

基于模糊 C 均值聚类和随机森林的短时交通状态预测方法

陈忠辉 凌献尧 冯心欣* 郑海峰 徐艺文

(福州大学物理与信息工程学院 福州 350116)

摘要: 交通拥堵长期以来是城市面临的主要问题之一, 解决交通拥堵瓶颈刻不容缓。准确的短时交通状态预测有利于市民预知交通出行信息, 及时采取措施避免陷入拥堵困境。该文提出一种基于模糊 C 均值聚类(FCM)和随机森林的短时交通状态预测方法。首先, 利用一种新颖的融合时空信息的自适应多核支持向量机(AMSVM)来预测短时交通流参数, 包括流量、速度和占有率。其次, 基于 FCM 算法分析历史交通流, 获取历史交通状态信息。最后, 利用随机森林算法分析所预测的短时交通流参数, 得到最终预测的短时交通状态。该方法在融合时空信息的同时采用随机森林算法应用于短时交通状态预测这一全新的研究领域。实验结果表明, FCM 对历史交通状态的评估方式适用于不同的高速路和城市道路场景。其次, 随机森林比其它常见的机器学习方法具有更高的预测精度, 从而提供实时可靠的短时交通出行信息。

关键词: 短时交通状态预测; 随机森林; 模糊 C 均值聚类; 自适应多核支持向量机

中图分类号: TP393; TP391

文献标识码: A

文章编号: 1009-5896(2018)08-1879-08

DOI: 10.11999/JEIT171090

Short-term Traffic State Prediction Approach Based on FCM and Random Forest

CHEN Zhonghui LING Xianyao FENG Xinxin ZHENG Haifeng XU Yiwen

(College of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, China)

Abstract: Traffic congestion is a problem faced by cities, and it is urgent for solving this issue. Accurate short-term traffic state prediction is benefit for citizens to know the traffic information in advance, and take the measures in time to avoid the congestion. In this paper, a short-term traffic state prediction approach is proposed based on Fuzzy C-Means (FCM) clustering and Random Forest. Firstly, a novel Adaptive Multi-kernel Support Vector Machine (AMSVM) which incorporates the spatial-temporal information is used to predict the short-term traffic parameters, including the volume, the speed and the occupancy. Secondly, the historical traffic data are analyzed based on FCM algorithm, and the historical traffic state information is got. Lastly, the Random Forest (RF) algorithm is utilized to analyze the predicted short-term traffic parameters, then the final predicted short-term traffic state is obtained. This method incorporates the spatial-temporal information as well as applying the Random Forest to a new research field of short-term traffic state prediction. The experimental results demonstrate that the evaluation method of historical traffic state based on FCM is suitable for both freeway and urban road scenarios. Besides, the Random Forest has higher prediction accuracy than other common machine learning methods, thus providing the short-term traffic information timely and reliably.

Key words: Short-term traffic state prediction; Random Forest (RF); Fuzzy C-Means (FCM) clustering; Adaptive Multi-kernel Support Vector Machine (AMSVM)

1 引言

交通拥堵是现代城市面临的一个严重问题, 不

可避免地导致经济损失、出行时间耗费并加剧环境污染。智能交通系统(ITS)的核心功能在于交通流诱导与控制, 通过发布实时有效的交通出行信息, 诱导司机选择最佳出行路径, 避免向拥堵区域的进一步集聚, 可有效节约出行时间和最大化路网通行效率。而实时交通出行信息依赖于短时交通状态的精准预测。

目前, 国内外业已提出多种交通状态的判别方法。Krause 等人^[1]以交通流量和速度作为输入变量,

收稿日期: 2017-11-20; 改回日期: 2018-04-16; 网络出版:

*通信作者: 冯心欣 fxx1116@fzu.edu.cn

基金项目: 国家自然科学基金(61601126, 61571129, U1405251), 福建省基金(2016J01299)

Foundation Items: The National Natural Science Foundation of China (61601126, 61571129, U1405251), The Foundation of Fujian Province (2016J01299)

提出一种基于模糊逻辑的交通状态识别算法,采用专家经验法将交通运行状态分为 6 个等级。巫威眺等人^[2]选取占有率和流量作为特征变量,将交通状态划分为畅通、拥挤和拥堵 3 级,提出基于 BP 神经网络的交通状态判别方法。张亮亮等人^[3]考虑了流量、速度和占有率等交通流参数对状态划分所起作用的不同,提出一种基于参数权重模糊 C 均值聚类的交通状态划分方法。Kong 等人^[4]采用模糊综合评价方法,根据交通流特性分配多个指标的权重,进而确定交通状态的划分。Deng 等人^[5]采用 K 均值聚类方法将交通数据划分成不同状态类别,然后利用多类支持向量机估计实时交通状态。Oh 等人^[6]提出一种基于 K 最近邻算法的顺序搜索策略的交通状态预测方法,取得了比传统单级搜索方法更好的性能。Ricardo 等人^[7]提出一种交通模式的动态分类方法,分为离线和在线两个阶段,实现了对日常交通模式的有效识别。Chen 等人^[8]采用长短期记忆网络方法并结合多种开源的异构数据来预测交通状态,取得了可观的预测效果。高林等人^[9]将随机森林分类算法用于交通状态识别,利用袋外数据估计模型的泛化性能,为交通状态预测提供了一种新思路。由于交通状态的随机性和不确定性,模糊方法和神经网络等人工智能算法被广泛应用于交通状态识别的研究中。然而,模糊算法容易陷入局部最优且对初始参数较为敏感;神经网络也普遍存在欠拟合或过拟合等问题,极大限制了算法的泛化性能。此外,这些方法往往基于对历史数据的分析和评估,以此来划分交通状态,不能满足用户对短时交通出行信息的需求。

不同于以上方法,本文首先利用一种新颖的融合时空关联信息的自适应多核支持向量机(AMSVM)来预测短时交通流参数,包括流量、速度和占有率。然后,采用 FCM 方法对历史交通流数据在 3 维空间上进行聚类分析,并从聚类结果中提取历史交通状态信息。同时,利用已知状态标签的历史数据训练一种随机森林模型,进一步分析所预测的短时交通流量、速度和占有率数据,得到最终预测的短时交通状态。本文方法可充分挖掘交通流参数与交通状态之间的隐含关系,融合时空信息的同时将 FCM 和随机森林方法应用于短时交通状态预测这一全新的研究领域,在保证预测精度和效率的同时,提供实时可靠的短时交通状态信息。

2 融合时空信息的短时交通流预测

本文方法的系统总体模型如图 1 所示,其说明了 AMSVM、FCM 和随机森林三者 in 短时交通状态

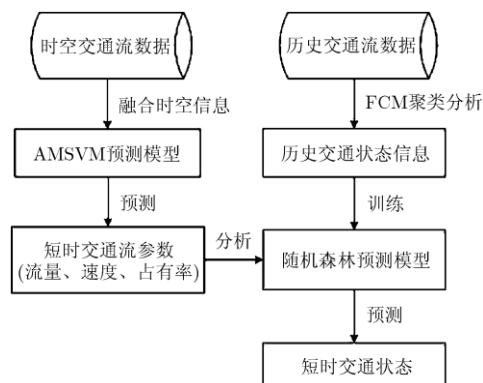


图 1 系统总体模型

预测方法上的关系和实现流程,其中首要环节是短时交通流参数预测。本文的先期工作发表于文献[10]和文献[11]中。为了获取表征交通状态的交通流特征参数并充分考虑时空相关性,利用一种基于文献[10]中提出的 AMSVM 的短时交通流预测方法。模型参数对交通流预测结果有着至关重要的影响,因此采用文献[11]中的自适应粒子群算法(APSO)来优化 AMSVM。

目标路段在下一时段(以固定的分钟数为一个时段)的交通流参数主要与对应时段的历史平均值、前几个时段的实时值以及空间关联路段交通流实时值有关。为了准确反映交通流变化规律对预测值的影响,在构建预测模型时,以目标路段的历史平均交通流数据、目标路段与空间关联路段在前几个时段的实时交通流数据作为输入变量。此外,定义回溯系数 N 表示利用前 N 个时段的交通数据进行实时预测。首先输入历史平均数据作为训练样本,利用 APSO 算法对 AMSVM 进行参数寻优;然后输入实时的时空数据训练优化后的 AMSVM,并预测下一时段的交通流参数。图 2 为完整的短时交通流预测模型,具体步骤如下:

步骤 1 初始化 APSO 环境变量,将待求参数当作粒子的位置向量;

步骤 2 输入目标路段历史平均交通流量数据到训练模块,通过 AMSVM 进行映射和回归,并利用 APSO 算法更新模型参数;

步骤 3 按照步骤 2 迭代更新,直到样本的训练误差满足精度限制,此时输出最优参数到预测模块;

步骤 4 输入目标路段与空间关联路段在前 N 个时段的实时交通流量数据到预测模块,然后在最优参数条件下训练 AMSVM,并预测输出下一时段的流量值;

步骤 5 针对速度和占有率,按照相同的方式

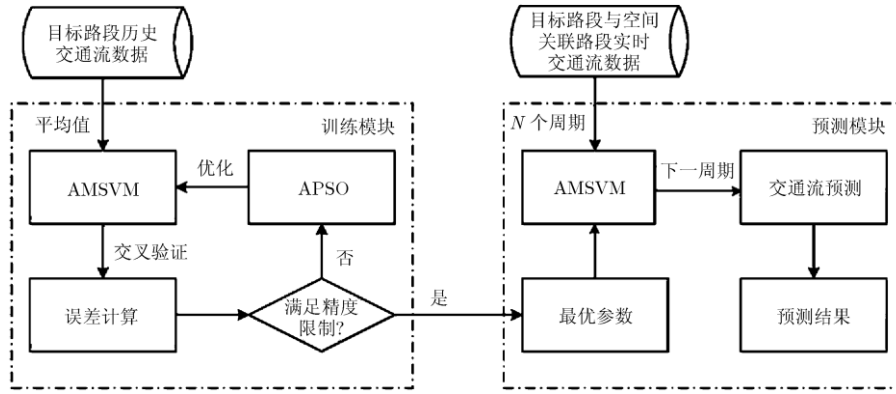


图 2 AMSVM 短时交通流预测模型

重复步骤 1-步骤 4，最终输出目标路段在下一时段的交通流量、速度和占有率预测值。

3 基于 FCM 的历史交通状态划分

如何根据短时交通流量、速度和占有率数据来判别下一时段的交通状态，需要进一步挖掘出两者之间所隐含的映射关系。由于预测得到的短时交通流参数是下一时段的不同特征变量，可采用监督学习中的分类方法来预测对应的短时交通状态。而监督学习方法的关键在于获取训练样本集中特征变量所对应的标签信息。因此，本文采用 FCM 算法对历史交通流数据进行聚类分析，从聚类结果中获取所需的状态标签。

3.1 FCM 原理

FCM 是一种无监督的模糊聚类算法，用于将高维空间分布的数据划分成特定的类别^[12]。它通过优化目标函数得到每个样本点对所有类中心的隶属度，从而决定样本点的类属以达到自动对样本数据进行分类的目的。

给定样本集 $X = x_1, x_2, \dots, x_n$ ，FCM 的思想是将 X 划分为 k 个模糊簇 $C = C_1, C_2, \dots, C_k$ ，目标函数为

$$J_m \mathbf{U}, X, C = \sum_{i=1}^n \sum_{j=1}^k \mathbf{u}_{ij}^m \|x_i - c_j\|^2 \quad (1)$$

其中， c_j 是第 j 个模糊簇 C_j 的聚类中心， \mathbf{u}_{ij} 表示样本点 x_i 属于簇 C_j 的隶属度， $\mathbf{U} = [\mathbf{u}_{ij}] \ 1 \leq i \leq n, \ 1 \leq j \leq k$ 表示隶属度矩阵； $m \ m \geq 1$ 为模糊加权指数，用来控制隶属度的影响。 \mathbf{u}_{ij} 满足约束条件^[13]：

$$\left. \begin{aligned} \sum_{j=1}^k \mathbf{u}_{ij} &= 1, & 1 \leq i \leq n \\ \mathbf{u}_{ij} &\in 0, 1, & 1 \leq i \leq n, \ 1 \leq j \leq k \\ 0 < \sum_{i=1}^n \mathbf{u}_{ij} < n, & 1 \leq j \leq k \end{aligned} \right\} \quad (2)$$

FCM 算法的具体步骤为：

步骤 1 初始化聚类个数 k ，模糊加权指数 m ，迭代次数 b_{\max} ，迭代终止阈值 ε ，以及隶属度矩阵 \mathbf{U} ；

步骤 2 计算模糊聚类中心：

$$c_j = \frac{\sum_{i=1}^n \mathbf{u}_{ij}^m \cdot x_i}{\sum_{i=1}^n \mathbf{u}_{ij}^m}, \quad j = 1, 2, \dots, k \quad (3)$$

步骤 3 更新模糊聚类的隶属度矩阵 \mathbf{U} ：

$$\mathbf{u}_{ij} = \left[\frac{\left(\sum_{t=1}^k d_{ij} / d_{it} \right)^{2/m-1}}{\sum_{t=1}^k \left(\sum_{t=1}^k d_{ij} / d_{it} \right)^{2/m-1}} \right]^{-1} \quad (4)$$

其中， d_{ij} 表示样本点 x_i 与聚类中心 c_j 的欧氏距离。

步骤 4 比较隶属度矩阵 $\mathbf{U}^{(b)}$ 和 $\mathbf{U}^{(b+1)}$ ，如果 $\|\mathbf{U}^{(b+1)} - \mathbf{U}^{(b)}\| \leq \varepsilon$ ，则说明目标函数 $J_m \mathbf{U}, X, C$ 已达到极小值，迭代终止；否则返回步骤 2，继续迭代。

3.2 FCM 评估历史交通状态

将目标路段的历史交通流数据当作训练样本，同时令样本集中的每个样本点 $x_i = x_{i1}, x_{i2}, x_{i3}$ ，其中 x_{i1} ， x_{i2} 和 x_{i3} 分别代表历史流量、速度和占有率数据。利用 FCM 算法对样本集在 3 维空间上进行聚类分析，按照 3.1 节所述的步骤进行训练，得到每个样本数据相对于每个聚类中心的隶属度，从而确定每个样本的所属类别。然后进一步分析交通流参数的变量特征，为每个类别赋予唯一的标签，表示该模糊簇能表征某一具体的交通状态。

4 随机森林模型的短时交通状态预测

目前，随机森林应用于交通领域的研究较少，如何利用其学习交通流参数特征并实现对短时交通状态的预测，是一个值得深入研究和探讨的问题。

4.1 随机森林算法

随机森林(Random Forest, RF)是一种利用多个树学习器进行分类和回归预测的组合算法^[14]。它

是一种有监督的集成学习方法，具有良好的抗噪声能力，效率高且可并行处理，不易陷入过拟合，且具有较强的泛化性能。

随机森林以 K 个决策树(Decision Tree, DT)为基分类器，进行集成学习后得到一个组合分类器。当输入待分类样本时，随机森林输出的分类结果由每个决策树的分类结果投票决定。随机森林包含两大随机化思想^[15]：

(1)Bagging 思想：从样本集 X 中有放回地随机抽取 K 个与 X 同样大小的采样集 $T_k = x_1, x_2, \dots, x_n$ (约占总体样本的 63.2%)，每个采样集 T_k 构造一个决策树；

(2)特征子空间思想：在对决策树的每个节点进行分裂时，从当前节点的属性集合中随机抽取一个属性子集，再从这个子集中选择一个最优属性用于划分。

4.2 随机森林预测模型

本节在前两节的基础上利用随机森林算法来学习交通流参数所表征的短时交通状态，以便尽可能地提高预测结果的准确性。图 3 所示为随机森林预测模型，待分类样本 $x_i = x_{i1}, x_{i2}, x_{i3}$ ，其中 x_{i1} ， x_{i2} 和 x_{i3} 分别代表目标路段在下一时段的交通流量、速度和占有率预测值，不同形状和颜色的标志代表不同的交通状态类别。给定一个待分类样本，每个决策树会计算该样本落入的叶节点处的训练样本中不同类别的百分比，以此确定该样本的分类结果^[16]；然后对随机森林中所有树的分类结果进行投票表决，输出目标路段在下一时段交通状态的最终预测结果。

5 实验结果与分析

为了验证本文所提出方法的可行性，选取 PeMS 系统的交通流数据进行分析^[17]。分别研究美国加州高速路和城市道路场景的交通状况，研究区域如图 4 和图 5 所示，图中红色标志表示目标预测点。高速路位于圣华金县，目标点经纬度为(37.97425°N, 121.247481°W)，时间为 2011 年 4 月 11 日；城市道路位于圣克拉拉县，目标点经纬度为(37.323066°N,

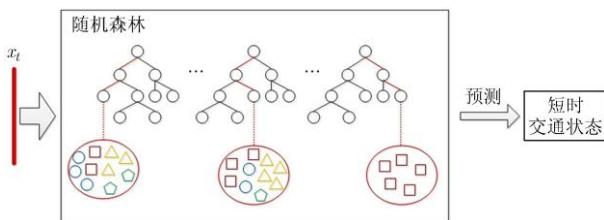


图 3 随机森林预测模型

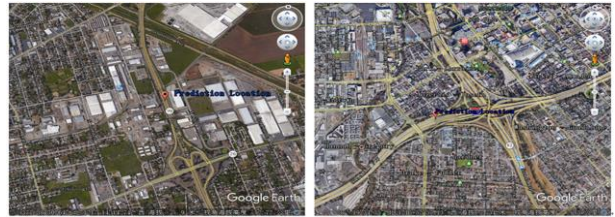


图 4 高速路场景

图 5 城市道路场景

121.896538°W)，时间为 2016 年 11 月 7 日。交通流取决于相邻道路段和区域通过交叉路口或立交桥而来的车辆。通过分析这些数据，预测目标点在对应时间(均为星期一)的短时交通状态。

5.1 短时交通流预测结果

首先利用 AMSVM 预测目标路段的短时交通流参数，分别对流量、速度和占有率进行单独预测。以 5 min 为一个时段统计这些参数，将一天 24 h 的数据划分成 288 个样本，同时令回溯系数 $N = 3$ 。训练模型前，对数据进行归一化处理，以减少训练时间、提高模型效率。将目标路段在前一周的工作日即星期一到星期五的历史平均交通流数据作为训练样本，利用 APSO 算法对模型进行参数寻优；然后输入目标路段及其空间关联路段在当前时刻之前 15 min 的实时交通流数据训练优化后的 AMSVM，并预测下一时段(未来 5 min)的交通流参数。按照第 2 节所述具体流程，分别得到高速路和城市道路的短时交通流量、速度和占有率的预测结果，如图 6 和图 7 所示。以高速路作为说明，表 1 显示其短时交通流参数的部分代表性预测值。

5.2 历史交通状态评估

利用 FCM 算法对目标路段在之前一周工作日

表 1 高速路短时交通流参数的部分预测值

时间	流量(veh/5 min)	速度(km/h)	占有率(%)
⋮	⋮	⋮	⋮
01:00	23.06	66.56	0.74
01:05	24.33	66.68	0.75
01:10	24.49	66.92	0.75
⋮	⋮	⋮	⋮
12:00	221.98	55.43	7.29
12:05	239.58	55.95	7.60
12:10	228.80	56.70	7.56
⋮	⋮	⋮	⋮
17:50	315.00	66.09	8.29
17:55	292.00	65.80	7.79
18:00	303.50	66.10	7.93
⋮	⋮	⋮	⋮

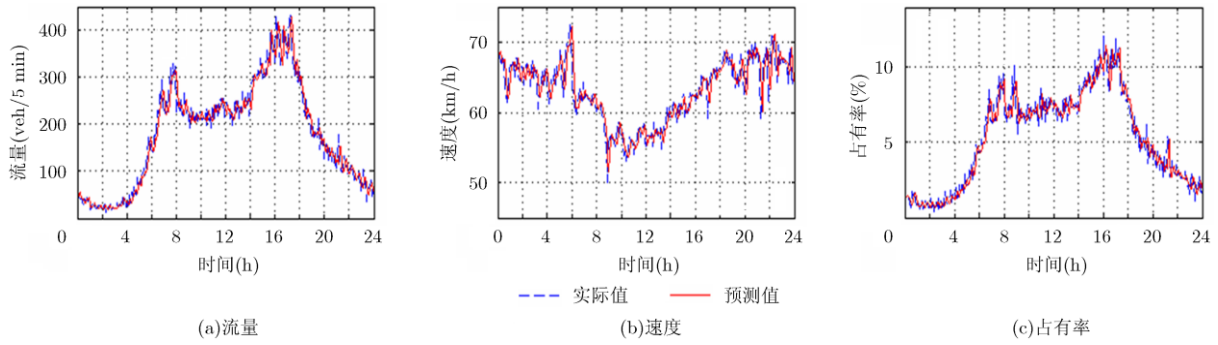


图 6 高速路短时交通流参数的预测结果

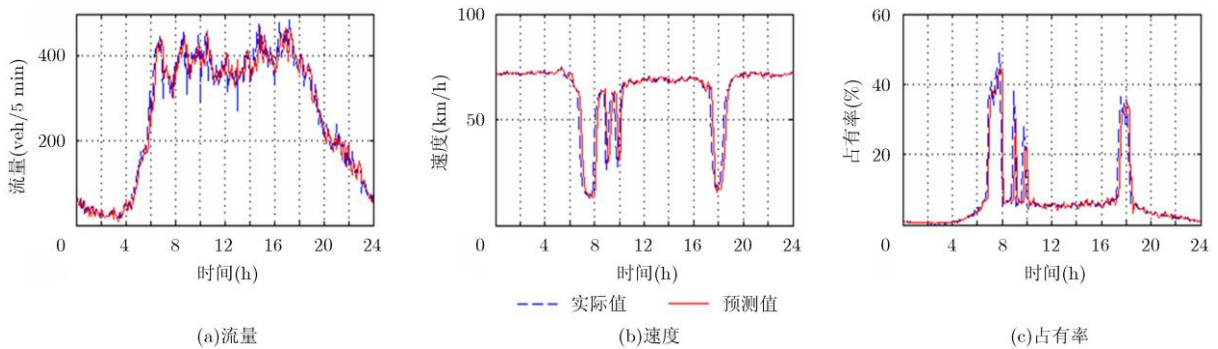


图 7 城市道路短时交通流参数的预测结果

的历史交通流数据进行聚类分析。由于不同道路场景下交通状况各异,交通流变量表现为不同的特征。本文将城市道路交通状态划分为 6 个等级,与文献[1]给出的等级划分建议相符。由图 4 和图 5 可看出,城市道路的路网结构比高速路更复杂;图 6 和图 7 也显示,城市道路交通流数据的实际流量和占有率最高值均比高速路上的对应数据要大,同时速度最低值比高速路上的对应数据要小。据此可将高速路交通状态划分成 5 个等级,而城市道路多一级的原因是它可以达到比高速路更程度的交通拥堵状态。显然,等级划分并不是绝对的,可以根据实际情况做出相应的改变。为此,FCM 对高速路和城市道路的交通流数据分别设定不同的初始参数。高速路初始参数设定为:聚类个数 $k=5$,模糊加权指数 $m=2$,迭代次数为 2000,终止阈值为 10^{-5} ;城市道路设定为:聚类个数 $k=6$,模糊加权指数 $m=2.5$,迭代次数为 2000,终止阈值为 10^{-5} 。聚类前对数据进行标准化处理,随着目标函数值的收敛,隶属度矩阵达到稳定。每个样本点会根据到各个聚类中心的隶属度来确定所属类别,即自动划分为隶属度最大的一类。分别得到高速路和城市道路历史交通流数据的聚类结果,如图 8 和图 9 所示。聚类结果将 1440 个样本数据划分成不同类别,分别用不同的颜色表示。

本文中交通状态等级越高表示相应的交通拥堵程度越高。进一步分析图 8 中不同类别交通流数据的变量特征。绿色部分数据整体上具有比较高的速度值,同时在 5 个类中的流量值和占有率值最低,因此判定为状态“1”;靛色部分数据的速度值和绿色部分相近,但流量值和占有率值相对高一些,判定为状态“2”;黄色部分数据速度值较前两者低,同时流量和占有率值比前两者高,可判定为状态“3”;粉色部分数据的流量值和占有率值在整体上与黄色部分相近,但速度值比它低,判定为状态“4”;红色部分数据速度值和“3”、“4”相近,并且在 5 个类中具有最高的流量值和占有率值,因此判定为状态“5”。状态“1”到“5”代表高速路交通状态的不同等级。同理,对图 9 中不同类别的数据进行分析,可将城市道路交通状态等级划分为“1”到“6”。最后得到用于训练随机森林预测模型的历史数据所对应的交通状态。

5.3 短时交通状态预测结果与分析

在已有实验结果的基础上,利用随机森林算法分别预测高速路和城市道路的短时交通状态。随机森林超参数设置如下:100 个基决策树,最优划分属性的评判指标为 Gini 值,属性子集个数为属性集总数的开方,节点最小分裂的样本数为 2。利用图 8 中已知状态的历史数据训练随机森林模型,然后对

图 6 所示的短时交通流参数预测值进行分析, 进而得到高速路短时交通状态的预测结果。同理, 利用图 9 中已知状态的历史数据训练随机森林模型, 并对图 7 所示的短时交通流参数预测值进行分析, 从而得到城市道路短时交通状态的预测结果。

为分析预测结果的准确性, 需要知道预测当天不同时段的实际交通状态。因此, 再次借助 FCM 算法分析这一天的实际交通流数据。为保证结果的客观、真实性, 将高速路实际交通流数据与图 8 中的历史交通流数据融合在一起, 并按照 5.2 节所述的方式重新进行聚类分析, 由此得到高速路实际交通状态, 该部分聚类结果如图 10 所示。同理, 城市道路实际交通状态及对应的聚类结果如图 11 所示。图 12 和图 13 所示分别为高速路和城市道路所预测的短时交通状态与实际交通状态的比较。

为充分说明随机森林算法预测短时交通状态的优势, 选取常见的机器学习分类方法如决策树、梯度提升决策树(GBDT)、K 最近邻(KNN)、支持向量机(SVM)和逻辑回归(LR)与其进行预测性能对比。两种场景下的预测性能对比分别如表 2 和表 3 所示。

由表 2 可知, 随机森林对高速路短时交通状态的预测精度达到 89.58%, 相比另外 5 种常见的机器学习分类方法预测精度高 0.69% 以上。由表 3 可知, 随机森林对城市道路短时交通状态的预测精度为 86.11%, 比另外 5 种方法预测精度高 2.08% 以上。

表 2 不同方法对高速路短时交通状态的预测性能对比

算法	RF	DT	GBDT	KNN	SVM	LR
准确率 (%)	89.58	88.54	88.19	88.89	88.54	86.81

表 3 不同方法对城市道路短时交通状态的预测性能对比

算法	RF	DT	GBDT	KNN	SVM	LR
准确率 (%)	86.11	83.68	84.03	84.03	82.29	76.74

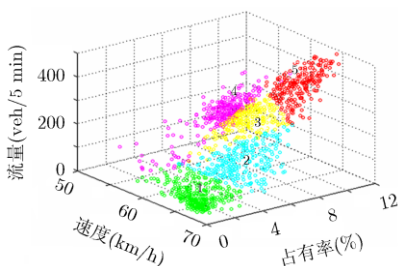


图 8 高速路历史交通流数据的 FCM 聚类结果与状态划分

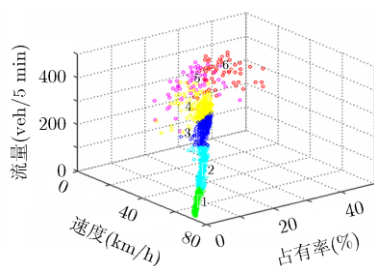


图 9 城市道路历史交通流数据的 FCM 聚类结果与状态划分

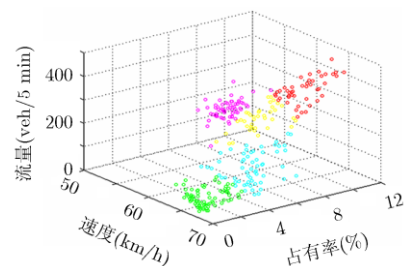


图 10 高速路预测当天实际交通流数据的 FCM 聚类分析

结果表明, 在相同超参数设置下的随机森林预测模型对不同道路场景的短时交通状态均取得更好的预测性能, 并且在城市道路场景下的优势更加明显, 说明其可以适应交通流复杂的随机变化特征。

由图 12 和图 13 可知, 预测的短时交通状态与实际交通状态具有良好的拟合度, 从而证明本文方法预测短时交通状态的优越性和可行性。进一步分析可知, 星期一当天高速路在白天期间(6:00~18:00)交通状态等级为 3~5, 并且在 7:30~8:00, 15:00~18:00 左右达到最高等级, 表示处于繁忙或拥堵状态; 晚上期间交通状态等级基本为 1~2, 表示处于畅通状态。另一方面, 城市道路在 7:00~10:00, 17:00~19:00 左右交通状态等级为 4~6, 且在此期间达到最高等级; 其交通状态之间的变化相比高速路更有规律性, 早晚高峰期的拥堵状况也更为明显, 这是由城市道路交通流的变化特征所导致的。此外, 高速路场景下的系统总运行时间为 408.5631 s, 城市道路为 384.6460 s。两种场景下平均每个时段的运行时间(除以 288 个时段)分别为 1.4186 s 和 1.3356 s, 均远小于 5 min, 可以满足短时交通状态预测对实时性的要求。实验结果符合实际交通状况的一般规律, 可及时为用户提供所需的短时交通状态信息。

6 结束语

准确的短时交通状态预测是缓解城市交通拥堵状况的有力举措。本文提出利用 FCM 和随机森林方法预测不同道路场景的短时交通状态。首先考虑交通流的时空特性并通过 AMSVM 预测短时交通流参数, 然后借助 FCM 算法评估历史交通状态, 建立随机森林预测模型后进一步分析短时交通流参数, 以此得到最终预测的短时交通状态。实验结果表明, FCM 对历史交通状态的评估方式适用于不同的高速路和城市道路场景, 且合理有效。其次, 随机森林比其他常见的机器学习方法有更高的预测精度和泛化性能, 从而可为用户提供实时可靠的短时交通出行信息。本文方法在充分挖掘交通流参数与交通状态隐含关系的同时, 旨在探索短时交通状态

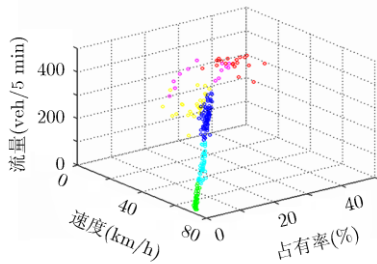


图 11 城市道路预测当天实际交通流量数据的 FCM 聚类分析

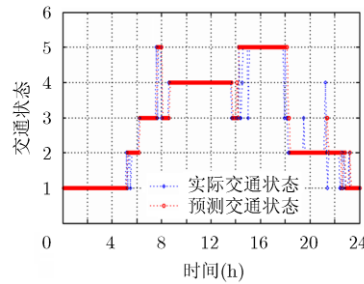


图 12 高速路预测的短时交通状态与实际交通状态的比较

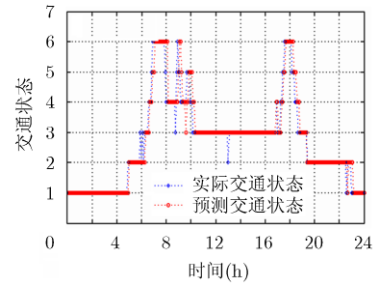


图 13 城市道路预测的短时交通状态与实际交通状态的比较

预测这一全新的研究领域，为智能交通系统的研究和应用提供了一种新思路。

本文方法有可待改进的地方，如可对 FCM 算法做相应的改进，以便获得更准确的历史交通状态信息。此外，可尝试对更多区域和道路(如国内城市)的交通流数据进行分析，从而进一步证明本文方法优越的预测性能和普适性。

参考文献

- [1] KRAUSE B, ALTROCK C, and POZYBILL M. Intelligent highway by fuzzy logic: Congestion detection and traffic control on multi-lane roads with variable road signs[C]. Proceedings of the 1996 5th IEEE International Conference on Fuzzy Systems, New Orleans, USA, 1996, 3: 1832-1837.
- [2] 巫威眺, 靳文舟, 林培群. 基于 BP 神经网络的道路交通状态判别方法研究[J]. 交通信息与安全, 2011, 29(4): 71-74. doi: 10.3963/j.ISSN 1674-4861.2011.04.016.
- WU Weitiao, JIN Wenzhou, and LIN Peiqun. The method of traffic state identification based on BP Neural Network[J]. *Journal of Transport Information and Safety*, 2011, 29(4): 71-74. doi: 10.3963/j.ISSN1674-4861.2011.04.016.
- [3] 张亮亮, 贾元华, 牛忠海, 等. 交通状态划分的参数权重聚类方法研究[J]. 交通运输系统工程与信息, 2014, 14(6): 147-151. doi: 10.16097/j.cnki.1009-6744.2014.06.022.
- ZHANG Liangliang, JIA Yuanhua, NIU Zhonghai, *et al.* Traffic state classification based on parameter weighting and clustering method[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2014, 14(6): 147-151. doi: 10.16097/j.cnki.1009-6744.2014.06.022.
- [4] KONG Xiangjie, XU Zhenzhen, SHEN Guojiang, *et al.* Urban traffic congestion estimation and prediction based on floating car trajectory data[J]. *Future Generation Computer Systems*, 2016, 61(C): 97-107. doi: 10.1016/j.future.2015.11.013.
- [5] DENG Chao, WANG Fan, SHI Huimin, *et al.* Real-time freeway traffic state estimation based on cluster analysis and Multiclass Support Vector Machine[C]. 2009 International Workshop on Intelligent Systems and Applications, Wuhan, China, 2009: 1-4. doi: 10.1109/IWISA.2009.5073027.
- [6] OH S, BYON Y J, and YEO H. Improvement of search strategy with K-Nearest Neighbors approach for traffic state prediction[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2016, 17(4): 1146-1156. doi: 10.1109/TITS.2015.2498408.
- [7] RICARDO G R, MARIA L L G, and MARIA S R. An approach to dynamical classification of daily traffic patterns [J]. *Computer-Aided Civil and Infrastructure Engineering*, 2017, 32(3): 191-212. doi: 10.1111/mice.12226.
- [8] CHEN Yuanyuan, LÜ Yisheng, LI Zhenjiang, *et al.* Long short-term memory model for traffic congestion prediction with online open data[C]. 19th IEEE International Conference on Intelligent Transportation Systems, Rio de Janeiro, Brazil, 2016: 132-137. doi: 10.1109/ITSC.2016.7795543.
- [9] 高林, 刘英, 盛子豪. 随机森林算法在交通状态判别中的应用 [J]. 实验技术与管理, 2017, 34(4): 43-46. doi: 10.16791/j.cnki.sjg.2017.04.012.
- GAO Lin, LIU Ying, and SHENG Zihao. Application of Random Forest algorithm to traffic state identification[J]. *Experiment Technology and Management*, 2017, 34(4): 43-46. doi: 10.16791/j.cnki.sjg.2017.04.012.
- [10] 冯心欣, 凌献尧, 林焯婷, 等. 可优化的自适应多核支持向量的短时交通流预测方法[P]. 中国专利, 106971548A, 2017.
- FENG Xinxin, LING Xianyao, LIN Yeting, *et al.* Optimized adaptive Multi-kernel Support Vector Machine for short-term traffic flow prediction[P]. China Patent, 106971548A, 2017.
- [11] LING Xianyao, FENG Xinxin, CHEN Zhonghui, *et al.* Short-term traffic flow prediction with optimized Multi-kernel Support Vector Machine[C]. 2017 IEEE Congress on Evolutionary Computation, Donostia-San Sebastian, Spain, 2017: 294-300. doi: 10.1109/CEC.2017.7969326.
- [12] ZHU Guangyu, CHEN Jianjun, and ZHANG Peng. Fuzzy C-means clustering identification method of urban road traffic state[C]. 12th International Conference on Fuzzy Systems and Knowledge Discovery, Zhangjiajie, China, 2015:

- 302-307. doi: 10.1109/FSKD.2015.7381958.
- [13] 吴启顺, 蔡晓禹, 蔡明. 基于 FCM 快速路交通状态判别加权指数研究[J]. 科学技术与工程, 2017, 17(6): 289-295.
WU Qishun, CAI Xiaoyu, and CAI Ming. A study of weighting exponent in expressway traffic state estimation based on Fuzzy C-means[J]. *Science Technology and Engineering*, 2017, 17(6): 289-295.
- [14] BREIMAN L. Random Forests[J]. *Machine Learning*, 2001, 45(1): 5-32. doi: 10.1023/A:1010933404324.
- [15] 董师师, 黄哲学. 随机森林理论浅析[J]. 集成技术, 2013, 2(1): 1-7.
DONG Shishi and HUANG Zhexue. A brief theoretical overview of Random Forests[J]. *Journal of Integration Technology*, 2013, 2(1): 1-7.
- [16] ZHOU Zhihua and FENG Ji. Deep Forest: Towards an alternative to Deep Neural Networks[C]. 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 2017: 3553-3559.
- [17] CALTRANS PEMS. Traffic flow database[OL]. <http://pems.dot.ca.gov/?dnode=VDS&content=loops&tab=dettimeseries&station id=1017510>, 2016.
- 陈忠辉: 男, 1960 年生, 教授, 硕士生导师, 研究方向为群智感知网络、车联网、无线通信网络.
- 凌献尧: 男, 1992 年生, 硕士生, 研究方向为车联网海量数据分析、机器学习、深度学习.
- 冯心欣: 女, 1983 年生, 副教授, 硕士生导师, 研究方向为参与式感知网络的激励机制设计、认知无线网络的动态频谱共享、博弈论和市场理论.
- 郑海峰: 男, 1978 年生, 副教授, 硕士生导师, 研究方向为无线传感网络、群智感知网络、压缩感知、机器学习.
- 徐艺文: 男, 1976 年生, 副教授, 硕士生导师, 研究方向为群智感知车联网的数据挖掘技术、H.265/HEVC 编码技术.