

## 2013 高教社杯全国大学生数学建模竞赛

### 承 诺 书

我们仔细阅读了《全国大学生数学建模竞赛章程》和《全国大学生数学建模竞赛参赛规则》（以下简称为“竞赛章程和参赛规则”，可从全国大学生数学建模竞赛网站下载）。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛章程和参赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛章程和参赛规则，以保证竞赛的公正、公平性。如有违反竞赛章程和参赛规则的行为，我们将受到严肃处理。

我们授权全国大学生数学建模竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

我们参赛选择的题号是（从 A/B/C/D 中选择一项填写）：\_\_\_\_\_A\_\_\_\_\_

我们的参赛报名号为（如果赛区设置报名号的话）：\_\_\_\_\_

所属学校（请填写完整的全名）：\_\_\_\_\_

参赛队员（打印并签名）：1.\_\_\_\_\_刘美含\_\_\_\_\_

2.\_\_\_\_\_胡若妍\_\_\_\_\_

3.\_\_\_\_\_安彤\_\_\_\_\_

指导教师或指导教师组负责人（打印并签名）：\_\_\_\_\_

（论文纸质版与电子版中的以上信息必须一致，只是电子版中无需签名。以上内容请仔细核对，提交后将不再允许做任何修改。如填写错误，论文可能被取消评奖资格。）

日期：\_\_\_\_\_年\_\_月\_\_日

---

赛区评阅编号（由赛区组委会评阅前进行编号）：

大美mixer  
简书>>

<https://www.jianshu.com/u/84f0ebbbac87>  
Github>><https://github.com/lmh760008522>  
CSDN>>[https://blog.csdn.net/qq\\_34746896](https://blog.csdn.net/qq_34746896)

## 2013 高教社杯全国大学生数学建模竞赛

### 编 号 专 用 页

赛区评阅编号（由赛区组委会评阅前进行编号）：

赛区评阅记录（可供赛区评阅时使用）：

评阅人										
评分										
备注										

全国统一编号（由赛区组委会送交全国前编号）：

全国评阅编号（由全国组委会评阅前进行编号）：

# 车道被占用对城市道路通行能力的影响

## 摘 要

本文针对车道被占用对城市道路通行能力的影响问题,建立了实际通行能力变化模型和考虑车道通行能力不一致的实际通行能力变化模型,解决了事故所处横断面实际通行能力变化过程的描述问题。建立了车辆变道影响力量化模型,解决了交通事故所占车道不同对该横断面实际通行能力影响差异的分析问题;建立了基于交叉验证法的支持向量机回归分析模型,解决了分析排队长度与事故横断面实际通行能力、事故持续时间、路段上游车流量间的关系问题;建立了基于支持向量机的动态预测模型,解决了车辆排队长度到达上游路口的时间预测问题。

针对问题一,建立了实际通行能力变化模型和考虑车道通行能力不一致的实际通行能力变化模型,解决了事故所处横断面实际通行能力变化过程的描述问题。首先,对视频内容进行分析并提取车辆通行数据,并剔除错误数据、对缺失数据进行插值拟合;其次,分析道路通行能力的影响因素,并建立实际通行能力变化模型初步判断事故所处横断面实际通行能力变化过程;然后,考虑车道通行能力不一致的因素,建立考虑车道通行能力不一致的实际通行能力变化模型,进一步研究问题。最后,对比两个模型的结果并对模型进行检验。最终得到事故所处横断面道路通行能力总体上呈先减少再平缓最后上升的状态;改进后的模型较贴近符合实际,精确度更高。

针对问题二,建立了车辆变道影响力量化模型,解决交通事故所占车道不同对该横断面实际通行能力影响差异的分析问题。首先利用 K-S 检验方法对上游车流量进行正态分布校验;其次对两个视频中的横断面实际通行能力进行显著性检验,初步判断视频一与视频二中的实际通行能力差异是否显著;最后建立车辆变道影响力量化模型,验证了影响差异的主要因素为交通事故所占车道不同。最终得到车道二三被占的影响大于车道一二被占的结论。

针对问题三,建立了基于交叉验证法的支持向量机回归分析模型,解决了分析排队长度与事故横断面实际通行能力、事故持续时间、路段上游车流量间的关系问题;首先引入支持向量机回归分析的原理和算法步骤;其次利用交叉验证法确定参数并选择合适的核函数;然后将统计数据的前 20 组作为训练集,后 7 组作为测试集创建并训练 SVR 回归模型;最后进行仿真测试并计算均方误差检验模型的准确性。最终得到交通事故所影响的路段车辆排队长度与事故持续时间之间的关系最为密切,与路段上游车流量关系较密切,与事故横断面实际通行能力的相关性较小。模型具有非常好的泛化能力。

针对问题四,建立了基于支持向量机的动态预测模型,解决了车辆排队长度到达上游路口的时间预测问题。首先增加考虑车祸横断面距上游路口的距离因素,重新创建训练集和测试集;其次采用默认的 RBF 核函数并得到最佳的惩罚因子和核参数中的方差;然后对车辆排队长度进行预测并预测数据加入历史数据集重新训练支持向量机;最后对模型进行检验。最终得到经过 4.5min 后,车辆排队长达 155.71 米;模型预测结果较为可信的结论

**关键词:** 插值拟合 显著性检验 交叉验证法 支持向量机 动态预测

## 一、问题重述

车道被占用是指因交通事故、路边停车、占道施工等因素，导致车道或道路横断面通行能力在单位时间内降低的现象。由于城市道路具有交通流密度大、连续性强等特点，一条车道被占用，也可能降低路段所有车道的通行能力，即使时间短，也可能引起车辆排队，出现交通阻塞。如处理不当，甚至出现区域性拥堵。

车道被占用的情况种类繁多、复杂，正确估算车道被占用对城市道路通行能力的影响程度，将为交通管理部门正确引导车辆行驶、审批占道施工、设计道路渠化方案、设置路边停车位和设置非港湾式公交车站等提供理论依据。

视频 1（附件 1）和视频 2（附件 2）中的两个交通事故处于同一路段的同一横断面，且完全占用两条车道。请研究以下问题：

1. 根据视频 1（附件 1），描述视频中交通事故发生至撤离期间，事故所处横断面实际通行能力的变化过程。
2. 根据问题 1 所得结论，结合视频 2（附件 2），分析说明同一横断面交通事故所占车道不同对该横断面实际通行能力影响的差异。
3. 构建数学模型，分析视频 1（附件 1）中交通事故所影响的路段车辆排队长度与事故横断面实际通行能力、事故持续时间、路段上游车流量间的关系。
4. 假如视频 1（附件 1）中的交通事故所处横断面距离上游路口变为 140 米，路段下游方向需求不变，路段上游车流量为 1500pcu/h，事故发生时车辆初始排队长度为零，且事故持续不撤离。请估算，从事故发生开始，经过多长时间，车辆排队长度将到达上游路口。

附件 1：视频 1

附件 2：视频 2

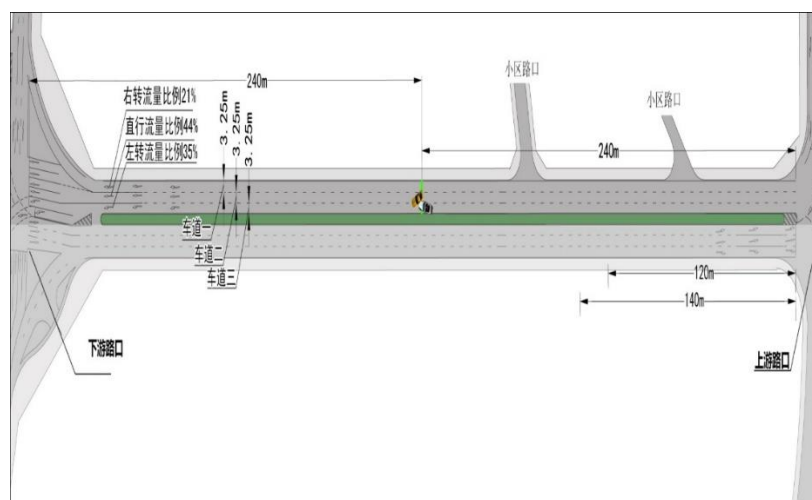
附件 3：视频 1 中交通事故位置示意图

附件 4：上游路口交通组织方案图

附件 5：上游路口信号配时方案图

注：只考虑四轮及以上机动车、电瓶车的交通流量，且换算成标准车当量数。

附件 3

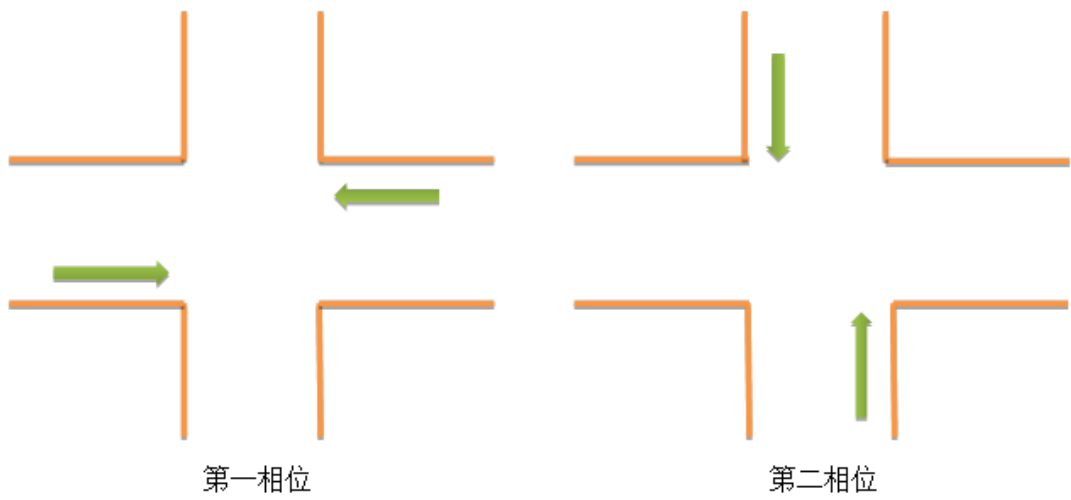


视频 1 中交通事故位置示意图

附件 4



附件 5



相位时间均为 30s，黄灯时间为 3s，信号周期为 60s

相位时间=绿灯时间+绿闪时间（3s）+黄灯时间

上游路口信号配时方案

## 二、问题分析

### 2.1 问题一的分析

描述视频中交通事故发生至撤离期间横断面实际通行能力的变化过程，就是要得出交通事故发生至撤离期间通过车流量的变化情况<sup>[1]</sup>。由此，我们考虑统计该段时间内通过的车辆数，以此来反映道路的通行能力。通常这种情况下我们需要考虑到不同车型对道路通行能力造成的不同影响，因此，我们引入不同的车型转换系数将其他车型都转换为标准车型<sup>[2]</sup>，评估不同类型的车辆对道路通行能力的影响程度。基本通行能力的计算方法通常有观测车头间距法、观测车头时距法和观测速度密度法。为了方便我们的统计，此处采用观测速度密度法，即抽样调查事故发生前在此路段的行驶时间，得到车辆经过此路段的平均速度。除此之外，道路的宽度、车流量、包括车辆在行驶过程中的超车变道行为都对实际通行能力造成一定的影响，以此对模型进行改进。结合视频统计所需数据，描述事故所处横断面实际通行能力的变化过程。通过分析道路通行能力的变化过程，我们能够更准确的得到在事故发生前和事故发生后通行能力的变化并初步判断其原因。最后，通过比较分析改进前后的实际通行能力对模型进行检验，判断模型的合理性。

### 2.2 问题二的分析

问题二要求我们分析说明同一横断面交通 所占车道不同对实际通行能力的影响差异，就是要求我们对视频 1 和视频 2 中道路的实际通行能力进行显著性检验，通常进行显著性检验的方法有 t 检验、方差检验、单因素分析等。在这里我们使用较为简单合理的 t 检验对两段视频中道路的实际通行能力进行差异性分析，并通过对视频的观察和显著性检验的结果初步推测产生差异的原因可能为事故发生时所占车道的差异。接着，对我们推测的原因进行探讨。考虑到允许通过的车道的差异，即道路上的车辆极有可能存在大量变道行为<sup>[3]</sup>，我们引入变道系数对实际通行能力的差异具体分析，判断车辆变道与实际通行能力的关系。

### 2.3 问题三的分析

问题三是整个问题的重点，要求分析视频 1 中交通事故所影响的路段车辆排队长度与事故横断面实际通行能力、事故持续时间、路段上游车流量间的关系。此类问题通常可用排队论模型、回归模型、交通流模型<sup>[4]</sup>、波动模型、差分方程模型等，但是基于视频中的车辆排队长度与上游车流量、事故发生路段等多方面因素都有关，且各个车道的排队机率也不尽相同，具有很大的随机性，是一个很难确定的复杂非线性函数<sup>[5]</sup>，排队长度与影响因素之间很难用一个确定的数学模型来表达，另外由于对于小样本事件模型的参数也难以确定，其预测性能存在各种不足，而 SVR 在解决小样本和高维非线性问题具有特别的优势。因此我们采用支持向量机的回归预测方法对车辆排队长度与其他因素之间的关系进行预测。采集该算法所需数据得出回归函数，利用

交叉验证法对模型参数寻优,并通过其平均绝对误差对模型的精度进行检验。

## 2.4 问题四的分析

问题四要求估算从事故发生开始车辆排队长度将到达上游路口的时间。目前,对车辆排队长度的估算模型,主要分析方法包括排队论、概率论、随机过程、累计曲线、冲击波等,然而这些模型主要从排队长度的计算方法上进行研究,由于车辆排队系统是一个高度复杂的非线性系统,车辆排队系统受多种因素影响具有不确定性和随机性,针对这些特征,神经网络预测模型成为一种有力的研究工具,然而神经网络存在训练样本大、过学习、易陷入局部最小以及网络结构的选取缺乏理论依据等缺陷。尤其在小样本的情况下神经网络等传统方法通常得不到充分的训练,使得性能不稳定,预测能力不强,因此效果并非十分理想。支持向量回归(Support Vector Regression, SVR)是基于 VC 维理论和结构风险最小化原则的函数拟合工具,在解决小样本、非线性以及高维模式识别等问题具有特别的优势,在很大程度上克服了“维数灾难”和“过学习”等问题,已经广泛应用于非线性时间序列预测。车辆排队预测模型不同于一般的时间序列预测问题,不能单纯从时间序列角度或者对交通产生影响的空间因素进行预测,本文从时空因素角度,综合考虑对车辆排队产生影响的多种因素,建立了支持向量回归机的动态回归模型对未来时刻车辆排队长度进行预测,并验证了算法的有效性。

## 三、基本假设

- 1、不考虑行人对交通道路的影响;
- 2、不考虑天气因素对道路通行能力造成的影响;
- 3、不考虑城市交通基础设施对道路通行能力造成的影响;
- 4、横断面的拥堵只由交通事故造成,并不由下游引起;
- 5、不考虑交警指挥对通行能力的影响;
- 6、不考虑上下班高峰期的影响。

## 四、符号说明

$C_B$	基本通行能力
$l_0$	最小安全间距
$\varphi$	纵向附着系数
$V$	平均车速
$C_p$	实际通行能力
$f_{HV}$	交通组成修正系数
$f_w$	车道宽度修正系数
$\alpha_i$	车道的修正系数
$\alpha_{变}$	车辆变道系数



$T$	车辆经过事故横截面的时间
$x_i$	输入列向量
$y_i$	输出值
$\Phi(x)$	非线性映射函数
$\varepsilon$	线性不敏感损失函数
$K(x,y)$	核函数
$N_{nsv}$	支持向量个数
$E$	均方误差
$R^2$	决定系数
MAE	平均绝对误差

## 五、模型的建立与求解

### 5.1 模型准备

本题以城市道路中任意一条单向 3 车道为研究对象,通过对道路监管部门对该单向 3 车道视频的收集,我们现进行如下分析。

首先,通过附件三,我们发现专右、直行、专左车道的流量比例分别为 0.21、0.44、0.35,且三条车道中没有既可以转向又可以直行的车道。事故发生地点距上游路口 240m,之间存在 2 个小区路口,且全段路线不存在公交站台,在事故发生至撤离区间内,不考虑车辆中途靠边停车对通行能力造成的影响。

其次,我们对附件四和附件五进行分析。根据附件四交叉路口的禁行标志,我们得到进入交通事故发生路段的车流只可能来自于图中上方专右车道的右转车辆和右方直行车道的直行车辆,且上方专右车道的右转车辆在行驶过程中不受色灯信号的控制。对于附件五,第一、二相位的放行时间均为 30s,信号周期均为 60s。

接着,我们对题目所给的 2 段视频进行分析。一方面,通过 2 段视频我们发现,在同一位置发生事故的车辆分别占了 3 条车道其中的 2 条,通过以 10s 为一个时间段,我们对 2 段事故视频进行车辆数据收集。在视频 1 中,事故发生时间段为 16:42:32- 17:01:20,通过数据取样得到 95 个时间段;在视频 2 中,事故发生时间段为 17:30:17-18:03:31,通过数据取样得到 182 个时间段。在事故发生至撤离的时间,2 段视频分别持续了 18:48s 和 34:14s,且视频中的片段并不完整,存在部分数据缺失的现象,数据缺失率都接近 5%,因此我们通过插值拟合的方法对这些缺失的数据进行补充处理。另一方面,该时段的车辆组成均为小型车、大型车和非机动车。

最后,通过对现实情况的认知,我们大致认为小型车的长度为 4m,公交车即大型车的长度为 8m。由于车道中的车型不尽相同,我们引入车型折算系数<sup>[6]</sup>,即将视频中的所有车型统一为标准车当量数。车型折算系数如表 1 所示:

表 1 车型折算系数

车型	折算系数
小型车	1
中型车	1.5

大型车	2
非机动车	0.7

由此，我们可以根据表 1 的数据将全部车型转换为标准车当量数，方便接下来的计算研究。

## 5.2 问题一模型的建立与求解

准确描述视频中交通事故发生至撤离期间事故所处横断面实际通行能力的变化过程，就要求我们要全面考虑现实中影响道路通行能力的因素。通过对这些影响因素的具体分析，计算各个时段事故所处横断面的实际通行能力，观察其变化过程，进而得出结论。

### 5.2.1 模型的建立

#### 1、道路通行能力概述

道路通行能力是道路能够疏导或处理交通流的能力。我国定义道路通行能力为在一定的时段，理想的道路、交通、控制和环境条件下，道路的一条车道或断面期望能通过人或车辆的合理的最大小时流率<sup>[7]</sup>。为了更好的对道路通行能力进行研究，我们着重考虑道路通行能力的影响因素。

#### 2、道路通行能力的影响因素

路段实际通行能力的影响因素包括道路条件、交通条件和交通外环境等。

道路条件是指道路的几何条件，包括道路平纵面线形、断面形式、交叉口、其它设施的布设情况。本题中的道路条件即为车道的数量、各个车道的宽度等。

交通条件是指使用路段的交通流特性，包括行驶车速、交通构成、交通流向等。包括车辆的组成、车辆的分布规律等等。

交通外环境是指交管部门或交通控制设备对车辆和行人在道路上通行以及其它与交通有关的活动所制定的带有疏导、禁止、限制或指示性质的具体措施。由于路段中不存在交警的指挥对道路通行能力的影响，因此，我们将交通外环境因素忽略不计。

#### 3、基于通行能力一致性模型的建立

根据通行能力的性质和使用要求，通行能力分成基本通行能力和实际通行能力。

##### (1) 基本通行能力的定义

基本通行能力是指在理想的道路、交通、控制和环境条件下，公路设施在 4 级服务水平时所能通过的最大小时交通量，即理论上所能通行的最大小时交通量<sup>[8]</sup>。其公式可以写为：

$$C_B = \frac{3600}{h_t} = \frac{3600}{l_0 / \frac{V}{3.6}} = \frac{1000V}{l_0} (pcu/h)$$

其中， $C_B$ 为道路的基本通行能力( $pcu/h$ )； $h_t$ 为饱和连续车流的平均车头时距(s)； $V$ 为行驶速度(km/h)； $l_0$ 为最小安全间距(m)。

##### ① 最小安全间距的定义

行驶车辆之间的最小安全间距为：

$$l_0 = l_{反} + l_{制} + l_{安} + l_{车} = \frac{V}{3.6}t + \frac{V^2}{254\phi} + l_{安} + l_{车} (pcu/h)$$

其中,  $l_{\text{反}}$  为驾驶员在反应时间内车辆行驶的距离(m);  $l_{\text{制}}$  为车辆制动距离(m);  $l_{\text{安}}$  为停车时的安全车间距(m);  $l_{\text{车}}$  为车辆本身长度(m);  $V$  为行驶车速(km/h);  $\varphi$  为纵向附着系数。在通常的城市道路设计范围内,  $l_{\text{安}}$  取  $2\text{m}^{[9]}$ ;  $l_{\text{车}}$  可采用小型车的车身长度  $4\text{m}$ ; 驾驶员的反应时间  $t=1\text{s}$ 。

## ② 行驶速度的定义

该小型车的行驶速度应是交通事故发生之前的自然速度, 为了便于研究, 可以通过随机抽样取得视频 1 中小型车在发生事故之前时所行驶的距离和时间, 并计算得到在发生事故之前小型车的平均行驶速度。

由纵向附着系数  $\varphi$  与车速  $V$  的关系可以得到  $\varphi$  值, 两者之间的关系如表 2 所示:

表 2 纵向附着系数  $\varphi$  与车速  $V$  的关系

$V(\text{km/h})$	120	100	80	60	50	40	30	20
$\varphi$	0.29	0.3	0.31	0.33	0.35	0.38	0.44	0.44

通过计算车速, 根据表 2 得到  $\varphi$  值。将所有数据代入, 解得基于视频 1 情况下的道路通行能力。

## (2) 实际通行能力的定义

实际通行能力是指在设计或评价某一具体路段时, 根据该设施具体的公路几何构造、交通条件以及交通管理水平, 对不同服务水平下的服务交通量(如基本通行能力或设计通行能力)按实际公路条件、交通条件等进行相应修正后的小时交通量。其计算公式为:

$$C_p = C_B \times N \times f_w \times f_{HV}$$

其中,  $C_p$  为实际通行能力, 单位为  $(\text{pcu/h})$ ;  $N$  为可通过的车道数;  $f_w$  为车道宽度修正系数;  $f_{HV}$  为交通组成修正系数, 其计算公式为:

$$f_{HV} = \frac{1}{1 + \sum P_i (E_i - 1)}$$

其中,  $P_i$  为小型车、中型车和大型车占总的车辆的比例;  $E_i$  为各种车型与车流量之间的折算系数。

## 4、模型的改进——考虑车道通行能力不一致的实际通行能力变化模型

由于在城市道路中, 同方向的车道数一般不止一条, 在多车道条件下, 同向行驶的车辆会发生变道、超车和停车等交通行为。因此车道的通行能力由内侧向外侧依次递减。车道对路段通行能力的影响用修正系数  $\alpha_i$  表示, 依据调查数据我们得到第一、二、三车道的修正系数分别为  $0.85, 0.72, 1.00^{[10]}$ 。

另一方面, 交叉口对路段通行能力也有显著影响, 信号灯控制交叉口上游车辆通过交叉口进入下游路段, 导致路段实际通过的车辆数受交叉口影响显著, 尤其 交叉口间距越小, 这种影响越明显, 交叉口对路段通行能力的修正系数用  $\alpha_{\text{交}}$  表示。其计算公式如下:

$$\alpha_{\text{交}} = \frac{\text{交叉口之间无阻碍的行程时间(s)}}{\text{交叉口之间实际的行程时间(s)}}$$

$$= \frac{l/v'}{\frac{l}{v'} + \beta_1 \frac{v'}{2a} + \beta_2 \frac{v'}{2b} + \Delta}$$

其中,  $l$  为交叉口之间路段的距离;  $v'$  为道路设计车速;  $a$  为启动时平均加速度;  $b$  为制动时平均减速度;  $\Delta$  为车辆的等候时间, 这里我们取  $\Delta=20s$ 。由于本次研究考虑城市道路不同车型混行, 我们取  $a=0.6$ ,  $b=1.5$ <sup>[11]</sup>。

由此, 我们对道路的实际通行能力做出改进, 得到其计算公式如下:

$$C'_p = \sum_{i=1}^3 (C_i \times N \times \alpha_i) \times f_w \times k \times f_{HV} \times \alpha_{交}$$

其中,  $\alpha_i$  为车道对路段通行能力影响的修正系数;  $k$  为车流量, 其计算公式为  $k=n/L$ ,  $n$  为车辆总数,  $L$  为路段长度;  $C_i$  为三种车道的通行能力, 直行车道、专左车道、专右车道的计算公式分别为:

$$C_s = \frac{3600}{\text{信号周期}} \times \left( \frac{\text{相位的绿灯时间} - \text{第一辆车绿灯亮后的启动时间}}{\text{直行车辆经过停车线的平均时间}} + 1 \right) \times \beta_s$$

$$C_l = \frac{\sum C_s}{1 - \beta_l - \beta_r} \times \beta_l$$

$$C_r = \frac{\sum C_s}{1 - \beta_l - \beta_r} \times \beta_r$$

其中, 我们认为信号周期为  $60s$ , 相位的绿灯时间为  $24s$ , 第一辆车绿灯亮后的启动时间为  $2.3s$ , 直行车辆经过停车线的平均时间为  $2.5s$ , 根据附件三我们可得出,  $\beta_s = 0.44$ ,  $\beta_l = 0.21$ ,  $\beta_r = 0.35$ 。由此, 计算改进后的道路实际通行能力。

### 5.2.2 模型的求解

#### (1) 基本通行能力的求解

##### ① 行驶速度的求解最小安全间距的求解

通过随机抽样取得视频 1 中小型车在发生事故之前正常行驶的时间结果如表 3 所示:

表 3 正常行驶的时间

车辆	1	2	3	4	5	6	7	8	9	10
时间/s	25	28	30	27	32	31	25	27	28	26

根据表 3 中小型车在发生事故之前正常行驶的时间, 我们得到其行驶的平均速度为:

$$V = \frac{L}{(T_1 + T_2 + \dots + T_{10})/N}$$

其中,  $L$  为上游路口到事故现场的距离;  $V$  为平均速度;  $T_i$  为车辆的行驶时间;  $N$  为观察的车辆数。根据表 3 中的数据, 我们求车辆的平均速度为  $30.97km/h$ 。

##### ② 最小安全间距的求解

根据求得的平均速度为  $30.97km/h$ , 我们根据表 2, 得到其纵向附着系数  $\varphi$  为  $0.44$ 。将  $V = 30.97km/h$ ,  $\varphi = 0.44$ ,  $l_{安} = 2m$ ,  $l_{车} = 4m$ ,  $t = 1s$  代入, 求得:

$$l_0 = 23.18m$$

由此，我们得到基本通行能力：

$$C_B = 1336.07 \text{pcu/h}$$

### (2) 实际通行能力的求解

根据 $C_p$ 的计算公式，N 为可通行的车道数，事故发生时段 N=1，其他时段

N=3； $f_w$ 为车道宽度修正系数，此处取 0.94<sup>[11]</sup>；我们得到：

$$C_p = C_B \times N \times f_w \times k \times f_{HV}$$

$$= 1336.07 \times N \times 0.94 \times k \times f_{HV}$$

由 $f_{HV}$ 的计算公式，我们以 30s 为一周期将视频 1 中不同时间段下交通组成修正系数代入上式，得到事故所处横断面实际通行能力情况表 4 所示，其中时段 1-7、38-40 为视频 1 中没有发生事故的时段，时段 8-37 为事故发生时段：

表 4 事故所处横断面实际通行能力情况

周期	$C_p$	周期	$C_p$	周期	$C_p$	周期	$C_p$
1	3775.308	11	1231.85	21	1280.077	31	1190.683
2	3409.808	12	1298.351	22	1246.137	32	1216.999
3	2697.097	13	1268.847	23	1265.067	33	1239.996
4	4260.641	14	1240.972	24	1290.454	34	1130.815
5	4285.336	15	1222.101	25	1225.314	35	1147.174
6	3428.832	16	1281.105	26	1176.041	36	1186.564
7	3504.405	17	1304.254	27	1171.066	37	1053.335
8	1034.452	18	1211.885	28	1215.213	38	3212.844
9	1135.767	19	1292.107	29	1153.858	39	3213.752
10	1166.088	20	1278.033	30	1195.039	40	3436.82

将表 4 的数据整理，如图 1 所示：

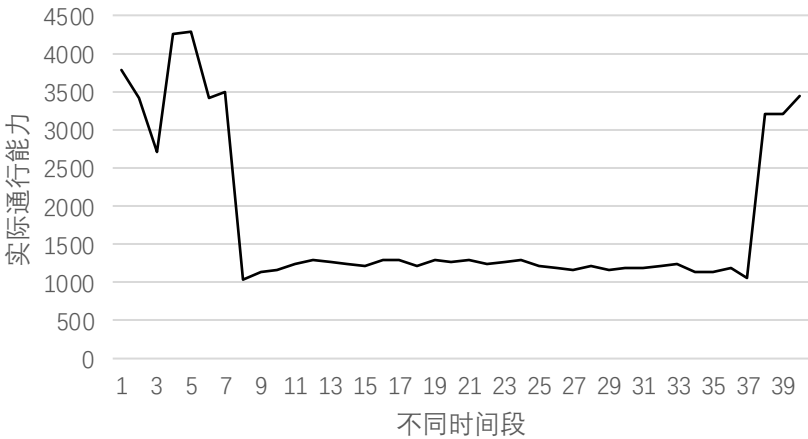


图 1 不同时间段实际通行能力的变化情况

通过图 1 我们发现，在事故发生至撤离的时段内，该路段的实际通行能力是较为平缓的，其实际通行能力都维持在 1000pcu/h 以上，较事故发生之前其通行能力有大幅度的减小。在事故撤离之后，其道路通行能力恢复原水平。由此可见，事故的发生对道路通行能力有较大的影响，其道路通行能力总体上呈先减少，再趋于平缓，最后上升的状态。

### (3) 改进模型的求解

参考国外相关规范及国内相关研究成果，我们首先得到不同速度下对应的

$\alpha_{交}$  的取值如表 5 所示：

表 5 不同速度下对应的 $\alpha_{交}$ 的取值

V(km/h)	60	50	40	30	20
$\alpha_{交}$	0.55	0.52	0.60	0.68	0.64

由此，我们得到 $V = 30.97\text{km/h}$ 的情况下的 $\alpha_{交}$ 为 0.68。

其次，根据三种车道通行能力的计算公式，代入数据得到 $C_s = 229.152$ ， $C_l = 109.368$ ， $C_r = 182.28$ 。

最后，根据之前的交通组成修正系数 $f_{HV}$ ，我们得到改进后道路的实际通行能力如表 6 所示：

表 6 改进后道路的实际通行能力

周期	$C_p$	周期	$C_p$	周期	$C_p$	周期	$C_p$
1	3775.308	11	46.20425	21	65.0275	31	102.9785
2	3409.808	12	53.78117	22	63.75713	32	96.05857
3	2697.097	13	54.75383	23	86.83837	33	93.58398
4	4260.641	14	60.5555	24	85.76144	34	90.28545
5	4285.336	15	54.62801	25	98.27655	35	98.17116
6	3428.832	16	45.60245	26	97.42956	36	109.8598
7	3504.405	17	47.73257	27	95.95123	37	84.86658
8	55.94013	18	66.30747	28	105.1	38	3212.844
9	65.24473	19	70.8144	29	102.945	39	3213.752
10	58.38761	20	75.39522	30	118.6953	40	3436.82

将表 6 的数据进行插值拟合如图 2 所示：

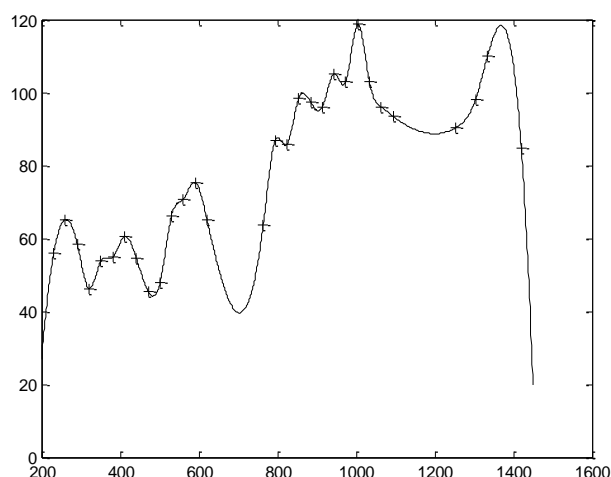


图 2 改进后不同时间段实际通行能力的变化情况

通过图 2 我们发现，模型经过改进后计算得到的不同时间段实际通行能力维持在 20-900pcu/h 之间，其变化的范围较大，且整个事故发生至撤离时段其道路实际通行能力的变化幅度较小。在事故发生之前和事故撤离后的时段，由于受上游路口信号灯的控制，曲线波动的幅度较大，具有一定的随机性。

### 5.2.3 模型的检验

将实际通行能力变化模型和改进的模型插值拟合所得该路段的实际通行能力进行比较，情况如图 3 所示：

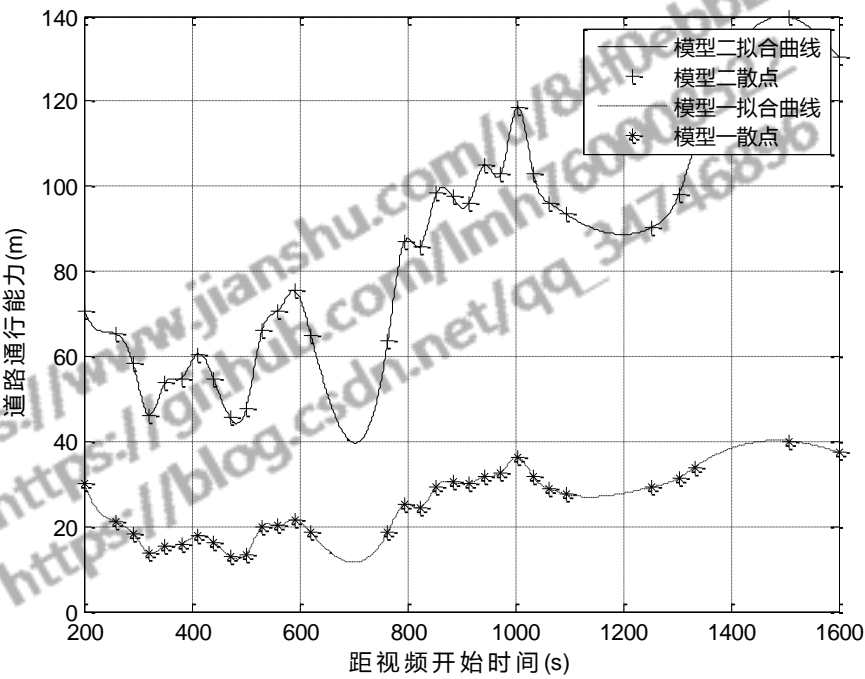


图 3 实际通行能力比较

通过图 3 的对比我们发现，由于模型而考虑了不同车道、上游车流量和交叉路口对该路段实际通行能力的影响，我们发现其实际通行能力较之前所得结果明显偏小。从图中我们可以看出，事故发生后，道路的通行能力有所波动。在事故发生初期，由于车速较稳定，道路的通行能力变化范围较小，大致维持在 10-30pcu/h；事故发生中后期，容易造成交通的拥堵，引发车辆排队的现象导致车道通行能力会有所下降。在事故撤离之后，车道逐渐恢复以往通行能力，呈现平稳态势。而在模型一中，由于没有考虑到影响道路通行能力的外环境，拟合出的通行能力曲线的变化程度较大，较为不稳定。且模型一在事故发生时段的中后期道路通行能力呈波动上升的趋势，并不完全符合实际情况。由此可见，模型二较贴近符合实际，精确度更高。

### 5.2.4 问题小结

本文建立了实际通行能力变化模型和考虑车道通行能力不一致的实际通行能力变化模型，解决了事故所处横断面实际通行能力变化过程的描述问题。首先，对视频内容进行分析并提取车辆通行数据，并剔除错误数据、对缺失数据进行插值拟合；其次，分析道路通行能力的影响因素，并建立实际通行能力变化模型初步判断事故所处横断面实际通行能力变化过程；然后，考虑车道通行能力不一致的因素，建立考虑车道通行能力不一致的实际通行能力变化模型，进一步研究问题。最后，对比两个模型的结果并对模型进行检验。最终得到事故所处横断面道路通行能力总体上呈先减少再平缓最后上升的状态；改进后的模型较贴近符合实际，精确度更高。

### 5.3 问题二的模型建立与求解

分析说明同一横断面交通事故所占车道不同对该横断面实际通行能力影响的差异,首先可以对两个视频中的横断面实际通行能力进行显著性检验,初步判断视频一与视频二中的实际通行能力差异是否显著。通过观察我们得知,两个视频中事故发生的车道不同会引起实际通行能力的变化。因此,我们在问题一的基础上建立车辆变道影响力量化模型,对车辆变道与实际通行能力之间的具体关系进行分析。

#### 5.3.1 模型的建立

##### 1、显著性检验

根据问题一中考虑车道通行能力不一致的实际通行能力变化模型的算法,我们对视频2中的事故发生过程进行处理,得到事故发生时段内事故横截面的实际通行能力的变化。与上题中得到的视频1中的结果进行显著性检验。

(1)显著性分析前提:

利用显著性分析的方法对事故横截面的实际通行能力样本数据进行处理、分析需要满足以下前提:

- ① 参与统计的样本数据必须来源于同一总体;
- ② 样本数据分布基本符合正态分布规律,即样本概率密度函数应满足正态分布函数<sup>[12]</sup>其计算公式如下:

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

其中,  $x$  为样本值;  $f$  为样本在该值出现的概率密度;  $\mu$  为样本均值;  $\sigma$  为样本标准差;  $\mu$  和  $\sigma$  的计算公式如下:

$$\mu = \frac{1}{n} \sum_{i=1}^n X_i$$
$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2$$

##### (2) 非参数检验

样本的非参数检验采用 Kolmogorov-Smirnov 检验(K-S 检验)方法进行<sup>[13]</sup>。在统计学中, Kolmogorov-Smirnov 检验基于累计分布函数,用以检验两个经验分布是否不同或一个经验分布与另一个理想分布是否不同。K-S 检验比较样本数据的累计频数分布和特定理论分布,若两者间的差距小于要求值,则推论该样本取自该分布。单样本 K-S 检验可以将一个样本的实际频数分布与正态分布、均匀分布、泊松分布、指数分布进行比较。样本的非参数检验过程如下:

首先建立假设:

$$H_0: F(x) = F_0(x)$$

其中,  $F_0(x)$  为连续型分布函数,对于正态分布  $F_0(x)$  为:



$$F_0(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

检验真伪的统计量:

$$D_n = \sup_{-\infty < x < \infty} F_n(x) - F_0(x)$$

其中,  $F_n(x)$  为容量是  $n$  的样本的经验分布函数。

$$F_n(x) = \begin{cases} 0 & x < X_{(1)} \\ \frac{i}{n} & X_{(i)} < x < X_{(i+1)} \\ 1 & x < X_{(n)} \end{cases}$$

其中,  $X_{(1)}, X_{(2)}, \dots, X_{(i)}$  是样本  $(x_1, x_2, \dots, x_n)$  的次序统计量。

K-S 检验是在样本的每个次序统计量  $X_{(i)}$  内, 求样本经验分布函数与假设的分布函数之间的偏差中最大的一个, 即:

$$d_i = \max \left\{ F_0(X_{(i)}) - \frac{i-1}{n}, \frac{i}{n} - F_0(X_{(i)}) \right\}, i = 1, 2, \dots, n$$

求得的  $n$  个  $d_i$  中最大的一个, 就是 K-S 检验统计量  $D_n$  的取值。

K-S 检验的检验规则为:

$D_n > D_{(n,\beta)}$  时, 拒绝原假设  $H_0$ ;

$D_n < D_{(n,\beta)}$  时, 接受原假设  $H_0$ 。

其中,  $D_{(n,\beta)}$  为假设检验 K-S 检验的临界值,  $D_{(n,\beta)}$  可通过查表获取, K-S 检验的显著性水平  $\beta$  一般取 0.05。

如样本数据 K-S 检验结果为正态分布, 假设  $\frac{x-\mu}{\sigma} = u$ , 则公式可变为:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

即  $u$  服从标准正态分布即  $u \sim N(0,1)$ 。在这里我们只需要考虑曲线的单边。

参考 GB/T4883-2008<sup>[14]</sup>, 通过对离群值的分析要求, 样本的显著性差异条件确定如下:

$$\frac{x - \mu}{\sigma} > u_{1-\alpha}$$

其中,  $\alpha$  为样本的显著性水平, 一般取 0.01。分布情况如图 4 所示:

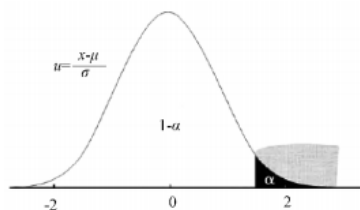


图 4 标准正态分布置信区间图

经查 U 分布表获取在显著水平 $\alpha$ 值下的置信区间上限值为 2.58。

### (3) 显著性检验

如果样本数据全部符合正态分布，那么我们利用 t 检验对道路实际通行能力进行显著性检验。以视频 1、2 分类，分别对第一个和第二个视频的相关数据进行 t 检验。若 t 检验中 Levene 检验的 Sig 值小于 0.01，则我们可以认为两组数据的假设方差不相等。在这种情况下观察均值方程的 t 检验中的假设方差不相等的 Sig 值，若 Sig 值大于 0.05，则我们可以认为两组数据无显著性差异。其他情况下我们则可以认为两组数据存在显著性差异。

#### 2、车辆变道影响力量化模型的建立

在车辆行驶道路上，驾驶员由于受到周边环境信息的刺激会采取变换车道的措施。如本题中的事故完全占用两条车道，则在该车道行驶的车辆必须变道以通过事故发生截面。在视频 1 中，事故发生完全占用第二、三两条车道，因此，原本在第二、三车道上游路口行驶的车辆必须变换到第一车道才能正常通行。由此，我们引入车辆变道系数 $\alpha_{\text{变}}$ ，来说明车辆变道对道路实际通行能力造成的影响。

其计算公式为：

$$\alpha_{\text{变}} = \sum t_i \times p_i$$

其中， $t_i$ 为变道时间； $p_i$ 为车辆数所占百分比。

### 5.3.2 模型的求解

#### 1、显著性差异分析结果

##### (1) 样本校验

利用 K-S 检验方法对样本进行正态分布校验。校验时，一般用 $D_n$ 和 n 反推接受原假设显著性水平 $\beta$ 值。若 $\beta > 0.05$ ，表明样本接受原假设，服从正态分布。将视频中所给数据分为 2 组，分别为视频 1 和视频 2。利用 K-S 检验方法对样本进行正态分布校验，将其校验结果的最大 $\beta$ 值整理如表 7 所示：

表 7 正态分布校验

	视频 1	视频 2
$\beta$ 的最大值	0.091	0.062

通过表 7 我们发现，校验结果的 $\beta$ 值分别为 0.091、0.062，都大于 0.05，即我们认为样本全部服从正态分布。

##### (2) 显著性检验

由于样本数据全部符合正态分布，接着我们利用 t 检验对道路实际通行能力进行显著性检验，结果如表 8 所示：

表 8 显著性检验结果

		方差方程的 Levene 检验		均值方程的 t 检验			
		F	Sig.	t	df	Sig.(双侧)	均值差值
var001	假设方差相等	7.701	.007	2.907	78	.005	849.96778
	假设方差不相等			2.907	68.790	.005	849.96778

通过表 8 我们可以看出，Levene 检验的 Sig 值均大于 0.01，则我们观察其假

设方差相等的 Sig 值。通过观察我们发现，假设方差相等的 Sig 值为 0.05，不大于 0.05，即我们可以认为两段视频道路实际通行能力的差异有显著性差异。

由此，我们也可以初步判断，两段视频道路实际通行能力的差异主要是因为事故发生时所占车道的不同而造成的。

## 2、车辆变道影响力量化模型的求解

首先，我们考虑视频 1 中的情况。设车辆数较少未持续堵车时，车道一标准当量的车通过事故横截面的平均时间为  $t_0$ ；当发生持续堵车时，其他两个车道的车辆要通过变道通过车道一，令  $\Delta t$  表示从车道二变道至车道一使得通行时间的增加值， $\Delta t_1$  表示从车道三变道至车道二使得通行时间的增加值。因此，从车道二经过事故横截面的时间为  $t_0 + \Delta t$ ，

车道三经过事故横截面的时间为  $t_0 + \Delta t + \Delta t_1$ 。设  $t$  时刻总车辆数为  $B$ ，再由附件三中各个车道流量比例可知，所有车辆经过事故横截面的时间为：

$$T_1 = 0.21 \times B \times t_0 + 0.44 \times B \times (t_0 + \Delta t) + 0.35 \times B \times (t_0 + \Delta t + \Delta t_1)$$

其经过事故横截面的时间均值为：

$$\begin{aligned} \bar{T}_1 &= \frac{0.21 \times B \times t_0 + 0.44 \times B \times (t_0 + \Delta t) + 0.35 \times B \times (t_0 + \Delta t + \Delta t_1)}{B} \\ &= 0.21t_0 + 0.44(t_0 + \Delta t) + 0.35(t_0 + \Delta t + \Delta t_1) \end{aligned}$$

为了简化模型，我们可以认为  $\Delta t = \Delta t_1$ ，所以：

$$\bar{T}_1 = t_0 + 1.14\Delta t$$

同理，我们得到视频 2 中车辆经过事故横截面的时间均值为：

$$\bar{T}_2 = t_0 + 0.86\Delta t$$

由此我们可以看出，由于事故发生所占车道的差异，导致正在行驶的车辆会采取变道的措施保证车辆的顺利通过，又因为各个车道流量比例的差异，使得车辆从不同车道变道的时间不同，导致其经过事故横截面的总时间的差异。因此，道路上同一横断面实际通行能力不同的原因是交通事故发生时所占的车道不同。

### 5.3.3 问题小结

本文建立了车辆变道影响力量化模型，解决交通事故所占车道不同对该横断面实际通行能力影响差异的分析问题。首先利用 K-S 检验方法对上游车流量进行正态分布校验；其次对两个视频中的横断面实际通行能力进行显著性检验，初步判断视频一与视频二中的实际通行能力差异是否显著；最后建立车辆变道影响力量化模型，验证了影响差异的主要因素为交通事故所占车道不同。最终得到车道二三被占的影响大于车道一二被占的结论。

### 5.4 问题三的模型建立与求解

分析视频 1 中交通事故所影响的路段车辆排队长度与事故横断面实际通行能力、事故持续时间、路段上游车流量间的关系，是一个多因素相关性分析的问题。车辆排队长度由于受到上游车流量、事故所在路段等众多因素的影响，有较大的随机性，因此本文选择支持向量机进行回归预测，再根据其平均绝对误差的大小判断其精度。

### 5.4.1 模型的建立

支持向量机(SVM)是以统计学理论为基础,基于 VC 维理论和结构风险最小化原理的一项新的数据挖掘技术,并借助最优化方法,非常成功的处理回归问题,即支持向量回归机(SVR)。

道路发生事故情况下车辆排队长度与上游滞留的车辆数、到达车流量、上游路口信号灯控制周期时长、信号灯绿信比、此刻道路实际通行能力甚至排队路段的长度都有很大关系,因此车辆排队长度具有时空特性<sup>[15]</sup>。由于下游交通流方向需求不同,导致上游各车道分配到的车辆数不同,因此各车道排队机率也有区别,这必然导致车辆抢道行驶等随机因素发生,而车道间交换律由于是随机的不易确定。因此车辆排队长度受到多重因素的影响,其数学模型是一个很难确定的复杂的非线性函数。排队长度与影响因素之间很难用一个确定的数学模型来表达,另外由于对于小样本事件模型的参数也难以确定,其预测性能存在各种不足,而 SVR 在解决小样本和高维非线性问题具有特别的优势。

#### 1、支持向量机回归预测原理

支持向量机的回归预测也是针对线性可分和线性不可分的情况进行分析。根据需要嵌入到高维空间来划分。如果对于一个线性函数,能将样本完全的分开,就说这个样本就是线性可分的,否则就是线性不可分。支持向量机的回归预测里主要用到的算法是不敏感函数以及核函数算法。对于线性回归的情况的预测,在开始时,约束条件不可实现时,需要引入变量,并且要把支持向量机要伸到回归的问题基础上的时候,那么就需要引入不敏感函数,根据不敏感函数可以保证某点的预测值是无损失的。

#### 2、支持向量机回归预测的算法

SVM 应用于回归拟合分析时,其基本思想不再是寻找一个最优分类面使得两类样本分开,而是寻找一个最优分类面使得所有样本离该最优分类面的误差最小,如图 5 所示:

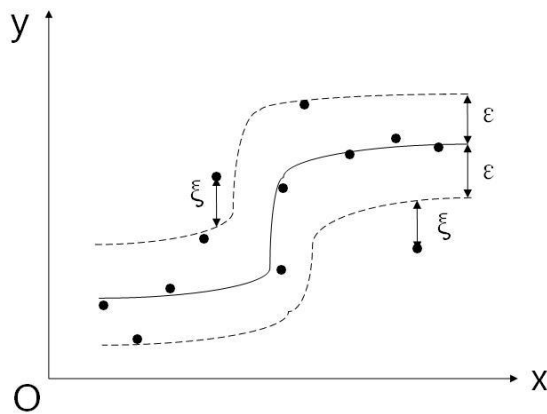


图 5 SVR 基本思想示意图

设含有  $i$  个训练样本的训练集样本对为  $\{(x_i, y_i), i = 1, 2, \dots, l\}$ , 其中,  $x_i (x_i \in R^d)$  是第  $i$  个训练样本的输入列向量,  $x_i = [x_i^1, x_i^2, \dots, x_i^d]^T$ ,  $y_i \in R$  为对应的输出值。

设在高维特征空间中建立的线性回归函数为:

$$f(x) = \omega \Phi(x) + b$$

其中 $\Phi(x)$ 为非线性映射函数。

定义 $\varepsilon$ 线性不敏感损失函数：

$$L(f(x), y, \varepsilon) = \begin{cases} 0 & |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & |y - f(x)| > \varepsilon \end{cases}$$

其中， $f(x)$ 为回归函数返回的预测值； $y$ 为对应的真实值，表示若 $f(x)$ 与 $y$ 之间的差别小于 $\varepsilon$ ，则损失等于 0，如图 6 所示：

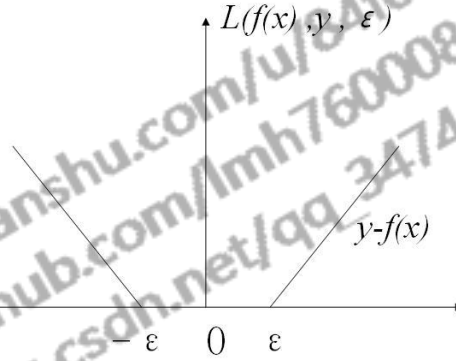


图 6  $\varepsilon$ 线性不敏感损失函数

类似于 SVM 分类情况，引入松弛变量 $\xi_i, \xi_i^*$ ，并将上述寻找 $\omega, b$ 的问题用数学语言描述出来，即：

$$\begin{cases} \min \frac{1}{2} |\omega|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ s. t. \begin{cases} y_i - \omega \Phi(x_i) - b \leq \varepsilon + \xi_i \\ -y_i + \omega \Phi(x_i) + b \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0 \end{cases} \end{cases}$$

其中， $C$ 为惩罚因子， $C$ 越大表示对训练集误差大于 $\varepsilon$ 的样本惩罚越大， $\varepsilon$ 规定了回归系数误差要求， $\varepsilon$ 越小表示回归函数的误差越小。

进行求解时，我们引入 Lagrange 函数，并转换为对偶形式：

$$\begin{cases} \max_{\alpha, \alpha^*} \left[ -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_{i=1}^l (\alpha_i + \alpha_i^*) \varepsilon + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i \right] \\ s. t. \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_i^* \leq C \end{cases} \end{cases}$$

其中， $K(x_i, x_j) = \Phi(x_i) \Phi(x_j)$ 为核函数。经过求解我们得到的最优解为 $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_l], \alpha_i^* = [\alpha_1^*, \alpha_2^*, \dots, \alpha_l^*]$ ，则有：

$$\omega^* = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \Phi(x_i)$$

$$b^* = \frac{1}{N_{nsv}} \left\{ \sum_{0 \leq \alpha_i \leq C} \left[ y_i - \sum_{x_i \in SV} (\alpha_i - \alpha_i^*) K(x_i, x_j) - \varepsilon \right] + \sum_{0 \leq \alpha_i \leq C} \left[ y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) K(x_i, x_j) + \varepsilon \right] \right\}$$

其中， $N_{nsv}$  为支持向量个数。于是，我们得到回归函数为：

$$\begin{aligned} f(x) &= \omega^* \Phi(x) + b^* = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \Phi(x_i) \Phi(x) + b^* \\ &= \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b^* \end{aligned}$$

其中，只有部分参数 $(\alpha_i - \alpha_i^*)$ 不为零，其对应的样本 $x_i$ 即为问题中的支持向量。从上式可以看出，SVR 最终的函数形式与 SVM 相同，其结构与神经网络的结构较为类似，如图 7 所示：

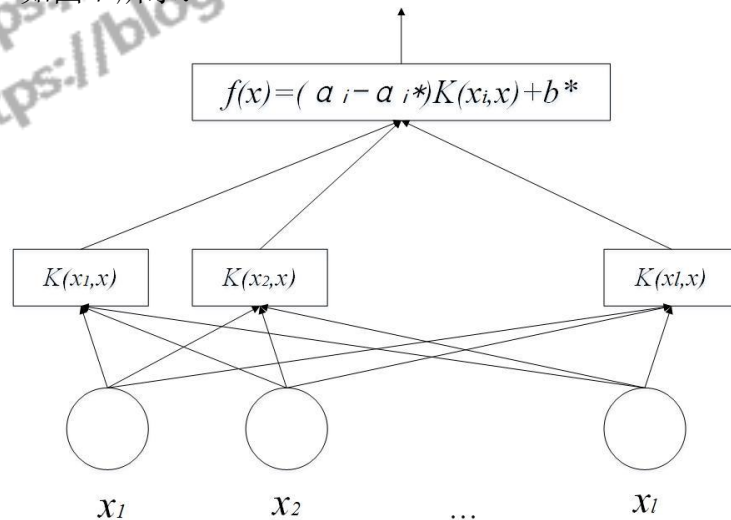


图 7 SVR 的结构

其中，输出是中间节点的线性组合，每个中间节点对应一个支持向量。

### 3、核函数

核函数在线性支持向量机回归算法和非线性核函数支持向量机回归算法中起着重要的作用，它不仅解决非线性问题以及克服维数灾难问题，而且还可以代替高维特征空间中的内积运算，可以避免高纬度运算的复杂性。根据统计学理论，只要对称函数  $K(x, y)$  满足 Mercer 条件就可以作为核函数。在这里我们利用高斯径向基核函数，其基本公式为：

$$K(x, y) = \exp\left(-\frac{|x - y|^2}{\sigma^2}\right)$$

支持向量机的回归预测从线性到非线性转换就是通过核函数，并且核函数就是一种映射，选择不同的核函数会产生不同的结果。

### 4、支持向量机回归的具体步骤

依据问题描述中的要求，实现支持向量机回归模型的建立大致可分为如下步骤：

### (1) 产生训练集/测试集

与 SVM 分类中类似，为了满足相关函数调用格式的要求，产生的训练集和测试集应进行相应的转换。训练集样本的数量及代表性要求与其他方法相同。

### (2) 创建/训练 SVR 回归模型

考虑到归一化、核函数的类型、参数的取值对回归模型的性能影响较大，因此在设计时需要综合衡量。

### (3) 仿真测试

实现 SVR 回归模型的仿真测试，返回的第一个参数为对应的预测值，第二个参数中记录了测试集的均方误差 E 和决定系数  $R^2$ ，具体计算公式如下：

$$E = \frac{1}{l} \sum_{i=1}^l (\hat{y}_i - y_i)^2$$

$$R^2 = \frac{(l \sum_{i=1}^l \hat{y}_i y_i - \sum_{i=1}^l \hat{y}_i \sum_{i=1}^l y_i)^2}{(l \sum_{i=1}^l \hat{y}_i^2 - (\sum_{i=1}^l \hat{y}_i)^2)(l \sum_{i=1}^l y_i^2 - (\sum_{i=1}^l y_i)^2)}$$

其中， $l$  为测试集样本个数， $y_i (i = 1, 2, \dots, l)$  为第  $i$  个样本的真实值； $\hat{y}_i (i = 1, 2, \dots, l)$  为第  $i$  个样本的预测值。

## 5.4.2 模型的求解

### 1、导入数据

根据题目要求，输入列向量  $x_1, x_2, x_3$  分别为事故横断面实际通行能力、事故持续时间、路段上游车流量，输出值  $y$  为路段车辆排队长度。

由此，以 30s 为一个周期统计事故横断面实际通行能力、事故持续时间、路段上游车流量的具体数据，结果如表 9 所示：

表 9 各时段的输入数据

上游车流量	事故持续时间 (s)	通行能力	排队长度
3	90.00	58.38761	0
19	120.00	46.20425	0
0	150.00	53.78117	0
2.8	180.00	54.75383	0
11.4	210.00	60.55555	0
1.7	240.00	54.62801	12
16.8	270.00	45.60245	0
10.4	300.00	47.73257	20
14.4	330.00	66.30747	12
6.8	360.00	70.8144	24
16.8	390.00	75.39522	20
6.4	420.00	65.0275	20
19.4	472.00	63.75713	16
2	502.00	86.83837	36
21.5	532.00	85.76144	16
5.4	562.00	98.27655	44
19.7	592.00	97.42956	32

1	622.00	95.95123	28
19.7	652.00	105.1	40
3.4	682.00	102.945	60
16.8	712.00	118.6953	44
0	742.00	102.9785	52
18.2	772.00	96.05857	40
3	802.00	93.58398	44
11.4	942.00	90.28545	48
13.7	1012.00	98.17116	36
4	1042.00	109.8598	60

通过表 9 发现共 27 个样本。我们将前 20 个样本作为训练集，后 7 个样本作为测试集。为了不失一般性，我们采用随机的方法产生训练集和测试集。

## 2、将训练集和测试机的数据归一化

由于数据集中各个变量的差异较大，不属于同一个数量级，因此，在建立回归模型之前我们对数据进行归一化。

## 3、寻找最佳 c 参数和 g 参数

由于核函数类型及模型参数对模型的性能影响较大，因此需要选取最佳核函数以及参数组合。这里我们采用默认的 RBF 核函数，利用交叉验证方法得到最佳的惩罚因子  $c=32$  和 RBF 核参数中的方差  $g=0.0884$ 。得到回归函数的结果为：

$$y = 0.5725x_1 + 6.1796x_2 + 4.7273x_3 - 0.1693$$

其中， $x_1, x_2, x_3$  分别为事故横断面实际通行能力、事故持续时间、路段上游车流量，输出值  $y$  为路段车辆排队长度。由此我们可以看出，事故持续时间所占权重最大，为 6.1796，即交通事故所影响的路段车辆排队长度与事故持续时间的关系最为密切，与路段上游车流量关系较密切，与事故横断面实际通行能力的相关性较小。

车辆排队长度的真实值与预测值的对比如图 8 所示：

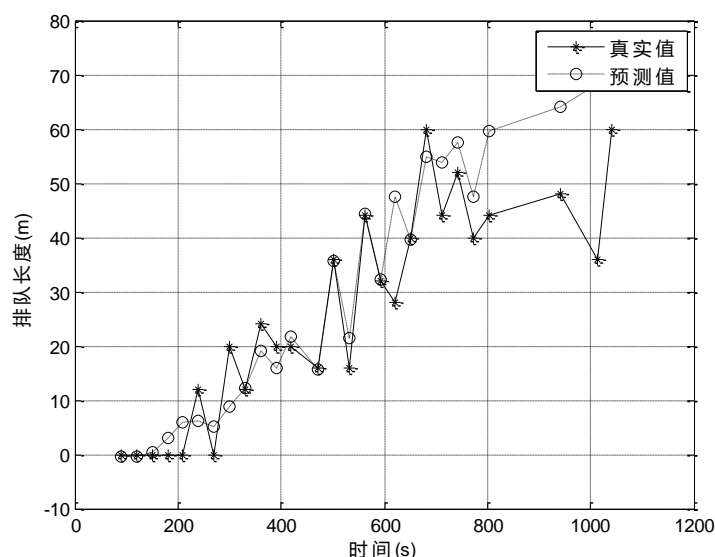


图 8 车辆排队长度真实值与预测值

通过图 8 我们发现，与真实值相比，我们预测出的排队长度的波动程度较平缓，且预测值与真实值都在时段内呈波动上升趋势。在事故发生初始时段，预测



值与真实值的误差较小；当事故发生一段时间之后，两者的差值较大且真实值逐渐低于预测值。产生这种现象的可能原因是道路上游的车辆受交叉路口信号灯的影响较大，绿灯时段车流量较多导致堵车更为严重。且由于上游车速较快而事故发生路段行车速度较慢，排队长度极有可能随着时间的增加而缓慢增长。

### 5.4.3 模型的检验

利用函数返回的均方误差  $E$  和决定系数  $R^2$ ，可以对所建立的 SVR 回归模型的性能进行评价。若性能没有达到要求，则可以通过修改模型参数、核函数类型等方法重新建立回归模型，直到满足要求为止。

为了综合评价模型泛化性能的优劣，我们选用平均绝对误差 MAE 评价指标进行评价，其计算公式为：

$$MAE = \frac{1}{n} \sum_{i=1}^n |z_i - z_i^*|$$

MAE 表示预测值与实测值的误差程度。

对数据集进行仿真预测，我们最后得出 MAE 的值为 0.5873，其训练集和测试集的均方误差分别为 0.0400 和 0.3143，决定系数分别达到 0.8773 和 0.2388，这表明所建立的 SVR 回归模型具有非常好的泛化能力。

### 5.4.4 问题小结

本文建立了基于交叉验证法的支持向量机回归分析模型，解决了分析排队长度与事故横断面实际通行能力、事故持续时间、路段上游车流量间的关系问题；首先引入支持向量机回归分析的原理和算法步骤；其次利用交叉验证法确定参数并选择合适的核函数；然后将统计数据的前 20 组作为训练集，后 7 组作为测试集创建并训练 SVR 回归模型；最后进行仿真测试并计算均方误差检验模型的准确性。最终得到交通事故所影响的路段车辆排队长度与事故持续时间的关系最为密切，与路段上游车流量关系较密切，与事故横断面实际通行能力的相关性较小。模型具有非常好的泛化能力。

## 5.5 问题四的模型建立与求解

为了合理的预测车辆排队长度到达上游路口的时间，首先要求得车辆排队长度与其他各影响因素的大致关系。其次，通过 SVM 动态学习过程对样本的数据进行训练学习，并随着样本的增加实时更新新的学习内容。最后，通过对预测数据进行反归一化，得到预测结果。

### 5.5.1 模型的建立

估算从事故发生开始时，车辆排队长度到达上游路口的时间在本质上是要求我们对车辆排队长度的变化进行分析并预测。而车辆排队长度受到多种因素的影响，如车辆的速度、信号灯时间周期的长短、上游的车流量等等，是一个动态的过程。因此，为了动态跟踪和预测系统的特征参数，当系统产生新的数据时，需要将此新信息加入系统的预测模型，并相应地改变模型参数。但是，随着数据的增加，

核函数的维度也随之增加,最终耗尽系统资源。因此,SVM 的动态训练和学习过程应包括增加样本和减少样本两个过程,利用样本增减迭代方法进行。

设新增加一个样本 $(x_{N+1}, y_{N+1})$ ,并将其加入原训练样本集 $\{x_k, y_k\}_{k=1}^N$ ,则样本为 N 的现有模型与样本为 N+1 的新模型的增量关系为:

$$A_{N+1}\alpha_{N+1} = Y_{N+1}$$

其中,

$$\begin{aligned} A_{N+1} &= \begin{bmatrix} A_N & \beta \\ \beta^T & c' \end{bmatrix}, A_N = \begin{bmatrix} 0 & 1^T \\ 1 & \Omega + c^{-1}I \end{bmatrix} \\ \alpha_{N+1} &= \begin{bmatrix} \alpha_N \\ \alpha_{N+1} \end{bmatrix}, Y_{N+1} = \begin{bmatrix} Y_N \\ y_{N+1} \end{bmatrix}, Y_N = \begin{bmatrix} 0 \\ Y \end{bmatrix} \\ \beta &= [1, K(x_1, x_{N+1}), \dots, K(x_N, x_{N+1})] \\ c' &= c^{-1} + K(x_{N+1}, x_{N+1}) \end{aligned}$$

在线增量训练算法的目标是有效地更新逆矩阵 $A_{N+1}^{-1}$ ,根据文献<sup>[16]</sup>,我们得到:

$$A_{N+1}^{-1} = \begin{bmatrix} c' - \beta^T A_N^{-1} \beta \\ -\beta^T A_N^{-1} \end{bmatrix}^{-1} \begin{bmatrix} A_N^{-1} \beta \\ -1 \end{bmatrix} \times \begin{bmatrix} \beta^T A_N^{-1} - 1 \\ 0 \end{bmatrix} + \begin{bmatrix} A_N^{-1} & 0^T \\ 0 & 0 \end{bmatrix}$$

随着新样本的持续增加,计算出来的逆矩阵 $A^{-1}$ 的维数也不断增加。为了避免这些问题的发生,在增加样本之后,必须进行相应的减小样本过程。

为了避免传统样本减少,在此过程中重新计算矩阵的逆,并将 $A_{N+1}^{-1}$ 更新为逆矩阵 $A_N^{-1}$ ,可将下式作为更新规则来剪除N+1个样本中的第k个样本:

$$\alpha_{ij} \leftarrow \alpha_{ij} - \alpha_{kk}^{-1} \alpha_{ik} \alpha_{kj}$$

其中,  $i, j = 1, 2, \dots, N$  且  $i, j \neq k$ ;  $\alpha_{ij}$  表示  $A_{N+1}^{-1}$  中第 i 行第 j 列的元素; k 表示第 k 个支持向量被移除。SVM 的动态学习过程如图 8 所示:

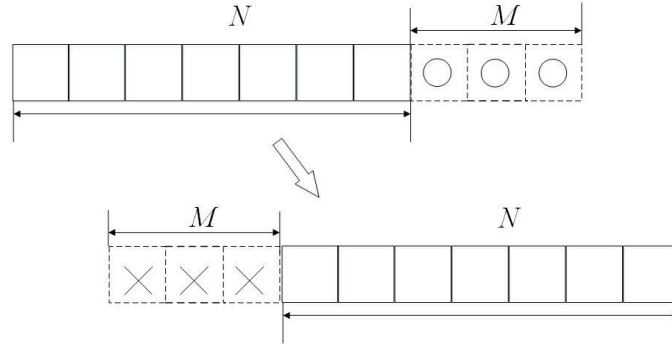


图 9 SVM 动态学习过程

动态 SVM 模型的建立过程如下:

- (1) 对训练数据进行归一化,使数据的均值为 0,方差为 1,  $R$  为归一化后训练集的变化范围之和;
- (2) 通过求解得到优化参数  $\omega$  和  $b$ 、 $c$  和  $g$ ;
- (3) 当有新的样本数据输入时,以样本增加前求得的参数  $c$  为起始点,求得新样本下的  $c$ , 完成更新;
- (4) 将新样本值和更新后的  $c$  代入计算,计算增加样本后的逆矩阵  $A_{N+1}^{-1}$ ,通过计算求得减少样本后的逆矩阵  $A_N^{-1}$ ,以完成 SVM 的增减学习,并计算新的  $\alpha$  和  $b$  值,训练出一个新的 SVM;
- (5) 当新样本数据持续输入时,重复(3)和(4),动态调整 SVM 的模型

参数,完成 SVM 的动态训练和动态预测。通过对预测数据进行反归一化,得到预测结果。

### 5.5.2 模型的求解

运用构建的动态 SVM 模型对问题四的情景进行模拟和求解。首先增加车祸横断面距上游路口的距离这一因素条件,其次对确定最优参数并用第三问的数据对支持向量机进行训练。然后将横断面至上游路口距离更改为 140m,由于交通信号灯的周期不变,因此取时  $t=30$ ,即每隔三十秒进行一次预测;上游路口车流量在 25pcu/min 的基础上,绿灯时到达 xpcu,根据视频观察,此时下  $x \in [24,25]$ ,红灯时到达  $(25-x)$ pcu;考虑到小区路口由两个变为一个,因此从小区出行车辆大概每分钟 1-2 辆,由此下  $x \in [23,27]$ ;

算法步骤为:

步骤一:导入增加车祸横断面距上游路口的距离因素的数据集 A,并将数据归一化;

步骤二:确定参数。采用默认的 RBF 核函数,利用交叉验证方法得到最佳的惩罚因子  $c=32$  和 RBF 核参数中的方差  $g=0.0884$ 。

步骤三:得到下一个时间结点的输入数据,并预测出车辆排队长度 L;

步骤四:若  $L < 140$ ,则将产生的归一化数据加入历史数据集 A,  $t=t+30$ ,并返回步骤二;否则跳出循环。

最终我们得到,经过 4.5min 后,车辆排队长达 155.71 米。

### 5.5.3 模型的检验

利用问题三的方法,每更新一次参数值,就对已知数据进行仿真预测计算均方误差;最终得均方误差如表 10 所示:

表 10 仿真预测均方误差和决定系数

	30	60	90	120	150	180	210	240
均方误差	0.05	0.14	0.08	0.09	0.12	0.10	0.07	0.08
决定系数	0.876	0.955	0.904	0.912	0.949	0.972	0.932	0.938

表 10 表明所建立的 SVR 回归模型具有非常好的泛化能力,适合解决此问题,由此方法得到的结果较为可信。

### 5.5.4 问题小结

本文建立了基于支持向量机的动态预测模型,解决了车辆排队长度到达上游路口的时间预测问题。首先增加考虑车祸横断面距上游路口的距离因素,重新创建训练集和测试集;其次采用默认的 RBF 核函数并得到最佳的惩罚因子和核参数中的方差;然后对车辆排队长度进行预测并预测数据加入历史数据集重新训练支持向量机;最后对模型进行检验。最终得到经过 4.5min 后,车辆排队长达 155.71 米;模型预测结果较为可信的结论。

## 六、模型的评价与推广

### 6.1 模型的优点与缺点

模型二的优势在于动态全面的考虑了外在条件对于道路通行能力的影响,这得益于车道宽度修正系数、交通组成修正系数等的引入,使得该模型评价能够全面、系统和实际的反映道路的运行状况,从而为规划决策者提供数据支持。模型三的 svm 回归预测是把线性的回归转为非线性,需要将内机核函数转换为高维空间的非线性映射,计算的复杂性取决于支持向量机的数目,不是样本空间的维数。所以,从一定层面上避免了维数灾难的问题

本文的缺点在于模型的建立过程中并没有考虑天气、地形地貌等现实情况,不能完全贴近实际生活;也没有考虑到其他时段道路的交通需求与通行能力的变化,对于道路的全年的运行情况不能做全面系统的评价。

### 6.2 模型的推广

问题三中采用的支持向量机的回归预测近年来在很多的方面得到应用,并且有很多的预测方法值得去探讨。例如在疾病预测、天气预测、市场预测、股价预测、实时业务预报等方面都有广泛的应用。这些让我们看到它的前景的可观。而本文建立的基于车辆排队长度的模型可以推广到医院挂号排队长度对医院服务质量、患者满意度的影响;或者银行排队长度对银行办事效率的影响等等。

### 参考文献

- [1]王璨.冯炜.城市道路路段通行能力计算方法探讨[J].华东公路.2016(1):109-110
- [2]丁宏飞.刘博.秦政.李演洪. 城市快速路交通事件影响范围研究[J].中国安全科学学报.2016(10):152-153
- [3]许伦辉.倪艳明.罗强.黄艳国. 基于最小安全距离的车辆换道模型研究[J].广西师范大学学报(自然科学版).2011(4):2-4
- [4]华雪东.王伟.王昊. 考虑自适应巡航车辆影响的上匝道系统混合交通流模型[J].物理学报.2016(8):1-4
- [5]孙培.刘凯.杨本朝. 基于排队论的城市交通问题分析[J]. 黑龙江交通科技.2015(11):173-174
- [6]茹红蕾.城市道路通行能力的影响因素研究[D].上海:同济大学,2008
- [7] CJJ37-2012.城市道路工程设计规范[S].北京:人民交通出版社, 2012
- [8]杨哲铭.刘君蒙.赵聪.王玲. 车道被占用对城市道路通行能力的影响[J]. 华北理工大学学报(自然科学版).2017(39):76-77
- [9]王法.吉岩.李力.胡坚明.张毅.靳雪翔. 基于马尔科夫链间距模型的静态车队车距分布解释[D].北京:清华信息科学与技术国家实验室, 2015
- [10]陈宽民.严宝杰.道路通行能力分析[M].人民交通出版,2011
- [11]王璨.冯炜.城市道路路段通行能力计算方法探讨[J].华东公路.2016(1):110-111
- [12]E.Bartezzaghi,R.Verganti.Managing demand uncertainty through order overplanning,International Journal of Production Economics,2015(40):107-120
- [13]DeJong R M,Amsler C,Schmidt P.A Robust Version of the KPSS Test Based on

Indicators. Journal of Econometrics, 2014, 127: 311-333

[14] GB/T 4883-2008, 数据的统计处理和解释 正态样本离群值的判断和处理[S]

[15] 顾燕萍, 赵文杰, 吴占松. 最小二乘支持向量机鲁棒回归算法研究[J]. 清华大学学报(自然科学版). 2015(4): 397-399

[16] 潘宇雄, 任章, 李清东. 动态贝叶斯最小二乘支持向量机[J]. 控制与决策. 2014(12): 2298-2299

大美mixer  
简书>>

Github>>

CSDN>>

<https://www.jianshu.com/u/84f0ebb22c87>  
<https://github.com/lmh760008522>  
[https://blog.csdn.net/qq\\_34746896](https://blog.csdn.net/qq_34746896)

## 附录

### 第一问附录

```
a=load('C:\Users\美含\Desktop\shumo\13\支撑材料\1data.txt');
b=a(:,1)';%时间, 自变量
b0=200:1:1600;
c=a(:,2)';%C, 因变量,改进后
d=a(:,3)';%改进前
pp1=spline(b,c,b0);%改进
pp2=spline(b,d,b0);%改进前
plot(b0,pp1,'k-',b,c,'k+',b0,pp2,'k-',b,d,'k*');
grid on
h=legend('模型二拟合曲线','模型二散点','模型一拟合曲线','模型一散点')
set(h,'FontSize',8)
xlabel('距视频开始时间(s)')
ylabel('道路通行能力(m)')
```

### 第三问附录

```
clear all
%load concrete_data.mat
attributes=load('C:\Users\ÃÃ\Desktop\shumo\13\testData.txt');
%n = randperm(size(attributes,2));
n=size(attributes,2);
time=attributes(3,:);
num_train=20;
num_x=4;
p_train = attributes(1:num_x,[1:num_train]);
t_train = attributes(num_x+1:end,[1:num_train]);
p_test = attributes(1:num_x,[num_train+1:end]);
t_test = attributes(num_x+1:end,[num_train+1:end]);
x=load('C:\Users\ÃÃ\Desktop\shumo\13\testYC.txt');
[pn_train,inputps] = mapminmax(p_train);
pn_train = pn_train';
pn_test = mapminmax('apply',p_test,inputps);
pn_test = pn_test';
[tn_train,outputps] = mapminmax(t_train);
tn_train = tn_train';
tn_test = mapminmax('apply',t_test,outputps);
tn_test = tn_test';

[xn,predict]=mapminmax(x);
xn=xn';
yn=zeros(1,size(x,2));
```

```

[c,g] = meshgrid(-10:0.5:10,-10:0.5:10);
[m,n] = size(c);
cg = zeros(m,n);
eps = 10^(-4);
v = 5;
bestc = 0;
bestg = 0;
error = Inf;
for i = 1:m
    for j = 1:n
        cmd = ['-v ',num2str(v),' -t 2',' -c ',num2str(2^c(i,j)),' -g ',num2str(2^g(i,j)),' -s 3 -p 0.1'];
        cg(i,j) = svmtrain(tn_train,pn_train,cmd);
        if cg(i,j) < error
            error = cg(i,j);
            bestc = 2^c(i,j);
            bestg = 2^g(i,j);
        end
        if abs(cg(i,j) - error) <= eps && bestc > 2^c(i,j)
            error = cg(i,j);
            bestc = 2^c(i,j);
            bestg = 2^g(i,j);
        end
    end
end
cmd = ['-t 2',' -c ',num2str(bestc),' -g ',num2str(bestg),' -s 3 -p 0.1'];
model = svmtrain(tn_train,pn_train,cmd);
w=model.SVs'*model.sv_coef
b=-model.rho

[Predict_1,error_1] = svmpredict(tn_train,pn_train,model);
[Predict_2,error_2] = svmpredict(tn_test,pn_test,model);
%[Predict_3,error_3] = svmpredict(yn,xn,model)
predict_1 = mapminmax('reverse',Predict_1,outputps)
predict_2 = mapminmax('reverse',Predict_2,outputps)
%predict_3 = mapminmax('reverse',Predict_3,outputps)
result_1 = [t_train predict_1]
result_2 = [t_test predict_2]
result=[result_1;result_2];
%result_3 = [xn predict_3]
(error_1+error_2)/2
%error_3
figure(1)
plot(time,[t_train;t_test],'k-*',time,[predict_1;predict_2],'ko');
grid on

```

```

legend('ÖæÊµÖµ','ÔæÖµ')
xlabel('Ê±/4 ä(s)')
ylabel('Å Å ¶Ó æ¶Ê(m)')

```

#### 第四问附录

```

%% 4Å Å ¶Ó æ¶ÊÖæ
yn=0;% × ðÅ Å ¶Ó æ¶ÊÛÃ ×
t=30;
while yn<140

    % ÖÖÐÂÑµÁ ·ðÁ ¿»ú
    if t~=30
        x0=[x;yn];
        attributes=[attributes x0];
        p_train=attributes(1:num_x,:);
        t_train=attributes(end,:);
        [t_train0,outputs] = mapminmax(t_train');
        % Ñ ÖÖ×¼Ñc ²Êý/g ²Êý
        [c,g] = meshgrid(-10:0.5:10,-10:0.5:10);
        [m,n] = size(c);
        cg = zeros(m,n);
        eps = 10^(-4);
        v = 5;
        bestc = 0;
        bestg = 0;
        error = Inf;
        for i = 1:m
            for j = 1:n
                cmd = ['-v ',num2str(v),' -t 2',' -c ',num2str(2^c(i,j)),' -g ',num2str(2^g(i,j)) ,',' -s 3 -p
0.1'];

                cg(i,j) = svmtrain(t_train,p_train,cmd);
                if cg(i,j) < error
                    error = cg(i,j);
                    bestc = 2^c(i,j);
                    bestg = 2^g(i,j);
                end
                if abs(cg(i,j) - error) <= eps && bestc > 2^c(i,j)
                    error = cg(i,j);
                    bestc = 2^c(i,j);
                    bestg = 2^g(i,j);
                end
            end
        end
        end
        % ½ ÑµÁ SVM

```



```

cmd = ['-t 2','-c ',num2str(bestc),'-g ',num2str(bestg),'-s 3 -p 0.01'];
model = svmtrain(t_train,p_train,cmd);
end

% 交叉验证
if mod(t,60)==0 % 每隔60次交叉验证
    x=[120;randi([1 5],1,1);t;1287.88632*0.94*(1+0.1*randn)*(0.12+0.05*randn)*0.68*0.9];
else % 每隔60次交叉验证
    x=[120;randi([23 27],1,1);t;1287.88632*0.94*(1+0.1*randn)*(0.12+0.05*randn)*0.68*0.9];
end

attribute=attributes([1:num_x,:]);
attribute=[attribute;x'];
[train,puts]=mapminmax(attribute);
train=train';
[Predict_3,error_3] = svmpredict(yn,train(end,:),model);
yn = mapminmax('reverse',Predict_3,outputps);
t=t+30;
yn=Predict_3;
end

```