BC	B
A	A	ABCD	B	B	BC	A
A	A	AD	A	B	CD	B
A	A	ABD	B	B	BC	A
A	A	ABCD	B	B	BC	A
B	A	BC	B	B	BCD	B
A	A	AD	B	A	B	B
B	A	BC	B	B	D	B
A	A	A	A	A	A	A
A	C	BC	C	D	D	C
B	A	A	B	C	D	B
B	A	BC	B	B	D	C
A	A	ABCD	B	A	CD	B
A	A	AB	D	C	BC	B
B	A	CD	A	A	A	A
A	A	AB	A	A	B	B
A	A	BD	B	B	BCD	B
B	A	B	B	B	A	B
A	A	A	B	A	BC	B
B	A	CD	B	B	CD	B
A	A	AB	B	B	B	B
A	A	ABCD	B	A	BD	C
A	A	ABC	B	A	C	A
A	A	ABD	A	C	BCD	B
A	A	ABCD	B	B	BCD	A
B	A	AB	B	B	BCD	A
A	A	D	C	C	B	B

表 2：司机数据集 D1

A	A	A	C	B	B	B				
B	A	D	B	C	C	B				
A	A	C	A	A	D	B				
A	A	D	A	A	A	A				
B	A	D	B	B	BC	B				
A	A	CD	B	B	C	B				
A	A	C	B	B	BD	B				
E	C	D	B	B	BC	A				
E	B	BC	B	C	BC	C				
A	A	E【没有】	B	A	B	B				
B	A	AD	B	A	CD	B				
B	A	B	B	A	BC	C				
A	A	CDE【无】	B	B	AE【wu】	B				
A	A	BC	A	B	B	C				
A	A	ABD	B	C	B	B				
A	A	AD	B	B	BC	A				
B	A	A	B	C	A	A				
B	A	ABCD	B	A	ABCD	B				
B	A	C	B	B	D	B				
B	A	AB	B	A	BD	B				
A.15-20	A.学生	A.乘车环境好，服务A.几乎没有	A.几乎没有	A.几乎没有	A.没有顾虑	A.十分看好，有市场前景				
B.21-25	B.服务领域	B.调度最近，等待B.偶尔需要	B.偶尔有	B.偶尔有	B.服务质量可能与B.中立态度，看未来发展					
C.26-30	C.生产领域	C.派单模式，被拒有C.需要	C.会遇到	C.会遇到	C.价格标准难以确定C.不看好，安全性，服务质量参差不齐等问题难以解决					
D.31-40	D.其他	D.价格更优惠D.经常需要	D.经常遇到	D.经常遇到	D.手机的定位不准D.其他					
E.41-50		E.其他			E.其它					
F.51-60										

表 3：司机数据集 D2

司机年龄	司机是否专职	司机是否只用Uber一款打车软件	对Uber奖励政策满意度	每天接单量	对个性化接送感兴趣程度	是否支持个性化专车
B	B	A	B	C	B	支持
C	B	B	B	D	D	不支持
C	A	B	B	C	D	支持
D	B	A	C	D	D	不支持
C	A	B	C	B	D	不支持
C	B	A	C	D	A	支持
D	B	B	B	C	C	不支持
C	B	A	C	D	D	支持
C	B	B	B	C	B	不支持
B	B	A	B	C	B	支持
B	A	A	B	C	B	支持
C	B	A	B	B	B	支持
D	A	B	A	D	B	支持
C	B	A	B	C	B	不支持
C	B	A	A	B	B	支持
B	B	A	C	C	D	不支持
C	B	A	C	D	D	支持
A: 12-18	A: 是全职司机	A: 不，我只是Uber司机	A: 非常满意	A: 30单以上	A: 非常有兴趣	
B: 18-25	B: 是兼职司机	B: 我也同时是多款打车软件的专 车司机	B: 基本满意	B: 20-30单	B: 有一些兴趣	
C: 25-35			C: 不太满意	C: 10~20单	C: 没兴趣	
D: 35-50				D: 10单以下	D: 看公司的奖励政策	
E: 50以上					E: 不清楚	

(三) 将数据集 D 分类，筛选出候选属性

1.模型构建-决策树思想：

决策树算法是一种逼近离散函数值的方法。它是一种典型的分类方法，首先对数据进行处理，利用归纳算法生成可读的规则和决策树，然后使用决策对新数据进行分析。本质上决策树是通过一系列规则对数据进行分类的过程。

2. 模型建立：

利用信息增益模型：

(1)定义信息

$$Info(D) = -\sum_{i=1}^m P_i \log_2 (P_i)$$

参数注释：

①m 表示数据集 D 中类别 C 的个数

②P_i 表示 D 中任意一个记录属于 C_i 的概率

③ Info(D) 表示将数据集 D 不同的类分开需要的信息量

(2)根据属性 R_i 的分类对数据集 D 进行分组 ($D_1, D_2, D_3 \dots D_k$)

$$Info_R(D) = \sum_{j=1}^k \frac{|D_j|}{|D|} \times Info(D_j)$$

结果表示将数据集 D 不同的类分开还需要的信息量

(3)计算两个信息量之差：

$$Gain(R) = Info(D) - Info_R(D)$$

(4)选出信息量之差最大的 4 个属性，作为决策树的分支，建立决策树

3.对于用户数据集 D_1 ，计算各属性信息量

$$Info_{V_1}(D_1) = 0.5837$$

$$Info_{V_2}(D_1) = 0.5537$$

$$Info_{V_3}(D_1) = 0.5952$$

$$Info_{V_4}(D_1) = 0.5889$$

$$Info_{V_5}(D_1) = 0.5434$$

$$Info_{V_6}(D_1) = 0.5908$$

$$Info_{V_3}(D_1) > Info_{V_6}(D_1) > Info_{V_4}(D_1) > Info_{V_1}(D_1) \\ > Info_{V_2}(D_1) > Info_{V_5}(D_1)$$

所以影响 Uber 用户支持个性化专车服务的属性决定程度次序为：

用户选择乘坐 Uber 的原因 > Uber 用户对个性化专车项目的态度 > 用户使用 Uber 的打车情况 > 用户年龄 > 用户的职业领域 > 用户对个性化专车的顾虑

4.对于司机数据集 D_2 ，计算各属性信息量

$$Info_{R_1}(D_2) = 0.0118$$

$$Info_{R_2}(D_2) = 0.3471$$

$$Info_{R_3}(D_2) = 0.6459$$

$$Info_{R_4}(D_2) = 0.9696$$

$$Info_{R_5}(D_2) = 0.9591$$

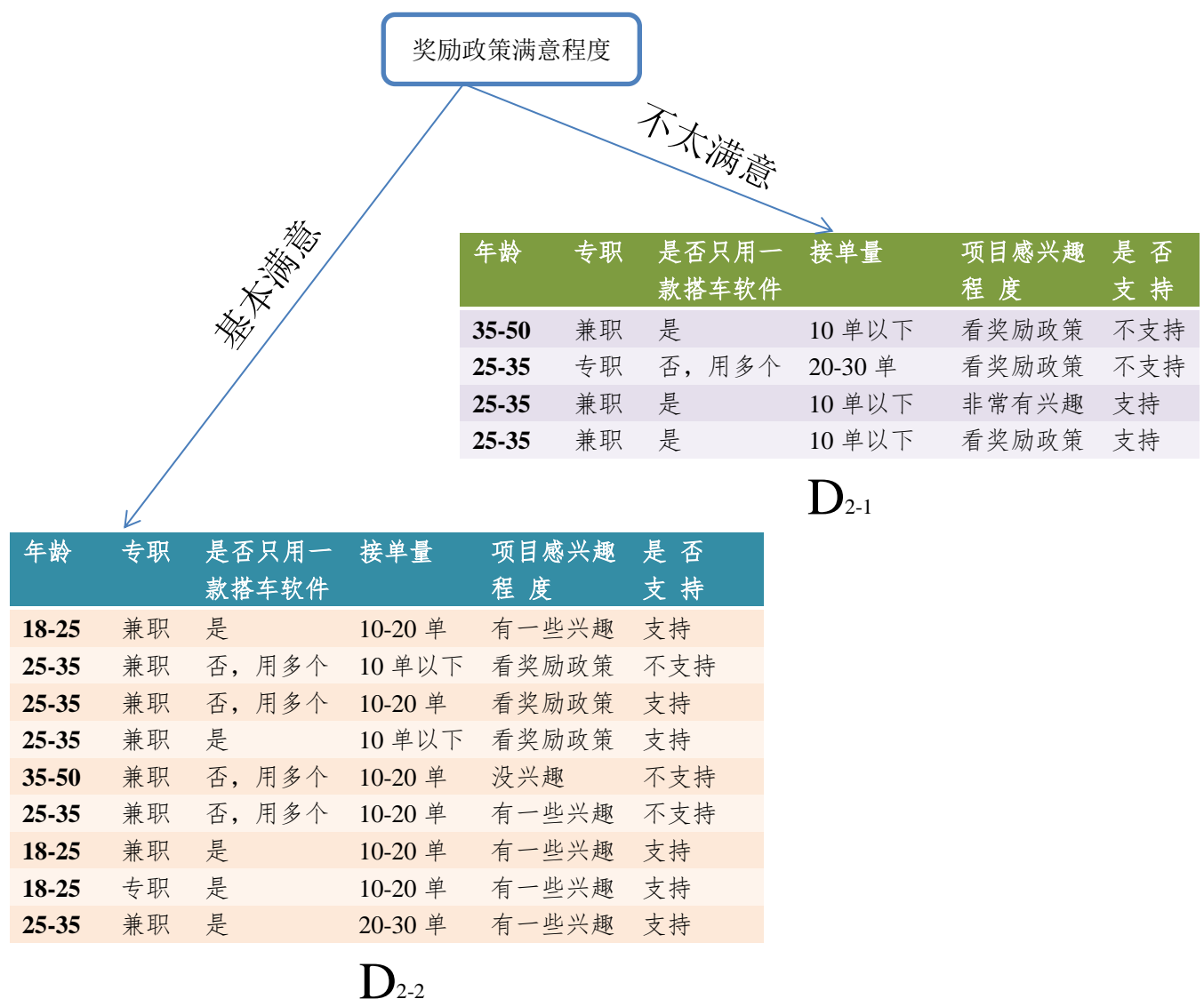
$$Info_{R6}(D2) = 0.7054$$

$$Info_{R4}(D2) > Info_{R5}(D2) > Info_{R6}(D2) > Info_{R3}(D2) \\ > Info_{R2}(D2) > Info_{R1}(D2)$$

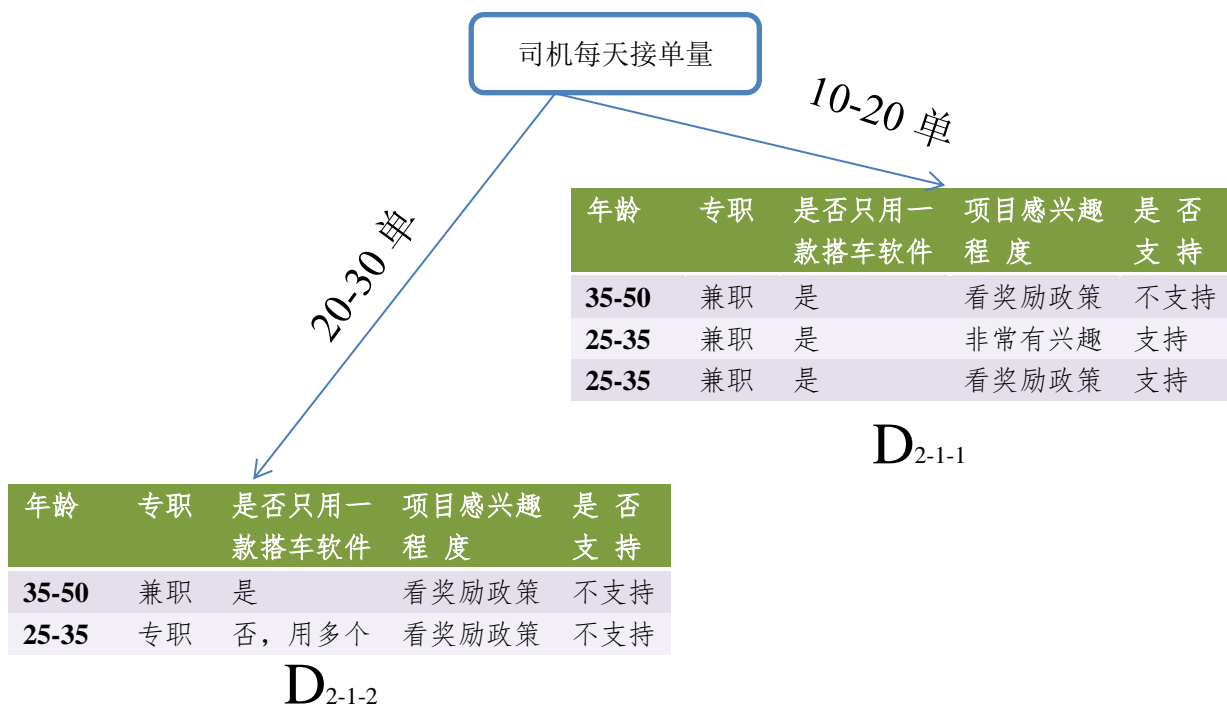
所以影响 Uber 司机支持个性化专车服务的属性决定程度次序为：

司机对 Uber 奖励政策的满意程度 > 司机每天接单量 > 司机对个性化专车接送感兴趣程度 > 司机是否只用一款搭车软件 > 司机是否专职 > 用户年龄

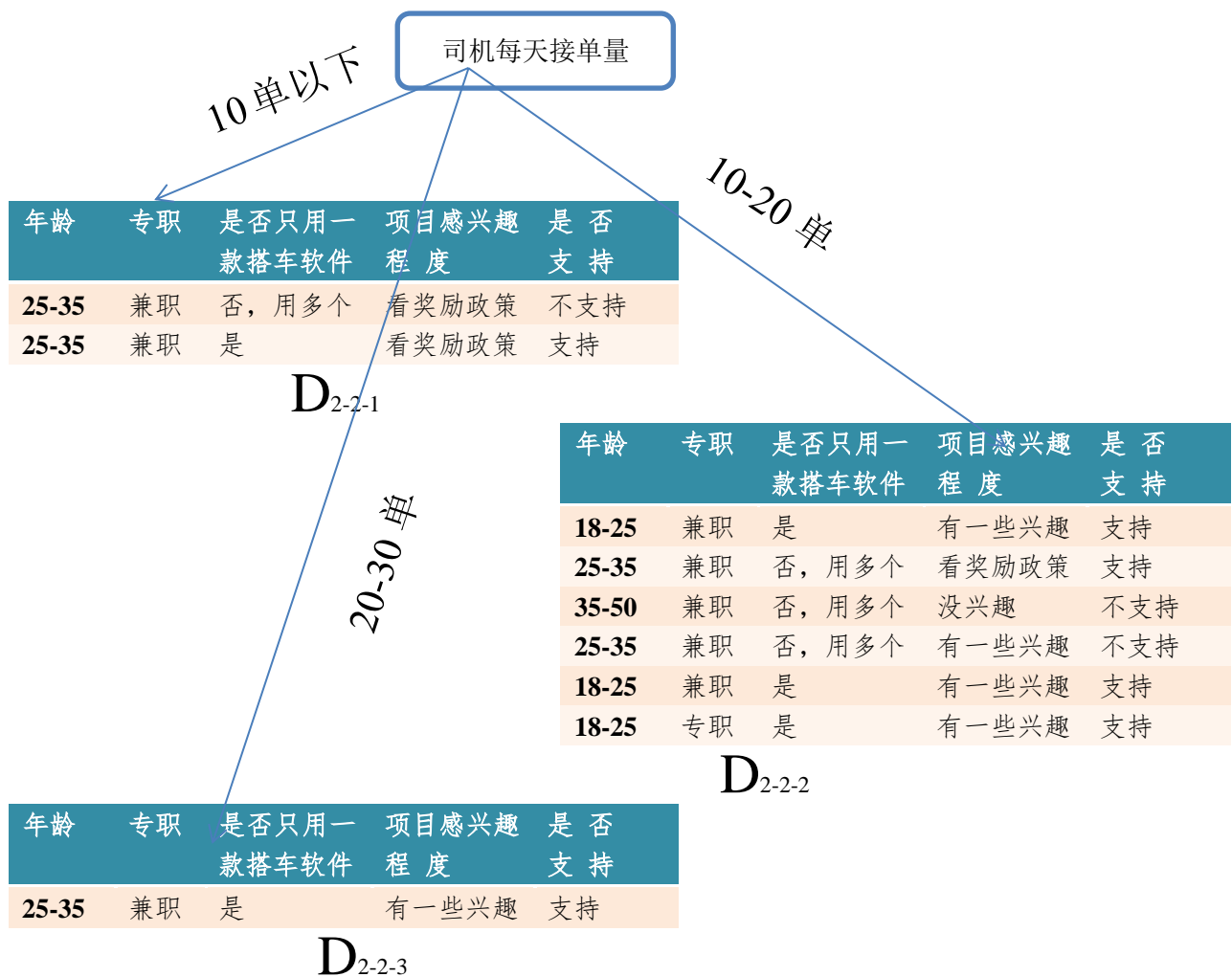
5.建立司机决策树



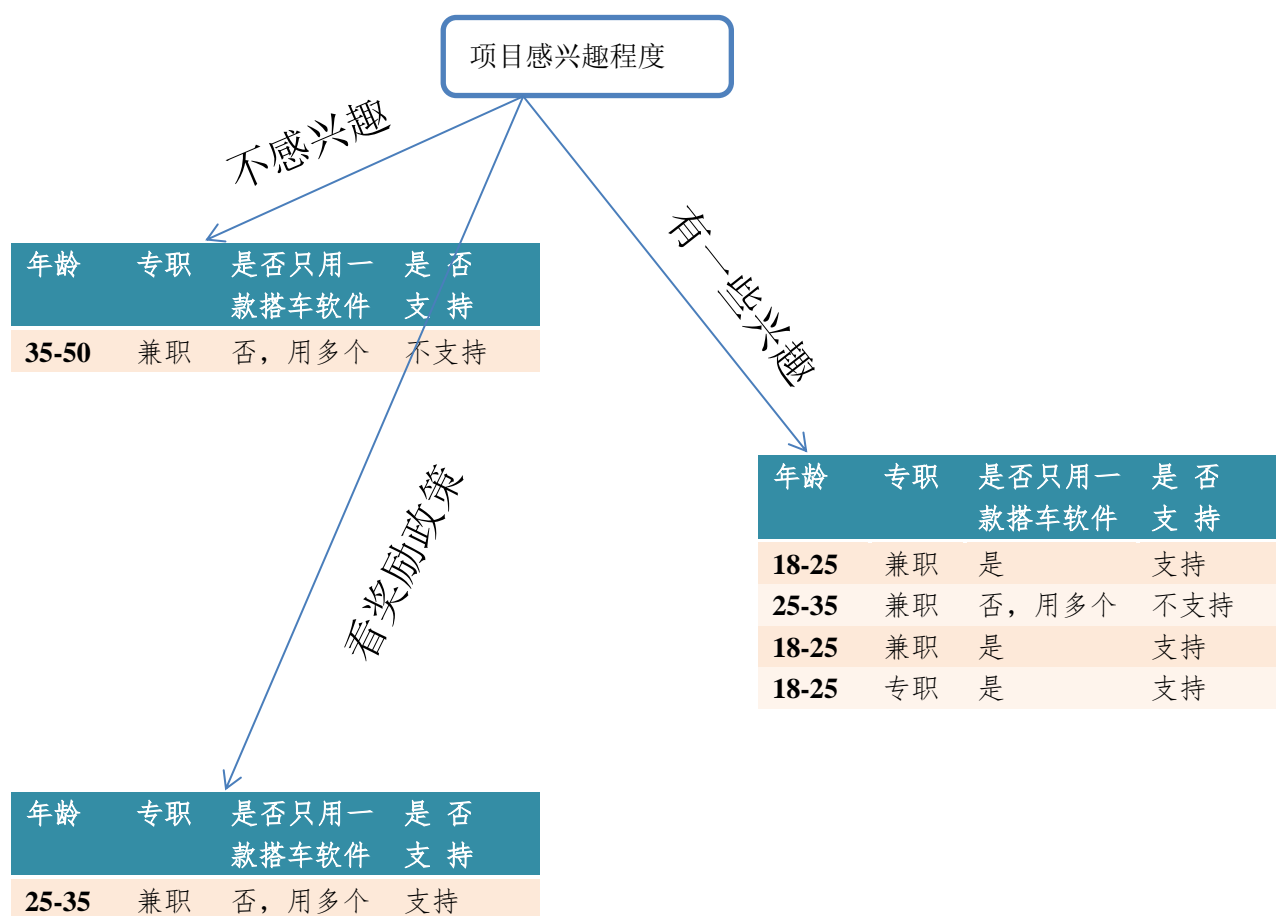
对于数据集 D_{2-1} ,



对于数据集 D_{2-2} ,



对于数据集 D_{2-2-2}



五、优化模型建立及建议

根据用户数据分析结果来看信息量的值介于 0.543 和 0.596 之间,波动相对而言较小,其中用户年龄,选择 uber 的原因,使用 uber 的情况和用户的态度有着较为重要的作用,而选择 uber 的原因又是重中之重,这个模型反应出来的结果契合了之前的分析——用户选择 uber 主要是为了方便,而个性化服务又针对私人将 uber 的服务向方便客户又推进一步。因此个性化服务是有群众基础的,大部分的用户是愿意尝试的,进一步根据数据而言,在推广私人化服务的前期,将用户定位为 20——40 岁且使用 uber 频率较高,原因是这类人群不仅容易接受新事物而且较为繁忙。

根据 uber 司机数据结果分析看来,其信息量的值介于 0.01 和 0.96 之前波动不定,很明显波幅很大,其中两项指标在 0.95 以上,分别是公司奖励政策和每天的接单量。接下来是司机的兴趣方面信息量为 0.7054,相比下降了 0.15,降幅明显,表明其已不是最主要的指标,回到数据本身分析,公司的奖励政策和每天接单量直接决定了 uber 的司机的收入,而其余数据并不能,因此对于司机而言,其收入是绝对的决定因素。所以只要能为司机提供可观的收入,就会有一定数量的司机去做个性化服务。在推广私人化服务时,应当面向全部 uber 司机,但是在选择 uber 司机接单的时候,应当选择评分较高的司机,因为这样容易给用户带来良好的体验,反过来也能促进 uber 司机的收入。

在对 uber 司机的访谈中,我们发现年龄相对较大的司机,对于接送人这方面的个性化服务表现出了明显的拒绝,原因也是相当一致——安全没有保障,他们司机本身无法承担发生事故后的责任。此外,部分兼职司机,对个性化服务不感兴趣,他们基本上是为了减

少养车压力而成为 **uber** 司机的，对这些特殊的服务并不感兴趣。不同的个性化服务，在选择司机方面也是有所不同的，这也是值得注意的地方。

六、 结语

通过用户决策树对数据集 **D** 进行合理的分类，使各个属性类别更好更直观的反映出这一类别的用户对 **Uber** 个性化专车服务的支持情况，从而将支持 **Uber** 个性化专车服务的用户人群与不支持 **Uber** 个性化专车服务的用户人群分隔开，实现用户精准定位。

通过司机决策树对数据集 **D** 进行合理的分类，分类出愿意接受个性化专车服务的司机类别，在 **Uber** 线上平台调度司机完成个性化专车服务的订单时，优先选用这些司机。

通过精确定位个性化专车服务需求人群和愿意提供个性化专车服务的司机人群，使最好的司机去满足用户的个性化用车需求，从而达到用户的最佳体验。