

基于数据挖掘技术的交通流预测模型^①



邓 晶, 张 倩

(南京工程学院 计算机工程学院 南京 211167)

通讯作者: 邓 晶, E-mail: dengj@njit.edu.cn

摘 要: 本文针对交通数据挖掘领域的交通流预测问题进行研究和实现. 主要对数据挖掘技术应用于交通流数据的特征选择和交通流预测模型的建立提出算法. 在对采样数据进行清洗后, 以分类与回归决策树作为基学习器, 采用梯度提升决策树进行回归拟合, 计算出交通数据的特征重要度. 并以此重要度作为自适应特征选择的依据. 其次, 采用聚类算法对选取后的特征数据进行聚类分析, 缩小样本大小的同时, 同类数据更加相似. 最后, 以实时数据匹配相应聚类作为训练数据集, 使用经过人工鱼群算法优化参数后的支持向量机进行交通流预测. 本文结尾通过实验数据论证本文所提出的算法和模型.

关键词: 数据挖掘; 交通流预测; 特征选择; 梯度提升决策树; 支持向量机; 人工鱼群算法

引用格式: 邓晶, 张倩. 基于数据挖掘技术的交通流预测模型. 计算机系统应用, 2019, 28(7): 114-120. <http://www.c-s-a.org.cn/1003-3254/6969.html>

Traffic Flow Forecasting Model Based on Data Mining Technology

DENG Jing, ZHANG Qian

(School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, China)

Abstract: In this study, traffic flow forecasting in the field of traffic data mining is studied and implemented. This paper presents an algorithm for feature selection of traffic flow data and establishment of traffic flow prediction model based on data mining technology. After cleaning the sampled data, the classification and regression decision tree are used as base learners, and the gradient lifting decision tree is used for regression fitting to calculate the characteristic importance of traffic data. The importance is used as the basis of adaptive feature selection. Secondly, the clustering algorithm is used to cluster the selected feature data, which reduces the sample size and makes the similar data more similar. Finally, real-time data matching and clustering are used as training data sets, and support vector machine is used to predict traffic flow after parameters optimization by Artificial Fish Swarm Algorithm (AFSA). At the end of this paper, experimental data are presented to demonstrate the proposed algorithm and model.

Key words: data mining; traffic flow forecasting; feature selection; gradient lifting decision tree; Support Vector Machine (SVM); Artificial Fish Swarm Algorithm (AFSA)

随着城市规模和交通出行需求的不断增长, 人们对交通预测服务的需求越来越强烈. 目前, 在智能交通流预测领域, 主要采用基于非线性理论的预测模型、基于神经网络的预测模型、多模型融合的预测模型等. 近年来, 国内一些学者采用混沌时间序列建立的预测

模型具有很强的鲁棒性和自适应性. 如一元或多元混沌时间序列交通流预测模型. 但城市道路是一个庞大的复杂的网络^[1]. 每条道路都会受其他道路影响. 一些模型采用协整分析, 但未能考虑本身的自回归因素. 影响交通的因素非常多, 而且存在着非线性的相互作用.

① 收稿时间: 2018-12-30; 修改时间: 2019-01-18, 2019-01-28; 采用时间: 2019-01-31; csa 在线出版时间: 2019-07-01

支持向量机 (SVM) 是近十几年出现的一种先进的数据挖掘方法。SVM 通过非线性变换将原始空间的数据样本投影到新的高维空间, 非常适合解决非线性、高维度的数据。由于交通流数据具有数据量大、维度高、非线性等特征, SVM 采用结构风险最小化的学习方式可以在小样本的学习环境下, 采用凸优化的学习方式, 有效地避免过拟合和局部最优, 从而使交通流预测性能大大提高。对于样本量较大的交通流数据, 本文在对数据进行清洗和特征选择后, 按照不同的属性对数据进行聚类分析, 同一类别内的数据相似度更高, 且数据样本量相对减小, 使用支持向量机回归进行预测能够获得更好的预测效果。最后的实验数据也证明了这一点。

1 基于梯度提升决策树的自适应特征选择算法

1.1 问题的提出与总体解决思路

随着交通规模的不断扩大, 影响交通状态的因素也越来越多。特征选择是从复杂的交通影响因子中选出有效的特征作为交通流预测模型的数据输入。可以降低数据计算量, 提高交通预测系统运行效率和运行准确率。决策树算法在分裂形成决策树的过程中, 天然地对特征进行了选择, 但单棵的决策树很不稳定, 且仅仅只能对特征重要度进行计算, 并不能直接选择出合适的特征集合。因此, 把集成学习方法和决策树算法相结合, 提出一种基于梯度提升决策树的自适应特征选择方法。其中迭代生成的若干决策树综合计算得出了特征的重要度, 解决了单棵决策树的局限性。通过二次下降法选择合适特征, 实现对高维度交通流数据的自适应选取。

1.2 单棵 CART 树中特征的基尼指数和分裂次数

分类与回归决策树 (CART) 既可以进行分类又可以用于回归。既可以处理离散数据又可以处理连续数据。CART 在进行分类时, 使用基尼指数选择最优特征。决策树作为基学习器, 在它的分类与回归过程中计算出每棵树的基尼指数以选择最优特征, 同时确定该特征的最优二值分割点, 从决策树的根节点开始依次选择能使分割后样本集基尼指数最小的特征和值。

对于给定的一个样本 D , 假设确定类别个数, 如果样本 D 根据特征 F 的某个值 f , 其被分隔成 D_1 和 D_2 两部分, 则在特征 F 的条件下, 样本 D 的基尼指数

可以定义为:

$$Gini(D, A) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (1)$$

1.3 梯度提升决策树 (GBDT) 模型

梯度提升决策树 (GBDT) 是一种迭代的决策树算法。它通过构造一组弱学习器, 即多个单颗的 CART 树, 并把这些弱学习器的结果累加起来作为最终的预测输出^[2]。每个分类器在上一轮分类器的残差基础上进行训练。训练的过程是通过降低偏差来不断提高最终分类器的精度。其模型可表示为:

$$f(x) = \sum_{m=1}^M f_m(x) \quad (2)$$

其中, M 为 CART 树的个数, $f_m(x)$ 表示决策树, 是通过拟合损失函数的负梯度得到的。算法中需要利用回归树对损失函数的负梯度进行拟合, 有关参数如下:

设训练数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, 其中 $x_i \in \chi \subseteq R_n$, χ 为输入空间, $y_i \in \gamma \subseteq R$, γ 为输出空间。如果将输入空间 χ 划分为 J 个互不相交的区域 R_1, R_2, \dots, R_J , 并在每个区域上确定输出的常量, 那么树可以表示为:

$$T(x, \otimes_m) = \sum_{j=1}^J \gamma_j I : x \in R_j \quad (3)$$

其中, 参数 $\otimes = \{(R_1, r_1), (R_2, r_2), \dots, (R_J, r_J)\}$, 表示树的区域划分和各区域上的常数。 J 是回归树的复杂度及叶子结点的个数。

单棵决策树的梯度提升算法采用的是前向分布算法。首先确定初始提升树 $f_0(x) = 0$, 第 m 步的模型为:

$$f_m(x) = f_{m-1}(x) + T(x, \otimes_m) \quad (4)$$

利用梯度提升算法使得损失函数最小, 从而确定下一棵决策树的参数 \otimes_m

$$\otimes_m = \arg_{\otimes_m} \min \sum_{i=1}^N L(y_i, f_{m-1}(x_i) + T(x_i; \otimes_m)) \quad (5)$$

1.4 基于梯度提升决策树的自适应选择算法

在梯度提升决策树训练完成得到最终模型后, 迭代生成了 M 棵 CART 树, 同时计算了每个特征的基尼指数和分裂次数。因此定义每个特征 x_i 的全局重要度 $I(x_i)$ 如下:

$$I(x_i) = \frac{\sum_{j=1}^M Gini(x_i)_j}{M} + division\pi(x_i) \quad (6)$$

其中, $Gini(x_i)$ 为特征 x_i 经过 0~100 归一化处理后的基尼指数, $\sum_{j=1}^M Gini(x_i)$ 为特征 x_i 在 M 棵 CART 数中基

尼指数的均值, $division(x_i)$ 为特征 x_i 经过 0~100 归一化处理后的总分裂次数.

在得出特征重要度基础上, 为了能够合理地选择

有效特征集合, 设计了一种二次下降法, 提出了一种基于梯度提升决策树的自适应选择算法. 该算法流程图如图 1

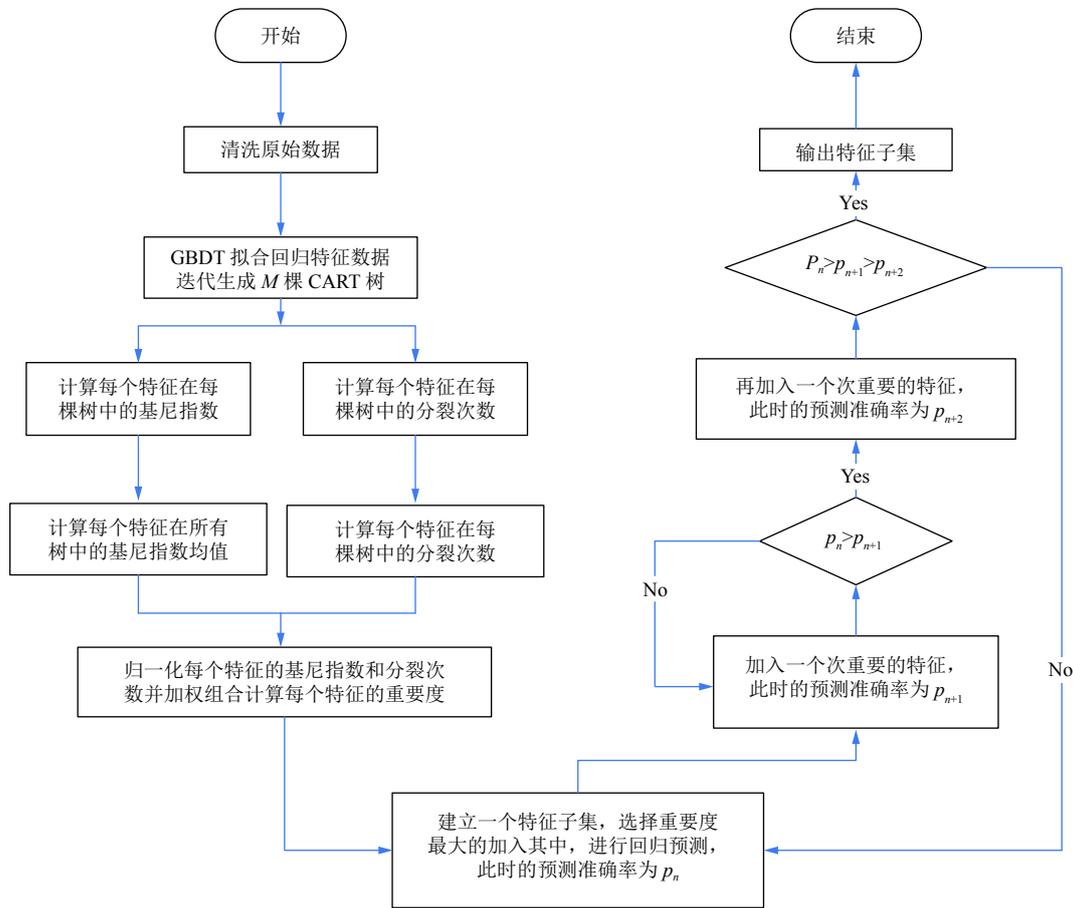


图 1 梯度提升决策树自适应选择算法

如图 1 所示, 如果连续两次加入次重要的特征进入特征子集后, 预测准确率都不会提高, 则结束算法, 输出选择后的有效特征集合, 此为二次下降法.

2 基于聚类分析和支持向量机回归的交通流预测

2.1 总体思路

交通流数据具有非线性、高维度的特性, 支持向量机回归算法在处理小样本、非线性、高维度数据时, 具有很好的预测效果. 因此, 我们采用支持向量机作为核心预测算法. 但交通流数据样本数量比较庞大, 可能会导致支持向量机的训练速度缓慢, 且预测容易产生偏差. 针对这一问题, 我们首先采用 K-means++ 聚类算法将数据按照交通流量、天气、节假日等属性进行聚

类得到若干类别. 同一类别内的数据相似度高, 数据规模也相对较小. 然后, 通过实时数据匹配相应类的数据进行样本训练, 再使用支持向量机进行预测能够获得更好的预测效果. 在支持向量机模型参数的选择的问题上, 采用人工鱼群算法对支持向量机回归的参数进行优化, 采用合适的参数能够进一步提高预测精度和推广能力.

2.2 交通流数据的聚类分析

在交通流数据聚类分析中, 我们采用 K-means++ 聚类算法. 传统的 K-means 算法需要人为确定聚类中心, 不同的聚类中心可能导致完全不同的聚类结果. K-means++ 在 K-means 算法基础上, 对初始聚类中心的选择进行了优化. 它的核心思想是选择的初始聚类中心之间的距离要尽可能远^[3].

对于给定的数据集 $D_n = \{x_1, x_2, \dots, x_n\}$, 选择其中一个 x_i 点作为第一个初始聚类中心 x_i , 计算每一个数据点与已选择的最近的聚类中心 x_i 之间的距离 $d(x_i)$, 计算每一个数据点作为聚类中心的概率 $P(x_i)$

$$P(x_i) = \frac{d(x_i)^2}{\sum_{x_j \in D} d(x_j)^2} \quad (7)$$

传统的聚类算法需要不断地重复聚类过程, 来得最优 K 值, 这需要耗费大量的时间. 而交通流预测需要实时快速地处理数据, 故使用轮廓系数 SC 来快速确定 K 值. SC 指标表示的是聚类后各样本之间的紧密程度和各类之间的离散程度^[4]. 同一样本之间距离越小, 不同类之间的距离越大, 即 SC 的值越大, 表示聚类效果越好.

$$SC(i, j) = \left| \frac{b(i, j) - d(i, j)}{\max(b(i, j), d(i, j))} \right| \quad (8)$$

式中, $SC(i, j) \in [0, 1]$. $b(i, j)$ 表示第 i 类中的第 j 个样本与其它每一个类中样本平均距离的最小值. $d(i, j)$ 表示第 i 类中的第 j 个样本与第 i 类中其它样本的平均距离.

为了评价整个数据集中所有样本的聚类情况, 计算所有样本的 SC 指标平均值 $\overline{SC}(K)$, 即可求出最佳聚类数 K .

2.3 支持向量机回归模型

假设训练样本数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, 其中 $x_i \in R^m$ 为数据样本点, 其属性的分量称为样本的特征. $y_i \in \{+1, -1\}$ 为数据样本类别标签, m 为数据样本的特征维数, n 为训练样本的个数. 使用核函数变换后的非线性支持向量机回归预测函数如下

$$f(x) = \omega^T \cdot \varphi(x) + b = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (9)$$

其中, $\varphi(x)$ 为原输入样本输入值转化为高维空间样本中的输出值. α^* 为支持向量机的非线性优化数学模型最优解. $K(x_i, x)$ 为核函数. 偏移量 b 为:

$$b = \frac{1}{N_{NSV}} \left\{ \sum_{0 \leq \alpha_i < c} \left[y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) K(x_j, x_i) - \varepsilon \right] + \sum_{0 \leq \alpha_i^* < c} \left[y_i - \sum_{x_j \in SV} (\alpha_j - \alpha_j^*) K^*(x_j, x_i) + \varepsilon \right] \right\} \quad (10)$$

其中, ε 为非负松弛变量, C 为惩罚因子.

在非线性支持向量机模型中, 核函数的选取是最关键的. 常用的核函数包括线性核函数、多项式核函数、高斯径向基核函数、Sigmoid 核函数等, 通过对

比, 本文采用高斯径向核函数进行支持向量机回归训练.

2.4 人工鱼群算法优化支持向量机参数

支持向量机回归中参数的选择对模型的预测效果有着十分重要的影响. 即损失函数 ε 、惩罚因子 C 、高斯径向基函数 δ , 如何组合这三个参数成了支持向量机回归不可避免的问题. 由于参数选择的复杂性, 目前还没有确定的理论来指导, 而群体智能算法在参数优化问题上有很好的效果. 本文使用群体智能化算法中的人工鱼群算法对支持向量机的参数进行寻优. 人工鱼群算法对初值和参数不太敏感, 具有自适应搜索空间的能力, 且鲁棒性强, 收敛速度快. 更重要的是, 能够避免陷入局部极值而失去全局最优.

在人工鱼群算法中包括三种主要的算子: 聚群算子、追尾算子和觅食算子^[5]. 这三种算子是算法的核心, 并且决定算法的性能和最优解搜索的准确度.

人工鱼群算法中每条人工鱼首先分别试探执行聚群算子和追尾算子, 通过目标函数计算其适应度是否得到改善. 将适应度改善较大的算子作为该条人工鱼的更新算子; 否则执行觅食算子; 若这条人工鱼达到觅食算子的最大尝试次数后, 适应度依旧没有改变, 则在自己周围随机移动一个位置. 所有人工鱼执行上述操作后, 最终人工鱼群集结在几个局部最优解周围, 且全局最优极值区域周围也有较多人工鱼集结.

2.5 基于 K-means++ 聚类算法和支持向量机回归的交通流预测

本文研究的交通流数据包括交通流量、车速、占有率、日期、天气、交通时间等. 在分析各方面因素后, 设计了一种基于 K-means++ 聚类算法和支持向量机回归的交通流预测模型. 该模型的流程图如图 2.

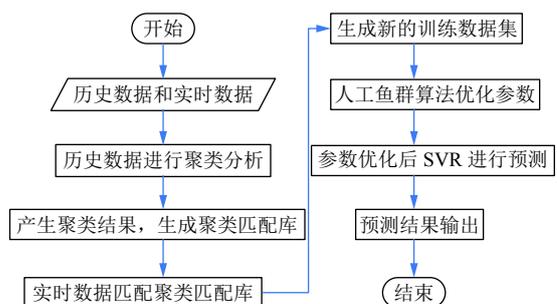


图 2 基于聚类分析和支持向量机回归的交通流预测模型

3 实验结果和分析

3.1 数据来源与评价指标

本文实验数据来源于美国加州交通局的 PEMS 系统. 在 PEMS 系统中, 随机选取一个时段的数据. 该段数据已经被处理成 5 分钟一个间隔的交通流数据. 包括车辆流数、车速及车道占有率等. 另外, 引入天气因子, 对天气情况特征做出说明; 引入时间类别, 对普通工作日、节假日、重大节日等作出区分.

针对本文的最终目的是为了准确地进行回归预测, 我们引入了广泛应用于回归问题的评价指标, 分别为平均绝对百分比误差 (MAPE) 和均方跟误差 (RMSE). 其中, MAPE 表征了预测值偏离实际值的程度. RMSE 表征了误差大小以及误差分布的离散程度.

3.2 实验一: 基于梯度提升决策树的自适应特征选择

为了验证本文提出算法的有效性, 选取了随机森林的特征选择算法和基于 CART 决策树的特征选择算法与本文的梯度提升决策树算法进行对比实验.

首先, 对数据进行清洗和去噪. 之后选取编号为 11、12、13 的三个路段从 2017 年 12 月 14 日到 12 月 20 日之间的交通流数据; 编号 14 路段作为预测路段. 以此构造特征变量集合. 其中, ch1 到 ch21 共 21 个特征, 包括编号 11 到 14 路段现在、5 分钟前、十分钟前的交通流量, 编号 11 到 14 路段现在车道的平均占有率, 以及天气状态和节假日状态. 将该特征集作为原始训练数据集, 应用于本文的特征选择算法和对比的算法模型中.

表 1 显示了梯度提升决策树的特征选择算法和其它两种算法对数据进行特征选择后的结果.

表 1 特征选择结果

特征选择算法	选择的特征
基于梯度提升决策树的特征选择算法	ch1、ch13、ch2、ch14、ch21、ch10、ch4、ch15、ch18
基于随机森林的特征选择算法	ch1、ch2、ch3、ch4、ch7、ch10、ch13、ch14、ch15 ch19、ch21
基于 CART 决策树的特征选择算法	ch1、ch2、ch3、ch4、ch5、ch7、ch10、ch11、ch12 ch13、ch14、ch15、ch17、ch18、ch19、ch20、ch21

从表 1 可以看出, 基于梯度提升决策树自适应选择算法选择出了 9 个特征, 而另外两种比较算法选出的特征则要多一些. 为了得到最终的结论, 使用三种算法选出的特征分别构造训练集和测试集, 分别将其放

入 ARIMA 模型、BP 神经网络模型和支持向量机模型进行交通流预测. 使用 MAPE 指标和 RMSE 指标对预测结果进行评判, 得到表 2、表 3 的结果.

表 2 特征选择的 RMSE 效果对比

特征选择算法	ARIMA	BP 神经网络	支持向量机回归
梯度提升决策树特征选择算法	79.82	49.8	46.2
随机森林特征选择算法	84.3	53.42	52.93
CART 决策树算法	93.29	63.34	60.13
未经选取的特征	94.23	65.82	61.92

表 3 特征选择的 MAPE 效果对比

特征选择算法	ARIMA	BP 神经网络	支持向量机回归
梯度提升决策树算法	0.0976	0.0731	0.0705
随机森林特征选择算法	0.1083	0.0852	0.0861
CART 决策树算法	0.1292	0.0983	0.0914
未经选取的特征	0.1331	0.1042	0.0986

从表 2 中可以看出, 在对特征集进行提取前后, RMSE 值的区别都很明显. 即使是在三种算法中特征选取效果较差的 CART 决策树算法, 相对于未经选择的原始特征, 预测模型的 RMSE 值也有明显下降. 而使用梯度提升决策树自适应算法的 RMSE 值最低, 预测效果更明显. 从表 3 中看出, MAPE 的变化情况和 RSME 的类似. 综上, 在进行交通流预测之前进行特征选取是很必要的, 可以有效地降低 RMSE 和 MAPE 的值. 而本文提出的基于梯度提升决策树的自适应特征选择算法相较于其它两种算法, 性能有明显提升, 能够选取更好的特征组合, 实现对数据更好的描述, 最终使得交通流预测模型的预测精度更高.

3.3 实验二: 基于聚类分析和支持向量机回归的交通流预测

3.3.1 聚类分析

本文抽取了 2017 年 12 月的交通流数据进行聚类分析. 12 月一共有 31 天, 将其中的 2 日、11 日、18 日的交通流数据抽出, 作为测试样本进行验证. 剩下的 28 天数据进行聚类分析. 根据最佳聚类数计算方法, 即聚类数应该在 2 到 \sqrt{n} 之间. n 为数据样本数 28. 则最佳聚类数取值为 2 到 5 之间. 对 12 月份 28 天交通流量历史数据进行聚类, 采用轮廓系数来评价聚类效果, 不同 K 值对应的聚类结果和轮廓系数的实验数据如表 4 所示.

表4 不同K值对应的聚类结果和轮廓系数

聚类数	聚类结果	轮廓系数
K=2	{4、5、6、7、12、13、14、19、20、21、26、27、28} {1、3、8、9、10、15、16、17、22、23、24、25、29、30、31}	0.452
K=3	{3、9、10、16、17、23、24、25、30、31} {1、4、5、6、12、13、19、20、21、22、26、27、28、29} {7、8、14、15}	0.683
K=4	{23、24、25} {3、9、10、16、17、30、31} {1、4、5、6、12、13、19、20、21、22、26、27、28、29} {7、8、14、15}	0.796
K=5	{1、4、5、12、13、19、21、22} {3、6、20、26、27、28、29} {23、24、25} {7、8、14、15} {9、10、16、17、30、31}	0.703

由上表可知,当 $K=4$ 时,聚类效果最好.即最佳聚类数 K 为4,聚类结果为四个簇.第一类为圣诞节及连接的周末;第二类是周末;第三类是工作日且天气良好;第四类为普通工作日且天气状况差.由此可见,聚类后的每个类的特征都比较相似,类间特征差别比较明显.之后将预测的日期匹配到相应的聚类簇中,使用簇内数据作为训练样本,得到更好的预测效果.

3.3.2 支持向量机交通流预测

本文对2日、11日、18日部分时段进行交通流预测.计算预测时段前三个小时交通流量数据与四个簇类样本中心的欧式距离,得到某一个聚类簇的欧式距离最小,并将其归为一类.将这一簇内的数据作为训练样本数据集.

以下以11日的预测为例,说明实验结果.

当预测11日的交通流量时,计算得到与第三个聚类的欧式距离最小.选取第三个类中14天的数据为训练样本.使用人工鱼群算法计算得到最优参数组合 $(\epsilon, C, \delta)=(0.74, 69, 0.26)$.使用优化的参数组合建立预测模型得到预测结果.与传统的SVR预测结果和实际值的对比如图3

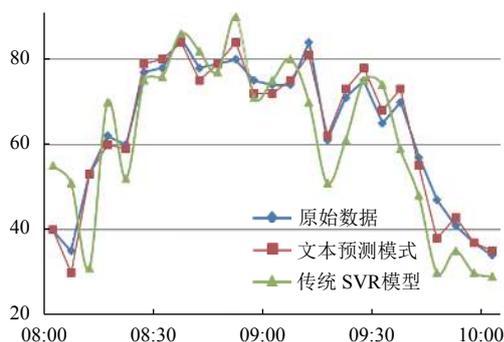


图3 交通流预测结果

为了验证本文提出的预测算法,将其与传统的支持向量机回归算法、ARIMA预测算法、BP神经网络预测算法进行对比.在相同的实验环境下,不同算法的MAPE和RMSE指标如表5.

表5 不同算法的MAPE和RMSE对比

预测算法	MAPE	RMSE
传统支持向量机算法	0.0734	48.6
聚类分析和支持向量机回归算法	0.0512	31.24
ARIMA预测算法	0.0956	76.78
BP神经网络预测算法	0.0741	50.08

3.3.3 实验结果分析

由图3可以看出,基于聚类分析和支持向量机回归预测模型要优于传统的支持向量机回归预测模型.其预测曲线与原始实际值拟合的更好,变化趋势也与原始实际值基本一致.虽然传统的支持向量机回归算法在某些点与实际值更接近,但是其预测的变化趋势与实际值并不一致.特别是,在交通流变化幅度较大时,基于聚类分析和支持向量机回归预测模型对交通流的拟合程度比传统的支持向量机回归算法明显要好.这表明本文提出的预测模型使用聚类对数据区分后进行训练,并对参数进行优化,可以有效的削弱随机数据导致的误差,从而有助于提升短期交通流预测的准确性.

由表5可以看出,ARIMA算法预测精度最差.传统的支持向量机和BP神经网络算法预测精度基本相似.但支持向量机还是略胜一筹.两种SVR预测模型都优于两种经典的预测算法.而基于聚类分析和支持向量机回归预测模型的两项指标又优于传统的SVR.其中MAPE降低了30.2%,RMSE降低了35.7%.在以上四种预测模型中,本文提出的基于聚类分析和支持向量机回归交通流预测模型的预测精度是最优的.

4 结束语

本文使用梯度提升决策树自适应特征选择算法,在对交通流数据进行特征选择后,通过聚类分析和支持向量机回归算法建立预测模型,实现对非线性、高维度交通流数据的有效预测.经实验验证,优于传统的预测模型且更接近于实际交通流数据.

由于时间和水平有限,本文的研究内容有待于进一步拓展.主要有以下几点:

由于影响交通流的因素很多,本文只考虑了其中重要的若干因素,需要考虑将更多的影响因子加入到模型中.

梯度提升决策树通过不断迭代弱学习器生成强学习器.由于弱学习器之间存在依赖关系,难以并行训练数据.因此耗费时间比较长.接下来应进一步研究如何优化学习过程,以期满足交通流预测的实时性.

人工鱼群算法对 SVR 参数寻优收敛速度较慢,可以研究其与其它寻优算法结合,以达到更好的参数优化效果.

参考文献

- 1 李青. 城市交通流预测关键技术研究[硕士学位论文]. 绵阳: 西南科技大学, 2015.
- 2 Son J, Jung I, Park K, *et al.* Tracking-by-segmentation with online gradient boosting decision tree. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 3056–3064.
- 3 Arthur D, Vassilvitskii S. k-means++: The advantages of careful seeding. Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms. New Orleans, LA, USA, 2007: 1027–1035.
- 4 朱连江, 马炳先, 赵学泉. 基于轮廓系数的聚类有效性分析. 计算机应用, 2010, 30(S2): 139–141.
- 5 李晓磊, 薛云灿, 路飞, 等. 基于人工鱼群算法的参数估计方法. 山东大学学报(工学版), 2014, 34(3): 84–87.