文本复制检测报告单(全文标明引文)

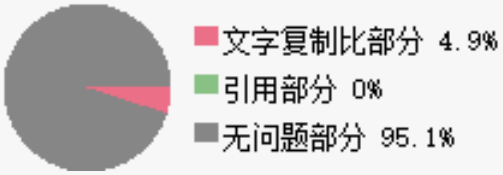
№:ADBD2018R_201109282223162018032515281683587886958 5　　检测时间:2018-03-25 15:28:16

检测文献:　90728733015288505_张金飞_城市交通路口短时流量预测2
作者:　张金飞
检测范围:　中国学术期刊网络出版总库
中国博士学位论文全文数据库/中国优秀硕士学位论文全文数据库
中国重要会议论文全文数据库
中国重要报纸全文数据库
中国专利全文数据库
互联网资源(包含贴吧等论坛资源)
英文数据库(涵盖期刊、博硕、会议的英文数据以及德国Springer、英国Taylor&Francis 期刊数据库等)
港澳台学术文献库
优先出版文献库
互联网文档资源
图书资源
CNKI大成编客-原创作品库
个人比对库
时间范围:　1900-01-01至2018-03-25

检测结果

总文字复制比：4.9%　　跨语言检测结果：0%
去除引用文献复制比：4.9%　　去除本人已发表文献复制比：4.9%
单篇最大文字复制比：3.5%（基于支持向量回归的短时交通流预测方法研究与应用）

重复字数：	[537]	总段落数：	[1]
总字数：	[10853]	疑似段落数：	[1]
单篇最大重复字数：	[384]	前部重合字数：	[142]
疑似段落最大重合字数：	[537]	后部重合字数：	[395]
疑似段落最小重合字数：	[537]		



指　标：☐疑似剽窃观点　☑疑似剽窃文字表述　☐疑似自我剽窃　☐一稿多投　☐疑似整体剽窃　☐过度引用　☐重复发表
表格：0　　脚注与尾注：0

（注释：无问题部分　文字复制比部分　引用部分）

1. 90728733015288505_张金飞_城市交通路口短时流量预测2　　总字数：10853

相似文献列表　文字复制比：4.9%(537)　疑似剽窃观点：(0)

1	基于支持向量回归的短时交通流预测方法研究与应用 武琼(导师：王夏黎) -《长安大学硕士论文》- 2016-04-28	3.5%（384） 是否引证：否
2	基于支持向量机的高速公路物流量预测研究 杨健(导师：杨维平) -《昆明理工大学硕士论文》- 2017-04-01	0.7%（79） 是否引证：否
3	基于高斯过程回归的锂电池数据处理 叶婧(导师：张三同) -《北京交通大学硕士论文》- 2016-04-02	0.5%（58） 是否引证：否

原文内容　红色文字表示存在文字复制现象的内容; 绿色文字表示其中标明了引用的内容

交通流数据分析与预处理

根据本文所要研究的短时交通预测的内容是以交通流作为研究对象，因此实验所需要的数据中应包含这一数据量，同时对于交通流的影响因素如速度、占有率、天气状况、突发事件等等这些数据都是短时交通预测所需要的数据，同时本文是针对城市交通进行短时交通流预测研究，因此为满足以上实验条件以及最终获取到的合肥市一交通路段上的交通数据，数据中主要有

流量、速度以及占有率等数据，数据特征虽然不是面面俱到，但是已经可以基本满足实验所需。但在获取到的交通数据当中必不可免的会有一些异常数据，这些异常数据的来源可能是由于检测器在某个阶段产生了故障或者失灵也可能是在数据传输的过程当中碰巧遇到设备的故障等等原因，因此对原始交通数据不能直接用于预测，必须首先对数据进行预处理工作。

3.1交通流基本理论

道路交通主要是由道路上形形色色连续流动的人流和车流所构成的，行人以及车辆在路面上不断行驶就会呈现一定的特征出来，交通流就是由车流和人流所共同形成的。然而在本文中研究的“主人公”仅仅是车辆形成的车流，因此在文章的后续研究中都是将车流作为交通流量，交通流状态就是指连续不断的交通流运行所表现出来的特性。

不同类型的车辆、不同的旅游目的地以及不同的出行目的都会促使交通流发生不同程度的改变；而且天气条件、道路状况、交通情况都会对交通流产生很大改变，所以根本不好用很清晰恰当的物理量来表达这都是因为交通流的混乱特性造成的。

虽然是有这些情况存在，经过不断地研究分析通过大量的实验观察，还是会发现在一条件下交通流的状态变化也并不是无迹可寻还是有一定的规律存在。实际情况中我们是可以来用我们所熟知的交通流状态的特征参数去表示存在的这些规律和交通特征。对于交通流参数我们是可以根据研究和观察分析的不同对象分别从宏观和微观这两个方面来对其进行表示。在研究对象仅仅就是单一的车辆主体的时候，我们使用微观的参数进行表示，微观参数包括像车头间距以及车头时距等[36]。而当我们将交通流这样个整体作为我们的研究分析对象的时候，肯定是从宏观的角度用宏观相关参数进行表示，这些参数是我们比较熟悉的例如：速度、密度、交通流量、占有率及排队长度等，这当中都是将密度、速度、交通流量作为交通流三要素[37]。

在本文后面的实验分析当中都是从宏观的层面来进行研究分析的，在这章下面主要介绍流量、速度、占有率这三个基本参数。

（1）流量---也叫交通量。它指的是在一个时间单位内通过某个路面断面的车辆数量，而且会由于时间以及地点的不一样变化也会非常大。短时交通流通常指的是在15min内道路断面上通行的车辆总数。在本文的研究中选取的对象即为流量，选取的时间间隔设定在10min,也就是每隔10min道路上某个断面路口总共通过的车辆数目。

（2）速度。它指的是研究对象车辆在一个时间单位内所行驶过的距离。事实上的交通流系统结构纷繁复杂，这其中不但包含许许多多的车辆，而且不同的车辆之间的特征差异也是很大的，故而我们所说的交通流的速度其实是指所有车辆的一个平均速度。时间平均速度（Time Mean Speed，TMS）以及区间平均速度（Space Mean Speed，SMS）都是包括在交通流的平均速度之中。

在实验的观测时间段内，路过所在实验路段断面内的所有车辆的瞬时速度的平均值即为时间平均速度，详细计算公式如（3-1）所示：

式（3-1）中，即为时间平均速度，n代表实验时间道路断面内观测到的所有车辆数目。

在实验进行的观察时间区间里面，所观察道路路段的长度与行驶过此路段的所有车辆平均行驶时间的一个比值即为区间平均速度，详细计算公式如（3-2）所示：

式（3-2）中，即为区间平均速度，L代表观察路段的长度，n观察路段断面上行驶的车辆数目，表示的是第i辆车辆行驶通过观察路段所花费的时间。

（3）密度。从理论上讲，密度指的是在单位长度上的道路路面中瞬间行驶过的机动车的数量。不过事实上，在实际情形下很难再瞬间去测量通过的车辆数目，所以退而求其次，就用车辆占有率来取代密度表示，占有率包括时间占有率和空间占有率[38]。

空间占有率表示的是在某一时间段内，直白的说就是观察的道路断面上行驶过的所有机动车车辆的长度总和除上这段断面的长度得到的值。空间占有率可以更清楚的显示道路断面上的占用状况，事实上要想求得所有机动车的长度总和还是个比较困难的事情。空间占有率的详细计算公式如（3-3）所示：

式（3-3）中，即为空间占有率，L表示所观察的道路断面距离长度，n表示观察路段上行驶过的机动车车辆总数，代表的是第i辆车的长度。

时间占有率通常指的是机动车辆在某一段时间中所占有车辆检测器的时间总和与用于检测的总时间的一个比值。如果检测断面上的机动车辆行驶的比较快则占用的检测器时间必然会很少，那么时间占有率就会相对小，这样就代表此路段不会拥堵；如果断面上的车流量增加，很显然检测器就会被机动车辆占有很长的时间，则相应的时间占有率定会变大，此时代表着此路段比较拥堵。时间占有率的详细计算公式如（2-4）所示：

式（3-4）中，即为时间占有率，T表示检测器的检测总时间，n代表检测时段内行驶过的机动车辆总数目，代表着第i辆车行驶过检测器的时间大小。

3.2实验数据来源

论文的仿真数据来源于交通运输部出行云中的合肥市示范区黄科路口相关数据。本文选取的数据为其中的微波检测数据。微波检测数据表名为DT_LANE_REPORT_H，包括编号、设备类型、设备编号、采集时间、时间占有率等多个字段信息，选取的时间为2016年6月30日至2016年7月1日共两天的历史数据，数据详情请见下表。

3.3交通流数据预处理

在本文实验中所采用的交通数据源最初是通过微波检测器收集而来的，而事实上在真正的工程应用里面，会因为例如交通数据采集设备的一些固有缺陷或者是在数据传输过程中传输设备发生的故障以及其他外界因素的干扰等等不同的原因，往往就导致了这些检测器收集到的原始交通数据不是有很高的精度问题。但是在这些检测器上收集到的原始的交通原数据是进行短时

交通预测模型里面的一个数据基础，因此它对这个模型的可信度和有效度是直接产生了不可低估的影响作用。这样的话对于从检测器收集到的第一手交通数据存在着质量问题是不能直接放到模型中使用的。那么既然如此，理所当然的必须对从检测器收集到的交通数据进行必要手段的预处理工作，将预处理之后的交通数据才能导入到最后的模型去试验。因而对收集起来的交通数据进行预处理工作是进行建模预测的第一步，这一步必不可少，是整个过程中很重要的一环。

在获取到的交通数据当中必不可免的会有一些异常数据，这些异常数据的来源可能是由于检测器在某个阶段产生了故障或者失灵也可能是在数据传输的过程当中碰巧遇到设备的故障等等原因。异常数据可以将其简单分为数据缺失和数据错误两类。

（1）对异常数据中的数据缺失的处理方法。

本文使用的数据来源是通过微波检测器采集得到的，它的采集时间间隔也是固定的，所以本文的数据采集间隔为1min，这样从理论上讲，一天24小时会采集到1440条交通流数据，然而实际得到的交通流数据因为数据的赘余或者缺失等原因，一个检测器在一天收集到的真正的数据量是围绕在1440这个数字上下波动的。对解决数据缺失的常用到的方法像有加权平均、历史平均法等等[39]。在本文使用到的处理方法是加权平均，具体实施步骤如下：

第一步：拿到缺失数据前一天而且是同一时间点的历史数据；

第二步：拿到缺失数据在当前时间点的前一时刻的实测值，然后将、使用加权平均法，即可恢复得到缺失数据，详细公式如（3-5）：

式（3-5）中，代表第k天t时间点的修复得到的修正值，代表加权因子，表示t-1时间点的实际数据及历史数据在丢失时间点数据进行修正的时候所拥有的影响的比重大小，第k天t-1时间点的实际数值用表带，第k-1天t时间点的历史数据由表带。

（2）对交通数据中产生错误数据的处理方法。

理论上来说，我们是可以进行定义一个合适的阈值，通过这个阈值来筛选交通流数据中的错误数据，因为在一定的时间区间内的各个交通流参数的取值会分布在一个合理的范围中，那么就可以把落在阈值范围外的数据定性为错误数据。而且我们在已有的交通流理论和大量的实践基础上可以给这些交通流参数的大概取值范围给确定下来，流量、速度、占有率这些交通流参数的合适阈值范围设定如下：

a.流量

通过交通路口上的微波检测器收集到的交通数据中交通流量q折算后的合理范围如式（3-6）：

式（3-6）中：

C代表道路路段上每小时所能通行的机动车数即路段的通行能力（veh/h）；

T代表微波检测器收集交通数据的时间间隔，本文中为1min;

代表对交通流量的一个修正权重，范围设定在1.3～1.5。

b.平均速度

平均速度v也是通过路口微波检测器采集所得，范围设定如式（3-7）：

式（3-7）中：

—代表设计速度,本文中取80km/h;

—代表修正的权重值，范围设在1.3～1.5。

c.占有率

在路口中微波检测器收集到的占有率数据为时间占有率Occ,其范围设定：

3.4交通数据归一化

在本文中，所做的归一化其实就是将通过微波检测器采集到的交通数据按照一定的计算标准，使用某种具体的方法将其规范化到一个合适的范围中。进行归一化的目的在于得到的交通数据参数不唯一，数据量大，数值差异也比较大，将其统一到一个合理的范围中去后使得后续的运算更加方便和规整。为了能很好地去除由于数据的太大差异性给实验带来预测结果的干扰，而且本身交通数据的变化幅度就比较大，因此在实验之前必须对得到的交通数据进行一个归一化的处理操作，且归一化后的数据范围映射在[0,1]之间。本文的主要实验操作都是在MATLAB2015a基础上进行的，此环境本身也自带了多种进行归一化的函数方法，本文在实验中利用了常用的mapminmax函数方法对数据进行归一化的操作。Mapminmax方法函数的具体使用格式如式（3-9）：

式（3-9）中，Y表示的即为归一化操作后得到的数据矩阵，Z代表的则是记录了相关信息的一个结构体，X表示的即为需要归一化的初始数据矩阵。

对某一个交通数据的具体计算原理如式（3-10）：

3.5本章小结

在针对城市交通里面的短时交通流预测，同时在获取到能够应用于这一前提条件的交通数据情况下，本章首先阐述了关于交通流的基础知识，并且表明本文实验的短时交通流量预测的研究对象为交通流，同时重点介绍了三个基本参数：流量、速度、占有率；其次交代了本文的实验数据来源为交通运输部出行云中的合肥市示范区黄科路口相关数据以及数据的相关特征；最后主要交代了对获取的交通数据的分析和预处理，分别有异常数据的处理方法、错误数据的处理方法和数据最终的归一化。这章的知识主要为本文后面的算法模型搭建提供一个数据基础。

第4章基于支持向量回归机的短时交通流预测

上一章主要是针对研究的内容获取到的交通原数据进行了预处理工作，本章主要是利用预处理后的交通数据选取合适的预

测模型。如何从这些大量的交通数据中发现潜在的规律出来这对算法的要求很高，也有很多的算法可以用来进行预测研究，最通俗的莫过于线性回归方法，对其结果容易理解，计算上也不复杂，但是对非线性的数据拟合不好，因此对于在交通上的预测很明显不切实际；决策树回归理论也是常用的一类回归算法，这种策略认为数据中的复杂关系可以使用树结构进行概括，使用树来对预测值分段，但是这种方法不太适用于时间序列数据，当数据量大的时候效率比较低不能满足短时交通流预测实时性的要求；自回归积分滑动平均模型（Autoregressive Integrated Moving Average Model,ARIMA）属于时间序列算法，很多专家学者研究过该方法在短时交通流上的应用，发现ARIMA适应于在交通数据量很大，且交通量在呈现出周期性变化时，模型预测效果会很好，这对短时交通来说并不适应，因为短时交通本身存在的大难题之一就是交通规律性很弱；卡尔曼滤波曾经也是深受大家喜爱的方法，在短时交通中也曾进行过研究，该方法将交通系统利用状态空间的模型来模拟并且采用递推算法的思想进行交通流预测，算法本质其实属于线性估计模型，所以对于处理短时交通中时常会遇到很大的波动交通特征时算法性能就会降低；混沌理论是交通中研究考虑的热点，混沌理论可以在非常复杂的系统中找出其存在的规律性，已经有很多研究者已经证明在交通中由于其影响因素很多且很多为非线性的关系，因而交通系统中存在着显然的混沌特征，并且使用混论理论对短时交通预测也取得好的效果。总而言之，对于如何从获取到的交通数据中挖掘出规律性出来进而可以完成较好的预测，有很多的算法可以使用，但优缺点不一，也并没有所谓最好的算法能够通用解决交通中所有的问题，本文在总结前人进行的各种研究结果加之对各种方法分析之后并根据自身获取到交通数据情况认为支持向量回归机算法比较适合用于在短时交通流的预测研究。

支持向量回归机（Support Vector Regression，SVR）是演变自SVM的思想发展出来的，把对分类的理念延伸到回归当中的问题中去，而且也在很多工程实践中得到了很好地应用[40]。尤其在非线性的情况，数据具有多维特征，SVR算法将数据反映到更高维数据空间，使得原本线性不可分问题通过这种另辟蹊径的方法变得容易了。这是得益于SVR中核函数的“善良”秉性，使得通过这样的空间变换之后计算量没有变得想象中那么复杂，这一点是非常难得的。在短时交通中，交通流具有规律性弱、不确定性强等非线性的问题，而SVR本身在解决小样本、非线性等问题有着很好的优势，且SVR结构不是很复杂，易实现，不会陷入局部解，所以比较锲合在短时交通上的工程应用。

4.1支持向量机理论

4.1.1支持向量机原理

SVM算法最初是用在解决线性可分的问题上并基于此慢慢发展而来，其理论思想通过图4-2展示出来的二维平面进行理解。按图中所示有两种类型的数据样本存在，中间的一条为分类线H，、是和H这条分类线平行的，而且这两条是与H距离最近的平行线，与之间的距离称为分类间隔；当这个分类线H可以将两类样本数据正确分离并且分类间隔最大的时候此时的分类线为最优[41]。

令分类线为(w)+b = 0并对其做标准化，对能够进行线性可分的数据样本集合S = {(,)}使其满足式（4-9）：

式（4-9）中，= {-1,+1}，，w H，b R。能够使分类间隔w / 2的值最小或者是2 / w的值最大同时也可以将训练数据样本(,)进行正确的分到所属的类别的这种分类面被称之为最优分类面。SVM中的支持向量就是位于、线上的数据样本点。支持向量机的核心思想在于要实现主要的泛化能力，这是要使得分类间隔最大化去实现的。所以，通过线性可分为条件来建立最优超平面将求最大化分类间隔问题转为二次规划为题[42]:

通过对式（4-10）采用Lagrange乘子法进行求解，可以将二次规划问题再次转为解Lagrange函数鞍点的问题[43]：

式（4-11）中，表示拉格朗日乘数且 0。据KKT条件若要取得最优解必须满足式（4-12）：

对式（4-11）进行求解，有唯一的解。并将设定为得出的最优解，那么可以得出：

在公式（4-13）中，即代表所需要的支持向量。表示分类间隔阈值，它可以利用上式（4-12）得出。最终可以得出需要的最优分类面：

以上对最优分类面的求解是对线性可分的情况下进行的，不能用于非线性的问题，若需要解决非线性的问题，需要引入核方法，可以将非线性下的问题由低维空间反映到高维的特征空间，然后同样的即可解出最优分类面。基于泛函数的理论，若核函数K(,)能够达到Mercer定理的要求，那么就有某一变换空间的点与其点积对应。因此，想在引入核函数K(,)后不增加计算复杂度成功的解决非线性转变成线性后的分类，需要着重考虑选取理想的核函数，此时最优的分类函数方法为：

4.1.2支持向量回归机原理

支持向量回归机针对的是线性以及非线性的预测回归问题，眼下SVR存在两种分别是SVR和SVR，文章使用的是SVR进行短时交通的预测。SVR的理论思想为：在拥有m个样本数据{（），（），…，（）},代表输入样本，代表输出样本。首先，对原始的输入样本数据利用非线性变换将其投影到高维的空间中，将非线性回归变成线性回归。则在投影到的高维特征空间中，我们可以建立出一个最优超平面完成分类。

在公式（4-16），b代表了偏移量。输入数据样本经过一个非线性变换的映射变换到高维的特征空间中去之后，达到了将非线性回归问题转为线性回归问题的求解目的，将不敏感损失函数定义为式（4-17）：

SVR问题目的就是企图找到一个合理的f()能够使得E(w)取得一个最小值：

在式（4-18），线性权重量用w = 表示；C代表惩罚参数，值过大会造成过拟合，值太小分类性能就会很差。

在对问题的求解的时候我们引入了两个非负的松弛变量和，于是得到（4-18）等价的对偶问题：

上式中，和分别代表的是和这两个引入的松弛变量所对应的拉格朗日乘子，K()即为合适的达到Mercer定理要求的核函数。

因此对于测试样本数据x所求的输出可按是（4-20）进行预测：

从上式可以得出，支持向量回归机与神经网络在形式上比较相像，支持向量回归机的结构如下图所示：

在SVR的推导计算后，最后SVR的分类函数是通过将输入的未知的数据向量与每个支持向量的内积，因此对SVR回归计算的复杂度仅仅是由支持向量的个数来决定的。

4.1.3松弛变量

在解决实际的样本数据中，会有各种原因产生使得很多时候不是所有的问题都可以像线性可分那样简单，在一些情况下会有不能进行线性分离的点出现，这些数据点可以称之为“离群点”。

对数据样本中，我们将间隔最小的点设定其间隔为1，同时加入新的变量为松弛变量ξ，则式（4-9）可改写为如下：

公式（4-21）中，引入的变量ξ≥0，通过约束条件进行计算是允许其小于1的，然而若有一数据点间隔小于1的话，实际上是很难被准确分类，则目标函数可写成式（4-22）：

对（4-22）式中，C即为惩罚参数，其值的变化可以用于减轻错误分类样本的惩罚；表示分类超平面训练数据的偏差值，若数据样本点即为离群点的时候，有ξ>0，相对应的非离群点有ξ=0。

通过上述的研究分析，当在有离群点引入ξ之后，（4-10）可进一步变化得出如下公式：

对（4-23）做对偶变换：

得出拉格朗日方程，并给出分类决策函数：

4.1.4核函数

线性问题利用超平面分类可以很好解决，对非线性的情况则行不通。如图4-4所示，将坐标轴中横轴（X轴）a、b之间即红色部分视为一个正类别，a和b两端的黑色部分视作一个负类。那么可以发现通过线性方法无法将两类分开，但是在（4-4）图中的蓝色曲线可以将两类别完美分开。

很明显，用线性的方法函数是无法对已有的数据样本进行区分的，所以引入核函数的方法。对核函数的方法思想前文也曾提及就是通过一种规则变换映射使得低维空间样本数据反映到高维，通过在高维的特征空间中利用线性方法进行区分。因此，在使用SVM或者SVR处理非线性情况，选择一个合适的核函数是解决问题的关键。经常使用的核函数有以下四类：

(1)线性核：K(x,)= (x,)

(2)多项式核：K(x,)=

(3)径向基核：K(x,)=exp(-)

(4)S形核：K(x,)=tanh(v(x)+C)

4.2基于支持向量回归预测模型研究

4.2.1模型参数分析

使用SVR模型算法，对核函数、特征空间以及非线性变换之间有相应的对应关系。对核函数的选择以及对应核函数中参数的选法对非线性变换也会有影响，从而影响到算法复杂度。对如何选择核函数上当前并没有很好地统一标准，在总结前人大量的实践中本文采用的是高斯径向基核，它有以下特点：

（1）高斯核的适应范围很广，若选择了合适的参数，就可以有较宽的收敛域并且会得到很优良的性能以及很好的学习能力，适用于任一分布的数据样本，原始数据即可使用非线性反映到高维空间中。

（2）核函数中参数的个数决定了算法模型选择的复杂度，高斯核与多项式核或是多层感知器核比较，它的参数量最少，使用更为方便简单。

经过上述分析对比，本文实验在搭建SVR算法模型使用的核函数即为高斯径向基核。则对SVR算法模型预测短时交通，影响其预测效果的关键参数为：惩罚参数C、不敏感损失系数及核参数。

a.核参数对算法模型的影响

核参数值的大小对SVR性能的影响很大，因为对支持向量的相关程度和训练数据的分布特性都是被核参数决定。若的值偏大则表明支持向量间的相关性越强，模型推广能力会由于的增加反而会减弱，则预测精度得不到保证；若的值偏小则支持向量间的相关性会很弱，模型的复杂性加强则推广能力无法得到保障。

b.惩罚参数C对算法模型的影响

惩罚参数C是主要来调控置信区间范围和学习机的经验风险在数据子空间的比例大小，希望使得算法结构能达到结构风险最小和很好的推广能力。算法模型的健壮性、复杂性以及在对那些在管道域外的样本点数据的惩罚程度的大小都是由C决定。因此，在已有的数据空间中，C值若太大，则说明此时算法的推广性就会很差因为算法模型相对于数据的拟合度偏高，出现了“过学习”现象；C值若太小说明对惩罚经验误差轻，训练数据的误差会变大，这也就会造成“欠学习”现象。对任一数据空间总会存在一个最优的惩罚参数C可以使得SVR具有很好地推广性能，但若C越过了某范围，则对SVR算法模型的推广力和降低经验风险方面失去意义。

c. 不敏感损失系数

在算法中，样本数据中存在的不敏感区域的范围、支持向量个数甚至模型推广性能都会受到不敏感损失系数的影响。如若取值过于小，必然会使SVR模型复杂度增加，预测效果会得到提高然而求解算法模型的时间会增加很多，支持向量的数目也会增加，那么就会出现“过拟合”，此时模型在推广能力上会大打折扣；相反如果值过大，导致“欠拟合”，模型的推广能力也会大打折扣。

通过对上述参数的研究分析可知，想要SVR算法模型有很好的预测效果就必须合理选择C、、这三个参数值。

4.2.2网格法选取模型参数

网格法的大体思想是通过一定的规则把算法模型的参数取值范围分割成若干个小区域，接着通过计算出参数取值的所有可能组合并且计算出相应的目标误差，然后进行一番比较选择出在该取值范围中目标值最小时对应的参数组合。这种选取参数的方法在理论基础上使得得出的解为范围中的一个全局最优解，可以较少发生重大的误差。

对SVR算法使用高斯核的并通过网格法调参的步骤如下：

（1）将惩罚参数C、核参数及不敏感系数分别按可能的取值范围通过固定的搜索补偿步长使用网格法进行划分，并且他们的取值范围应当在2的指数空间上这样方便离散化搜索。

（2）在通过（1）的基础上选出所有可能的参数之间的组合通过交叉验证的方法求出均方误差并比较，选出均方误差最小的参数组。若在后续的搜索时发现有某一组的均方误差与已经得出的最小误差相近并且其C的值还要更小时，则原来的参数组合进行更新。

4.2.3仿真实验

实验是在单机的情况下进行的，使用的是win10操作系统，CPU为英特尔的Core i7-6700HQ ,显卡为GTX965M，内存为8GB，硬盘位1TB，采用的仿真实验平台为Matlab 2015a，整个实验都是基于以上环境下进行的。

因为本文为短时交通流预测因此将交通流量作为实验预测对象。实验交通数据是之前已经预处理好的数据样本。样本从2016年6月30日到2016年7月1日，数据检测周期为1min，预测间隔为10min，选取其中08:00—22:00期间的数据每天共有79组数据分别进行训练和预测。结合使用的支持向量回归的短时交通流预测模型进行预测，BP神经网络在短时交通预测中的身影随处可见属于使用比较广泛的一种神经网络结构，实验将和BP神经网络进行一个对比，分析SVR算法在对短时交通预测的价值。预测结果与实际记录值得结果对比图如下：

4.2.4实验评价指标

为了可以更好地量化分析模型的预测性能本文中主要采用平均绝对误差（Mean Absolute Error,MAE）、均方误差（Mean Square Error,MSE）、均等系数（Equal Coefficient）[44]。

平均绝对误差：

均方误差：

均等系数：

其中，为t时刻模型预测值，N为预测时段长度，为t时刻交通流实际测量值。MSE反应误差分布情况，值越小,说明预测模型描述实验数据具有更好的精确度，预测效果越好。EC反映预测值和实际测量值之间的拟合程度，值越大越接近于1，表示预测效果越好。

4.2.5实验结果分析

对实验所得结果分别计算SVR模型和BP神经网络的平均绝对误差、均方误差和均等系数三个指标，统计如表（4.1）和（4.2）所示：

通过图4-5观察发现SVR模型和BP神经网络对短时交通预测的结果差不多，进一步通过分析两者计算出的指标发现SVR对于短时交通的应用要略优于BP神经网络。分析预测结果指标：SVR和BP网络的均等系数EC指标都要大于0.9，而且SVR要略优于BP算法，SVR的MAE和RMSE值分别为6.8439和2.4513，这两个值都比相应的BP神经网络算法要小，表明SVR模型预测精度更高。通过对比分析可得，SVR算法对于应用在短时交通预测中是值得研究的，有实际的应用价值。

4.3本章小结

针对预处理后的交通数据进一步要做的就是采用合适的算法可以比较充分的挖掘出数据中的规律性来。因此本文在分析对比多种回归预测方法之后，最终根据交通流的数据特点以及本文获取的交通数据情况选择了支持向量回归进行短时交通流的预测研究。在此基础上，介绍了关于支持向量机和回归机的理论思想，同时介绍了支持向量机中的核函数以及算法中涉及的主要参数核参数、惩罚参数和不敏感损失参数，分析了几个参数对模型的影响；然后使用获取到的交通数据建立了SVR模型，并且和使用神经网络建立的预测模型进行了对比，通过交通模型中常用的评价指标对结果进行了分析得出使用SVR算法模型和神经网络的预测效果差不多，表明使用支持向量回归算法针对短时交通流进行预测是一种可行的、有效的方法，值得进行进一步的研究。

5.指标是由系统根据《学术期刊论文不端行为的界定标准》自动生成的。

6.红色文字表示文字复制部分;绿色文字表示引用部分。

7.本报告单仅对您所选择比对资源范围内检测结果负责。

8.Email：amlc@cnki.net http://e.weibo.com/u/3194559873 http://t.qq.com/CNKI_kycx