参赛编号:本 2021010
选题: B (A 或 B)　参赛赛道: 本科生（本科生、专科生或研究生）

杭州电子科技大学数学建模校赛

题　　目　基于排队论和多种评价模型的共享出行匹配模式

摘　　要:

共享出行作为一种新的出行模式，以其迅捷方便作为特点深受大众喜爱。但是在这个过程中，我们遇到车辆与顾客之间如何匹配，以及如何让乘客得到最为舒适的乘车体验的难题。本文就基于排队论和多元评价模型处理单人就近匹配，拼车方式的问题，并且从多方面进行改进提升乘客的体验感，使得乘客的平均等待时间减少，拼车时的体验感提升等。

针对问题一，本文在 DBSCAN 构图的基础上建立了排队论模型，将匹配选单的时间作为排队时间，将等待的时间作为服务时间，车辆离开即认为服务结束，新出现的乘客是下一个排队对象。同时结合实际情况，将匹配的模式分为峰期和谷期，通过确定峰期和谷期内的各个参数数据最终得到的策略方案为：谷期的候车时间为 1 分钟，预估排队时间为 2 分钟；峰期的候车时间为 3 分钟，预估排队时间为 14 分钟。

针对问题二，首先，我们用三种形式表示出拼车的可能性，然后通过最小化的乘客等待时间，拼车的价格估计以及乘客的社会体验三方面进行多因素评价分析，最终通过得分系统计算出拼车的分数为 64.96 分，并且相比于不拼车的模式，拼车会显的更加快捷方便并且价格优惠。

针对问题三，面对一个开放性的动态策略，我们先对匹配区域进行网格化的估值以及区域聚类，并结合距离和时间建立了以下的四个评价指标：1、价格越高，匹配度越高 2、当前位置价值越大，匹配度越低 3、未来位置价值越大，匹配度越高 4、接驾里程，隐形表达，越大则预计送达时间越大，衰减系数越小，匹配度越低。同时改进拼车存在的问题，并通过模糊综合评价模型结合数据求出此体系下的不足点并加以改进，最终得到的动态策略总结成六点，具体内容见原文。

针对问题四，根据《2020 网约车出租车出行调研报告》中的数据显示，目前网约车在价格控制、司机收入、用户投诉数量等方面依旧有巨大的缺陷。因此，我们将从用户满意度模型、司机接单综合改进模型、平台宏观收益模型以及社会福利模型四个模型进行综合性改进，通过对这四者的目标函数以及权重的刻画，从而得到较好的社会总体福利以及用户体验。

关键词：乘客-车辆模型，排队论，多因素评价模型，模糊综合评价，价值体系

目录

一、问题重述

1.1 问题背景

共享出行是指人们无需拥有车辆所有权，以共享和合乘方式与其他人共享车辆，按照自己的出行要求付出相应的使用费的一种新型交通方式。包括以打车软件、共享单车为代表的一大批创新模式。当下滴滴的专车、快车等业务线已经非常的流行与普遍，那么，他们又是如何进行配单的呢 ？我们如何让用户再订单的时候就能与最近的几辆出租车配对，如何解决多用户少出租车的问题？如何能够实施拼车的问题，让有限的时间获得更大的利润？如何解决实时出现的新的用户需求和空闲司机之间的匹配关系？如何从总体的角度将客户的体验做到最佳，都是我们需要考虑并且解决的问题。

1.2 问题提出

根据题目背景以及附件信息，解决以下问题：

从各个角度进行分析，如何将用户与最近的几辆出租车进行匹配，使得大部分乘客的候车时间都能平衡在一个合理的区间。

在第一问的前提下，能够实现乘客之间的拼车，如何确定司机的接单情况，确定行进路线以及人员安排，并且使得乘客的总体体验更加良好。

在一个会随机产生新的客户和空车司机的环境中，运用第一问和第二问的结论，对其进行动态的配对策略制定。

从社会总体福利或者用户的体验亦或是从其他角度对配对策略进行一个可行的方案，使得用户的体验提升。

二、问题分析

2.1 问题一的分析

首先，我们用 DNSCAN 的构图思想去解决乘客与车辆之间的关系，保证以乘客为核心点的图中非核心点能够与其直达。其次，因为车的数量大，顶替速度快，我们可以在大范围内假设离开的车会有新的车辆来补充，这样，一个顾客与司机位置都变化的问题就转化成了以车为参考系（服务台），人为变量的一个模式，便可以用排队论的思想去进行解决。将匹配选单的时间作为排队时间，将等待的时间作为服务时间，车辆离开即认为服务结束。这样便能够指定最为合适的一个半径范围以及等待时间和必要的移动距离。

2.2 问题二的分析

拼车的本质就是在排队论的情况下，能够一个服务平台同时服务两个人，但是服务的时间改变，服务的成本改变。同时，为了让乘客获得更好的体验，我们从最小化的乘客等待时间以及最佳乘客体验两个方面对其进行分析。同时，对于三种不同情况下的拼车路线，要逐个的进行分析，最后与第一问进行比较，从平均等待时间，车费，路上的行驶时间三方面进行评价，最终得出结论。

2.3 问题三的分析

从平台的角度去制定动态的人车配对策略无疑要考虑到的方面更多，因为最终的目标是要在乘客感观好的情况下尽可能多的盈利。不仅仅按照最近距离配送，无距离限制配送等强假设条件。我们要根据配送的费用，出发点和目的地的价值评估，服务利用率，配送率以及拼车路线等方面进行总体的规划，每一次派单的决定也都在影响未来的司机分布。当下滴滴的专车、快车等业务线也已经在普遍使用智能派单模式，即从全局视角出发，由算法综合考虑接驾距离、服务分、拥堵情况等因素，自动将订单匹配给最合适的司机接单。

2.4 问题四的分析

问题四作为开放题，要求从社会福利和用户个体体验进行分析，那便围绕这个两点向外衍生，从用户满意度模型、司机接单综合改进模型、平台宏观收益模型以及社会福利模型四个模型进行综合性改进。对于其所要达到的目标函数以及权重做出相应的参数方程。最终再立足发展和利益的角度对总体方案写成研究报告。

三、模型假设

限定条件，平台只会和最近的若干辆出租车进行匹配，无极端不可控因素。

规则 A：快车司机不能接专车订单

规则 B：保证司机接单后不会通过限行限号区域

规则 C：为设定实时目的地的司机过滤不顺路区域

规则 D：为只听预约单司机过滤实时订单

规则 E：同一个订单只会发给一个司机一次

四、符号说明

在每个小题所需要的地方进行罗列

五、模型的建立和求解

5.1 问题一的模型建立和求解

5.1.1 运用 DBSCAN 的构图以及排队论建立服务台模型

对于区域的乘客与出租车之间的配送关系，我们选择借鉴 DBSCAN 的算法进行构图。因为此次我们的目标函数是乘客的等待时间达到最优，所以我们将乘客的位置作为核心点，将出租车司机的位置作为非核心点进行构图。首先，我们定义几个参数：

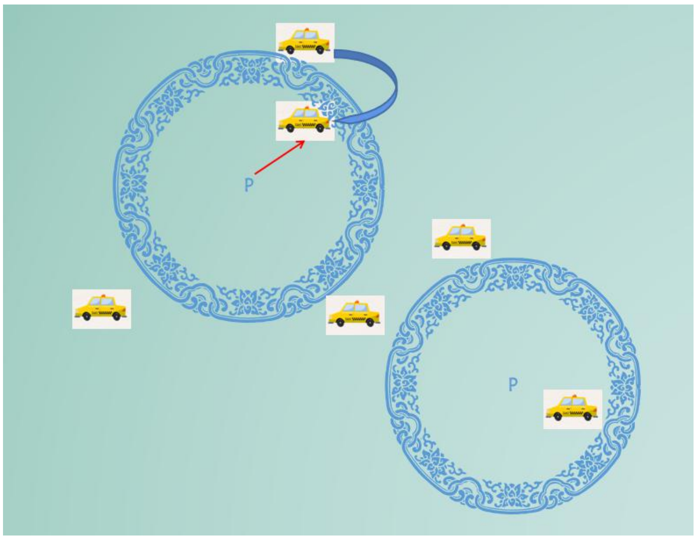
1、$\epsilon-$邻域：对于$X_j \in D$，其$\epsilon-$邻域包含样本集 D 中与X_j的距离不大于ϵ的子样本集，即 $\mathrm{N} \in (X_j) = \mathrm{X_i} \in \mathrm{D}|\mathrm{distance}(\mathrm{X_i}, \mathrm{X_j})| \leq \epsilon$，这个子样本集的个数记为$|\mathrm{N} \in (X_j)|$

2、核心对象：对于任一样本$X_j \in D$，如果其$\epsilon-$邻域对应的 $\mathrm{N} \in (X_j)$至少包含 MinPts 个样本，即如果$|\mathrm{N} \in (X_j)| \geq$MinPts，则$X_j$是核心对象。

3、密度直达：如果X_i在X_j的邻域$\epsilon-$邻域内，并且X_j是核心对象，那么称X_i可以由X_j密度直达，反之则不一定成立，除非X_i自己本身也是核心对象。

4、密度可达：多个密度直达，连接起来就是密度可达，密度可达也是不可逆的。对于X_i和X_j，如果存在样本序列$P_1, P_2, \ldots, P_T$，满足P_1=X_i，P_T=X_j，且P_{T+1}由P_T密度直达，则称X_j由X_i密度可达。此时序列中的传递样本$P_1, P_2, \ldots, P_{T-1}$均为核心对象，因为只有核心对象才能使其他样本直达。

5、密度相连：对于X_i和X_j，如果存在核心对象样本X_k，使得X_i和X_j均由X_k密度可达，则称X_i和X_j密度相连，简单来说就是可以由一个或多个核心对象连接起来。



5-1-1 接乘客示意图

因为题目中给出“用户与最近的若干辆出租车配对”，说明车的数量远远大于乘客的数量，所以，对于算法中的密度直达和密度可达都满足要求。同时我们考虑到在顾客接走后两者的位置会发生改变，以及现实中会实时产生新的乘客和司机，那么在如此大

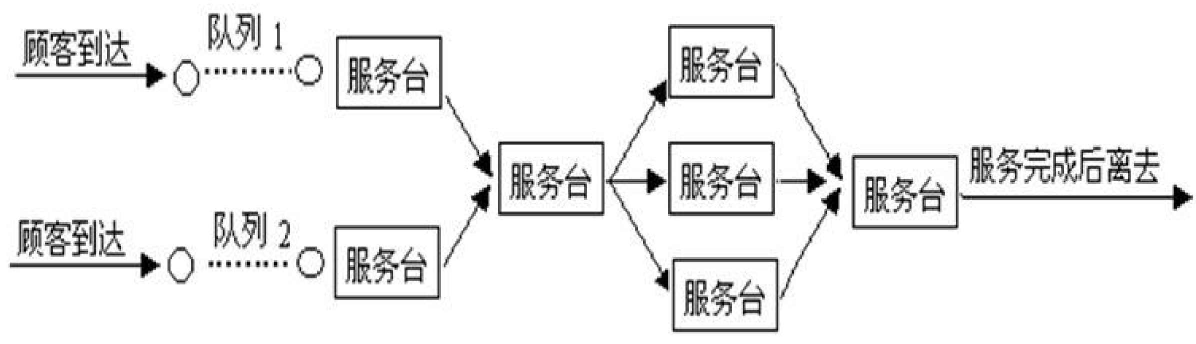
的一个服务系统中，我们可以近似看成：在旧的出租车开走后，会有一辆新的出租车顶替其原来的位置来保证“服务台”的位置大体上没有发生变化。而且又因为车的密度远大于乘客密度，所以，必定又乘客可以乘坐到出租车。

这样的情况就可以满足排队论的模型：我们将匹配选单的时间作为排队时间，将等待的时间作为服务时间，车辆离开即认为服务结束，新出现的乘客是下一个排队对象。

在实际情况中，如果遇到高峰期，会出现服务台的数量小于乘客的数量，这也就要求我们去建立一个平衡模型，让乘客的等待时间尽可能的少。因为题目中还要求可以让少数顾客的等待时间较长，所以在处理数据的时候，我们要将其形成一个平衡态。

我们先从峰期开始推算，最后通过忙期和闲期的总长度比所对应的长度比推算出概率，所以平均时间也会满足一个比值，下面就是排队论的一个总体建模过程：

符号	说明	单位
λ	平均到达率	人/分
μ	平均服务率	人/分
s	车辆个数	个
p_0	系统初始状态的概率	/
W_S	平均逗留时间	min
W_q	平均等待时间	min
L_s	平均队长	人
L_q	平均排队长	人



5-1-2 排队模型示意图

记$p_n = P\{N = n\}(n = 0,1,2, \ldots)$的乘客-车辆系统达到平衡状态后队长$N$的概率分布，注意到对个数为$s$的多服务台系统（即车辆），有

$$\lambda_n = \lambda, \quad n = 1,2, \ldots, s$$

和

$$\mu_n = \begin{cases} n\mu, \ n = 1,2, \ldots, s \\ s\mu, \ n = s, s+1, \ldots \end{cases}$$

记$\rho_s = \frac{\rho}{s} = \frac{\lambda}{s\mu}$，则当$\rho_s < 1$时，有

$$C_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!}, \ n = 1,2, \ldots, s \\ \dfrac{(\lambda/\mu)^s}{s!}\left(\dfrac{\lambda}{s\mu}\right)^n = \dfrac{(\lambda/\mu)^n}{s!\, s^{n-s}}, \ n \geq s \end{cases}$$

故

$$p_n=\begin{cases}\dfrac{\rho^n}{n!}p_0,\ n=1,2,\ldots,s\\ \dfrac{\rho^n}{s!\,s^{n-s}}p_0,\ n\geq s\end{cases}$$

其中

$$p_0=\left[\sum_{n=0}^{s-1}\frac{\rho^n}{n!}+\frac{\rho^s}{s!\,(1-\rho_s)}\right]^{-1}$$

上述两个公式给出了在平衡条件下系统中顾客数为n的概率，当$n\geq s$时，即系统中顾客数大于或等于服务台的个数，这是再来的顾客必须等待，因此记为：

$$c(s,\rho)=\sum_{n=s}^{\infty}p_n=\frac{\rho^s}{s!\,(1-\rho_s)}p_0$$

该式称为$Erlang$等待公式，它给出了顾客到达系统时需要等待的概率。对多服务台等待制排队系统，由已得到的平稳分布可得平均排队长L_q为

$$L_q=\sum_{n=s+1}^{\infty}(n-s)p_n=\frac{p_0\rho^s}{s!}\sum_{n=s}^{\infty}(n-s)\rho_s^{n-s}=\frac{p_0\rho^s}{s!}\frac{d}{d\rho_s}\left(\sum_{n=1}^{\infty}\rho_s^n\right)=\frac{p_0\rho^s\rho_s}{s!\,(1-\rho_s)^2}$$

或

$$L_q=\frac{c(s,\rho)\rho_s}{1-\rho_s}$$

记系统中正在接受服务的顾客的平均数为$\bar{s}$，显然$\bar{s}$也是正在忙的车的平均数，故

$$\bar{s}=\sum_{n=0}^{s-1}np_n+s\sum_{n=s}^{\infty}p_n=\sum_{n=0}^{s-1}\frac{n\rho^n}{n!}p_0+s\frac{\rho^s}{s!\,(1-\rho_s)}p_0$$
$$=p_0\rho\left[\sum_{n=1}^{s-1}\frac{\rho^{n-1}}{(n-1)!}+\frac{\rho^{s-1}}{(s-1)!\,(1-\rho_s)}\right]=\rho$$

该式说明，平均在忙的共享车个数不依赖于共享车的个数s，这是一个与现实符合的结论，由该式，可得到平均队长L_s为：

$$L_s=\textit{平均排队长}+\textit{正在接受服务的顾客的平均数}=L_q+\rho$$

对多服务台系统（多车体系），$Little$公式依然成立，即有

$$W_s=\frac{L_s}{\lambda},W_q=\frac{L_q}{\lambda}=W_s-\frac{1}{\mu}$$

5.1.2 参数的确定

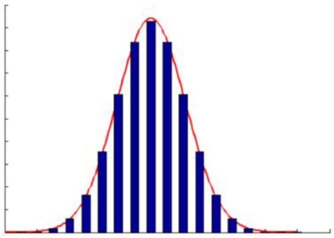
对于平均到达率和平均服务率，我们采用 LR 和 GBDT 模型，LR 模型比 GBDT 略好一些，北京市的准确率为 0.7822，AUC 是 0.8680；上海市的准确率是 0.7632，AUC 是 0.8470。

$$\begin{cases} \max\limits_{a_{ij}} E_{SR} = \dfrac{\sum_{i=1}^{N}[1-\prod_{j=1}^{M}(1-p_{ij})^{a_{ij}}]}{N}, \\ \text{s.t. } \forall j, \sum\limits_{i=1}^{N} a_{ij} \le 1, a_{ij} \in \{0, 1\}. \end{cases}$$

由数据表中我们知道总共给了 25000 个数据，根据区域密度计算，排除干扰点，我们得到在单位区域内由车辆个数 300 辆左右。

关于达到平衡状态后队伍长度为 0 的分布，我们分成峰期和谷期进行考虑。

在此，我们令峰期时队伍长度为 0 的概率为 0~0.1 的正态分布，谷期队伍长度为 0 为 0.7~0.9 的正态分布：



$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

5-1-3 概率分布函数基本关系

根据题意：接单响应时间最好控制在 2 分钟以内，候车时间最好不超过 10 分钟。

所以我们将W_s和W_q分别设为 2 分钟和 10 分钟，若后面最终得出的结果并不符合要求，通过对此数据进行一系列的调整来达到一个合理的值。

将所有的参数带入进行计算我们可以得到：

在峰期，平均的队长为 2.7 不符合要求。

在谷期，平均的队长为 0 符合要求。

通过查阅资料可得在高峰期，预估排队 15 分钟，实际排队要半小时，我们将候车时间和接单响应时间进行不断的调整，使得在高峰期的平均队长不超过 1.5。

令候车时间为 X，接单相应时间为 Y：

$$S.T.\begin{cases} X \leqslant 3, Y \leqslant 15 \\ X = \dfrac{L_s}{\lambda}, Y = \dfrac{L_q}{\lambda} \\ L_s = L_q + \rho \end{cases}$$

最终得到近似解，候车时间为 3 分钟，预估排队时间为 14 分钟。

在峰期和谷期的状态下，忙期和闲期的总长度比为ρ：$(1-\rho)$，又因为忙期和闲期时交替出现的，所以两者的长度和概率比也为ρ：$(1-\rho)$，即：

$$\frac{\bar{\beta}}{\bar{I}} = \frac{\rho}{1-\rho}$$

又因为在到达为 Poisson 流时，根据府直属分布的无记忆性和到达与服务相互独立的假设，容易成名从系统空间时刻起到下一个顾客到达时刻止的时间间隔要仍然服从参数为λ的负指数分布，使求得平均忙期为：

$$\overline{\beta}=\frac{\rho}{1-\rho}\times\frac{1}{\lambda}=\frac{1}{\mu-\lambda}$$

通过峰期的等待时间我们可以得到（不到 1 近似为 1），谷期的候车时间为 1 分钟，预估排队时间为 2 分钟。

可以观察到，谷期的平均时间还是符合要求的，如何解决峰期问题是关键性的。

所以我们最终的策略就是：

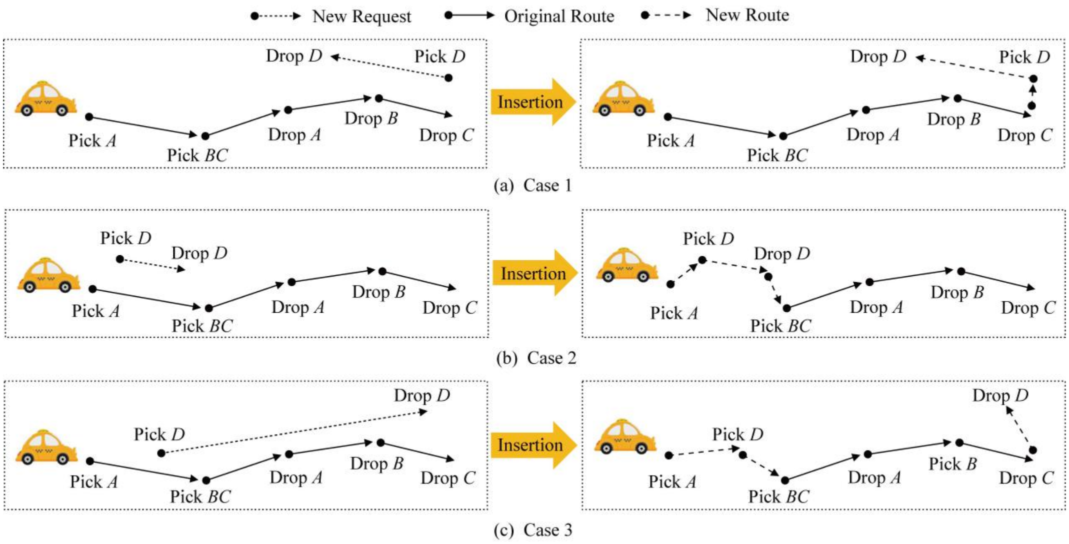
谷期的候车时间为 1 分钟，预估排队时间为 2 分钟

峰期的候车时间为 3 分钟，预估排队时间为 14 分钟

5.2 问题二的模型建立和求解

5.2.1 拼车模型的简介及其优化模型

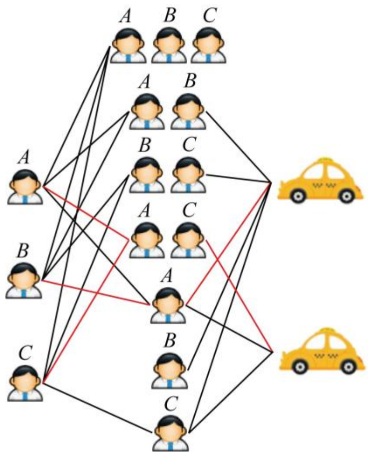
首先用简单的示意图来介绍拼车的具体方式：



5-2-1 拼车示意图

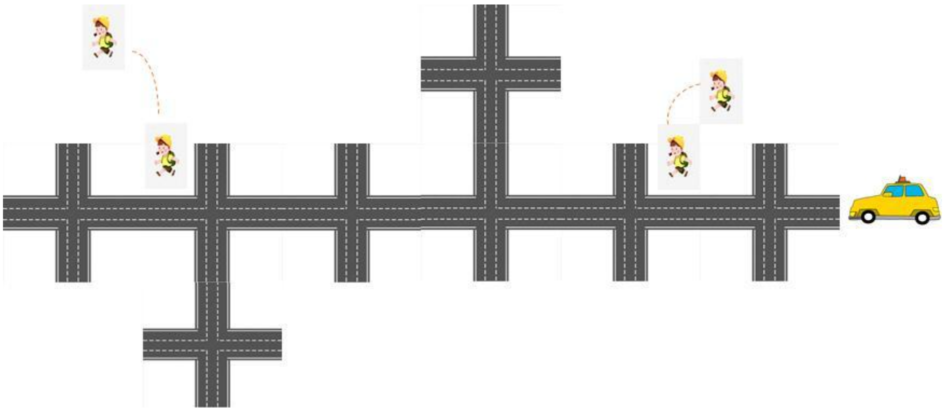
首先给定司机的当前行驶路线和 1 个新发布的订单,而插队操作旨在将这个新订单的起点和终点分别插入到原有路线中,从而获得新的路线并使得司机的行驶距离尽可能短. 对于这对起点与终点插入位置,插队操作共有如图 2 所示的 3 种情况:1、新订单的起点和终点相邻地添加至当前路线的末尾; 2 、新订单的起点和终点相邻地插入至当前路线的开始位置或中间位置; 3 、新订单的起点和终点不相邻地插入到当前路线的中间位置

因为最终的目标是要让乘客获得最佳的体验,我们就基于乘客的角度进行路线规划:乘客往往需求尽量少绕路并且最短时间内到达终点（甚至对于司机和另一名乘客有要求）



5-2-2 顾客选择示意图

1、最小化的乘客等待时间



5-2-3 拼车的乘客上车点示意图

为了让乘客的等待时间尽可能的短，在确定两人的位置后，我们要保证的是最先上车的乘客到上车点的位置要近，而第二位乘客的位置可以稍微长一些，以减少总的车辆的绕行路程，使得总体的时间更短。

这边我们先考虑两者都在城市的情况（较为复杂）

令乘客 1 此时的坐标点为(x_{11},y_{11})，离他最近的十字路口为(x_{12},y_{12})

令乘客 2 此时的坐标点为(x_{21},y_{21})，离他最近的十字路口为(x_{22},y_{22})

用Δ表示乘客到等待点的时间，用∇表示司机接到乘客的时间（必要的时候考虑掉头，在此处简化模型，不进行赘述）

$$\text{乘客一的等待时间}t_1\text{：}\nabla\frac{3.6D_1}{25}-\Delta|(x_{11},y_{11}),(x_{12},y_{12})|$$

$$\text{乘客二的等待时间}t_2\text{：}\nabla\frac{3.6D_2}{25}-\Delta|(x_{21},y_{21}),(x_{22},y_{22})|$$

考虑在郊区的情况，不需要考虑掉头，找位置的情况，在路边随时停车、上车即可

此时得到两人的平均等待时间 T 为：

$$2T=a\times t_1+(1-a)\times t_2$$

对于第一个上车人和第二位上车人的权重不同，在下面进行具体的分析。

2、最佳乘客体验

为提高乘客在大规模拼车中的出行体验，考虑乘客与乘客，司机与乘客之间的关系。此处，我们通过定义社会效用模型来表示乘客间是否适宜共享形成，是否具有相同偏好等社交关系。然后，再以此作为标准之一去进行路线规划。具体如下，乘客 r 和司机 c 之间的社会效用函数：

$$\mu(r,c)=\alpha\mu_v(r,c)+\beta\times\mu_r(r,c)+\gamma\times\mu_t(r,c)$$

其中，函数μ_v表示乘客 r 和司机 c 的社交因素指标，函数μ_r表示乘客 r 和司机 c 所搭载的其他乘客间的社交因素指标。

最终我们达到的目标就是要在平均等待时间短的情况下使乘客的体验更好。

5.2.2 综合评估拼车的价格，等待时间和乘客的社会体验

根据实践经验，我们分别对于这三个指标指派一个隶属度函数，用模糊综合评价对曲线进行评分。以下就是三个隶属度函数：

$$A_1=\begin{cases}1, x\le a\\ \dfrac{1}{1+\alpha(x-a)^{\beta}}, x>\mathrm{a}\end{cases}$$

5-2-2-1 柯西分布的偏小型函数

$$A_2=\begin{cases}1, x<a\\ \dfrac{b-x}{b-a}, a\le x\le b\\ 0, x>b\end{cases}$$

5-2-2-2 梯形分布的偏小型函数

$$A_3=\begin{cases}0, x\le a\\ 1-e^{-\left(\frac{x-a}{\sigma}\right)^2}, x>a\end{cases}$$

5-2-2-3 正态分布的偏大型函数

然后我们需要得到的是乘客所支付的平均价格，等待的平均时间，以及对社会体验的满意程度，按照以上三个隶属函数的模型进行打分，并赋予 4:4:2 的权重比，进而判断总体体验。

令最后的得分为$Points$，三个隶属度函数的值分别为$Price, Time, Feel$，计算最后的得分公式为:

$$Points=\frac{0.4Price+0.4Time+0.2Feel}{3}$$

时间段	日均载客收费	日均等待时间	社会体验的满意程度
0:00——1:00	17.08	2	46
1:00——2:00	15.86	5	82
2:00——3:00	12.84	6	74
3:00——4:00	11.12	8	62

4:00——5:00	14.48	7	74
5:00——6:00	19.18	4	92
6:00——7:00	21.36	1	60
7:00——8:00	21.72	2	50
8:00——9:00	22.58	8	40
9:00——10:00	21.56	6	50
10:00——11:00	21.82	2	84
11:00——12:00	21.74	6	98
12:00——13:00	20.34	3	75
13:00——14:00	21.26	15	36
14:00——15:00	22.62	2	91
15:00——16:00	22.22	8	72
16:00——17:00	21.90	6	68
17:00——18:00	21.76	2	62
18:00——19:00	20.42	5	67
19:00——20:00	22.66	2	83
20:00——21:00	22.58	10	40
21:00——22:00	21.08	12	50
22:00——23:00	18.58	4	43
23:00——24:00	16.72	8	82

5-2-2-4 北京市一天不同时间段的基本数据

计算得平均值$x_1 = 19.72833, x_2 = 5.583333, x_3 = 65.875$
代入隶属函数可得$Price = 0.946, Time = 0.631, Feel = 0.094$
所以得到：

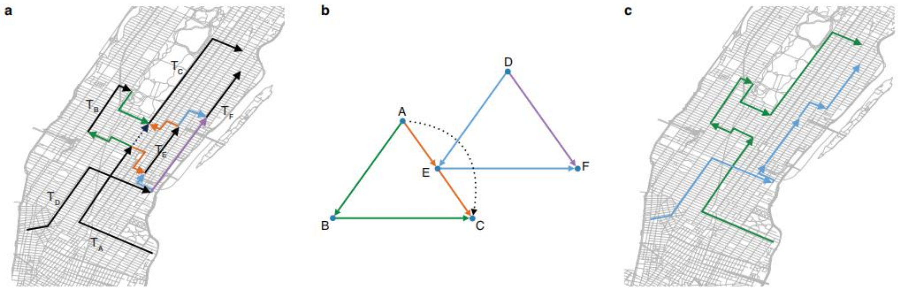
$$Points = \frac{0.4Price + 0.4Time + 0.2Feel}{3} \times 100 = 64.96$$

总体来说，如果以价格作为最主要的评价标准，那么拼车再节约时间的同时节省了费用，能让乘客获得更好的体验。至于社会体验的满意程度，这个并不作为决定性因素，对于总体的评价不产生巨大的影响。

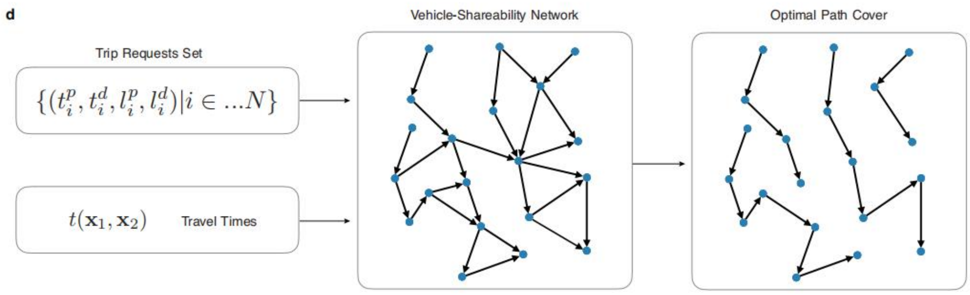
5.3 问题三的模型建立和求解

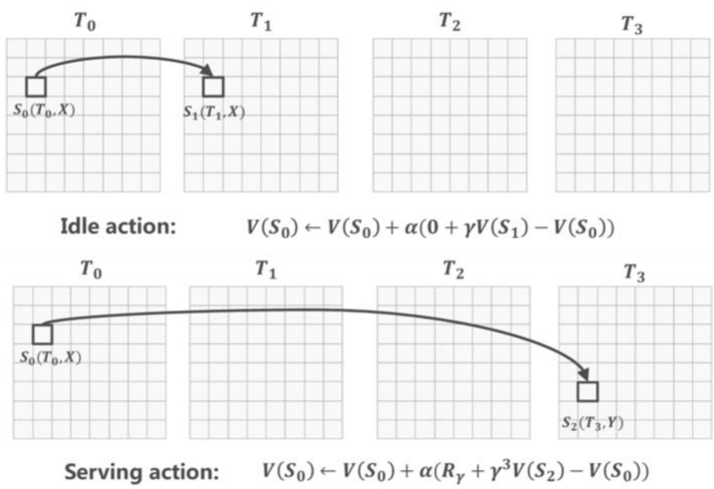
5.3.1 对于平台动态人车配对策略的价值体系构建

第三题中为平台指定动态策略，而平台一般在意的是订单完成的数量，从而达到的平台的总盈利额。



再例如上图中的城市路线规划的时候我们发现，路线虽然说可长可短，但是总会经过一定的区域转换，而在无数种不同的选择当中，最终我们要根据一定的标准将其达到最优目标：

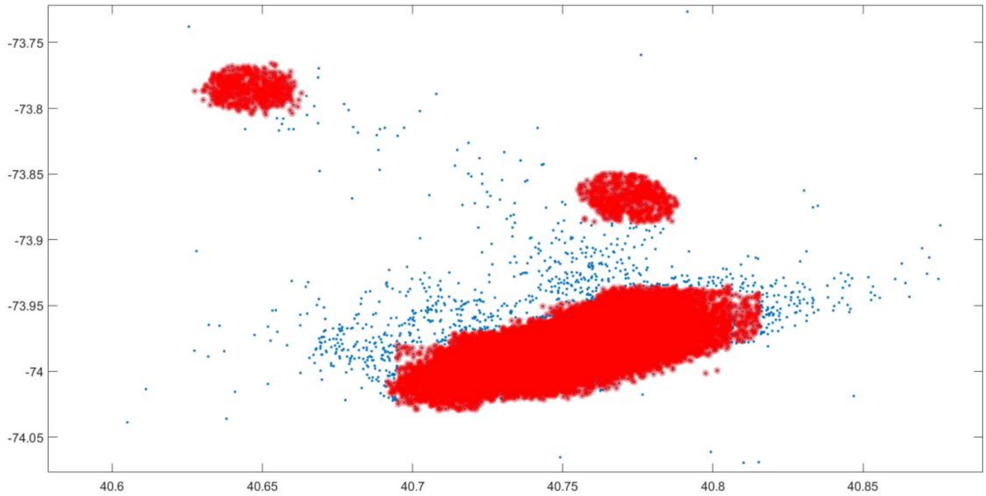




5-1-1 时空网格价值图

在不同的区域，其乘客的数量也必定不一样，那么获得利润的可能性也必然不一样。而在接送乘客的过程中，必然会出现区域转移的现象，所以要先对每块区域进行估值。最终我们要实现的目标是，尽可能多的车辆留在价值高的区域，对于偏离高价值点的运输，进行特殊处理。

首先，我们将出租车的位置在平面直角坐标系中进行刻画，如图红色的区域为出租车流量密集点，说明这些区域的价值较大。



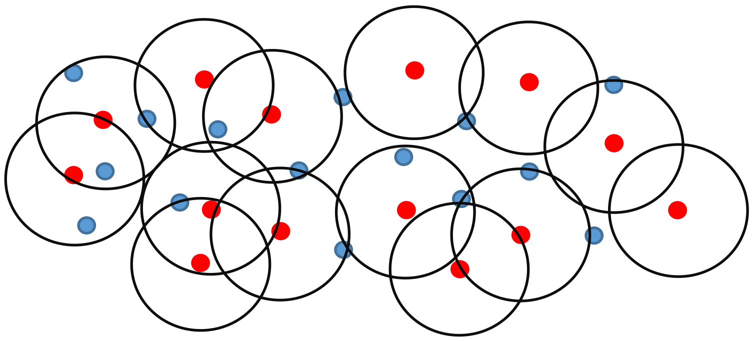
5-1-2 出租车位置散点图

我们将任意时刻任意空间位置离散化为时空网格，根据派单记录（含参加调度但无单的司机）计算该时空网格到当天结束时刻的预期收入。

动态规划思路：将总的坐标系刻画为网格。假设总共有时刻区间为[0, T)；先计算 T-1 时刻的所有网格的预期收入（此时未来收入为 0，只有当前收入），其本质就是计算当前收入的均值；然后计算 T-2 时刻的所有网格的预期收入；...；以此类推，这样的话，就可以计算出每个时空网格到当天结束时刻的预期收入。

其次，题目要求将乘客与最近的出租车进行配对，简图如下所示：

红色的小点为出租车，蓝色的小点为乘客，黑色的圆圈为 DFS 半径区域，也就是信号感应区域。



5-1-3 车辆匹配简图

根据以上的几点要求，我们提出以下的匹配标准，并使用以下公式描述订单和司机之间的匹配度：

$$A_\pi(i,j) = \gamma^{\Delta t_j} V(s'_{ij}) - V(s_i) + R_\gamma(j)$$

1、价格越高，匹配度越高

2、当前位置价值越大，匹配度越低

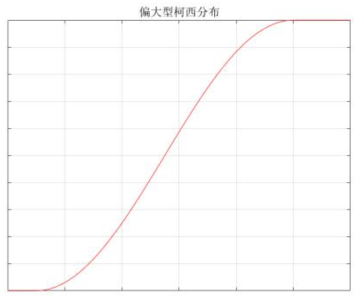
3、未来位置价值越大，匹配度越高

4、接驾里程，隐形表达，越大则预计送达时间越大，衰减系数越小，匹配度越低

接下来，通过模糊综合评价的模型来确定四个指标的隶属度函数。

1、价格与匹配度之间的关系：偏大型的柯西分布

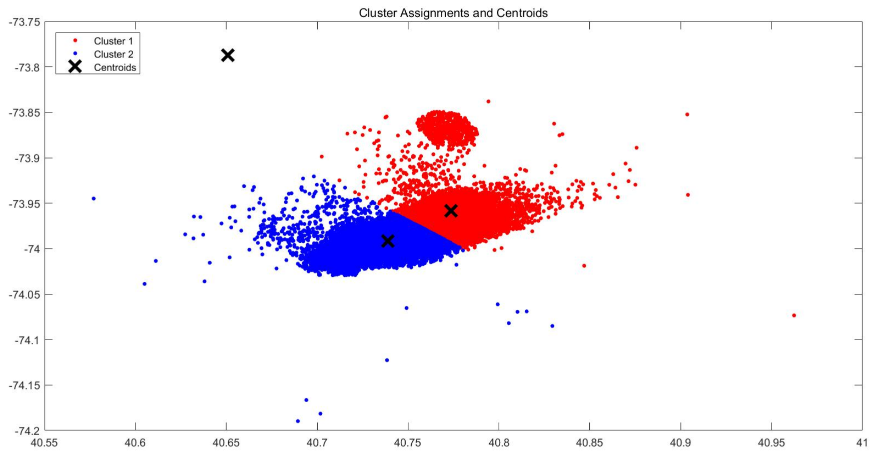
$$\underset{\sim}{A}(x)=\begin{cases}0, & x\leqslant a\\ \dfrac{1}{1+\alpha(x-\alpha)-\beta}, & x>a\end{cases}\quad(\alpha,\beta>0)$$



2、当前位置价值-未来位置价值与匹配度之间的关系以及跨区域成本

对于刚才地图中的点，我们先进行三层的聚类分析，得到下图：

因为三丛大密度点中，右上方与面积最大的一块区域之间的距离更近，所以要对其进行单独的分析：

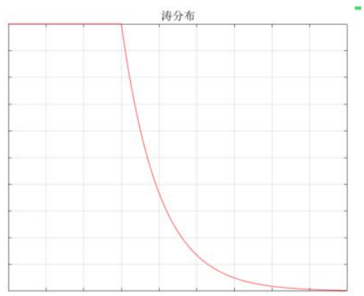


5-1-4 跨区域分界线

对于三个区域，除上图中蓝色和红色的区域外，我们将分界线的边缘 4%作为分解标准，若在接单的过程中发生了跨区域，便要进行减分策略。

对于远离中心的距离，我们设置惩罚函数：

$$A_1(x)=\begin{cases}100 & x\leqslant a\\ 100\exp[-k(x-a)] & x\geqslant a(k>0)\end{cases}$$



3、预计时间与奖励机制

我们对于乘客等待的前两分钟固定为并要等待时间，从两分钟开始后，随着时间的不断增长，惩罚的数值也就越大。这里我们采用梯度降分法：

$$\underset{\sim}{P(x)}\begin{cases}5 & T\leqslant 2\\ 4 & 2<T\leqslant 4\\ 3 & 4<T\leqslant 6\\ 2 & 6<T\leqslant 8\\ 1 & 8<T\leqslant 10\end{cases}$$

5.3.2 不允许拼车下的策略展示

一般情况下，就像目前市面上所有的打车软件都采用最近距离的匹配方式。但是再显示场景中，不但要考虑有实时的乘客和司机的产生，还有一些价值低，位置偏离中心区域的坏点产生，此时，我们要借鉴“滴米”的做法，去实现乘客匹配的全图性，防止产生大量的司机拒载。对于司机来说，行驶里程多、道路状况好的优质单会扣除滴米，而行驶里程较少、道路状况拥堵的“劣质单”的司机则会奖励滴米。我们最终将滴米算入综合的评价系统中。

5.3.3 允许拼车的策略调整

对于拼车，目前存在几个不合理的地方：

1、不能让乘客找到最近的上车点，让乘客提前到达候车点

2、城市路况较为发杂，存在驾驶不能左转，不能掉头，所以不能让乘客在最近的中途点上车

3、如果单纯的按照最近线路让乘客在中途点上车，往往会忽略乘车时间，费用，司机性别等更多的因素。

在此，我们将乘车时间 T，乘车费用 M，车辆最大容量 V 等几个重要因素考虑进来。我们建立以下的拼车具体步骤流程：

符号	说明
D	用户路线到司机发布路线的最近距离
[Tsu, Teu]	用户要求的时间区间
[Tsd, Ted]	司机所发布的时间区间
Mu	用户可接受的最高费用
Md	司机发布的此路线的拼车费用

1、计算距离匹配程度 ED

$$ED=\frac{1}{\sqrt{2\pi}}e^{-\frac{\left(\frac{D}{2}\right)^2}{2}}$$

2、计算时间区间的匹配程度 ET，设 $\mathrm{L}=\min(\mathrm{Teu}-\mathrm{Tsu},\mathrm{Ted}-\mathrm{Tsd})$，设[Tsu, Teu]

和[Tsd, Ted]的交集时间长度为L_1

$$ET=\frac{1}{\sqrt{2\pi}}e^{-\frac{\left(\frac{L-L_1}{L}\right)^2}{2}}$$

3、计算费用的匹配程度 EM，令 $\mathrm{M}=\mathrm{Mu}-\mathrm{Md}$，如果拼车的类型为长途拼车且 $\mathrm{M}>50$，或者拼车类型为市内拼车且 $\mathrm{M}>\mathrm{Ms}$，那么 Ms 可以按照城市大小以及物价情况在适当的范围内自行调整，当拼车类型为长途拼车且 $\mathrm{M}\leq 50$

$$ET=\frac{1}{\sqrt{2\pi}}e^{-\frac{\left(\frac{M}{50}\right)^2}{2}}$$

如果拼车类型为市内拼车且 $\mathrm{M}\leq\mathrm{Ms}$

$$ET=\frac{1}{\sqrt{2\pi}}e^{-\frac{\left(\frac{M}{Ms}\right)^2}{2}}$$

4、最终用户路线与司机路线的匹配程度我们用下面的参数方程进行描述，再将具体的数据进行分析后得到每个变量的具体值，最终发放给乘客信息的高低排名就以此作为标准。

$$E=a\times ED+b\times ET+c\times EM$$

5.3.4 具体策略的指定与成型

$$\begin{cases} arg_z=\underset{\sim}{A}(x_1)-A_1(x_2)+P(x_3) \\ \varpi=\begin{cases} -\left(\dfrac{b-x}{b-a}\right)^k & a<x<b \\ 0 & b\leqslant x\leqslant c \\ \left(\dfrac{x-c}{d-c}\right)^k & c<x<d \end{cases} \\ E=0.4\times\dfrac{1}{\sqrt{2\pi}}e^{-\frac{\left(\frac{D}{2}\right)^2}{2}}+0.4\times\dfrac{1}{\sqrt{2\pi}}e^{-\frac{\left(\frac{L-L1}{L}\right)^2}{2}}+0.2\times\dfrac{1}{\sqrt{2\pi}}e^{-\frac{\left(\frac{M}{50}\right)^2}{2}} \end{cases}$$

联立进行求解，其中：

1、x_1与 M 可以视作近似相等

2、x_2与 D 之间存在概率关系，我们用正态分布进行刻画

3、x_2与滴米之间存在关系，联系区域图进行分析

4、x_3与 L 呈正比关系

5、保证空载率要尽可能的低

对出租车资源配置的评价为多因素的，本文主要分级考虑时间、空间及乘客满意度、司机满意度的综合评价。对此，我们设计一二级指标并进行求解：

$$U=\{u_1,u_2\}=\{\text{时间分布，空间分布}\}$$

$$U_1=\{u_{11},u_{21}\}=\{\text{乘客满意度，司机满意度}\}$$

$$Q=\{4,3,2,1\}=\{\text{合理，较合理，较不合理，不合理}\}$$

$$E=\begin{pmatrix}E_1\\E_2\\\cdots\\E_j\end{pmatrix},E_i=[e_{ij}]_{k*n}$$

$$B_m=F\odot\begin{pmatrix}F_1\odot E_1\\\cdots\\F_2\odot E_j\end{pmatrix}$$

$$\mu_m^*=\frac{\sum_{i=1}^{n}g(q_i)\cdot s_i^k}{\sum_{i=1}^{n}s_i^k}(m=1,2,3,4)$$

$$E_{11}=(0.1\ 0.3\ 0.4\ 0.2)\,E_{12}=(0.3\ 0.3\ 0.3\ 0.1)$$

$$E_{21}=(0.1\ 0.2\ 0.4\ 0.3)\,E_{22}=(0.1\ 0.1\ 0.5\ 0.3)$$

$$\mu_m^*=\frac{\sum_{i=1}^{n}g(q_i)\cdot s_i^k}{\sum_{i=1}^{n}s_i^k}=\frac{4\times0.156^2+3\times0.198^2+2\times0.414^2+1\times0.232^2}{0.156^2+0.198^2+0.414^2+0.232^2}=2.11792$$

我们发现在谷期，空载率相当的高，此时就要不断升高共享车的利用率。

首先提升滴米在各项指标中的比重，或者如果在空载率达到 50%的时候，对于顾客的请求必须安排一名司机去解决并给予相应的补偿措施。

动态人车的配对策略总体如下：

1、如果不拼车，在峰期中对于新出现的乘车需求点中的坏点，需平台在一定时间后强制派出车辆进行匹配，并给予相应的补偿，以解决严重的空载现象。

2、拼车的费用要随着运送距离不断的进行调整，峰期时的拼车不是那么认可，并且会酌情的升高费用以达到收入的稳定态。

3、在峰期时，尽量不选择跨区域的接单，对于预计时间过长的单由专门的指派车来完成（不计入滴米值）。

4、在打车软件的背景下，不考虑目标区域待乘乘客量，也不需考虑驾驶员个人经验值，按照具体情况具体执行，给乘客最好的体验。

5、对于空间时间中随机产生的乘客和司机，采用第一问的结论去进行匹配，以免出现利用率不高的情况。为后期可能会产生的峰期做准备。

6、在向用户显示多条匹配信息时，需要将每条匹配的发布路线和信息从高到底排列，具体得分按照上面的计算公式。

5.4 问题四的模型建立和求解

5.4.1 用户满意度模型

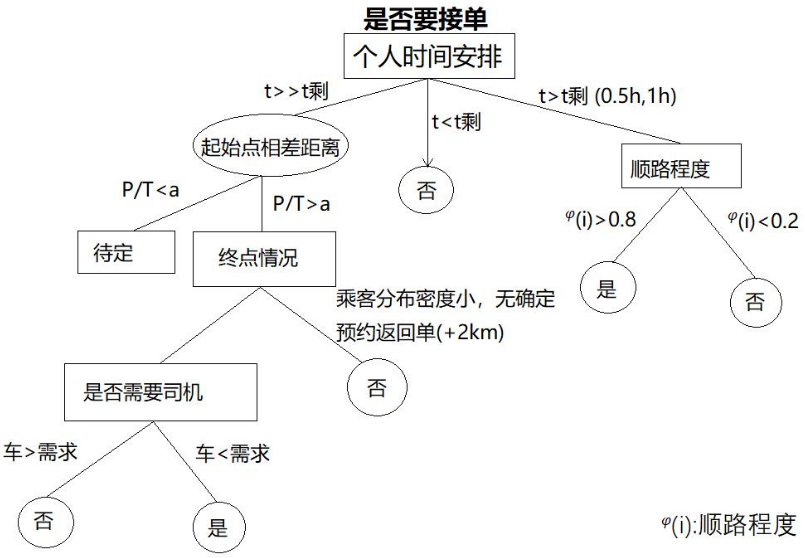
对于用户满意度模型，模型的目标层主要考虑价格最低以及等候时间最短的最优情

况。我们假设用户满意度为$\omega(i)$,价格变量为p,时间变量为t则$\omega(i) = p * a1 + t * a2$。由于在非极端情况下，汽车的平均速度是远大于人步行的速度，因此我们引入“推荐上车点”这一概念，即乘客步行至“推荐上车点”与司机会和。“推荐上车点”的引入能够使整个打车的过程减少与司机沟通上车点的时间，提升乘客与司机双方的旅途幸福指数。对于“推荐上车点”的坐标确定，主要采用聚类分析的策略。推荐上车点为乘客下单实际GPS精确坐标附近的历史乘客上车点的聚类结果结合POI点（非地理意义的有意义点，例如：医院，停车场）所得出的结果坐标。同时，推荐上车点的引入能有效影响价格的数值。对于“绕路”距离综合收益，可以化简为：

$$综合收益 = 乘客减少的路程 * 权重_1 - 司机接驾和去目的地增加的路程 * 权重_2。$$

5.4.2 司机接单综合改进模型

对于网约车司机，往往系统并不会显示接驾目的地的乘客情况；下班时间的接近以及该订单距离是否“值得”接单往往会成为司机接单时所烦恼的事务。而在实际情况中，对于驾驶中的司机，即使他们有“能力”通过地图和社群了解目的地状况等消息，也是极为危险的。因此，在司机接单综合改进模型中，我们通过Boosting Decision Tree（提升决策树）这一算法创建一个智能化接单推荐系统，使平台对于司机更为友好。决策树的价值在于，与人类做决策不同，人的决策往往具有一定的随意性，所做出的决策未必是最优的；而决策树模型则从数学理论上一定程度保证了模型结果的最优性。通过对特征变量的选择，我们首先创建分类决策树，如下：



5-4-1 分类决策树

通过CART分类与回归树经典算法进行决策，得到结果，不但能有效地提高司机的收益水平，还能提升司机的工作体验幸福指数。

5.4.3 平台宏观收益模型

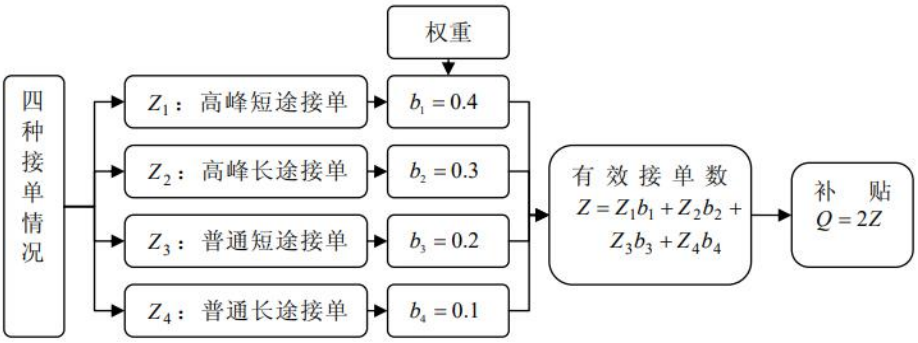
用户满意度的提高以及司机幸福度的提高，从宏观上来说与平台的收益呈正相关，因此，在该模型中，我们主要从企业收益的宏观目标函数以及高峰期价格变换两个方面进行分析。

$$企业收益的宏观目标函数 = 乘客(司机)等待时间 * 权重_1 + 订单完成率 * 权重_2$$

因此，积极的订单完成率在宏观角度来看极为重要；所以我们引入动态的平台价格，即“高峰期”价格以及非“高峰期”价格。因为司机的数量是有限的，因此在订单密度较大的情况下， 将价格提高可以有效地减少一些“坏单”的出现，从而增加总体的收益。所谓的“坏单”与“滴米”系统中的“劣质单”较为相近，即行驶里程较少、道路状况拥堵的订单。

5.4.4 社会福利模型

从各大主流论坛以及搜索引擎所显示的数据分析，无论是司机群体还是乘客群体对于“滴滴”等网约车软件持不满意的态度。平台抽成高，司机实际收益低，叫不到车，“低碳出行”等问题尤为严重，因此一个成熟可靠的补贴系统显得非常重要。对于司机收益低，为提高司机出车率以及接单率，我们针对高峰短途接单数，高峰长途接单数，普通短途接单数，普通长途接单数四种出租车接单数，由于网约车司机主要以获取更高的收入为目的，因此对这四种接单数设置不同的权重，根据出租车司机每天的四种不同接单数，计算出出租车司机一天的有效接单数， 按每单给网约车司机补贴 2 元的补贴政策，计算得到出租车司机一天的补贴金额然后直 接打入出租车账户，从而实现出租车收入提高，进一步缓解乘客出租车打车难的问题。



5-4-2 出租车司机接单补贴政策

对于“打不到车”这一现象，主要是由于高峰时段网约车供不应求，空载率较高所导致，对于高峰时段网约车车数不够的情况，应考虑施行拼车优惠策略。由于车流量比较大，部分上班地点具有聚合性，需要尽量发挥已载有乘客的网约车剩余载客资源，让网约车尽量载满乘客，提高载客率。拼车政策可以用积分的形式给车上原有乘客实施奖励，不仅对原有乘客起到激励作用，还可以给后来乘客免费乘车的机会，能够很大程度

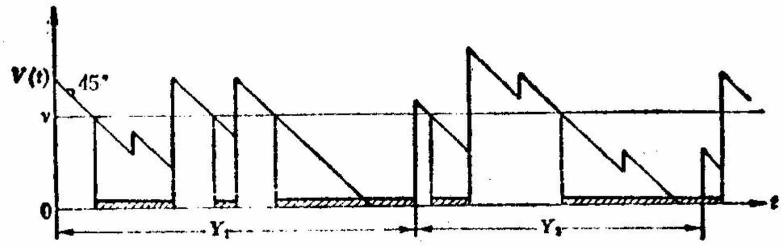
上调动人们拼车的积极性。

“低碳出行”是应该时刻思考的因素，对于一些科技园区如“张江科技园区”，“阿里软件园区”，网约车公司可以用“共享电动巴士”进行上班族的接送，地铁→公司的模式目前已经在部分地区开始实行。同时，我们设立“推荐下车点”，不但能节省乘客所需付款的费用，还能节约能源，减少二氧化碳排放。

六、模型的分析与检验

由于收到峰期的影响，乘客的数量与出租车的数量会收到很大的影响，此处检验在峰期和谷期中顾客的平均等待时间是否发生变化，以此来检验模型的稳定性。

对此，我们将其与转移概率的 Morkov 排队论进行比较分析。



我们发现，即使转移概率，也只是对某几个小段的服务速度产生了较小的影响。总体上还是呈现一个直线下降，又重新开始新的接单服务的过程，说明模型的稳定性极佳。

七、模型的推广与评价

7.1 模型的优点

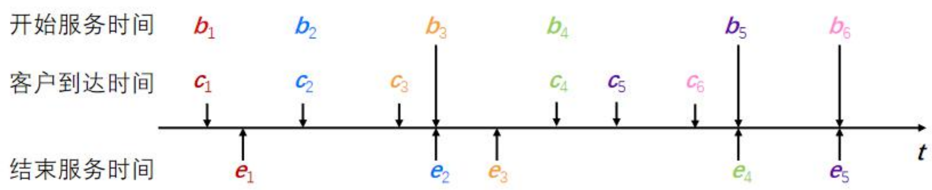
多因素评价模型从不同方面分析出租车资源的供求匹配程度，减少了单一评价标准造成的误差，为问题的分析提供了可靠的依据。

模型考虑的情况多样化，不仅从峰期谷期进行不同的策略制定。同时，根据多方面的指标对于乘客和司机都有一定的体验提升，更加的人性化。

7.2 模型的缺点

7.2.1 排队论模型的延迟弊端

排队论中，对于单一的一条服务链进行分析，假设乘客到来的时间间隔服从参数 0.1 的指数分布，服务的时间服从均值为 K，方差为 4 的正态分布，我们进行蒙特卡罗模拟会发现（此处我们令 K=10）：



100 个工作点中，滴滴每日平均服务的客户人数为: 42.78

100 个工作点中，滴滴每日客户的平均等待时间为: 30.3421

时间已过 11.584277 秒。

我们不难发现，随着时间的推移，乘客的平均等待时间会不断增长，这也是目前高峰期的时候共享车数量不足导致的等待时间长达半小时的真实写照。当一个到达发生时，虚延迟立即增加，增加量与新来的顾客的服务时间等长。

7.2.2 时空网络区域的价值转变性

在峰期和谷期，随着乘客的数量和共享车的不匹配性越来越明显，原先的离线时空网络区域的价值性会随之改变。所以对于结果可能会有写影响。

7.3 模型的推广

排队论既被广泛的应用于服务排队中，也可以广泛的应用于交通物流领域。在服务的排队中到达的时间和服务的时间都存在模糊性，然而有基于扩展原理又对模糊排队进行了一定的分析。然而在交通领域，可以非常好的模拟一些交通、货运、物流等现象。对于一个货运站建立排队模型，要想研究货物的一个到达形成的是一个复合泊松过程，每辆货车的数量为 W，而且不允许货物的超载，也不允许不满载就发车，必须刚刚好，这个还是一个具有一般分布装车时间的一个基本的物流模型。

八、参考文献

[1] 李新盛，谭雷.一种基于网络地图 API 的拼车系统路线发布.四川，2015.09.09：8-9.
[2] 徐毅，董咏昕，李未.大规模拼车算法研究进展.北京航空航天大学，背景，100191
[3] M.M. Vazifeh,P. Santi.Addressing the minimum flfleet promlem in on-demand urban mobility.Istituto di Informatica e Telematica del CNR,via G. Moruzzi 1,56124 Posa,Italy.
[4] 王潇潇.生灭过程的 M/M/S 等待制排队模型，多服务台模型.2019.05.02
[5] 星速云.滴滴派单算法_从算法模型思路到评估方案-详解.2019.09.28

九、附录

9.1 代码 1 排队论的变量推导

在排队论中，如果$N(t)$表示 时刻t系统中的顾客数，则$\{N(t), t \geq 0\}$就构成了一个随机过程。如果用“生”表示顾 客的到达，“灭”表示顾客的离去，则对许多排队过程来说，$\{N(t), t \geq 0\}$就是一类特殊的随机过程－生灭过程。

下面结合排队论的术语给出生灭过程的定义：

（1） 假设$N(t) = n$, 则从时刻t起到下一个顾客到达时刻止的时间服从参数为λ_n的负指数分布，$n = 0,1,2, \ldots$。

（2） 假设$N(t) = n$, 则从时刻t起到下一个顾客离去时刻止的时间服从参数为μ_n的负指数分布，$n = 0,1,2, \ldots$。

（3） 同一时刻只有一个顾客到达或离去

则称$\{N(t), t \geq 0\}$为一个生灭过程。

一般来说，得到$N(t)$的分布$p_n(t) = P\{N(t) = n\} (n = 0,1,2, \ldots)$是比较困难的，因此通常是求当系统达到平衡后的状态分布，记为$p_n, n = 0,1,2, \ldots$。

为求平稳状态的分布，考虑系统可能处的任一状态n。假设记录了一段时间内系统进入状态n和离开状态n的次数，则因为“进入”和“离开”是交替发生的，所以这两个数要么相等，要么相差为 1。但就这两种事件的平均发生率来说，可以认为是相等的。即当系统运行相当时间而达到平衡状态后，对任一状态n来说，单位时间内进入该状态的平均次数和单位时间内离开该状态的平均次数应该相等，这就是系统在统计平衡下的“流入=流出”原理。根据这一原理，可得到任一状态下的平衡方程如下：

0　　$\mu_1 p_1 = \lambda_0 p_0$

1　　$\lambda_0 p_0 + \mu_2 p_2 = (\lambda_1 + \mu_1) p_1$

2　　$\lambda_1 p_1 + \mu_3 p_3 = (\lambda_2 + \mu_2) p_2$

$\cdots$

n　　$\lambda_{n-1} p_{n-1} + \mu_{n+1} p_{n+1} = (\lambda_n + \mu_n) p_n$

$\cdots$

由上述平衡方程，可求得

0　　$p_1 = \frac{\lambda_0}{\mu_1} p_0$

1 $$p_2=\frac{\lambda_1}{\mu_2}p_1+\frac{1}{\mu_2}(\mu_1p_1-\lambda_0p_0)=\frac{\lambda_1}{\mu_2}p_1=\frac{\lambda_1\lambda_0}{\mu_2\mu_1}p_0$$

2 $$p_3=\frac{\lambda_2}{\mu_3}p_2+\frac{1}{\mu_3}(\mu_2p_2-\lambda_1p_1)=\frac{\lambda_2}{\mu_3}p_2=\frac{\lambda_2\lambda_1\lambda_0}{\mu_3\mu_2\mu_1}p_0$$

$\cdots$

n $$p_{n+1}=\frac{\lambda_n}{\mu_{n+1}}p_n+\frac{1}{\mu_{n+1}}(\mu_np_n-\lambda_{n-1}p_{n-1})=\frac{\lambda_n}{\mu_{n+1}}p_n=\frac{\lambda_n\lambda_{n-1}\cdots\lambda_0}{\mu_{n+1}\mu_n\cdots\mu_1}p_0$$

$\cdots$

记

$$C_n=\frac{\lambda_n\lambda_{n-1}\cdots\lambda_0}{\mu_{n+1}\mu_n\cdots\mu_1},n=1,2,\dots$$

则平稳状态的分布为

$$p_n=C_np_0,n=1,2,\dots$$

由概率分布的要求

$$\sum_{n=0}^{\infty}p_n=1$$

有

$$\left[1+\sum_{n=1}^{\infty}C_n\right]p_0=1$$

于是

$$p_0=\frac{1}{1+\sum_{n=1}^{\infty}C_n}$$

通过以上的三个式子可以推得正文中的C_n

9.2 代码 2 蒙特卡洛模拟模拟排队论的弊端

```
clear
tic   %计算 tic 和 toc 中间部分的代码的运行时间
day = 100;   % 假设模拟 100 点
n = zeros(day,1); % 初始化用来保存每点接待客户数结果的矩阵
t = zeros(day,1); % 初始化用来保存每点客户平均等待时长的矩阵
for k = 1:day
    i = 1;   % i 表示第 i 个客户，最开始取 i=1
    w = 0;   % w 用来表示所有客户等待的总时间，初始化为 0
    e0 = 0;   c0 = 0;     % 初始化 e0 和 c0 为 0
    x(1) = exprnd(10);   % 第 0 个客户(假想的)和第 1 个客户到达的时间间隔
    c(1) = c0 + x(1);   % 第 1 个客户到达的时间
    b(1) = c(1); % 第 1 个客户的开始服务的时间
    while b(i) <= 480   % 开始设置循环，只要第 i 个顾客开始服务的时间(时刻)小于
480，就可以对其服务（车辆每天工作 8 个点，折换为分钟就是 480 个分点）
        y(i) = normrnd(10,2); % 第 i 个客户的服务持续时间，服从均值为 10 方差为 4(标
准差为 2)的正态分布
        if y(i) < 1   % 根据题目的意思：若服务持续时间不足一分点，则按照一分点计
算
```

```
            y(i) = 1;
        end
        e(i) = b(i) + y(i); % 第 i 个客户结束服务的时间 = 第 i 个客户开始服务的时间 + 第 i 个客户的服务持续时间
        wait(i) = b(i) - c(i); % 第 i 个客户等待的时间 = 第 i 个客户开始服务的时间 - 第 i 个客户到达车辆的时间
        w = w + wait(i); % 更新所有客户等待的总时间
        i = i + 1; % 增加一名新的客户
        x(i) = exprnd(10); % 这位新客户和上一个客户到达的时间间隔
        c(i) = c(i-1) + x(i); % 这位新客户到达车辆的时间 = 上一个客户到达车辆的时间 + 这位新客户和上一个客户到达的时间间隔
        b(i) = max(c(i),e(i-1)); % 这个新客户开始服务的时间取决于其到达时间和上一个客户结束服务的时间
    end
    n(k) = i-1; % n(k)表示车辆第 k 天服务的客户人数
    t(k) = w/n(k); % t(k)表示该车辆第 k 天客户的平均等待时间
end
disp([num2str(day),'个工作日中，车辆每点平均服务的客户人数为: ',num2str(mean(n))])
disp([num2str(day),'个工作日中，车辆每点客户的平均等待时间为: ',num2str(mean(t))])
toc   %计算 tic 和 toc 中间部分的代码的运行时间
```

9.3 代码 3 价值区域聚类

```
[idx,C] = kmeans(X,3)
figure;
plot(X(:,1),X(:,2),'.');
title 'Randomly Generated Data';
figure;
plot(X(idx==1,1),X(idx==1,2),'r.','MarkerSize',12)
hold on
plot(X(idx==2,1),X(idx==2,2),'b.','MarkerSize',12)
plot(C(:,1),C(:,2),'kx',...
     'MarkerSize',15,'LineWidth',3)
legend('Cluster 1','Cluster 2','Centroids',...
       'Location','NW')
title 'Cluster Assignments and Centroids'
hold off
```