

文章编号: 2095-2163(2020)08-0148-04

中图分类号: U495

文献标志码: A

基于 K 均值聚类算法与 RBF 神经网络的交通流预测方法

张天逸, 孙毅然, 刘凡琪, 梁悦祺, 林永杰, 马明辉

(上海工程技术大学 机械与汽车工程学院, 上海 201600)

摘要: 随着汽车保有量的增加, 交通路网问题愈发严峻。针对现阶段城市交通流预测问题, 在分析道路交通车流量、平均速度的周期性波动规律及变化趋势的基础上, 融合 K 均值聚类算法(K-Means Clustering Algorithm)与 RBF 神经网络算法, 构建了一种基于多维时间序列的交通流预测模型。实验结论表明, 该模型预测结果与实际交通流数据拟合度高, 精准度达到 96.7%。该模型在理论研究与实际应用方面, 对提升交通控制效果具有重要意义。

关键词: 交通流预测; K 均值聚类算法; RBF 神经网络

Traffic flow prediction method

based on K-Means clustering algorithm and radial basis function

ZHANG Tianyi, SUN Yiran, LIU Fanqi, LIANG Yueqi, LIN Yongjie, MA Minghui

(School of Mechanical and Automobile Engineering, Shanghai University of Engineering Science, Shanghai 201600, China)

[Abstract] With the increase of car ownership, the problem of transportation network is more and more serious. For urban traffic flow prediction problem at this stage, in the analysis of road traffic flow and average velocity of periodic fluctuation and the change trend, a traffic flow prediction model based on multidimensional time series it introduced with the fusion of k-means Clustering Algorithm (K Means Clustering Algorithm) and RBF neural network Algorithm. The experimental results show that the model prediction fits the actual traffic flow data well, where the accuracy is 96.7%. The model is of great significance for improving traffic control effect in theoretical research and practical application.

[Key words] traffic flow prediction; K-Means clustering algorithm; radial basis function

0 引言

随着交通需求的快速增加, 路网交通问题愈发严峻, 采用智能交通控制方法解决上述问题, 逐渐成为交通领域研究的热点。交通预测数据作为交通控制方法有效实施的基础输入, 交通流预测结果的准确性和有效性对交通控制的可靠性有着重要的影响^[1-2]。

当前国内外学者针对交通数据预测方面进行了深入研究。为了更好地捕获交通流数据之间的特征, 教欣萍等人设计了一种融合高维交通流特征的短时交通流预测算法^[3]; 孔繁辉基于深度学习理论, 提出一种深度信念网络(DBN)算法优化传统 BP 神经网络预测模型^[4]; 针对交通预测数据的波动性问题, Mykola Guchenko 等人通过分析交通流预测数据, 提出了利用控制过程局部模型进行交通预测的方法^[5]; Lingru Cai 等人提出了一种结合重力搜

索算法(GSA)和 SVR 模型的混合交通流预测模型, 利用 GSA 搜索最优 SVR 参数, 实现对交通数据的预测^[6]; Liu 等人根据短时交通流的特点, 建立了基于 BP 神经网络的预测模型, 反映出短期交通流的变化规律和趋势^[7]; 徐先锋等人针对现有预测方法不能充分利用交通流时空特征的问题, 提出了一种结合卷积神经网络和双向长短时记忆网络的深度学习预测模型^[8]。

本文综合考虑交通系统一体化的科学性, 从底层数据基础角度出发, 分析交通数据间的时间关联特性, 融合 K 均值聚类与 RBF 神经网络算法, 提出了一种基于多维时间序列的交通流预测方法, 在理论研究与实际应用方面对提升交通控制效果具有重要的意义。

1 交通流特性分析

交通流数据具有很强的规律性, 通过分析交通

基金项目: 上海工程技术大学大学生创新项目(cx1901013)。

作者简介: 张天逸(1998-), 男, 本科生, 主要研究方向: 交通运输; 孙毅然(1999-), 男, 本科生, 主要研究方向: 交通运输; 刘凡琪(1998-), 女, 本科生, 主要研究方向: 交通运输; 梁悦祺(1999-), 女, 本科生, 主要研究方向: 交通运输; 林永杰(1998-), 男, 本科生, 主要研究方向: 交通运输; 马明辉(1989-), 女, 博士, 副教授, 硕士生导师, 主要研究方向: 智能交通控制。

通讯作者: 马明辉 Email: mhh789@126.com

收稿日期: 2020-02-05

流数据的特点,可以为交通流预测模型的建立提供参考依据。与常规时间序列预测数据不同,交通流数据具有以下特点:

(1) 交通流数据具有丰富的时间关联特征。在时间上,一个观测点可以在各个时刻持续收集交通流数据,形成一个随时间变化的交通流量序列。由于道路上车辆的聚集和发散是一个渐进过程,所以前一时刻的交通流量会影响到下一时刻的交通流量。

(2) 交通流数据具有周期特征。工作日和休息日的交通数据会随着出行人数的变化而对交通流预测产生较大影响,但在星期、月、年等周期性的时间内,在其它因素不变的情况下,不会有太大的变化。

(3) 交通流数据构成的时间序列会受其它因素影响。交通流量在相邻时间点之间会表现出更强的相关性。但是,因受到天气因素、道路施工等不可抗因素的影响,不能只通过时间序列的变化来收集交通流数据,应尽量采取对数据影响最小的时间序列。

2 交通流预测基础模型

2.1 RBF 神经网络模型

RBF 神经网络是一种常见的三层结构神经网络,主要包括输入层、隐含层及输出层,如图 1 所示。RBF 神经网络的作用原理,是将径向基函数(RBF)作为网络第二层隐含层的节点函数,以此构成隐含层空间。当数据被输入网络后,输入层会把数据传递给隐含层。经过隐含层节点函数计算之后,再将数据传递给输出层。通常而言,隐含层节点的计算函数是非线性的。当隐含层的节点数增加时,处理数据的次数也随之增加,使 RBF 网络得到的结果也就更加精确。但是,过多的节点数会减低网络的执行效率。第三层输出层的节点函数通常是线性的,其作用通常是对隐含层函数计算所得结果进行加权处理,将数据处理成方便输出,容易读懂的形式。

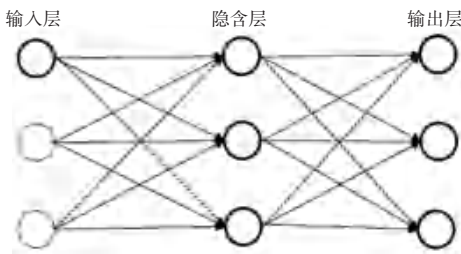


图 1 RBF 神经网络模

Fig. 1 RBF model

在 RBF 神经网络中,设输入层节点个数为 I , 隐含层节点数为 M , 输出层节点数为 N , 输入量为 x 。当 x 经输入到模型后,会经过 I 次传递。因此定义 x 为 I 维输入量。设输出量为 y , 同理 y 会经过 N 次输出,称 y 为 N 维输出量。一般来说,采用高斯激活函数作为隐含层节点的作用函数,该函数在 RBF 网络隐含层第 i 个节点输出为:

$$u_i = \exp\left[-\frac{1}{2\sigma^2}(x - c_i)^T(x - c_i)\right], \quad i = 1, 2, \dots, M. \quad (1)$$

其中, u_i 是隐含层第 i 节点的输出值; σ^2 是隐含层第 i 个节点的标准化常数; M 是隐含层节点的数目; x 是输入量; c_i 是隐含层第 i 个节点高斯函数的中间向量。由此可知,隐含层节点的输出范围是在 0 - 1 之间。当输入的样本距离节点中心越近时,输出值越大。

RBF 网络隐含层输出映射到输出层,其表达式为:

$$y_k = \sum_{i=1}^M w_{ki} u_i - \theta_k, \quad k = 1, 2, \dots, N. \quad (2)$$

其中, y_k 是输出层第 k 节点的输出量; w_{ki} 是隐含层到输出层的加权系数; θ_k 是输出层的阈值。

2.2 K-均值聚类算法

K 均值聚类算法是目前应用最为广泛的划分聚类算法。其算法具有原理简单、模型清晰、操作方便、计算快速等特点,可以大规模同时对多种类型的数据进行聚类,快速挖掘出数据中隐含的关系和结构。

K 均值聚类算法是判断基于数据到中心点的距离来区分数据的所属类别。其把 N 个对象划分成 k 个簇,用簇中对象的均值表示每个簇的中心点(质心),利用合适的距离计算公式,计算出数据与聚类中心的距离,将其划分到合适的聚类中。当所有数据聚类结束后,检查聚类中心是否已收敛,如果收敛则终止,否则将继续迭代。

3 融合交通流预测模型

城市交通流量预测是十分复杂的非线性问题。RBF 神经网络具有较强的非线性拟合能力,能够映射各种复杂的非线性关系,对非线性连续函数具有一致逼近性,可以进行大范围和高效的数据处理,现已广泛应用于交通预测。由于 RBF 神经网络中隐藏层单元的中心及半径需要预先确定,因此 K 均值聚类算法在 RBF 神经网络中心选取时,需要通过科学调整聚类中心,使得网络中心的选取更加精确。

本文提出的融合交通流预测模型核心为:RBF神经网络中心初始化,而后根据迭代求解均值中心,计算样本点到中心的距离,进而依据 K 均值聚类算法修正得到新的类内均值,确定基函数中心 C_i ;判断中心是否有变化,若没有变化,再用最小二乘法得到权值 w ;否则重新确定基函数中心 C_i ,直至不再变化。

由于本文是基于 k -均值聚类算法在 RBF 神经网络下的学习,所以定义一个目标误差函数:

$$E = \frac{1}{2} \sum_{i=1}^L \| d_i - y_i \|^2. \quad (3)$$

其中, d_i 为在 L 组输入样本 x_i 下的每组期望输出; y_i 为在 x_i 下的输出向量,如此设置可降低 RBF 神经网络误差值。

k 均值聚类算法的实现步骤:

- (1) 给定初始聚类中心 $C_i(0)$;
- (2) 运用欧式距离计算到每个类的距离:
 $d_i(u) = \| x_i(u) - C_i(u) \|, i = 1, 2, \dots, k. \quad (4)$
- (3) 求最小距离:

$$d_i(u) = \min d_i(u); \quad (5)$$

(4) 划分所有对象点到最近的聚类中心,标记为 $1, 2, 3, \dots$,结束时 $k = 0$;

- (5) 求聚类中心的均值,即新的聚类中心;
- (6) 检查集群中心是否已收敛 $\| C_{in} - C_i \| < 0.1$. 是则终止,否则迭代;

(7) 当 $k = N$ 时,达到算法的终止条件,聚类结果不再变化。刷新聚类中心,输出聚类后的数据。创建一个图形输出窗口,将更新后每个聚类中的数据分别绘制成图,使其可以直观地展示出聚类的分布。

输出结果 C_i 后,需要对 RBF 的权值进行调整。这里采用最小二乘法对隐含层和输出层进行权值调整。设 C_{max} 为聚类中心的最大值; t 为样本数; i 是隐含层的节点数,则公式为:

$$w = \exp\left(\frac{M}{C_{max}^2} \| x_t - c_i \|\right), \quad t = 1, 2, \dots, T, i = 1, 2, \dots, M. \quad (6)$$

4 实验验证

4.1 实验设计

为了验证本文提出交通流量预测模型的有效性,分别采用基于 RBF 神经网络算法的交通流预测模型和本文提出的基于 K -Means 聚类算法和 RBF 神经网络算法融合的交通流预测模型进行对比实验。通过 Matlab 软件编写程序实现上述算法,进行

仿真,将所得预测数据与实际数据进行比较,得出最终结论。

选择 8 月 24 日京珠高速公路获取的交通流量实测数据作为实验数据。将获取的大量实测数据根据时间序列划分为 48 组,并选取 10 组数据作为测试数据集,对提出的预测方法进行测试。所采用的部分交通流量数据如图 2 和表 1 所示。

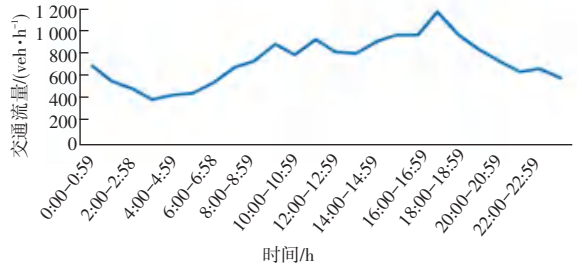


图 2 24 日交通流量分布图

Fig. 2 Traffic flow data on 24th

表 1 部分交通流量数据示例

Tab. 1 Partial traffic flow data

时间	交通流量/veh	时间	交通流量/veh
9:00-9:10	138	9:50-10:00	142
9:10-9:20	141	10:00-10:10	171
9:20-9:30	162	10:10-10:20	165
9:30-9:40	153	10:20-10:30	159
9:40-9:50	146	10:30-10:40	148

4.2 实验结果

通过 matlab 程序仿真,得到的 RBF 神经网络预测交通流效果如图 3 所示。由图中可以看出,交通流预测值曲线与实际值大致拟合,但在交通流数值出现较大波动时,输出曲线的响应时间偏慢,预测准确率在 62.7%。基于 K 均值聚类算法的 RBF 神经网络预测交通流效果如图 4 所示。由图可以看出预测曲线与实际值基本吻合,预测误差控制在 3%。

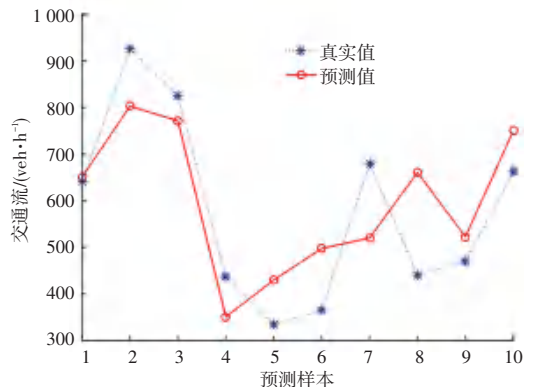


图 3 原预测结果

Fig. 3 Original forecast results

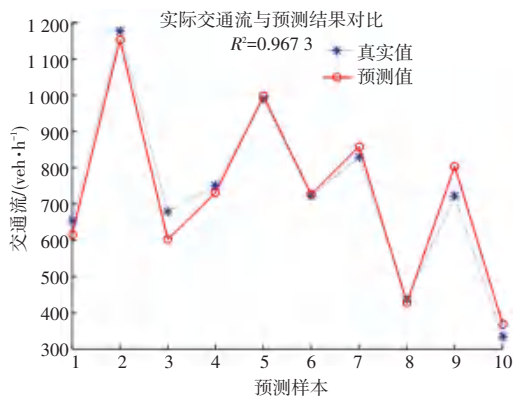


图 4 加入 K 均值算法的预测结果

Fig. 4 Forecast results with K-means

5 结束语

为了提高通过 RBF 神经网络预测交通流的准确度,本文运用 K 均值聚类算法选取 RBF 神经网络隐含层的中心向量,同时采用最小二乘法调整 RBF 神经网络的权值,保证参数选取合适。RBF 神经网络具有样本精度高,收敛速度快的特点,并且解决了神经网络局部极小的缺陷,在面对交通流预测这类较复杂的非线性问题时,能够达到良好的效果。从

(上接第 147 页)

类图像分割算法(简称为 ACS_FCM),该算法针对传统 FCM 聚类中心和隶属度矩阵生成的随机性,导致算法容易陷入局部最优值的缺陷,将改进的布谷鸟优化算法引入到 FCM 聚类过程中。实验表明,本文算法不仅分割效果优于 FCM,在算法运行时间上也有很明显的优越性。对于具有伪影、组织边界模糊的 MRI 脑部图像具有很好的分割效果。

参考文献

- [1] FITZPATRICK J M, SONKA M. Medical Image Processing and Analysis[J]. Image Rochester NY, 2000, 2(11):1-40.
- [2] NEUBERT A, YANG Z, ENGSTROM C, et al. Automatic segmentation of theglenohumeral cartilages from magnetic resonance images[J]. Medical physics, 2016, 43(10): 5370-5379.
- [3] CHOY S K, LAM S Y, YU K W, et al. Fuzzy model-based clustering and its application in image segmentation[J]. Pattern Recognition, 2017, 68:141-157.
- [4] XU X, ZHANG Z, XIE J, et al. Threshold image segmentation based on granular immune algorithm[C]// Control and Decision Conference, 2009:3512-3515.
- [5] KAIHUA W, TAO B. Optimal Threshold Image Segmentation Method Based on Genetic Algorithm in Wheel Set Online Measurement[C]//Third International Conference on Measuring Technology and Mechatronics Automation, Shanghai, 2011:799-802.
- [6] ZHONG Y, MA A, ZHANG L. An Adaptive Memetic Fuzzy Clustering Algorithm With Spatial Information for Remote Sensing Imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2014, 7(4):1235-1248.

预测结果可以看出,本预测模型是有效的,对城市道路交通疏导具有一定的参考和应用价值。

参考文献

- [1] 桂智明,李壮壮,郭黎敏. 基于 ACGRU 模型的短时交通流预测[J/OL]. 计算机工程与应用, 1-9.
- [2] Guillem Boquet, Antoni Morell, Javier Serrano, Jose Lopez Vicario. A variational autoencoder solution for road traffic forecasting systems; Missing data imputation, dimension reduction, model selection and anomaly detection[J]. Transportation Research Part C, 2020, 115.
- [3] 教欣萍,王江锋,陈磊,等. 基于 HALRTC 理论的短时交通流预测算法[J]. 山东科学, 2019, 32(6): 62-68.
- [4] 孔繁辉,李健. 深度信念网络优化 BP 神经网络的交通流预测模型[J]. 管理评论, 2020, 32(3): 300-306.
- [5] Mykola Guchenko, Natalya Sokhin, Anton Skalsky, et al. Research of Prognostic Abilities of Local Model of Controlled Process for Traffic Forecasting[J]. Transportation Research Procedia, 2020, 44.
- [5] CAI L, CHEN Q, CAI W, et al. SVRGSA: a hybrid learning based model for short-term traffic flow forecasting[J]. Intelligent Transport Systems, IET, 2019, 13(9): 1348-1355.
- [6] LIU S, FANG Z, ZHANG L. Research on Urban Short-term Traffic Flow Forecasting Model [C]//Journal of Physics: Conference Series. IOP Publishing, 2019, 1237(5): 052026.
- [8] 徐先峰,黄刘洋,龚美. 基于卷积神经网络与双向长短期记忆网络组合模型的短时交通流预测[J]. 工业仪表与自动化装置, 2020(1): 13-18.
- [7] 雷涛,张肖,加小红,刘侍刚,张艳宁. 基于模糊聚类的图像分割研究进展[J]. 电子学报, 2019, 47(08): 1776-1791.
- [8] FEI Z, HONG F. Research on medical image segmentation based on fuzzy C-means clustering algorithm[J]. computer engineering and applications, 2014, 34(4): 144-151.
- [9] 杨立才, 赵莉娜, 吴晓晴. 基于蚁群算法的模糊 C 均值聚类医学图像分割[J]. 山东大学学报(工学版), 2007, 37(3): 51-54.
- [10] DENG WQ, LI XM, GAO X, et al. A modified fuzzy c-means algorithm for brain MR image segmentation and bias field correction [J]. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 31(3): 501-511 May 2016. DOI 10.1007/s11390-016-1643-5.
- [11] YANG X S, DEB S. Cuckoo search via Levy flights [A]. In: 2009 World Congress on Nature & Biologically Inspired Computing [C]// Coimbatore, India; IEEE, 2009: 210-214.
- [12] WALTON S, HASSAN O, MORGAN K, et al. Modified cuckoosearch: a new gradient free optimisation algorithm [J]. Chaos, Solitons&Fractals, 2011, 44(9): 710-718.
- [13] Manoj Kumar Naik, Rutuparna Panda. A novel adaptive cuckoo search algorithm for intrinsic discriminant analysis based face recognition[J]. Applied Soft Computing, 2016, 38.
- [14] 胡加亮. 基于模糊聚类算法的图像分割研究[D]. 华东师范大学, 2019.
- [15] MEKHMUKH A, MOKRANI K. Improved Fuzzy C-Means based Particle Swarm Optimization (PSO) initialization and outlier rejection with level set methods for MR brain image segmentation [J]. Computer methods and programs in biomedicine, 2015, 122(2): 266-281.
- [16] XIE X L, BENI G. A validity measure for fuzzy clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(8): 841-847.