**摘 要**

在当今的智能交通控制和车辆诱导的研究领域里面，短时交通中的交通流量预测技术具有重大研究意义。而且在交通实际的工程运用当中，它的理论基础性在对智能交通领域亦是有非常重要的作用。利用交通数据来对交通流进行判断、研究分析和预测，进而可以对城市中智能交通的诱导出行等提供帮助，为需要的用户来提供选择最佳的出行路径，提高出行效率，让出行变得不再那么困扰。

短时交通流预测的时间跨度并没有一个非常标准的定义，通常是指基于获取到的交通数据针对未来15min内的预测，而且在交通控制和诱导中对提高实时性方面起着很大作用。智能交通系统中比较关键的一点就是希望对交通流实时、动态和精准地预测，以提高城市交通管理和运行效率，这也是为什么短时预测能够成为当前智能交通系统的重要研究内容的原因。另外短时交通流量的预测时间跨度相对较短，交通数据的变化有时并没有太强的规律，各种干扰噪声对交通流预测会产生较大的影响，这些无疑导致了短时交通流预测的发展在当前非常具有挑战性。

当前国内外的交通专家学者针对短时交通流不确定性较强和规律性较弱等特点所提出来的预测模型已有数十种之多。据所采用的预测原理大致可将其分为两类：一类属于数学模型的方法，如卡尔曼滤波模型、指数平滑模型、ARIMA模型等；另一类是基于非数学模型的方法，如支持向量回归机、非参数回归模型、神经网络模型等。其中的数学模型方法在构建和求解交通模型过程中难度较大，因此很难达到短时交通预测的要求；而对于非数学模型的预测方法相对来说实现要更简便，只要向模型里面喂足足够的历史数据，不需要去构建过于庞大冗余的预测模型，而且最终得到的结果也可以满足在智能交通系统中的需要。不可置否，对在短时交通流预测中所存在的不确定性较强和规律性较弱等一系列不可忽略的特点，利用非数学模型对历史数据进行挖掘训练，很难进一步提高短期交通流预测的准确性，尤其是在突发事件发生的一些情况下其预测精度会明显下降。为此本文将利用经过改进的粒子群算法配合支持向量回归算法完成历史数据的挖掘训练；通过分析待测路口上下游之间的时空关系，挖掘得出路口之间的时空关联性，将对历史数据的挖掘训练的预测与基于时空关联性的预测结合起来，同时利用BP神经网络，让两者进行优势互补、迭代加权构建出在线自我学习完善的短时交通流预测模型。

**关键词：**短时交通流量预测；支持向量回归机；粒子群；时空关联性；BP神经网络

**Abstract**

In the field of intelligent traffic control and vehicle-induced research, the prediction of traffic flow in short-term traffic is of great significance.Moreover, in the practical engineering application of transportation, its basic theory is also very important in the field of intelligent transportation.Using traffic data to judge the traffic flow, the research analysis and forecasting, thus intelligent transportation in the city can be induced travel help, such as in need of users to provide the best travel path, improve the efficiency of travel, make travel no more.

Time span of the short-term traffic flow forecasting is not a very standard definition, usually refers to the traffic based on access to the data for future forecasts within 15 min, and in the traffic control and guidance plays a big role to improve the real-time performance.In intelligent transportation system is the key point is hope of real-time, dynamic and accurate prediction of traffic flow to improve the efficiency of city traffic management and operation, which is why the short-term forecast to become the important content in the research of intelligent transportation system.Other short-term traffic flow prediction of a relatively short time span, the law of the change of traffic data sometimes is not too strong, all kinds of jamming noise will produce great influence for traffic flow prediction, these will undoubtedly lead to the development of the short-term traffic flow prediction is very challenging in the current.

Currently, there are dozens of prediction models proposed by traffic experts and scholars at home and abroad on the characteristics of short-term traffic flow uncertainty and weak regularity.According to the prediction principle, it can be divided into two categories: one belongs to the mathematical model, such as kalman filter model, exponential smoothing model, ARIMA model, etc.The other is based on non-mathematical model, such as support vector regression, non-parametric regression model, neural network model, etc.The mathematical model method is very difficult to construct and solve the traffic model, so it is difficult to meet the requirement of short-term traffic prediction.Mathematical model for the prediction method of implementation is relatively easier, as long as the fed enough historical data, the model do not need to build too large redundant forecasting model, and the final results can also meet the needs of intelligent transportation system.Cannot deny whether, in the short-term traffic flow prediction in the uncertainty of the strong and weak regularity and a series of characteristics that cannot be ignored, the mathematical model is used to analyse the historical data mining, it is difficult to further improve the accuracy of short-term traffic flow forecasting, especially in emergencies occur in some cases the prediction accuracy will be markedly reduced.This paper will use the improved particle swarm algorithm to support vector regression algorithm to complete the mining training of historical data.By analyzing the space-time relation between the upstream and downstream intersection under test, the temporal and spatial correlation between mining road, training of history data mining based on spatio-temporal correlation between prediction and forecast, at the same time using the BP neural network, the two complementary advantages, iterative weighted build online self learning improve short-term traffic flow prediction model.

**Key Words:** Short-term traffic flow prediction; Support Vector Regression; Particle Swarm Optimization; time-space correlation；BP Neural Network

**第1章 绪论**

1.1研究背景及意义

在经过改革开放的几十年高速发展过程中，我们国家的社会经济发展取得了举世瞩目的成绩，与此同时，随着大家生活水平的提高，汽车的数量在不断攀升，这也导致了现在城市中道路交通面临的压力也是越来越严重了。有相关的数据告诉我们机动车的数量在这几年增长相当之快，到前年的2016下半年，我国在全国范围类的机动车的数目保守估计在3.09亿，在这当中的汽车数量将近在1.96亿，占到63.43%的比例；全国的机动车驾驶员的人数有3.58亿，这里面的汽车驾驶员的数量就已经超过了2.9亿人，数字让人惊讶。跟2010年的机动车量2.07亿相比来说，这6年的时间增长了49.27%，增长的速度可以说是让人叹为观止。也正由于机动车数量的急剧增加，所以在当今的交通管理中交通拥堵、交通事故频发等这些问题也是现在最为棘手的交通难题。

针对上面分析的交通状况，已有的传统交通管理技术手段对当前这样的交通形势已经不能很好的来解决这些存在的交通问题了，可以说是捉襟见肘。“时势造英雄”，传统的手段既然不能适应了，对于当前的存在的交通问题，智能交通系统（Intelligent Traffic System，ITS）开始粉墨登场，发挥优势了。ITS对于目前的交通状况可以说是很受欢迎的一套完整解决方案，是交通管理者的“宠儿”。这套系统可以说是集百家之长，包含了很多先进的科学技术手段，例如:信息技术，计算机技术，数据通信技术，传感器技术以及大数据和人工智能等，将这些技术手段结合形成的这样一套智能的综合性的交通运输管理系统，并将其运用之于交通运输、服务控制的体系中去，通过运用这样的ITS系统可以更容易、更方便、全方位了解、掌控、管理城市交通的交通状况，进而达到能够实时、准确、高效的交通管理和诱导的作用。而且不仅是局限于理论水平，ITS在许多发达国家尤其是欧美国家已经很早就进行了深入的研究并且投入使用，在我国的各大城市也开始发展成一定规模。目前，ITS系统在很多国家的成功运用表明了它是解决当今交通难题的最实用的方案。

ITS研究领域中最注重的研究方向之一就是对交通诱导和交通控制的研究。能够达到一个实时、高效、准确的交通流预测是对交通诱导和控制的重要依据。交通流预测说白了就是对车流量的预测，根据历史交通时间序列数据即已有的过去交通数据以及现有的路面实时数据，将这些数据放在我们搭建好的算法模型来对待测路段或者路口进行下一时刻的车流量进行预测分析。对于交通中交通流预测，可以根据预测时间的跨度把它分为中长期预测和短时预测，其中短时交通流预测的时间跨度并没有一个非常标准的定义，通常是指基于获取到的交通数据针对未来15min内的预测，而且在交通控制和诱导中对提高实时性方面起着很大作用。智能交通系统中比较关键的一点就是希望对交通流实时、动态和精准地预测，以提高城市交通管理和运行效率，这也是为什么短时预测能够成为当前智能交通系统的重要研究内容的原因。另外短时交通流量的预测时间跨度相对较短，交通数据的变化有时并没有太强的规律，各种干扰噪声对交通流预测会产生较大的影响，这些无疑导致了短时交通流预测的发展在当前非常具有挑战性。

显而易见，在交通的管理和控制里面短时交通流量的预测有着举足轻重的作用。如若能对交通流信息可以做到准确实时的预测，我们就可以在大家出行的时候及时提供出实时且是有效的道路交通信息，节省了出行时间；另一方面；可以为城市交通管理也提供诸多便利，无形当中可以缓解交通拥堵，一定程度的减少了污染和节约了资源，使得城市交通管理也更加合理，这些都无疑会对城市交通的建设发展有着重大意义。

1.2国内外研究现状

在起初阶段，基于线性理论模型被运用在短时交通的预测当中，方法有如熟知的历史平均模型，卡尔曼滤波以及滑动平均这些常用的算法模型；紧接着从线性发展到非线性的理论模型，耳熟能详的有小波理论，非参数回归，混沌理论等模型；随后，基于智能理论的模型能够更好的反应交通的特点，提高预测精度而成为学术界研究的热点；当前由于ITS系统的出现，首先将交通数据进行必要的预处理，针对短时交通流预测研究的重点是将选择智能算法结合神经网络的特性来进行预测分析研究。

2004年，华冬冬等人通过将神经网络和系统相空间重构参数进行组合，以此来选取最优的神经网络输入模式。对遗传算法利用隔离小生境手段进行了优化，根据BP网络的特性，采用隔离小生境遗传算法来对神经网络进行优化，从而可以得出最好的神经网络隐层结构，最终构建相对合理的关于神经网络的预测模型。2005年，徐启华、了兆奎等人在短时交通流的预测上是采用了动态递归神经网络的算法模型，因为递归神经网络算法在动态记忆能力这个特征上比静态神经网络适用性具有绝对优势，所以最终对短时交通的预测结果还是比较精确的。2006年，田晶、杨玉珍等人在短时交通流的预测上使用了混沌时间序列算法和基于LM（Levenberg-Marquardt）的BP神经网络算法模型，从实验结果来看，两者之间的预测精度相差无几，效果差不了多少，但在实时性上混沌序列的算法模型则是表现更优越，交通流的预测结果会随着时间上的不断推移，预测效果总是比LM算法的BP神经网络要好一点。2006年，王均、关伟等人采用Kalman滤波的方法进行预测，在根据城市环路的实际交通运行特性，结合北京市三环路的实际交通数据来验证了模型的可行性，结果表明预测效果和实时性都比较理想。2007年，张玉梅、曲仕茹等提出来基于RBF网络和混沌算法的短时交通预测模型，先是在一小部分小数据集Lyapunav指数方法判断出交通流存在混沌的条件下，相空间重构一下交通流数据，建立RBF网络模型进行了仿真研究。2007年，范鲁明、贺国光等利用改进的K近邻非参数回归算法模型进行短时交通流量预测，对传统K近邻做了两方面改进：一方面是采用相关性理论选择状态向量；另一方面是聚类分析的变K近邻搜索算法。2008年，姚智胜、邵春福等人在短时交通流量预测上是采用主成分分析和支持向量机的方法，将两者结合，首先是对已有的诸多断面的交通数据进行一个主成分分析，得到需要的主成分数据序列而后以此即可来通过训练集训练支持向量机同时采用遗传算法进行参数的优化；最终提供给支持向量机输入数据得到主成分预测结果，并将其转换成断面的交通数据，即可预测短时交通流量。2008年，薛洁妮、史忠科等人基于混沌时间序列的分析方法，将实测数据进行相空间重构并分析了欧氏距离和均等系数，在此之上提出来了最邻近点的两步优化选择方法，同时运用局部多项式拟合对最邻近点逐渐逼近来得到预测公式，通过验证表明模型具有较高精度。2010年，承向军、刘军等人利用分形理论来进行短时交通的流量预测，通过G-P算法，对相空间中任意两点间的欧氏距离利用筛选法得到备选点的欧氏距离，加快计算效率，让2min内的交通预测实现变成了可能。2010年，罗向龙、牛国宏等人采用交通流量经验模态分解与神经网络的方法对短时交通进行预测，他们是利用EMD分解对交通流量进行分解得到不同的模态，然后使用神经网络将分解之后的流量分量在进行预测，把预测的值加到最后的预测结果，结果表明方法具有较高的预测效果。2011年，赵亚萍、张和生等利用基于最小二乘支持向量机的方法进行短时交通模型搭建，根据实际的交通流量数据考虑了星期数、时段数、天气等多个影响因素从时间和空间角度出发，模型适用性较好。2011年，王建、邓卫等采用基于贝叶斯网络多方法组合的算法模型，首先是搭建2中以上基本的预测交通模型，将这些模型预测的结果和实际结果进行离散处理，之后用贝叶斯来学习；进行联合推理得出基本模型预测结果组合后可能组合的预测结果的后验概率，将最大的作为预测值。2012年，于振洋等使用了小波消噪的神经网络短时交通预测模型，开始先将获得到的交通数据利用小波进行去噪处理，BP神经网络的输入个数采用关联维数方法来确定，最终仿真表明经过小波去噪的预测效果明显得到提高。2013年，贾显超、陈旭梅等在采用混沌理论的方法上对短时交通流量进行了多步预测，首先是对交通流量时间序列采用最大Lyapunov指数判别法来分析其混沌特征，进而相空间重构交通流量数据时间序列，然后再进一步利用加权一阶局域方法来构建出基于混沌理论的多步交通流量预测模型。2013年，Yangyan Xu、Qing-Jie Kong等人利用分类和回归树进行短时交通流量预测，提出了一种新的基于非参数模型的短期交通量预测方法，应用的非参数模型是分类和回归树（CART）模型。在应用程序中，CART模型首先将历史交通状态分类为丰富的类别。之后，建立对应于每个交通状态模式的线性回归模型。最后，该模型通过将当前状态向量聚类为最适合的历史模式和回归模型来预测短时交通。2013年，Ziwen Leng、Junwei Gao等人利用广义回归神经网络（GRNN）建立了基于GRNN的交通流量短时预测模型。 GRNN模型选择交叉验证算法对网络进行训练，以预测误差的均方根为网络评价准则确定平滑因子，并采用滚动预测的方法预测交通流量。与RBF和BP神经网络预测模型相比，GRNN具有更强的逼近能力和更高的预测精度。2013年，李军怀、高瞻等人采用指数平滑和马尔科夫链进行短时交通预测，将指数平滑理论与马尔可夫链进行相互结合，马尔可夫链用于解决指数平滑预测缩小预测区间的问题，并改善预先计算的每个状态中心和状态转移概率矩阵的加权状态。利用实测交通流模拟结果表明，该方法比传统的指数平滑方法具有更高的精度，适应性强。2014年，王娇、李军等人介绍了一种基于最小最大概率回归的短时交通流量预测模型，对于北京的某高速公路监控站的交通流数据和英国某地区实际测量交通流量数据，采用基于MPMR的预测模型，测试结果表明，模型可以较好地跟踪实际流量值，在相同条件下较传统预测模型具有更好的预测精度，验证了所提模型的有效性。2015年，Yongxue Tian、Li Pan等人提出了一种称为长时间记忆递归神经网络（LSTM RNN）的模型，该模型利用存储器块中的三个乘法单元动态地确定最佳时间滞后。使用Caltrans Performance Measurement System（PeMS）的数据集来构建模型，并将LSTM RNN与几种众所周知的模型（例如随机游走（RW），支持向量机（SVM），单层前馈神经网络（FFNN） ）和堆叠自动编码器（SAE）。结果表明，所提出的预测模型具有较高的精度和一般性。2015年，Haikun Hong、Wenhao Huang等人提出了一种基于KNN的新型三阶段框架来处理短期交通流量预测的上述问题。第一阶段，从整个交通网络中发现目标任务的相关起源站点和目标站点。然后，对于每个目标任务，在第二阶段中学习特定的距离度量。最后，在第三阶段建立一个扩展的多度量k-最近邻回归模型。现实世界交通数据集的实验结果表明，带有Lasso的多度量KNN模型优于传统的KNN模型，特征构造方法是有效的。2015年，Wenbin Hu、Liping Yan等人提出了一种混合的PSO-SVR预测方法，以获得较低的学习时间精度;该方法利用粒子群优化(PSO)搜索最优SVR参数。并找到一个更合适的SVR参数搜索算法,提出了三种策略来处理颗粒流出搜索空间,根据比较,其中一个策略可以使算法得到最优参数更迅速,称之为算法使用此策略算法的速度一样快。此外，为了处理原始数据中噪声对精度的影响，提出了一种基于历史短期流数据相似性的混合PSO-SVR方法。大量对比实验的预测结果表明，所提出的模型比其他最先进的算法能得到更精确的预测结果;而当包含噪声的数据时，具有历史动量的方法仍然值得精确的预测。2016年，Denwen Xia、Binfeng Wang等人在Hadoop平台的分布式建模的通用MapReduce框架中提出了一个名为STW-KNN的时空加权K-最近邻模型，以提高短时交通流量预测的准确性和效率。更具体地说，STW-KNN通过趋势调整特征来考虑交通流的时空相关性和权重，以优化包含状态向量，接近度量，预测函数和K选择的搜索机制。此外，STW-KNN在广泛采用的具有MapReduce并行处理范例的Hadoop分布式计算平台上实现，用于实时并行预测交通流量。2016年，Carl Goves、Robin North等人介绍了应用人工智能的结果，特别是人工神经网络（ANN），根据当前和历史交通信息，估计未来15分钟的交通状况。对于这项研究，英国高速公路事件检测和自动信号（MIDAS）系统的数据用于约20公里的英国曼彻斯特附近的M60，M62和M602高速公路建立短期预测模型。为了减少问题的复杂性，使用自动编码器成功地减少了模型的输入维数，最终模型显示出非常好的预测能力。2016年，Dong Wang、Jie Xiong等人采用简化的单层前馈网络（SLFN）结构，提出了一种新型的短时交通流量预测方法 - 集合实时序贯极限学习机（ERS-ELM）建议。通过对历史数据进行快速训练，并利用新到达的数据逐步更新模型。2017年，Bin Sun、Wei Cheng等人通过同时调整关于动态交通特性的所有参数来提高KNN预测精度。提出加权参数元组（WPT）根据流量动态计算加权平均值。对一年的实际数据进行全面的实验。结果表明，流动感知WPT KNN比手动调整的KNN以及极端梯度提升（XGB）和季节自回归综合移动平均（SARIMA）等基准方法表现更好。2017年，Cheng Anyu、Jiangxiao和Li Yongfu等人提出了一种基于混沌和支持向量机回归的多源交通流预测方法。预先将交通数据进行去噪处理，接下来将一维交通流信息重构到高维空间中，获取更多的交通信息和特征。然后利用支持向量回归模型对短时交通流进行预测。实验表明对SVM进行相空间重构得到的模型对短时交通流预测的准确性和及时性都得到了提升。

根据以上介绍，对短时交通流量预测的模型和方法研究可以概括为三个方面：一种是对交通的原始数据进行处理分析，重点是数据的特征提取和对缺失数据的处理；一种是在有海量的交通数据基础上，还需要其他相关的数据源，例如天气、温度、湿度甚至突发事件等数据，实际情况这些数据当前不容易能够拿到；另外一种就是将流行的神经网络和智能算法进行结合以期能够提高预测的精度、稳定性及实时性。

1.3论文主要内容及章节简介

本文内容共分六章，每章内容如下：

第一章 绪论。这章主要介绍了短时交通流预测的背景、意义也阐述了短时交通这个课题在 国内外的研究状况，最后介绍了本文研究的内容和章节安排。

第二章 交通流数据分析与预处理。这一章首先交代了交通流的一些基本理论，紧接着介绍了在交通流中所需要用到的三个最基本的衡量参数即流量、速度和占有率；然后交代了本文实验所用到的交通数据源的情况介绍数据的数据特征；最后是针对本文所有的交通数据所要用到的预处理方法。

第三章 基于支持向量回归机的短时交通流预测。在这章首先介绍了一些统计学的原理知识，在此基础上阐述了关于支持向量机和支持向量回归机的原理，进而建立了支持向量回归机的短时交通预测模型，表明利用SVR进行短时交通预测的可行性。

第四章 基于改进的PSO-SVR短时交通预测模型。在前文分析SVR的基础上，这一章首先利用传统粒子群算法来对SVR模型参数选择优化，并考虑到粒子群算法固有的缺陷，引入改进的方法，然后搭建改进后的PSO-SVR模型进行仿真并与传统的PSO-SVR及网格搜索的SVR比较，表明改进后模型的可行性。

第五章 基于时空关联性的改进PSO-SVR算法预测模型。仅仅依靠待测路口的历史数据进行预测，其精度难以提高，特别是在某些突发事件的情况下，预测精度会大打折扣。因此本文在改进PSO-SVR模型的基础上，考虑相邻路口周边的时空关联性，通过BP网络的特性，得到一个可以在线实时不断迭代修正的预测模型，提高模型预测效果，更加具有实用性！

**第2章 交通流数据分析与预处理**

2.1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。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。

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（2-1）所示：

= （2-1）

式（2-1）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（2-2）所示：

= （2-2）

式（2-2）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。

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（2-3）所示：

= （2-3）

式（2-3）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：

= （2-4）

式（2-4）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。

2.2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。

表DT\_LANE\_REPORT\_H

|  |  |  |  |
| --- | --- | --- | --- |
| **字段名** | **数据类型** | **键/索引** | **备注** |
| ID | NUMBER(15) |  | 编号ID |
| DETECT\_CLASS | VARCHAR2(2) |  | 设备类型  C:微波检测器 |
| DETECT\_ID | VARCHAR2(16) |  | 设备编号  （与路网的对应关系请见下文） |
| ROAD\_ID | NUMBER(15) |  | 路段编号 |
| COLLEC\_TIME | VARCHAR2(20) |  | 采集时间 |
| ROOM\_OCCUPANCY | VARCHAR2(10) |  | 时间占有率（单位%） |
| SPEED | NUMBER(3) |  | 车道平均速度（KM/H） |
| LANE | VARCHAR2(3) |  | 车道号 （从道路外侧到道路中央. 0：从左到右第1个车道 1：从左到右第2个车道 2：从左到右第3个车道 3：从左到右第4个车道） |
| VEHICLE\_CLASS | VARCHAR2(10) |  | 车辆类型 (0 =pedestrian 未定义车型 1 = pedestrian行人 2 = bike 自行车、摩托车、电动自行车 3 = passenger car客车、小客车 4 = truck 货车、卡车、搬运车;) |
| VOLUME | NUMBER(10) |  | 一个统计周期的车流量（辆） |

2.3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放到最后的模型进行试验。所以，对采集到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。

2.3.1异常数据修正

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。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（2-5）：

=+(1-) （2-5）

式（2-5）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（2-6）：

0 (2-6)

式（2-6）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;

B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（2-7）：

0v （2-7）

式（2-7）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,其范围设定如式（2-8）: 0Occ100% （2-8）

2.4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（2-9）：

[Y,Z] = mapminmax(X) (2-9)

式（2-9）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（2-10）：

 = + （2-10）

2.5本章小结

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；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3章 基于支持向量回归机的短时交通流预测**

学习和模仿能力在我们人类的整个繁衍、进化和发展的历程当中毫无疑问是我们可以生活在今天这样一个美好时代的很重要的原因。我们通过不断地对信息的收集、研究、分析和归纳来找到包含在数据背后的规律，有了规律就可以顺藤摸瓜通过这些规律来进行预测，从而发现我们原本不能直接观测得到的信息。渐渐地，我们希望计算机也可以像人的思维一样来模仿我们的这种学习思维方式，这就产生了机器如何学习的问题即机器学习的由来。这几年机器学习的一个研究大热点就是AI即我们所说的人工智能，它的一个重要的意义在于希望通过计算机来发现海量数据中存在的相关性，找出可用的有价值的信息，基于此我们进一步的来进行判断、分析和研究去预测未来的数据信息也就是“未卜先知”的能力。

这些年机器学习的快速发展得益于统计学为其提供了丰厚的理论支持和基础。但是传统的统计学和本文所要说到的统计学习理论（Statistical Learning Theory，STL）也是不尽相同，在本文中其实主要是针对在没有丰富的样本数据的情况之下来对机器学习开展的研究。支持向量机（Support Vetor Machine,SVM）的理论思想是可以通过某种方法将不属于同一类别的样本进行正确的归类，SVM就属于是本文所用的统计学理论下的一个完美产物。

支持向量回归机（Support Vector Regression，SVR）其实是借用了支持向量机的思想，将分类的思想上升到回归的问题，而且也在很多工程实践中得到了很好地应用。尤其在非线性的情况，数据具有多维特征，SVR算法将数据映射到高维空间，来解决在原始空间中线性不可分的问题。这是由于SVR中核函数的优良品质，这样的非线性扩展在计算量上并没有比原来复杂多少，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。

3.1统计学习理论

3.1.1学习问题的描述

统计学习理论是建立在一整套丰富的机器学习理论之上的，具有很扎实的理论。相比较于传统的统计学，STL可以在没有很充足的实验数据的情形下去分析机器学习的规律，这就提供了一个可以在原本没有丰富样本的情况下进行机器学习的统一参考的框架。

对机器学习问题其实就是在式（3-1）的基础上进行如下研究：

假设有n个独立同分布的数据样本：

,,…, （3-1）

通过对数据样本的研究分析，期望能够在{f(x,w)}函数里面，求解得出一个最优的f(x,)可以使期望风险R(w)即式（3-2）最小。

R(w)=L(y,f(x,w))dF(x,y) （3-2）

式（3-2）中，f(x,w)代表的是预测函数集，w表示广义上的参数，所有的函数集合都可以用{ f(x,w)}表示；通过f(x,w)对y预测所产生的损失为L(y,f(x,w))，损失函数会由于不同类型的学习问题有多种形式，像模式识别类型的问题损失函数的定义如式（3-3）：

L(y,f(x,w))= （3-3）

在对连续的回归估计问题上，损失函数如式（3-4）：

L(y,f(x,w)) = (3.4)

对于概率密度的估计问题，密度函数用p(x,w)表示，对损失函数的定义如式（3-5）：

L(p(x,w)) = -log(x,w) （3-5）

3.1.2经验风险最小化

使得期望风险最小才是我们进行机器学习的目的，然而我们手中能够利用的数据信息只有如样本数据（3-1）这般，这样我们希望式（3-2）的期望风险就没有办法进行计算，因此在传统的研究方法中就另辟蹊径，将样本来定义出一个经验风险，采用的是经验风险最小（Experience Risk Minimization ,ERM）准则。

= （3-6）

构造出来学习算法使得对式（3-2）的估计即期望风险的值能最小，式（3-3）的损失函数即经验风险对于模式识别类型的问题来说就是得出训练数据样本的错误率；式（3-4）的损失函数即经验风险对于函数回归估计问题来说就等同于训练误差；式（3-5）的损失函数即经验风险对于概率密度估计的问题可以视为最大似然函数方法。理论上来说，把期望风险最小化用ERM准则来进行替代也没有很充足的理论支持，仅仅是从直观上理解我们认为这是合理的。例如，若果当n的取值倾向于一个无穷大的时候，此时式（3-6）的结果就会和式（3-2）相差无几了，因此，对于许多的问题，他们的数据样本其实是有限的的，那么在这样有限的数据样本下通过ERM准则得出的最小经验风险就未必是真的最小。  
3.1.3结构风险最小化

基于VC维的理论基础，统计学习理论中推广性界分析探究了不同类型的方法函数集以及真实风险和经验风险的关系。实际的风险R(w)与经验风险两者存在至少1-的数学概率并满足式（3-7）、（3-8）两个关系式，这是对二分类问题总结得出的结论。

R(w) + (h/n) （3-7）

(h/n) = （3-8）

式（3-7）、（3-8）中，n代表的是数据样本的数目，h表示方法函数集合的VC维。

可以得出若h/l比较大的话，会使得学习机置信区间变大，那么就会使得真实风险增加进而会导致较大的误差产生。相反，若数据样本量很大，h／l会变得比较小，学习机的置信区间会缩小，此时ERM的值与实际值就会很接近。

若我们的数据样本是一个定值时，学习机的复杂度如果越小即VC维越小，则其置信范围必然会减小，这样检验风险与实际值的差异会缩小。那么当需要建立分类器就需要考虑不仅要尽可能小的经验风险还得把控好VC维，就可以减小置信范围。

最小化结构风险的思想：若想要达到一个最小化的结构经验风险，方法就是让式（3-7）中的和(h/n) 进行制约，一起趋向最小，而且得使置信区间达到最小也即是让h的值要尽能的小。

对于公式（3-7），当所需的训练样本数据量n有限的情况下，则可以通过和h两个参数来调整结构风险R(w)：

1. 机器学习的ERM由函数集合f(w,x)决定的，利用w可以对经验风险进行掌控。

（2）学习机的函数方法集合会对VC维的维度产生影响，首先对函数集合进行结构从而搭建出函数子结构与维度h关联，把VC维与函数集合联系在一起。

对方法函数集{f(w，)，w∈}（代表的是抽象参数集合）进行结构化采取的方法如下：

对函数集合，当中 = {f(w,),}，对任一函数集S里面的元素都会存在一个VC维且是有限的，并有h1≤h2≤h3…≤。对一组数据样本()，()，()，…()，在函数集里面利用结构风险最小化的思想选取出一个方法函数f(w,)以期得到ERM，并保证有最小的置信范围。

欠学习 过学习

风

险 真实风险的界

置信范围

经验风险

函数集

VC维

h1≤h2≤h3

图3-1 结构风险最小化示意图

3.2支持向量机理论

3.2.1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3-2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。

X X X

X X

X X

X Margin

H

图3-2 样本分类示意图

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（3-9）：

((w)+b) 1, i = 1,2,3…,m （3-9）

式（3-9）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

其中w H， ，b R，{}。训练样本(,)能被正确可分，并且让分类间隔2／w最大或w／2最小的分类面称为最优分类面，而作为支持向量的就是位于虚线上的训练样本点。而SVM的关键思想就是通过让分类间隔最大化来实现主导泛化能力的目的。

因此，以线性可分为条件构造最优超平面能转化为二次规划问题:

(3.10)

对式(3.10)运用Lagrange方法求解，则该二次规划问题可以转化为求解Lagrange函数鞍点的问题:

L(w,b,a) = 0.5(w.w) – (3.11)

式(3.11)中， 0，为Lagrange乘数。

由KKT定理可知，最优解还必须满足以下条件：

= 0 i = 1,2，…，m (3.12)求解式(3.11)，其存在唯一解。假定为最优解，则有：= (3.13)

式(3.13)中，为支持向量。为分类阈值，可通过条件式(3.12)来求解。最后能够求得最优分类面函数为：

f(x) = sgn{(w)+b} = sgn{} (3.14)

上述求解过程主要针对线性可分问题，不适应非线性问题的求解，但是可以通过一定的变换将低维空间中的非线性问题映射为高维特征空间中的线性可分问题，再求解最优分类面。由泛函理论可知，要是核函数K(,)满足Mercer条件，则它可以与某一变换空间中的点积相对应。所以，要在不增加计算复杂度的情况下完成非线性变换后的线性分类，则需要在最优分类面中选取理想的核函数K(,)。此时分类函数变为:

f(x) = sgn() (3.15)

3.2.2支持向量回归机原理

支持向量回归机能够解决非线性回归问题，目前主要有SVR和SVR两种方式，本文采用SVR来进行交通流预测。SVR的基本原理是：给定m个数据样本，其中,和分别为输入和输出样本。首先，通过非线性映射将原空间中的输入样本映射到M维特征空间中，然后完成线性回归。在M维特征空间中，构建分类超平面或逼近函数：

f(x) = + b (3.16)

式(3.16)中，b为偏移量。输入样本经过变换，可以通过在高维空间中求解线性回归问题以达到解决原空间中的非线性回归问题的目的，定义不敏感损失函数如下:

(,,f) = max{0,|| - } (3.17)

支持向量机回归问题就是找到合适的f()，让E(w)取最小值:

E(w) = 0.5(w) + C (3.18)

式(3.18)中，w = 为线性权重向量；C是函数回归模型复杂度与样本拟合精度二者的折中。

在求解过程中，引入非负松弛变量和，得到式(3.18)的等价的对偶问题：

(3.19)

式(3.19)中，和分别为两个松弛变量和对应的Lagrange乘数，K()是满足Mercer条件的核函数。

则测试样本x所对应的输出可以按下式进行预测：

f(x) = (3.20)

由式(3.20)可知，SVR逼近函数在形式上与神经网络类似，其输出是l个中间节点的线性组合，并且每个中间节点与一个支持向量相对应，如图3-3所示

输出f(x)

K()

K()

K()

f(x) =

,,…,为

支持向量

… 输入样

图3-3支持向量回归机结构图

由于最终的SVR逼近函数中只会包含输入未知向量分别与每个支持向量内积之后的线性组合，所以在回归时计算的复杂度完全取决于支持向量的数目。

3.2.3松弛变量

在实际操作过程中，由于种种原因，使得并非所有的问题都线性可分，少数情况下也会出现不可分的样本点，我们称之为离群点。

将样本点中间隔最小的那个点的间隔定为 1，我们便给它加上一个松弛变量 ξ，此时，公式(3.9)可以写为：

((w)+b) 1-ξ i = 1,2,3…,m (3.21)

上式中，松弛变量ξ≥0，则约束条件计算出来允许小于 1，但当某一点的间隔小于1，就意味着这些点无法被精确分类。相应的，目标函数我们可以写为：

min + C (3.22)

上式中，代表在广义分类超平面中训练样本的偏差，当样本点为离群点时，ξ>0；非离群点的松弛变量ξ=0，C为惩罚系数，其作用是避免对错分样本的过分惩罚。

基于以上分析，在引入松弛变量后，式(3.10)进一步写为：

(3.23)

进行对偶变换：

(3.24)

并建立 Lagrange 方程，得到决策函数：

f(x) = sgn() (3.25)

3.2.4核函数

对于非线性问题，超平面的分类方法无法解决。

如下图所示，我们把X轴（横轴）上a和b之间的点(即红色段部分)看作正类，a、b 两边的点(即黑色段部分)看作是负类。显而易见，线性函数无法将正类和负类正确的分开，但图(3-4)中的蓝色二次曲线却可以达到这一目的。

a b

显然，样本数据无法用线性函数来区分，我们引入K(x,)。核函数的基本思想是将低维空间的数据通过某种规则映射到高维空间，在低维空间接受输入向量，然后在高维空间通过线性回归计算出内积值。

在使用支持向量机处理问题时，合适的核函数是求解的关键。以下是常用的四类核函数：

(1)线性核函数：K(x, )= (x,)

(2)多项式核函数：K(x,)=

(3)径向核函数：K(x,)=exp(-)

(4)S形核函数：K(x,)=tanh(v(x)+C)

式(3.24)引入核函数后变换得到式(3.24)，决策函数为式(3.25)。

3.3基于支持向量回归预测模型研究

3.3.1模型参数分析

在SVR模型应用中非线性映射、特征空间H和核函数K存在一定的对应关系。核函数及其参数如何选取会影响到非线性函数映射，进而影响到数据子空间分布的维数，即复杂度。

目前，核函数的选取并无公认最好的指导原则，在实验和研究中最常用到的是高斯径向基核函数，它与其他核函数相比有以下特点：

(1)高斯径向基核函数是适应性极广的核函数，假若参数选择合适，就可以获得很强的学习能力和优良的性能，同时具有较宽的收敛域，在任意分布的样本中也是适用的，可将原空间输入数据通过非线性映射较好的转换到高维空间。

(2)模型选择的复杂度取决于核参数的个数，与多项式核函数和多层感知器核函数相比，高斯径向基核函数的参数最少，选择起来更为简单。

通过上述分析比较，本文在建构SVR模型时将选择高斯径向基核函数。影响SVR短时交通流预测模型性能的主要因素是：核宽度系数、惩罚因子C、不敏感系数。

a.核宽度系数的影响

训练样本数据的分布或范围特性以及支持向量间的相关程度都是由核宽度系数决定的，其值的大小将很大程度上决定SVR性能的优劣。较大的意味着支持向量之间相关程度也越强，其推广能力随着参数的增大而减弱，因此回归模型的精度难以得到保证；较小的意味着支持向量之间的关联会不紧密，学习机器复杂性较强，无法保证其推广能力。

b. 惩罚因子C的影响

在给定的数据子空间中学习机器经验风险和置信范围间比例是由惩罚因子C来调节的，以期获得学习机器的结构风险最小和推广能力最好。算法对管道区域外的样本数据的惩罚程度是由惩罚因子C来表征的，而且模型的鲁棒性和复杂性也是由其决定的。在给定的数据子空间中，C过小意味着惩罚经验误差的程度轻，训练误差相应会变大，这样就易出现“欠学习”；C过大意味着模型对数据的拟合度较高，而推广能力会较低，易出现“过学习”。每个数据子空间总是有一个使得SVR推广能力最佳的C，若C超出某一范围，对SVR的经验风险的降低和推广能力提高基本上没有任何作用。

c. 不敏感系数

不敏感系数能够决定拟合函数对数据样本的不敏感区域的大小、支持向量的数目以及模型推广能力的大小。要是不敏感系数的取值较小，则SVR模型会变得更加复杂，预测的精确度会变高，但是会让模型求解时间变长，支持向量的数目变多，从而导致“过拟合’现象，模型推广能力会降低：反之，易出现“欠拟合”现象，模型的推广能力也会比较低。

从以上分析可知，要想使SVR模型获得比较好的预测性能，必须合理地选用C、、的取值。

3.3.2网格法选取模型参数

网格法的基本思想是将需要选取的模型参数的取值区间，按照一定的规则划分为若干的小区间，然后计算出变量取值的所有组合以及各个组合所对应的目标误差，并通过比较择优选择出在该区间中目标值最小时所对应的参数组合。这种参数选取方法在理论上能保证所选取的解为在这个区域内的全局最优解，以达到避免重大误差的目的。

对以高斯径向基为核函数的SVR模型的网格法主要步骤如下：

(1) 对参数C、和的取值区间按照一定的搜索步长进行网格划分，同时将参数C、和的取值区间指定在2的指数范围区间内以便进行离散化查找。

(2) 在步骤(1)所划分的网格中选择每个可能的参数组合并且对所取的参数组合进行交叉验证，对交叉验证所求得的均方误差(Mean Square Error，简称MSE)进行比较并选出MSE取最小时所对应的参数组合。然而在网格搜索过程中，要是某一参数组合所对应的MSE近似等于当前最佳MSE，而且惩罚因子C取值小于当前最佳MSE所对应的惩罚因子C时，更新最佳参数组合。

3.3.3实验评价指标

实验误差的评价指标采用平均绝对误差MAE，均方误差RMSE，均等系数EC。

平均绝对误差：

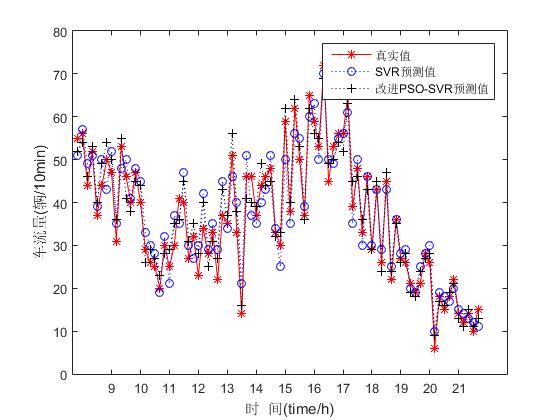
均方误差：

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。RMSE反应误差分布情况，值越小,说明[预测模型](http://baike.baidu.com/item/%E9%A2%84%E6%B5%8B%E6%A8%A1%E5%9E%8B)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。

3.3.4仿真实验

本文选择交通流的最常用参数流量作为预测对象。数据来源是前文中已经进行过预处理的预测样本。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，预测结果与实际记录值得结果对比图如下：



评价指标如下表:

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | MAE | RMSE | EC |
| SVR | 7.2845 | 2.7764 | 0.91356 |

从预测结果结果对比图和表中数据来看，采用网格搜索的方式可以达到一定的预测效果。网格算法以固定步长对参数空间的所有组合进行穷举实验，对于参数较少的情况该方法取得了很好的效果，但是在实际应用中该算法效率不高。因此本文将在下面对SVR模型参数的选择采取其他方法进行改进。

3.4本章小结

本章首先介绍了支持向量机的基础理论和机器学习的相关知识，其中包括了经验风险最小化，以及结构风险最小化的理论。

在此基础上，分析了支持向量机和回归机的原理，建立 SVR 的模型，并将预测结果与实测值进行对比，来证明支持向量回归模型是一种可行的、有效的预测模型。为了进一步提高模型预测效果可以接近实际应用，本文将在SVR基础上采取更适合在交通流预测中的参数选择方式进一步进行试验。

**第4章 基于改进的PSO-SVR短时交通预测模型**

4.1粒子群算法理论

4.1.1粒子群的起源

粒子群算法的创始人是 R.C. Eberhart和 J. Kennedy（1995），其模拟鸟类觅食过程中的迁徙和群聚行为，是一种被广泛应用的群体智能全局随机搜索算法。在搜索解空间时，粒子借助的飞行方向和距离主要是通过速度的变化来控制的，并同时也综合学习自我的历史经验，以及群体其他成员的经验，以达到不断调节自身所找到的局部最优解和整个种群当前找到的全局最优解的目的，在解空间中不断搜索直至迭代结束。

4.1.2粒子群优化算法思想

在粒子群优化算法(Particle Swarm Optimization,PSO)中，想象每个需要优化的问题都存在一个隐含的解，而这个解的位置隶属于 n 维搜索空间上的某一个点，以此称之为粒子（Particle），它启发于鸟类随机寻找食物的过程中，设想在固定范围内放置一块食物，每只鸟都事先不知其位置，但均可获知其离食物的距离，所有的鸟都可以当作一个被目标函数决定的适应值（fitness value），然后每只鸟都在给定方向和速率的情况下根据当前最优飞翔方法下寻找食物。对其结果研究得出，每只鸟虽然各自为政，但是整个鸟群组成的整体就类似在一个中心的控制下而发生的行为，即简单的个体规则运动组成了复杂的全局行为。

当若干个粒子组成的群体在一个n维空间进行搜索，该若干个粒子组成种群记为X = {， ，…， }，而每个粒子的位置 = { ，，…, }都代表问题的一个解。粒子进行随机移动改变自己的位置，随之不断的搜索新的解。带惯性权重的每个粒子都具有记忆能力，对搜索到的最优解进行存储并记作，而当整个粒子群经过搜索所记忆的最好位置，则为目前搜索到的最优解，记作。另外，当粒子移动时都具有各自的速度，记为V ={，，…， }，当各自的最优解和整个群体的最优解都搜索到后，每个粒子根据公式（4.1）来更新自己的速度。

(t+1) = (t) +rand()( - (t))+rand()( - (t)) (4.1)

(t+1) = (t) + (t+1) (4.2)

式中，(t+1) 表示第i 个粒子在t +1次迭代中第d维上的速度，为惯性权重，、为加速常数，rand()为0~1之间的随机数。此外，对粒子的速度进行上限设置记做，即当上式中(t+1) 时，(t+1) = ;(t+1) -时，(t+1) = -。

从式（4.1）（4.2）中的系数关系可以得出，粒子搜索的方向除了受到自己本身速度(t)的影响，还被两个距离因素所干涉，分别为自身搜索过程中的最优经历的途径 - (t)以及整个粒子种群的最优经历途径 - (t)，同时以上三个影响因素的地位分别由ω、、 这几个权重系数控制决定。

粒子群优化算法所体现出的最明显优点在于算法中所需调整的参数少，但是每一个参数都对算法有很大的影响，因此为了保证一定的精度和优化效率，这些参数的设置也至关重要。以下是粒子群优化算法中相对来说影响较大的参数设置一般规则及其作用：

（1）粒子的群体数量：一般取 20-40，根据实验经验通常 10 个粒子的搜索能力就可以完成目标，但由于某些实验需要解决某些特定或者复杂度高的问题，可以适当增加粒子的群体个数，甚至可以达到过百的数量。但应当注意的是，粒子种群数量的增加也会导致搜索区域范围的扩大，对于发现全局最优解来说会更容易，但相应的运行时间也会变长。

（2）粒子长度：取决于具体的优化问题。

（3）粒子范围：一般的来说，为了避免无谓的搜索，需要事先确定可能的搜索范围，缩小搜索空间。

（4）粒子最大速率：为保证粒子的搜索范围处于所定义的空间内，对粒子的最大速率进行上限设定，以此也达到了限制粒子在本次搜索中的最大移动长度的目的。

（5）适应度函数：通过采用适应值来评价粒子好坏的程度，而适应度函数可以对适应值进行判定。一般可取问题的目标函数作为优化需要的适应度函数，也可以根据需要进行适当的变换。

（6）学习能力：选择具有较好学习能力的系数可以适当的提高收敛速度，并且能够避免局部最优解的情况，通常对系数 、的取值保持在0到4 之间。

除了以上的基本参数，还包括几个重要参数，其中惯性权重ω ，它主要功能是令粒子保持好的运动惯性；速度调节参数 、，表示粒子趋于个体最优位置和群体最优位置的加速向权重。若ω =0 ，粒子的速度不具有记忆能力，粒子群失去辨识能力，由目前的极值解来决定其位置，粒子群将不具有聚集的能力；若 =0 ，粒子丧失认知能力，虽然粒子群的收敛速度增快，但是由于无认知，极有可能陷入局部极值；若 =0 ，粒子只具有认知能力，无团结、无组织性能，粒子群各自完成自己的搜索功能，有可能无法得到全局的最优解。由实验和经验的理论研究，绝对最优的参数是不存在的，而对于不同的问题，采用适合该问题的参数才能够得到收敛速度快和鲁棒性好的结果。

粒子群优化算法是一种依赖于概率搜索的算法，对优化函数以及约束条件的连续性等性能没有硬性要求，并且搜索得到的为全局最优解的概率较其他方法来说占优势，虽然该算法需要得到适应度函数，但与传统的一些遗传、演化算法相比较，其优点是非常显著的。

（1）算法具有很强的鲁棒性，因为不受到统一的命令支配，因此个别个体并不会影响全局。

（2）种群之间不直接通信，确保了系统的扩展性，当群体中个体数目增加时，额外的通信开销也较少。

（3）群体中每个粒子的执行动作简单容易理解，寻优时间短，因此算法整体也较简单。

4.2 PSO-SVR短时交通预测模型

4.2.1 PSO算法流程

Step1: 对粒子群的种群数量进行设定即 m ，并且给定每个粒子的初速度和初始位置，并且对各个相关参数也进行初始化设置；

Step2: 对每个粒子的适应值f()进行评价，并与它自身经过的最优点作比较，若较好则将该点作为第i 个粒子当前点的最优位置，所有粒子中的最优点进行比较后，较好的作为种群最优位置；

Step3: 根据公式（4.1）（4.2）对各个微粒的速度和位置进行调整，产生新的粒子群 X(t)；

Step4: 对新的种群 X(t) 中粒子新位置的适应值进行计算，分别与历史的个体及种群最优位置进行对比，若优于历史位置，则采用新的取代，反之不替换。

Step5: 当更新的所有结果满足设置的寻优结束条件时，即为所求的最优解，寻优终止；否则继续令t =t +1，转Step2 再次迭代。

粒子群算法的基本流程图如下所示：

算法开始

初始化粒子群

计算每个粒子

的适应度值

粒子适应度评价

计算粒子个体最优和群体最优

根据公式更新粒子的速度和位置，产生新种群

满足结束条件

N

Y

输出最优解，算法结束

图4-1 粒子群算法流程图

4.2.2 PSO-SVR模型流程

将原始数据进行去噪、归一化等预处理操作后，根据PSO 算法在设定的参数范围内，寻找最优参数；将最优参数用于对模型的训练，使用最优模型预测交通流数据。算法流程图如下:

原始交通数据

初始种群

输入(x,y)

数据预处理

更新粒子速度和位置

SVR模型

回归参数

计算个体适应度值

最佳适应度值判断

更新粒子个体最优和全局最优

N

Y

建立参数最优模型

预测

交通流

图4-2 PSO-SVR模型

4.2.3实验仿真

根据已有的交通流数据进行基于PSO-SVR算法的短时交通流预测，同时与传统的SVR模型预测进行对比。仿真结果如下和评价指标如图4-3和表4-1：

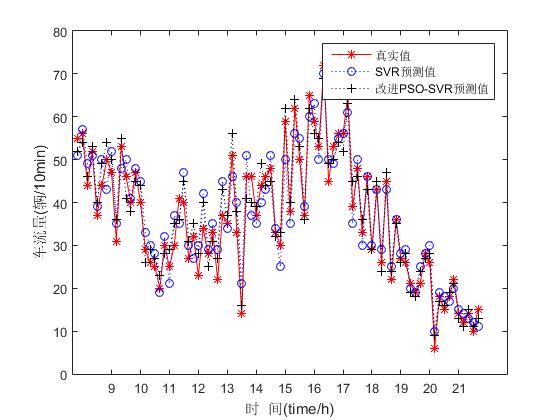


图4-3

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | MAE | RMSE | EC |
| SVR | 7.2845 | 2.7764 | 0.91356 |
| PSO-SVR | 5.6375 | 2.1298 | 0.93489 |

表4-1

根据对仿真结果的分析，从图4-3和表4-1可以得出：采用PSO-SVR算法得到的最优参数支持向量机的预测结果较SVR算法好，但是此方法任然可以改进以进一步提高预测精度。

4.3改进PSO-SVR算法预测模型

基本PSO算法的缺陷主要体现在容易发生早熟收敛的现象，全局最优的搜索能力不足，容易陷入局部最优，收敛速度也较慢，对于搜索到的最优质可能不够精确。

存在这种现象的主要原因是PSO算法在搜索过程中有着如下的特点：在ｎ维空间搜索前期，有着很高的收敛速度，能够快速的收敛到极值附近，但是随着粒子群的不断捜索，收敛速度会明显的减慢，粒子的速度降低到了0，粒子群没有了更新速度和位置的能力，此时的极值的位置也即此次捜索的最优解，但是这个最优解不一定是全局最优，有可能只是局部最优，粒子在此时己经收敛于这个局部最优解，不能够再捜索全局最优解。因此，PSO算法搜索过程的特点决定了该算法存在了一些不足，需要进行一些弥补和改进。

对粒子群的初始位置在之前算法中很多都是随机的，然而，实际在算法后续寻优过程是会受到粒子初始值的影响的。因此对初始化粒子群本文引入混沌搜索的方法，这样粒子在整个空间中均匀分布，使得算法收敛速度可以加快同时得到全局最优解的速度也得以提升。

4.3.1混沌搜索

混沌是自然界广泛存在的一种非线性现象。他看似混沌，却有着精致的内在结构，具有随机性、遍历性及规律性等特点，对初始条件极度敏感，能在一定范围内按其自身规律不重复地遍历所有状态，利用混沌运动的这些性质可以进行优化搜索。目前对混沌尚无严格的定义，一般将由确定性方程得到的具有随机性的运动状态称为混沌。

4.3.2改进PSO-SVR算法

将问题解的维数设为d维.再引入混沌搜索来初始化粒子的初始位置，详细操作为：第一步先生成每一个分量数在(0,1)之间的d维的随机向量=(,,…,),然后由Logistic方程对迭代，直到N个随机向量,,…,,将的各分量按下式投影产生混沌初始化序列：

， i=1,2,…N；j= 1,2,…,d （4.3）

粒子群=的适应度值由目标函数来计算, 粒子群的初始位置则从N个中选出n个较优的即可。

在文献中已经提出:粒子速度其实并不能很好地反映接近于最优位置的参照，粒子的收敛速度和精度反而可能因为向错误的方向搜索而降低。据此，对传统粒子群简化后的优化公式为：

（4.4）

最终粒子群的迭代公式在经过引进王振武[16]对粒子群的进一步改进方法之后如下：

β，其+ （4.5）

上式中加入了新的参数和随机因子,是高于所有粒子的平均值而且适应度值也要比优。每个粒子在算法寻优的过程中有、和三个一起向种群中粒子传递信息，从而可以得到更多的消息。β是新增的动量项，和粒子历史速度相关, β∈[0,1]为动量参数,可正可负。算法在寻优过程之中的震荡也因为新增的动量项而得到改善。

4.4基于改进的PSO-SVR短时交通预测结果

使用前文处理好的交通数据进行仿真实验。改进PSO-SVR实验步骤如下：

1)首先设定误差阈值以及迭代次数并为PSO中的各个参数、、、及β赋初值;

2) 根据适应度函数确定种群规模并用混沌搜索来初始化粒子种群即SVR的三个参数(C，);

3)选出最初全局极值和个体极值;

4) 根据所改进的粒子群公式更新粒子的位置，使用适应度函数计算适应度值，更新和;

5) 满足结束条件（寻优次数达到迭代值或者适应度值大于设定阈值）则寻优结束，返回参数(C，),否则转到4);

6)使用参数(C，)建立的SVR模型进行短时交通流预测。

模型预测结果如下：

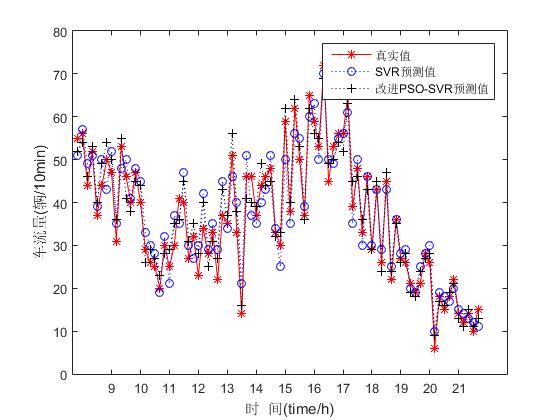


图4-4

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | MAE | RMSE | EC |
| SVR | 7.2845 | 2.7764 | 0.91356 |
| PSO-SVR | 5.6375 | 2.1298 | 0.93489 |

表4-2

根据对仿真结果的分析，从图4-4和表4-2可以得出：采用改进PSO-SVR算法得到的最优参数支持向量机的预测结果精度进一步得到了提升，而且对于短时交通流的预测更具有实际意义。

4.5本章小结

本章通过对粒子群优化算法的分析，提出粒子群优化支持向量回归机的方法，采用改进粒子群寻优算法，对支持向量回归机参数进行优化处理，最后基于改进PSO-SVR算法建模仿真，通过仿真结果表明，采用粒子群优化后的支持向量回归机模型收敛速度提高，同时预测精度也有所提高，该方法具有可行性。

**第5章 基于时空关联性的改进PSO-SVR算法预测模型**

5.1时空关联性分析

时空相关性分析所研究的主要是路口上下游之间随着时间变化的规律，反映出交通数据在时间和空间上的关联性。在城市交通中，交通流有很强的时空关联性。在时间上，交通流遵循着一定的时间序列变化规律；空间上，每个路口流量受上下游交通路口流量的影响也会呈现一定的相关性，下游路口的交通量可以根据上游路口的流量估计得出。



图5-1

本文将上图中的①号路口作为待测路口，研究它和其上游号路口之间的流量关系。由于是短时预测，本文首先根据两个上下游路口之间的历史数据挖掘出两者流量之间的最相关时间t.例如:需预测的是待测路口①在9点时的交通流量，则本文根据④号路口的历史数据主要包括交通流量，速度，占有率三个指标，挖掘出此路口9点之前这三个指标的数据，认为当这三个指标的数据与①路口9点时的数据最接近的即为最相关的时刻，假如分析得出时间为8:54,则号路口与①路口9点最相关时间t=6min.以此类推，可以挖掘出①路口每个时间点对应的号路口的最相关时间t。同时记录出号路口每个最相关时刻的流量，并将对应的流量与①路口对应流量做出百分比一起记录在数据表中。即该表保存着号路口与①路口最相关的时刻、最相关时间t以及对应的流量之比。

在找出号路口与①路口最相关时间t的基础上，挖掘出号路口交通数据中的速度与t之间的关系。因为检测器记录的数据为号路口在检测点的瞬间速度，与整个路段之间的行程速度会有误差，所以直接分析这些数据会有很大误差。但两路口之间的距离为一定值，所以本文根据检测器的速度和t来估算出其距离，这样可以得出每个t和速度对应的多组距离s.将得出的s每隔几个数据求一次平均值，得到数组,最后用和t推算出整个路段中车流的速度v.如此根据得出的最相关时间t以及v的数据，基于此可以挖掘出两者之间的函数关系为：

t= (k为常数) （5.1）

通过上式得出的关系，根据号路口有新的数据速度时，可以得出其t，然后可以在之前得出的表中找到对应的时间及流量百分比，从而得出①路口下一时刻的交通流量。方法预测结果如下：



图5-2

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | MAE | RMSE | EC |
| 时空相关性 | 7.2953 | 2.7858 | 0.90985 |

表5-1

5.2 人工神经网络发展

人类大脑的神经元数量约为 1.4\*个，是人们公认的最复杂的处理信息工具。在信息的处理方面人脑具有如下几个特点：

（1）能够同时大规模处理信息。能够在同一时间段并行处理大量信息，在情况复杂时做出判断。

（2）具有很强的容错性。人脑区别于其他生物的机理在于其能够类比、推理以及应用，能通过记忆来学习事物的规律性并且将结果储存在大脑中。

（3）自适应能力比较强。人处于非常复杂的社会环境中，大脑的自学习能力是我们适应自然社会的最大保障。

脑细胞也被人们称为神经元，脑神经元之间有着大量的连接通道，这样就构成了一个复杂连接的神经网络。神经网络的结构主要有三部分组成：

（1）树突。树突位于神经元细胞体的周围，树突接受刺激并且把刺激传导至神经元细胞体。

（2）神经元细胞体。是信息的载体，接收从树突传递来的刺激并经过加工处理之后，传递到轴突。

（3）轴突。神经元细胞体传递来的信号通过神经纤维末端的轴突再传递至其他神经元。

生物上的神经网络是以神经元为单位处理信息，然后对信息进行处理和存储。这种复杂的工作方式使得人的大脑有着计算机无法实现的智能化的优势。为了深入研究大脑的处理机制，科学家从模仿大脑生物神经元处理存储信息的过程入手，提出了一种类人脑的网络处理机制—神经网络模型，其在人工智能、大数据处理、智能化交通等方面表新出良好的智能性。

神经网络自身性质如下：

（1）非线性。神经元如果是非线性的，那么互相连接的非线性神经元也是非线性的。当本身输入的特征是非线性性质时，非线性神经元的这种性质就显得尤为重要。

（2）映射关系。有监督的学习方式是神经网络学习的一种重要途径。该方法的原理是给神经网络的训练集进行标记处理或者是对神经网络的权值进行修改。每个输入都是有一个之对应的输出。从搜集的数据集中随机选取一组样本由神经网络进行训练，网络就会自动赋予一个权值，不断修改直到误差达到我们设定的范围或者收敛次数达到预定值。

（3）自适应性。神经网络的自适应性表现在其能够对初始权值的设置进行调整。当把系统设置为随环境变化而变化时，网络的权值就会随着时间的变化而变化。

20 世纪40年代，科学家 Mc Culloch和W.Pitts提出了一种新型的神经网络模型，他们证明了这种神经网络可以无限逼近任意函数，在科学领域有着重要的应用，从而也开创了神经网络在人工智能领域应用的先河。20世纪60年代Minsky发表了《Perceptron》一书，详细阐述了感知机的原理和缺陷。1982 年，美国数学家J.J.Hopfield提出了 Hopfield神经网格模型，对整个网络稳定性条件给出了自己的判断条件。随后他又发表了关于连续型Hopfield神经网格模型的成果，对整个神经网络研究的进展做出了开创性的贡献。

1986年，Rumelhart和 Mc Celland 领导的科学家小组在《Parallel Distributed processing》一书中，首次细致阐述了BP神经网络的原理，给出了一个信号前向传递误差反向传播的BP神经网络模型。但是BP神经网络有个明显的缺陷，就是在误差反向传播时，利用的梯度下降法容易陷入局部极小值以及收敛速度较慢，需要训练的次数较多。所以学界里提出了很多改进型算法。主要有梯度下降法、附加动量法等。本文引用的LM算法也是训练BP神经网络的一种非线性优化算法，在实际中取得了良好的效果。迄今，BP 神经网络已被用于解决大量实际问题。

5.3人工神经网络的结构和学习

5.3.1 人工神经网络的结构

MP模型是人工神经网络所建立的第一种模型，此模型的运行机理为：神经元细胞接受刺激后产生两种反应分别是兴奋或者是抑制。生物神经网络中，神经元细胞之间突触的活动强度代表了神经元的连接程度。因而，在此模型中，连接神经元细胞之间的突触可以采用神经网络中连接神经元的权值来表示。相比于生物神经元，MP神经网络存在机理上的相似性。在接受外界刺激后产生兴奋，当这种兴奋程度达到一定的程度，即兴奋度在所取的阈值之上，那么就会产生相应的刺激；如果兴奋程度在阈值以下，那么其不会产生任何反应。下图为MP神经网络的结构示意图：前面的x代表输入，后面y代表输出，中间w代表连接神经元之间的权值。该神经网络模型能够进行或与非的逻辑判断，利用自身简单的结构来完成复杂的运算。这种模型的缺陷也是显而易见的，作为中间权值的w是固定不能更改的，无法通过训练得到一个最优的权值，这种动态的神经网缺乏的是认识规律与自学习能力。

y

图5-3 MP神经网络模型图

大多数神经网络的激活函数选用的是线性函数。此时网络的输出y就是输入的数值和权值之间的加权和。表达式如下：

y = – b = w – b （5.2）

其中b代表神经元的阈值。目前在神经网络中应用最广泛的函数称为S型函数，它是一条光滑的曲线，表达式如下：

f(x) = （5.3）

5.3.2人工神经网络的学习原则

学习这个概念是神经网络特有的重要概念，神经网络的训练是通过自适应性学习来体现的，生物神经网络之所以能够适应整个环境，就是因为能够不断学习。神经网络具有强大的自主学习能力，它自身存在的优良性能必须通过大量训练才能获得。通过训练之后的神经网络具有记忆功能，这些记忆功能就存在于连接每个神经元之间的突触也就是权值中。整个网络的训练过程，就是神经网络权值的调整过程，通过预设的输入，网络输出与期望输出误差的反向传播来达到改变权值的目的。通过调整神经网络的权值，改良整个系统的预测精度。人工神经网络常用的学习原则有Hebbian 学习原则、感知器学习原则和δ学习原则。

1. Hebbian学习原则

该原则源于突触修正的假设，当前后神经元同时兴奋时，传输信号的能力就会增强，如果前后神经元处于状态不一致的状态时，传输信号的能力就会减弱。也就是说当两个神经元处在同时兴奋的状态时，两神经元之间的权值连接程度就会增强。

在此学习规则下，学习信号便等同于神经元输出。假设现在神经元输入为x = ，权值为w(𝑡)，其输出变为y = f(∗ 𝑥)，则权值的调整量为

w(t) = δyx （5.4）

上面的公式表明，权值的调整量和输入输出的乘积有一定的数量关系，呈现正比的关系。这里我们可以把∆w(𝑡)看作是输入函数对于整个神经网络权值的影响因子。Hebbian原则需要预先设定一个权值饱和值的界定范围，以防止权值没有任何约束条件的增加。权值之间的调整公式为

w(t+1) = w(t) + δyx （5.5）

初始权值设置的一般原则是取在零附近的随机值，同时激活函数f可以取任意函数。

2.感知器学习原则

对于感知器学习原则，与Hebbian原则稍有不同，其学习信号为神经网络的期望输出与网络输出之间的差值。

对于每一个输入样本x，我们假设神经元的期望输出为a，实际的输出我们假定为y，则在感知器学习的规则中，权值之间的调整量为

w(t) = e(t)x （5.6）

其中我们将误差信号表示为

e(t) = a – y （5.7）

3.δ学习原则

δ学习原则又被人们称为最速下降法或者是梯度法。它是最常用的神经网络学习算法，并是一种有监督的学习算法。设神经元权值的误差和函数为E(𝑤)，神经网络当前权值设为w(𝑡)，那么神经网络权值的调整公式可以表达为：

w(t+1) = w(t) + w(t) （5.8）

对于每次的权值误差和函数E(𝑤)，我们需要得到E(w(𝑡 + 1)) < E(w(𝑡))，即误差和函数越来越小。

随后我们对误差和函数E(w(𝑡 + 1))研究其一阶泰勒展开式，可以得到如下公式：

E(w(t+1)) = E(w(t)) + w(t) （5.9）

其相似等于E(w(𝑡)) + ，这里，g(𝑡)表示的是E(w(𝑡))在t位置时的梯度。这就是梯度下降法的基本原理。为了保证目标函数值的准确性，我们一般把求解函数设定为一个线性函数。δ学习规则就是使得神经元权值w(𝑡)最终输出的误差值的和函数E(w(𝑡))最小，即达到收敛条件。我们设网络输出为a，则误差和函数的表达式：

E = = [– f] （5.10）

又根据梯度下降的原理可以得到：

w(t) = f()x （5.11）

其中神经网络的权值的初始值是默认的，一般取（0,1）之间的随机值。

5.4 BP神经网络

5.4.1 BP神经网络结构

误差反向传播算法（Error Back－propagation Training）简称 BP 神经网络。BP神经网络是一种前反馈性质的网络结构，其网络结构的主要特点是信号在神经元之间正向传输，得到网络输出以后与期望输出相比较，得到误差值，误差值再反向传播，用来改变初始权值和阈值，不断迭代直到所输出的误差和最小。在这个过程中，神经网络的权值和阈值通过误差的反向传播而不断调整。如图3-2所示，设定，, …分别是神经网络预先设定的输入特征值，，，表示神经网络的网络输出值，中间有若干隐含层称之为神经网络之间的连接权值。研究表明，这种简单的 BP 神经网络结构可逼近任意非线性的函数，是一种典型的函数映射之间的关系。

.

. . .

. . .

. . .

输入层 隐含层 输出层

图5-4 神经网络结构图

5.4.2 BP神经网络学习算法

（1）BP神经网络的正向传播原理

输入层中各神经元接收外部数据的输入并将其传递给中间隐含层神经元，其单元数由数据处理器构造形成的输入样本维数决定。设输入层的输入变量数为n个其中训练样本个数为m时，则输入变量可表示为：

= (，，…，)，k = 1,2，…，m （5.12）

中间隐含层对数据信息的变换，并依据数据信息变换能力的要求可设计单个或多个隐层结构。若以Ｐ表示隐层神经元个数，则其输入可表示为：

= - ，j = 1,2,…,p （5.13）

其中，为输入层与隐层神经元的连接权重；为隐层神经元的阈值；若以sigmoid函数f(x) = 为传递函数，则隐层神经元输出可表示为：

 = （5.14）

输出层接收最后一个隐层神经元信息的输出并做出进一步处理。若以q表示输出层变量数，则输入模式对应的输出向量为：

= (，，…，)， k = 1,2,…,m （5.15）

若以，分别表示隐层一输出层间神经元的连接权重、阈值，则神经网络的输出层单元的输入、输出分别为：

= - ，t = 1,2，…，q （5.16）

= （5.17）

（２）BP神经网络的误差反向传播原理

根据由输入层一隐含层一输出层的过程，基于第kＡ个学习训练样本，可以得到样本期望输出与实际输出间的误差与全局总误差E:

= （5.18）

E = （5.19）

在一定的学习率条件下，为使以梯度原理实现不断减小，可得到 = -，进而可推导出其调整量为：

= （5.20）

= ()t(1 - ) （5.21）

同理输出层阈值调整量，、隐层权重与阈值调整量、分别为： = （5.22）

= （5.23）

= （5.24）

= ()(1-) （5.25）

至此，BP神经网络的一个误差反传过程完成，整个模型完成一次训练学习。BP神经网络会不断地进行误差的反向传播，并依据推导出的调整值对输出层与隐含层的权值与阈值进行修正，直至预测值与实际值之间的偏差满足设定的精度要求。

5.5基于时空关联性的改进PSO-SVR算法模型

本文最终是将改进的PSO-SVR算法模型和得出的时空关联性相结合进行流量的最终预测。由于短时交通预测具有不确定性和规律性弱等特点，仅是依据PSO-SVR这样的非数学模型利用历史数据进行预测始终会有偏差，当路面发生一些突发情况，如某个时刻需对路口进行限流限速等，此时PSO-SVR模型的预测精度就会明显下降；所以，此时可以通过其相邻路口的交通流量即根据时空关联性来进行主要预测，从而可以得出更加准确的预测精度。因此将PSO-SVR模型与本文之前得出的时空关联性进行结合，优势互补，取长补短，可以更好的克服短时交通预测的不确定性和弱规律性，从而得出更好的预测效果。本文利用BP神经网络的特性，分别将PSO-SVR和时空关联性得出的流量作为BP神经网络的两个输入，经过隐层处理后传向输出层。如果BP网络的输出层与理想的输出有差别，就将误差通过隐藏层向输入层传递，如此可以使误差分给各层单元，并作为修正各权值的依据。实验采用3层BP神经网络，训练次数为1000，训练目标为0.0001，学习率为0.01。流程图如下：

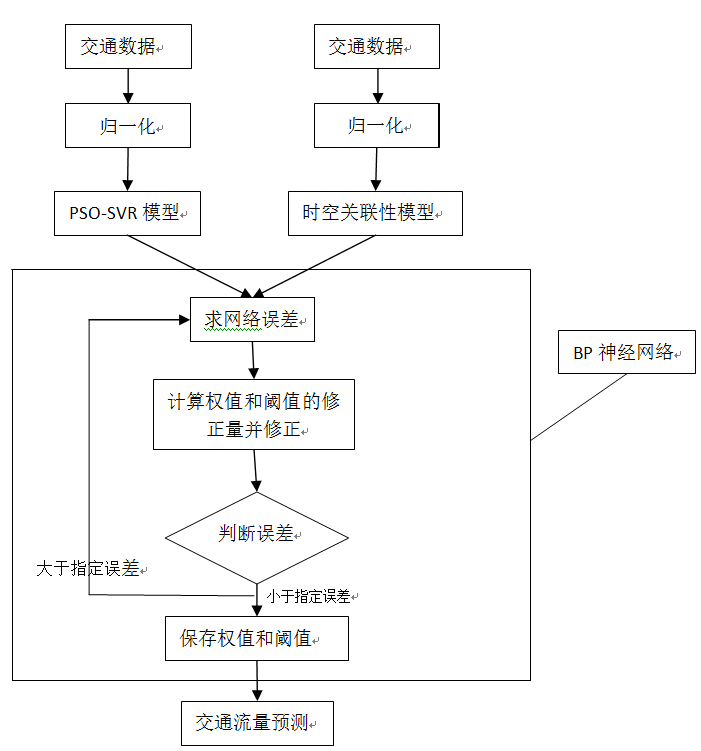


图5-5 算法流程图

预测结果如下所示：

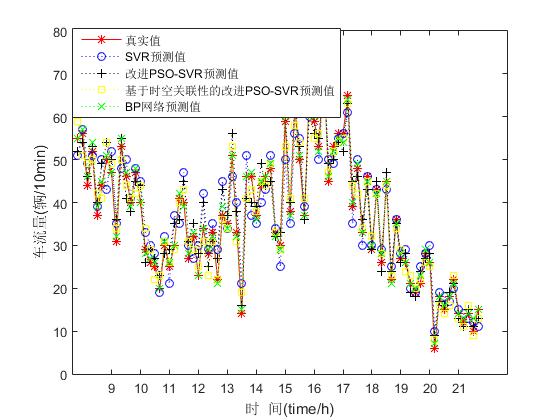


图5-6

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | MAE | RMSE | EC |
| SVR | 7.2845 | 2.7764 | 0.91356 |
| PSO-SVR | 5.6375 | 2.1298 | 0.93489 |
| BP神经网络 | 5.8439 | 2.4513 | 0.92464 |
| 时空关联的PSO-SVR | 4.9879 | 2.1066 | 0.96766 |

表5-2

结合图5-6和表5-2可以看出，基于时空关联性的改进PSO-SVR预测模型优于传统的SVR预测模型也优于单纯的BP神经网络预测模型，且具有较高的预测精度，也验证了此方法的有效性。

5.6本章小结

本章在改进PSO-SVR模型的基础上，考虑了交通中的时空影响，加入了时空相关性分析，通过分析周边路口交通流量对待测路口的影响，利用BP神经网络的特性整合改进的PSO-SVR模型，得到了可以进行不断修正实时预测的短时交通预测模型！

**第6章 总结与展望**

本文全面评述了当前交通流预测的研究现状，简要介绍了交通流三大参数，详细阐述了 SVM的基本原理，在已有的研究成果，将SVM理论运用到交通流预测中。在参数选择的问题中，使用了PSO 优化算法及改进的 PSO算法对参数的优化进行了系统的研究，最终构建了一个考虑时空关联性的，利用BP神经网络进行不断迭代修正实时预测预测的模型。本文主要开展了以下研究工作：

1、介绍了当前困扰着各大城市的交通拥堵问题及本课题在这一前提下所具有的研究意义，系统总结了当前的研究进展，为后续的建模提供决策支持。

2、根据交通流的特性，建立短时交通流预测模型。采用SVR作为主要研究模型，使用PSO 优化算法和改进的 PSO算法，建立参数优化的预测模型。将仿真结果与传统SVR的模型进行比较，实验结果表明，该方法能够满足当前研究对实时性和准确性的要求。

3、进行时空关联性分析。在短时交通流量预测中，交通流具有不确定性和规律性弱等特点，仅仅依靠待测路口的历史数据进行预测，其精度难以提高，特别是在某些突发事件的情况下，预测精度会大打折扣。本文将基于先验数据的非数学模型方法与实时数据的交通流量关联方法结合起来，采用改进的PSO-SVR方法对待测路口的时间序列进行训练学习，获得尽可能多的基于时间的流量关系。同时利用交通数据挖掘出待测路口于其它路口的时空关联性，利用相关上游路口的流量预测待测路口下一时刻的流量，将两者结合取长补短通过BP神经网络不断迭代在线修正两者权值直到误差足够小从而实时预测出最终的交通量。在实际应用中减少了人为干预的不良影响，在一定意义上实现了交通流预测的智能化。

**创新点**

综合分析了当前的研究现状，并对比了传统的SVR预测方法设计了短时交通流预测系统。主要创新点如下：

1、提出了一种改进PSO 的优化算法对SVR模型参数进行寻优。与以网格搜索参数的模型相比，采用改进算法优化参数，使得模型在实际应用中预测精度更高；与传统的 PSO 算法相比，改进的算法能够达到全局最优解，更能满足预测的实时性和准确性。

2、采用改进的PSO-SVR方法对待测路口的进行预测，同时利用交通数据挖掘出待测路口于其它路口的时空关联性，利用相关上游路口的流量预测待测路口下一时刻的流量，将两者结合取长补短通过BP神经网络不断迭代在线修正两者权值直到误差足够小从而实时预测出最终的交通量。