

高速公路行车时间估计及最优路径选择问题

1 问题复述

I

行车时间的估计对于旅行者来说非常重要。因此，有些美国高速公路安装了传感器。比如在圣安东尼奥（San Antonio）市在所有的双向六车道的高速路上都安装了传感器。但是车辆往往会不停的变换车道，为了简化问题我们可以忽略换道的影响，而只考虑一个车道的交通问题（如下图所示（参见原题），正方形代表传感器）。

1. 传感器可以每天 24 小时探测每个车辆的速度。每辆车的速度信息每 20 秒刷新一次记录。下表是一组真实数据（由于交通数据非常巨大，因此只记录了每 2 分钟间隔中最后 20 秒的数据，单位：英里/小时）。请分析高速公路上的路况特点（如：拥塞及其疏导。一般来说时速高于 50 英里/小时认为不存在拥塞问题。）如果车辆在时间 t 经过传感器，那么经过多久它通过第 5 个传感器？请设计一种算法来估计车辆的运行时间，并证明算法的合理性和精确性。如果路况信息每 20 秒（而不是每 2 分钟）刷新一次，那么这对你们的估计算法有影响吗？

在上面问题条件的基础上，如果传感器不仅能探测车辆速度，而且能探测单位时间的交通流量（如下表，流量的单位是：车辆数/每 20 秒）。这些信息是否有助于算法的合理性和精确性的提高？如果是，请重新设计你的算法。

II

第一张图（参见原题）是美国德克萨斯州圣安东尼奥市的地图。第二张图（参见原题）反映了圣安东尼奥市的路况信息。旅行者可以在车辆内置的导引系统中输入当前位置和目的地，系统会帮助选择路径并估计需要的时间。不幸的是，由于每一路段（两个十字路口之间的道路）的路况是随机的，系统不能很好地确定最优（最快）路线和可靠的时间估计。

你能在问题 1 的基础上改进这个系统吗？

1. 假设每段路的运行时间是互相独立的随机变量，请为系统设计一种算法来确定最优路线及时间估计。请明确你的“最优”是如何定义的。

2. 每一段路的行车时间依赖于其出发时间，并且行车时间之间具有相关性。为了考察相关性如何随时间变化，通常会建立一个和时间有关的协方差矩阵 $Cov(ij, kl)$ ， i, j, k, l 分别表示两端路的两个交叉口。用上图设计一个合理矩阵和算法来寻找最有路线。请定义清楚最优路线的含义。如果有 n 个交叉口，那么时间相关的函数 $Cov(ij, kl)$ 是一个 $n(n-1)/2$ 阶的矩阵每一行或列代表路段：

$12, ..., 1n, 21, ..., 2n, n1, ..., n(n-1)$ 。

III

上图（参见原题），粗线表示高速公路的不同方向，上面的表示从左到右，左边的表示从上到下。车辆通过十字路口时，可以到达其它连接的路段。图中共有 14 个路口。路口之间的距离如上表所示。请分别找出从路口 3 到路口 14，从路口 14 到路口 3 的最优路线和时间估计。旅行时间的条件同问题 1，每一段路的运行时间的期望和该路段长度成比例，方差和（路段长度） $^{2/3}$ 的倒数成比例，同时也和路段两头的连接的道路数量的乘积成正比。

2 基本假设

- (1) 认为道路上行驶的车辆只有一种，即不考虑由于车辆种类而造成路况及行车时间的影响；
- (2) 道路每个方向上只有一个车道；
- (3) 道路发生拥塞时能够有效疏导，即不会发生长时间的严重拥塞；

3 符号说明

L_i : 两个路口之间的一条道路；

l_i : 传感器 i 到 $i+1$ 的距离（英里）；

t_i : 车辆到达第 i 个传感器的时刻；

$\tau_i(t_k)$: t_k 时刻从第 i 个传感器出发，到达第 $i+1$ 个传感器的行车时间；

T_i : 道路在 t 时刻出发需要的行车时间；

u_{it} : t 时刻传感器 i 处的速度；

\bar{u}_i : t 时刻路段上探测点的平均速度，

4 模型背景及分析

道路交通状态是非线性的，而且车辆、速度、路况之间具有交互作用。其研究一般有二种模型：宏观、微观。宏观模型主要考察一些描述整体行为的变量，如单位时间流量、流量密度、平均速度等。而微观模型主要考察单个车辆的瞬时速度、车距等指标。从问题本身来看主要考察的是交通的宏观行为。

高速公路（highway）是国内较普遍的翻译，但在美国高速公路有另外一个单词 freeway，其实 highway 大致相当于我国的一级公路或主干公路。由于美国交通特别发达，而且没有中国这样多的收费站，因此高速公路（highway）是美国最繁忙的公路。其特点是：1. 上下班有明显的高峰期；2. 由于公路建设时间较长，很多车道需要维修，因此经常会有临时的车道封闭，从而造成交通拥挤的情况。

对高速公路行车时间的预测和当时的路况、天气、地形等因素有关，当拥塞现象比较明显时，行车时间的随机性和非线性更加显著，因此造成预测值可靠性非常差。

从旅行者的出行角度考虑，一般用四个指标来衡量，一是考虑期望的行车时间最短；二是考虑路线具有较高的可靠性，即发生严重阻塞的概率较小；三是考虑路程最短；四是考虑费用最低。而有时这些指标之间往往是互相矛盾的，需要在两者间做出平衡，在本文中我们把行车时间最短作为设计目标。

5 问题（I）的分析与建模

令 t_i 表示车辆到达第 i 个传感器的时刻， $\tau_i(t_i)$ 表示 t_i 时刻出发，从第 i 个传感器到第 $i+1$ 个传感器的行车时间。易知，从 t 时刻出发，到达各传感器的时刻分别为：

$$t_1 = t$$

$$t_2 = t_1 + \tau_1(t_1)$$

...

$$t_n = t_{n-1} + \tau_{n-1}(t_{n-1})$$

整个路段的运行时间为：

$$T_t = \sum_{i=1}^n \tau_i(t_i) \quad (1)$$

因为传感器之间的距离不大，因此可以认为速度线性变化，此时每一个路段的运行时间可以用(2)式计算：

$$\tau_i(t_i) = l_i / \frac{(u_{ii} + u_{t,i+1})}{2} \quad (2)$$

这个算法可用以下模型表示：

$$\begin{cases} T_t = \sum_{i=1}^n \tau_i(t_i) \\ \tau_i(t_i) = l_i / \frac{(u_{ii} + u_{t,i+1})}{2} \quad i = 1, \dots, n \\ t_1 = t \\ t_{i+1} = t_i + \tau_i(t_i) \end{cases} \quad (3)$$

针对题目中的数据，用 Matlab 计算出了从某一时刻出发从第 1 个传感器到第 5 个传感器的行车时间，结果如下图所示：

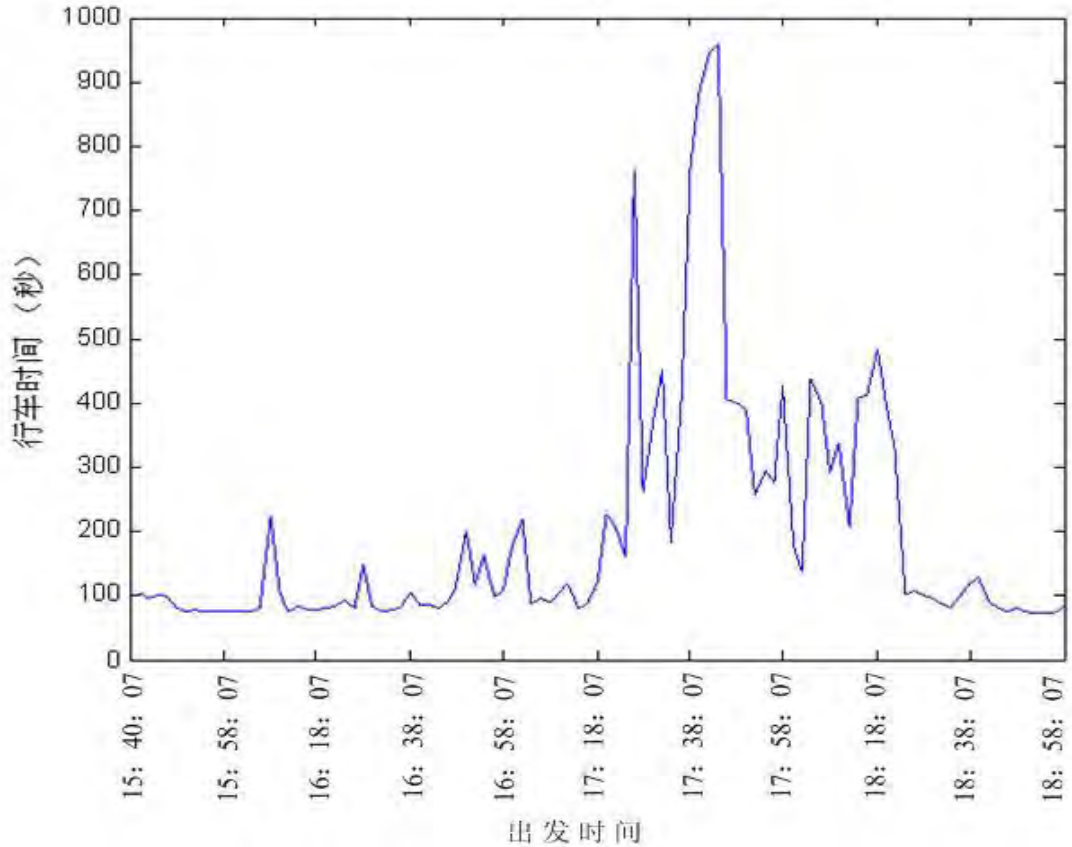


图1 不同出发时刻总行车时间的变化图

为了考察不同时刻出发运行时间的变差情况，我们把每 10 分钟的探测信息作为一组（共 20 组），然后计算每组的变差值 σ_i ，计算公式如下：

$$\bar{T}_{i,i+5} = \frac{\sum_{n=i}^{i+5} T_i}{5}$$

$$\sigma_i = \sqrt{\frac{\sum_{l=1}^5 (T_{i+l} - \bar{T}_{i,i+5})^2}{5}}$$

图 2 为总行车时间标准差随出发时刻不同的变化图；图 3 是每段路行车时间标准差随出发时刻不同的变化图。

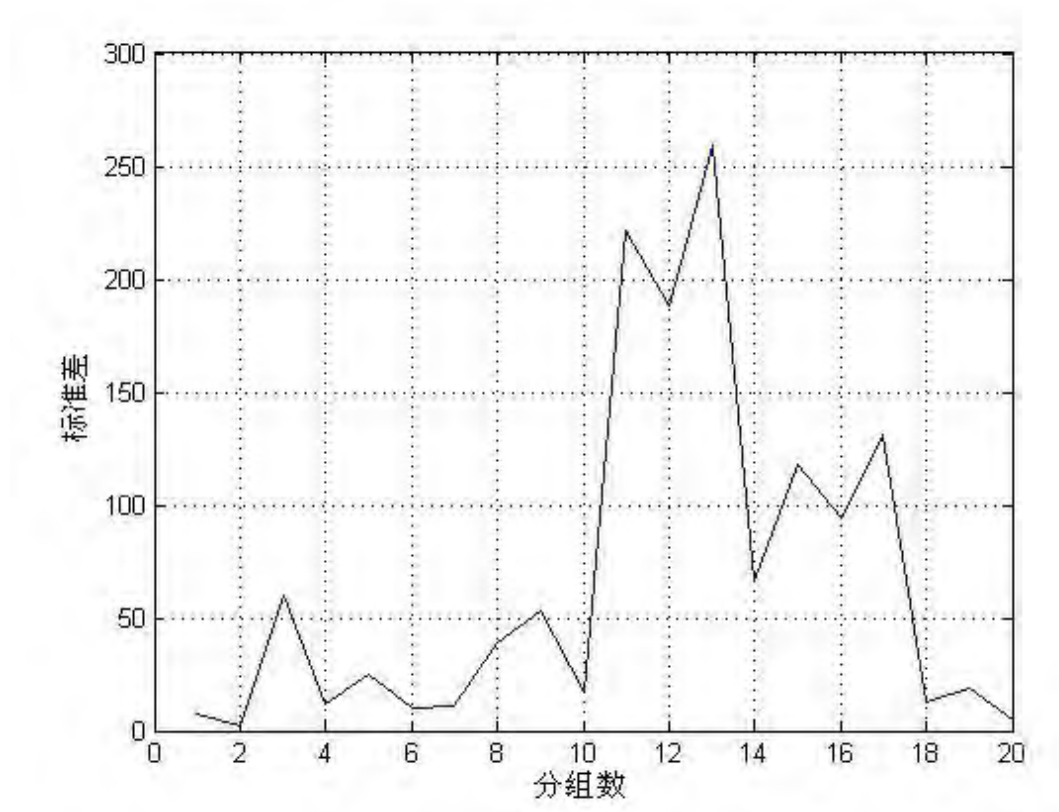


图2 不同时刻出发的总行车时间标准差

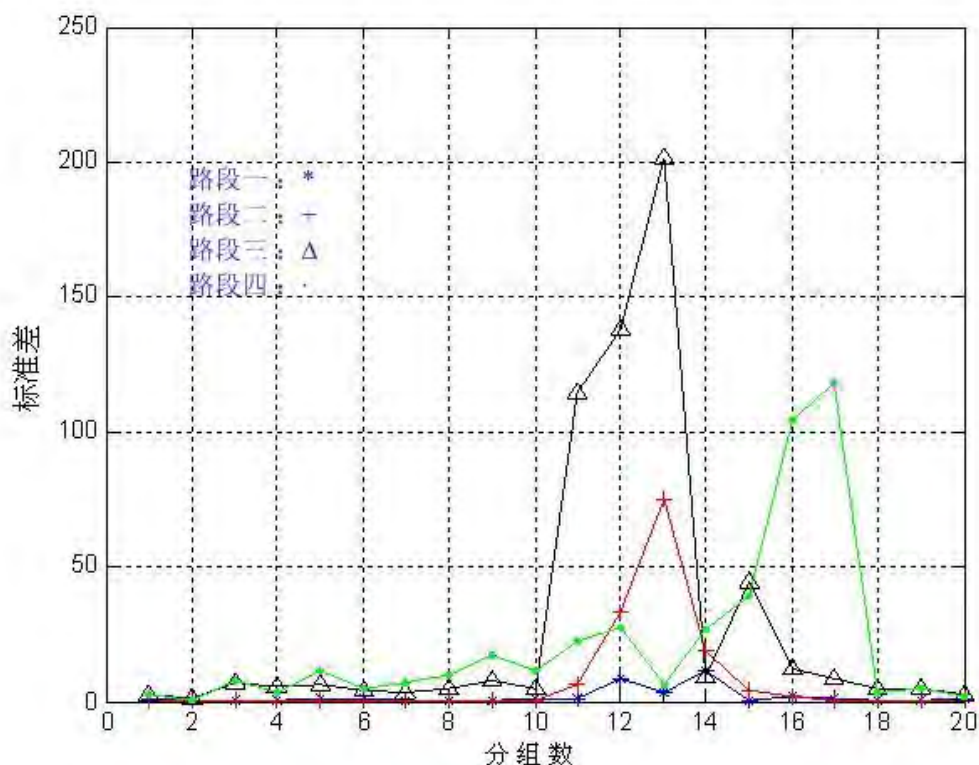


图3 不同路段不同时刻出发的行车时间标准差

从图中可以看出在 5:18:07PM (分组 11) 到 6:18:07PM (分组 17) 这段时间变差比较大, 这反映了下班时间道路比较拥挤的情况。路段三在高峰期的方差最大, 其次是路段四, 这说明这两段路的通行能力较差 (可能是本身的原因, 也可能是受交通信号、意外事故的影响)。

为了刻画不同时间出发, 所经历的路程上路况的变化, 我们定义了空间变差 σ_t^s , 它表示从 t 时刻出发, 路程上速度的变化, 计算公式如下:

$$\sigma_t^s = \sqrt{\frac{\sum_{i=1}^{n_i} (u_{ti} - \bar{u}_t)^2}{n_i}}$$

其中 u_{ti} 表示 t 时刻传感器 i 处的速度, \bar{u}_t 表示 t 时刻路段上探测点的平均速度, n_i 为探测点数量。针对题目中数据, 可得图形如下:

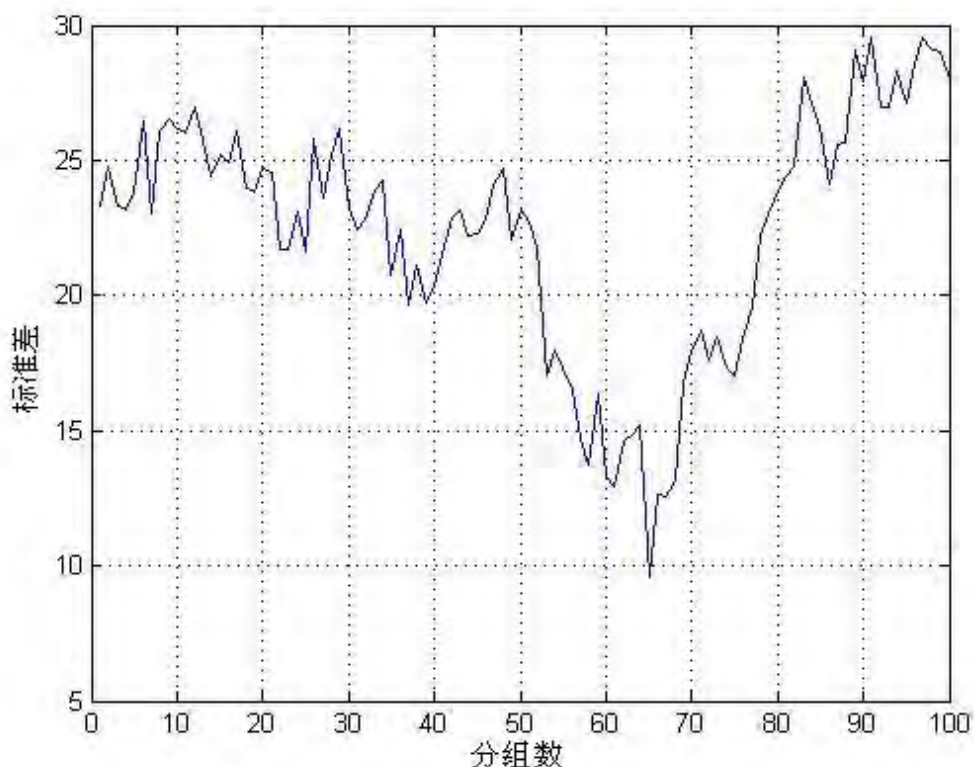


图4 不同时刻出发旅行者所经历的速度变差图

为了进一步分清道路所处的状态，可以用两个探测点的探测速度（英里/小时）来定义此段路所处的状态：

$$\begin{cases} u_{ti} \geq 50, u_{t,i+1} \geq 50 & \text{正常} \\ u_{ti} < 50, u_{t,i+1} \text{ 或 } u_{ti} > 50, u_{t,i+1} \leq 50 & \text{拥堵} \\ u_{ti} < 50, u_{t,i+1} < 50 & \text{严重拥堵} \end{cases}$$

$i = 1, 2, 3, 4$

此时，我们可以计算出每一段路在不同路况情况下每英里行车时间（秒/英里）的期望和标准差：

表1 不同路况下行车时间的期望和标准差

路段 路况	路段一		路段二		路段三		路段四	
	期望	标准差	期望	标准差	期望	标准差	期望	标准差
正常	59.8	26.0	57.3	25.2	58.4	25.9	60.0	29.1
拥堵	74.6	27.8	86.5	29.6	88.0	45.9	84.5	39.8
严重拥堵	132.0	38.4	392.1	168.7	536.0	347.9	283.4	225.2

从表中可以看出在正常情况下车辆的速度非常稳定，在拥堵的情况下变动要大一些，而在严重拥堵的情况下车辆的速度变化非常剧烈。

我们可以对两个传感器之间运行时间进行一种很简单的预测：用从 t_k 时刻开始和从 t_{k+1} 时刻开始走完两个传感器间距离的时间的平均值，预测从 t_{k+2} 时刻开始走完这段路所需要的时间。即：

$$\tau_i(t_{k+2}) = \frac{\tau_i(t_k) + \tau_i(t_{k+1})}{2}$$

然后根据公式 (3) 既可预测走完整段路的时间。下图是 4 个路段预测值和真实值的比较图。

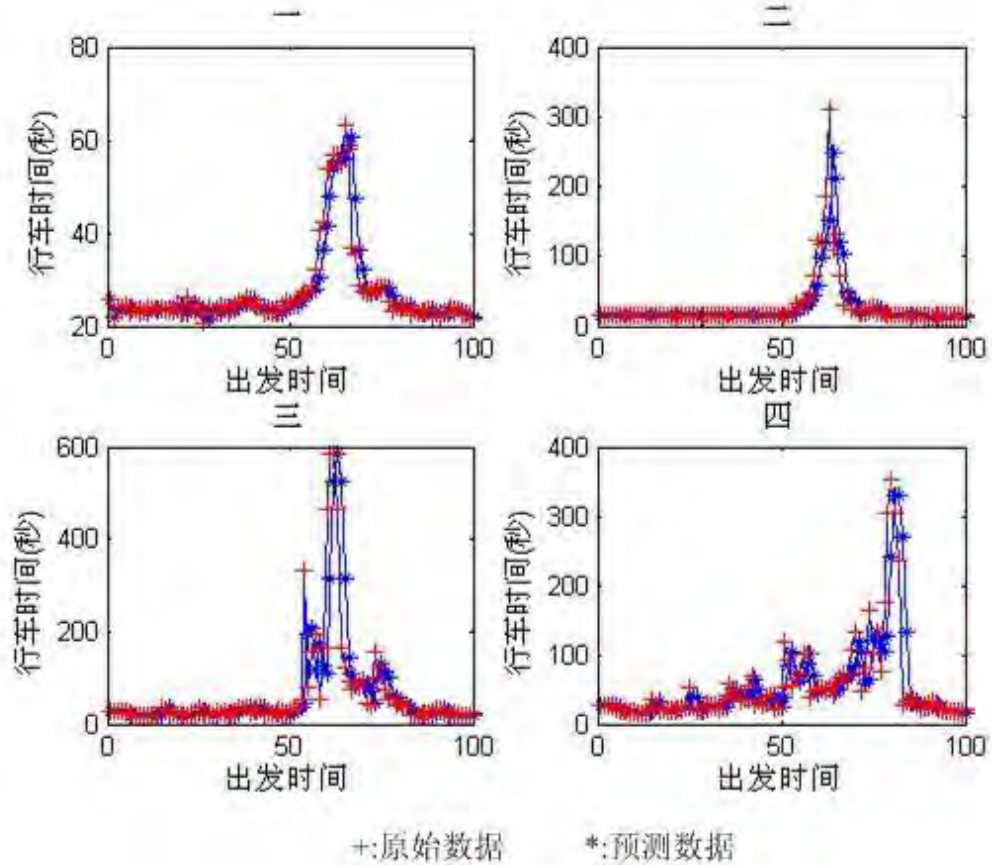


图5 预测值和真实值比较图

从图中可以看出虽然预测方法简单,但是大部分时间预测值和实际值非常接近;但是在路况拥堵的情况下,预测误差就比较大。这是因为模型考虑因素较为简单,无法反映交通流的不确定性与非线性特征,抗干扰能力差。

卡尔曼滤波(Kalman Filtering)是一种先进的控制方法,是一种基于线性回归的预测方法。其采用由状态方程和观测方程组成的线性随机系统的状态空间模型来描述滤波器,并利用状态方程的递推性,按线性无偏最小均方误差估计准则,采用一套递推算法对滤波器的状态变量作最佳估计,从而求得滤掉噪声的有用信号的最佳估计。卡尔曼滤波具有独特的优点:具有广泛的适应性,由于卡尔曼滤波采用较灵活的递推状态空间模型,既能处理平稳数据,也能处理非平稳数据;只要对状态变量作不同的假设,就可使其描述及处理不同类型的问题;模型具有线性、无偏、最小均方差性;模型便于在计算机上实现,且大大减少了计算机的存储量和计算时间,适于在线分析;预测精度较高。

设 $x(t)$ 为要预测的 t 时刻出发的行车时间, $\Phi(t)$ 是状态转移参数(通过历史数据获得), $w(t)$ 为随机干扰,服从参数为 $(\mu=0, \sigma^2=Q(t))$ 的正态分布。则状态转移方程为:

$$x(t) = \Phi(t-1)x(t-1) + w(t-1) \quad (4)$$

令 $z(t)$ 为区段运行时间的观测值（需要通过观测的速度计算出来）， $v(t)$ 为观测误差（在此表示用速度均值计算区段运行时间的误差），它服从参数为 $(\mu=0, \sigma^2=R(t))$ 的正态分布。因为只有区段运行时间一个参数，所以观测方程为：

$$z(t) = x(t) + v(t) \quad (5)$$

对任意 i, j ，有 $E[w(i)v(j)] = 0$ 。令 $P(t)$ 为预测方差。根据卡尔曼滤波理论可以得到如下的计算步骤：

①初始化：

$$t = 0, E[x(0)] = \hat{x}(0), E[(x(0) - \hat{x}(0))^2] = P(0)$$

②外推：

$$\hat{x}(t)_- = \Phi(t-1)\hat{x}(t)_+$$

$$P(t)_- = \Phi(t-1)P(t-1)_+ \Phi(t-1) + Q(t-1)$$

③卡尔曼增益矩阵：

$$K(t) = P(t)_- [P(t)_- + R(t)]^{-1}$$

④更新：

$$\hat{x}(t)_+ = \hat{x}(t)_- + K(t)[z(t) - \hat{x}(t)_-]$$

$$P(t)_+ = [I - K(t)]P(t)_-$$

⑤令 $t = t+1$ ，直到终止条件满足。

通过编制程序，用卡尔曼滤波的方法对时间进行预测，并把结果和上文提到的简单平均的方法对比，其偏差平方和明显降低，尤其在公路的高峰时段，比简单平均的方法，预测值更加平稳。

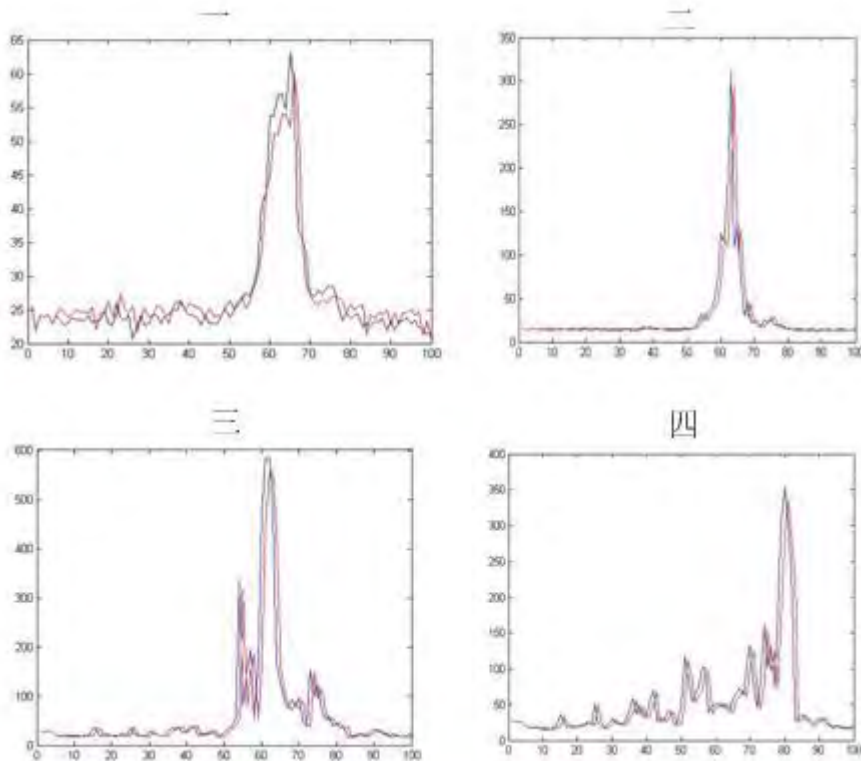


图6 卡尔曼滤波获得的预测值和真实值比较

如果要对 24 小时的行车情况进行预测，我们可以根据所给数据进行平移，在上班和下班高峰期，路况和车辆的运行速度应该是相似的，因此，我们可以把表中 16:00:07—>17:00:07 这一个小时的数据可以近似的认为是交通正常时车辆的运行情况。把表中 17:00:07—>18:00:07 这一个小时的数据近似的认为是上班或者下班高峰期的车辆运行情况。这样我们就可以得到一天 24 小时的路况预测信息。

6 问题（II）的分析与建模

（1）问题 1

根据上一部分的结果，我们可以把一条路 L_i 上的路况信息分为三种：正常、拥塞、严重拥塞，对应路段的个数分别是 n_{i1}, n_{i2}, n_{i3} 。通过上面的办法可以得到每一段路行车时间的期望和标准差，则整条路的期望和标准差为：

$$\mu_{Li} = \sum_{i=1}^{n_{i1}} \mu_i^n + \sum_{i=1}^{n_{i2}} \mu_i^c + \sum_{i=1}^{n_{i3}} \mu_i^h$$

$$\sigma_{Li} = \sqrt{\sum_{i=1}^{n_{i1}} \sigma_i^n + \sum_{i=1}^{n_{i2}} \sigma_i^c + \sum_{i=1}^{n_{i3}} \sigma_i^h}$$

其中 $\mu_i^n, \mu_i^c, \mu_i^h$ 分别表示在正常、拥塞、严重拥塞情况下路段 i 的行车时间的期望； $\sigma_i^n, \sigma_i^c, \sigma_i^h$ 分别表示在正常、拥塞、严重拥塞情况下路段 i 的行车时间的标准差。则道路 L_i 的行车时间服从参数为 $(\mu_{Li}, \sigma_{Li}^2)$ 的正态分布，对没有数据的路段可以取相邻路段参数的均值作为自己的参数。

如果资料不够充分，则可以对走完 L_i 所需要的时间做出三种估计：最乐观估计、最可能估计和最保守估计，分别对应于正常、拥塞、严重拥塞。可以将三种估计分别记为 a, c, b ，并假定取值为 a, c, b 的概率分别为 $\frac{1}{6}, \frac{4}{6}, \frac{1}{6}$ ，于是可以得到走完这段路的时间的期望可以用 $ET = \frac{a+4c+b}{6}$ 来估计。由切比雪夫不等式

$P\{|t_i - ET| \geq \varepsilon \sigma_i\} \leq \frac{1}{\varepsilon^2}$ 可知，当 $\varepsilon = 3$ 的时候，走完这段路所需要的时间的 t_i 与其期望时间 ET 的偏差超过走完该路段所需要的时间的均方差 σ_i 三倍的概率不大于 $\frac{1}{9}$ ，忽略偏差不计，令 t_i 的最小值 $a = ET - 3\sigma_i$ ， t_i 的最大值 $b = ET + 3\sigma_i$ ，则

$b - a = 6\sigma_i$ ，所以走完 L_i 所需要的时间的均方差可以用 $\sigma_i = \frac{b-a}{6}$ 来估计。

下面考虑某一条路，假设这条路总共可以分为 m 段，那么走完这条路所需要的时间期望 μ 的估计，可以表示为这 m 段路时间期望的和，走完这条路所需要的时间方差 σ^2 的估计，可以表示为这 m 段路时间方差的和。根据概率论的中心极限定理可知，走完这条路的时间渐进的服从正态分布 $N(\mu, \sigma^2)$ ，据此可计算得到在某一段时间内走完这条路的概率 P 。

$$P = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\tau-\mu)^2}{2\sigma^2}} d\tau = \Phi\left(\frac{t-\mu}{\sigma}\right)$$

根据以上讨论，我们可以将原问题转换为不确定的 PERT 网问题，优化目标定义为，在某一个概率 P 的条件下，走完某条路下的最短时间。

(2) 问题 2

1、随机过程的方法

因为每一段路的运行时间和出发时间具有相关性，因此我们可以将每一段路的运行时间看做一个随机过程。

走完一段路的时间记为 $X(t), t \in T$ 。对于每一个固定的 $t \in T$ ，随机变量 $X(t)$ 的分布函数与 t 有关，记为

$$F(x, t) \triangleq P\{X(t) \leq x\}, x \in R,$$

称它为随机过程的一维分布函数，一维分布函数能够刻画随机过程在各个个别时刻的统计特性。为了描述随机过程在不同时刻状态之间的统计关系，对任意 n 个不同的时刻 $t_1, t_2, \dots, t_n \in T$ ，引入 n 维随机变量 $(X(t_1), X(t_2), \dots, X(t_n))$ ，它的分布函数记为

$$F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) \triangleq P\{X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_n) \leq x_n\}$$

对于固定的 n ，我们称 $\{F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n), t_i \in T\}$ 为随机过程的 n 维分布函数族。当 n 充分大时， n 维分布函数族能够近似的描述随机过程的统计特性。显然 n 取得越大，则 n 维分布函数族描述随机过程的特性也越完善，通常有限维分布函数族，即 $\{F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n), n=1, 2, \dots, t_i \in T\}$ ，完全地确定了随机过程的统计特性。

给定一个随机过程，固定 $t \in T$ ， $X(t)$ 是一随机变量，它的均值与 t 有关，记为

$$\mu_X(t) \triangleq E[X(t)]$$

我们称 $\mu_X(t)$ 为随机过程的均值函数，均值函数 $\mu_X(t)$ 表示随机过程在各个时刻的摆动中心。

设任意 $t_1, t_2 \in T$ ，我们把随机变量 $X(t_1)$ 和 $X(t_2)$ 的二阶中心混合矩记作

$$C_{XX}(t_1, t_2) \triangleq \text{cov}[X(t_1), X(t_2)] = E\{[X(t_1) - \mu_X(t_1)][X(t_2) - \mu_X(t_2)]\}$$

称 $C_{XX}(t_1, t_2)$ 为随机过程的自协方差函数。

由多维随机变量数字特征的知识可知，自协方差函数是刻画随机过程在两个不同时刻的状态之间统计依赖关系的数字特征。

综上所述，我们可以用随机过程的自协方差函数来描述每一段路的运行时间和出发时间的相关性，建立一个运行时间和出发时间有关的协方差矩阵。

2、卡尔曼滤波方法

令 $x(t)$ 为 $\frac{n(n-1)}{2} \times 1$ 的列向量，代表系统中路段 $12, \dots, 1n, 21, \dots, 2n, n1, \dots, n(n-1)$

在 t 时刻出发的行车时间。此时卡尔曼方程为：

$$x(t) = \Phi(t-1)x(t-1) + w(t-1)$$

$$z(t) = H(k-1)x(t) + v(t)$$

此时的递推公式为：

$$\hat{x}(t)_- = \Phi(t-1)\hat{x}(t)_+$$

$$P(t)_- = \Phi(t-1)P(t-1)_+ \Phi(t-1) + Q(t-1)$$

$$K(t) = P(t) - H^T(t)[H(t)P(t) - H^T(t) + R(t)]^{-1} K(t) = P(k)$$

$$P(t)_+ = [I - K(k)H(k)]P(t)_-$$

可知 $P(t)$ 是维数为 $\frac{n(n-1)}{2} \times \frac{n(n-1)}{2}$ 的矩阵，它表示预测值和真实值的误差，因此可用它作为要求的协方差矩阵。

7 问题（III）的分析与建模

我们可以参考问题（I）的计算结果来给出期望和方差的比例系数的大致估计，第一问将路况分为三种不同的等级：正常、拥塞、严重拥塞。根据问题 III 提供的每一段路的运行时间和该路段长度成正比，方差和（路段长度）^{2/3} 的倒数成正比，同时也和路段两头的连接的道路数量的乘积成正比这两个条件，可以按照以下的方法来估计期望和方差的比例系数。

令路段 1 到路段 4 的长度分别为 l_1, l_2, l_3, l_4 。

在正常情况下的期望分别是 $E_{11}, E_{12}, E_{13}, E_{14}$ ，标准差分别是 $\sigma_{11}, \sigma_{12}, \sigma_{13}, \sigma_{14}$ 。

估算在正常路段期望时间与路程的比例系数为

$$k_{1e} = \frac{1}{4} \sum_{i=1}^4 \frac{E_{1i}}{l_i}$$

时间方差与（路段长度）^{2/3} 的倒数的比例系数为

$$k_{1\sigma} = \frac{1}{4} \sum_{i=1}^4 \sigma_{1i}^2 l_i^{2/3}$$

在拥塞情况下的期望分别是 $E_{21}, E_{22}, E_{23}, E_{24}$ ，标准差分别是 $\sigma_{21}, \sigma_{22}, \sigma_{23}, \sigma_{24}$ 。

估算在拥塞路段期望时间与路程的比例系数为

$$k_{2e} = \frac{1}{4} \sum_{i=1}^4 \frac{E_{2i}}{l_i}$$

时间方差与（路段长度）^{2/3} 的倒数的比例系数为

$$k_{2\sigma} = \frac{1}{4} \sum_{i=1}^4 \sigma_{2i}^2 l_i^{2/3}$$

在严重拥塞情况下的期望分别是 $E_{31}, E_{32}, E_{33}, E_{34}$ ，标准差分别是

$\sigma_{31}, \sigma_{32}, \sigma_{33}, \sigma_{34}$ 。

在严重拥塞路段期望时间与路程的比例系数为

$$k_{3e} = \frac{1}{4} \sum_{i=1}^4 \frac{E_{3i}}{l_i}$$

时间方差与（路段长度）^{2/3} 的倒数的比例系数为

$$k_{3\sigma} = \frac{1}{4} \sum_{i=1}^4 \sigma_{3i}^2 l_i^{2/3}$$

按照以上方法估计的比例系数可以估算出任意两个相邻路口 i, j 的时间期望和方差。

当相邻路口 i, j 之间的路段为正常路况的时候，其行驶时间的期望和方差分

别为 $\mu_{ij} = k_{1e} l_{ij}$ 和 $\sigma_{ij} = \sqrt{k_{1\sigma} \frac{1}{l_{ij}^{2/3}} n_i n_j}$ 。

当相邻路口 i, j 之间的路段为拥挤路况的时候，其行驶时间的期望和方差分

别为 $\mu_{ij} = k_{2e} l_{ij}$ 和 $\sigma_{ij} = \sqrt{k_{2\sigma} \frac{1}{l_{ij}^{2/3}} n_i n_j}$ 。

当相邻路口 i, j 之间的路段为严重拥挤路况的时候，其行驶时间的期望和方

差分别为 $\mu_{ij} = k_{3e} l_{ij}$ 和 $\sigma_{ij} = \sqrt{k_{3\sigma} \frac{1}{l_{ij}^{2/3}} n_i n_j}$ 。

其中 n_i, n_j 分别为与路段两端 i, j 相邻的道路数量。

按照以上的方法可以估算出车辆在任意两个路口之间行驶的时间期望和方差。如果一条路线由路段 $d_1, d_2 \dots d_{n-1}, d_n$ 确定的，并且这 n 个路段的行车时间期望和方差可以按照以上的方法估算为 $\mu_1, \mu_2 \dots \mu_{n-1}, \mu_n$ 和 $\sigma_1^2, \sigma_2^2 \dots \sigma_{n-1}^2, \sigma_n^2$ 为那么车辆在这条路线上的行驶时间期望 μ 和方差 σ^2 可以按照以下两个式子估算出来

$$\mu = \sum_{i=1}^n \mu_i; \quad \sigma^2 = \sum_{i=1}^n \sigma_i^2;$$

由此可以得到走完任意一条路线的时间期望和方差。

根据问题二定义优化问题的方法，我们同样可以将问题三优化目标描述为：在一定的概率条件下，车辆从路口 a 往路口 b 行驶所需要的最短时间。

根据表一给出的数据，按照以上给出的方法，可以给出每一个路段的时间期望和方差的估计如下：

表2 不同路段行车时间的期望和方差

行车方向	时间（秒）	行车方向	时间（秒）
1 —> 2	476.7239	2 —> 1	476.7239
2 —> 3	538.5214	3 —> 2	538.5214
1 —> 4	406.0981	4 —> 1	2317.7
2 —> 5	344.3006	5 —> 2	344.3006
3 —> 6	308.9877	6 —> 3	308.9877
4 —> 5	300.1595	5 —> 4	300.1595
5 —> 6	291.3313	6 —> 5	412.8241
4 —> 10	565.0061	10 —> 4	565.0061
4 —> 7	361.9570	7 —> 4	512.9026
7 —> 8	114.7669	8 —> 7	655
5 —> 8	500.3928	8 —> 5	353.1288

8 —→ 9	300.1595	9 —→ 8	300.1595
6 —→ 9	185.3926	9 —→ 6	185.3926
10 —→ 11	503.2085	11 —→ 10	503.2085
7 —→ 11	856.5	11 —→ 7	150.0797
8 —→ 12	167.7362	12 —→ 8	167.7362
11 —→ 12	141.2515	12 —→ 11	141.2515
9 —→ 13	132.4233	13 —→ 9	132.4233
12 —→ 13	361.9570	13 —→ 12	361.9570
10 —→ 14	326.6441	14 —→ 10	326.6441
11 —→ 14	476.7239	14 —→ 11	476.7239

根据表二所提供的数据,在概率为 0.9 的条件下,可以得到以下的优化结果:
从路口 3 到路口 14 的最佳路线为:

3—>6—>9—>8—>12—>11—>14

其中 $\mu_1 = 1580.25$ 秒, $\sigma_1 = 111.437$

需要的时间在 $\tilde{T}_1 = \mu_1 + \sigma_1(1 - z_{0.1})$ 以内, 即 $\tilde{T}_1 \leq 1723.45$ 秒。

从路口 3 到路口 14 的最佳路线为:

14—>11—>7—>8—>9—>6—>3

其中 $\mu_2 = 1536.11$ 秒, $\sigma_2 = 108.044$

需要的时间在 $\tilde{T}_2 = \mu_2 + \sigma_2(1 - z_{0.1})$ 以内, 即 $\tilde{T}_2 \leq 1674.95$ 秒。

8 模型的评价及误差分析

问题 (I) 的要求是在给定历史数据的基础上, 首先计算 t 时刻出发到达第 5 个传感器的时间, 然后给出预测时间的方法。由于大多数车辆 20 秒左右就可以通过这 5 个传感器, 因此采用两端速度均值来计算每个路段的行驶时间是可靠的, 然后根据递推公式可以较好的计算出总共需要的时间。

由于一般的预测模型在历史数据比较少的时候难以保证预测精度, 而在数据量大时, 往往存在计算量大、运行时间长等不足, 并且难以保证满足需要。本文的主要建模思想是根据距离车辆最近的观测点速度, 来预测下一时刻车辆在此路段的运行速度, 再根据预测速度预测走完此路段所需要的行车时间。该模型出于如下的考虑: 即相对比较远的路段, 旅行者更加关心离自己较近路段的速度预测结果。本文的模型能够在较低的历史数据量水平下, 以较低的计算量和较短的时间内完成预测速度的计算, 对于旅行者而言具有更高的使用价值。随着数据量的增大, 本文的模型具有较好的鲁棒性, 能够在同样的计算复杂度和时间复杂度下提供更高精度的预测, 从另一方面看, 即在同等精度下的预测路段更远。模型的不足之处在于模型的预测精度受观测点间隔的影响, 即如果历史数据的时间间隔较大, 会使得预测精度降低, 这实际上是由于所使用的插值预测方法的局限。另外对于交通高峰期行驶时间的巨涨落问题, 本文提出了采用卡尔曼滤波的方法来处理, 得到了比较满意的结果。

根据问题二所提供的条件, 我们将高速公路行车模型转换为非肯定型 PERT 网络模型, 运用问题一所提供的估算某一个路段行车时间的方法, 来估算 PERT

网络模型中每一个时序的期望和方差，最后将优化问题定义为，在满足一定的概率的条件下，完成给定的两点的行车时间的最小值。

问题三给定了一个高速公路行车的具体例子。根据问题三提供的估计期望和方差的方法，在结合问题一的数据结果，以及问题二所提出模型转换方式，即将高速公路行车模型转换为非肯定型 PERT 网络模型，来估计从路口 3 到路口 14，以及从路口 14 到路口 3 的最优路线和时间估计。

6 结论

本文研究了高速公路行车时间估计及最优路径选择问题。通过对原始数据的分析及定义相应的指标，详细考察了高速公路上的行车时间的变化规律，给出了不同路段在正常、拥塞、严重拥塞等 3 种模式下的行车时间的期望和方差，并进行了插值预测和卡尔曼滤波预测的有关讨论。

通过建立路段运行时间的分布函数，给出了一般路径的行车时间分布的密度函数。并把寻找最优路径问题转换为不确定的 PERT 网问题，给出了解算的方法。通过对路线行车时间函数的讨论，得出了不同路线间的协方差矩阵。

根据文中讨论的结果，针对一个具体的最优路径问题，给出了具体的求解过程，及其置信度。

参考文献

- [1] 运筹学教材编写组. 运筹学（修订版）[M]. 北京：清华大学出版社. 1990.3.
- [2] 最优化计算原理与算法程序设计. 粟塔山等著[M]. 长沙：国防科技大学出版社. 2002.12.
- [3] 城市交通流诱导系统理论与模型. 杨兆升著[M]. 北京：人民交通出版社. 1999.9.
- [4] 概率论与数理统计.（第二版）盛骤等著[M].北京：高等教育出版社.1989.8
- [5] 实用运筹学.魏国华等著[M].上海：复旦大学出版社.1987.2
- [6] 运筹学.（下册）蔡溥等著[M].北京：纺织工业出版社.1986.12