

基于核超限学习机群组算法的交通拥堵预测

邢一鸣^{1,2} 班晓娟¹ 刘旭¹ 尹航² 沈晴¹

(北京科技大学计算机与通信工程学院 北京 100086)¹ (沈阳航空航天大学工程训练中心 沈阳 110136)²

摘要 城市交通拥堵预测是智能交通系统研究的重要内容之一。交通运行状态具有高度不确定性和复杂性,目前已经有多种基于神经网络的预测技术被引入交通预测领域中。然而,传统的神经网络具有训练时间长、易陷入过拟合和局部最优等缺点,这严重阻碍了神经网络在交通预测领域的大规模应用。超限学习机是一种新型的单隐层前馈神经网络,具有泛化能力强、训练速度快、产生唯一最优解等诸多优点。基于超限学习机算法,文中提出了核超限学习机群组算法,此算法由多个超限学习机子模型组成,每个子模型只负责某一类样本的学习,该算法使每一类样本均能达到全局最优,整体可以获得比超限学习机更高的预测准确率。实验结果表明,单进程的核超限学习机群组算法比超限学习机的训练时间稍短,但前者的准确率较后者提高了8%;相比其他流行的机器学习算法,核超限学习机群组算法的训练速度快、预测准确度高;经过核超限学习机群组算法预测的结果与实际情况较为符合,可靠性高,具有很强的实用价值。

关键词 核超限学习机,群组,交通拥堵,预测,神经网络

中图分类号 TP399 **文献标识码** A **DOI** 10.11896/jsjx.191100507C

Traffic Congestion Prediction Based on Kernel Extreme Learning Machine Group Algorithm

XING Yi-ming^{1,2} BAN Xiao-juan¹ LIU Xu¹ YIN Hang² SHEN Qing¹

(School of Computer and Communication Engineering, University of Science and Technology, Beijing 100086, China)¹

(Engineering Training Center, Shenyang Aerospace University, Shenyang 110136, China)²

Abstract Prediction of urban traffic congestion is one of the important research contents of intelligent transportation system (ITS). At present, a lot of neural networks are introduced into the field of traffic forecasting and are widely used. However, the traditional neural network training is time-consuming, easy to fall into local optimal and over fitting. It has seriously hindered the large-scale application of neural network in the field of traffic forecasting. ELM is a new kind of single hidden layer feed-forward neural network, which has the advantages of fast training speed, strong generalization ability and unique optimal solution. In this paper, the new algorithm named KELM-Group was proposed, which is composed of multiple KELM sub-models. KELM-Group algorithm enables each class of samples to achieve the global optimum, and the overall prediction accuracy can be higher than that of ELM. The experimental results show that the KELM-Group algorithm is faster than other popular machine learning algorithms. The accuracy rate of KELM-Group algorithm is 8% higher than that of the ELM. The results predicted by the KELM-Group algorithm are more consistent with the actual situation, and have great practical value.

Keywords Kernel extreme learning machine, Group, Traffic congestion, Prediction, Neural network

1 引言

近年来,随着经济的快速增长和城市化进程的日益加快,城市人口和车辆不断增加,城市交通问题受到广泛关注,交通拥堵问题日益严重,交通事故多发频发,城市交通空间饱和度越来越高。因此,智能交通问题的研究进入了一个快速发展

期,智能交通已经成为交通领域的关注热点和研究重点。随着智能交通控制与诱导系统的发展,交通预测尤其是短期交通预测受到越来越多的关注,已经成为智能交通系统的核心研究内容之一。在交通流预测方面,目前已有多种预测模型和方法,如卡尔曼滤波法、时间序列法、历史平均法、线性回归法、非参数模型等。其中,很多预测方法都是对现

到稿日期:2018-11-07 返修日期:2019-01-22 本文受国家重点研发计划(2016YFB0700500),国家自然科学基金项目(61702036,61572075),国家航空科学基金项目(2015ZB54007),辽宁省教育厅科学研究项目(L201627)资助。

邢一鸣(1986-),女,硕士,讲师,主要研究方向为人工智能,E-mail:peachming@163.com;班晓娟(1970-),女,博士生,教授,CCF会员,主要研究方向为人工智能与人工生命理论研究、可视化技术、计算机动画与仿真技术研究、智能软件理论及应用,E-mail:banxj@ustb.edu.cn(通信作者);刘旭(1986-),男,博士生,主要研究方向为人工智能、流体仿真技术;尹航(1978-),男,博士生,副教授,主要研究方向为人工智能及大数据、重大设备健康管理;沈晴(1988-),男,博士生,主要研究方向为人工智能、智能软件理论及应用。

有模型的改进和提升^[1-7]。

道路交通状态具有复杂性、混合性和随机性,而神经网络具有学习复杂非线性系统的特性。因此,神经网络与道路交通状态高度契合,适用于交通预测领域。当前,许多基于神经网络的预测技术已经被引入到交通预测领域,并取得了丰硕的成果,在一定程度上解决了存在的城市交通问题,为改善城市交通状况发挥了积极作用。Smith 等于 1994 年首次将反向传播神经网络应用到短期交通流预测领域中^[8]。目前应用于交通预测领域的神经网络模型有:多层反馈神经网络^[9]、径向基 RBF 神经网络^[10]、反向传播算法^[11]、递归神经网络^[12-13]、时间延迟神经网络模型^[14]等。

然而,通过理论研究和实践应用发现,应用传统的神经网络解决交通预测这种复杂的非线性问题时,仍然存在输入数据量大、训练速度慢、模型复杂、易陷入局部最优等不足,在解决实际问题方面还有欠缺,使得神经网络在交通预测中还没有得到大规模的应用,对于解决交通拥挤与堵塞问题的作用还很有限。

Huang 等^[15]于 2006 年提出了超限学习机(Extreme Learning Machine, ELM),这是一种新型的单隐层前馈神经网络学习方法,在保留了学习复杂非线性系统的能力的基础上,还具有训练速度快、产生唯一最优解等诸多优点。因此,超限学习机算法非常适合于在具有大量数据的复杂和可变交通预测模型中使用。虽然 ELM 在很大程度上提高了训练效率,但在相同训练数据和模型参数的情况下,输入层和隐含层参数的随机分布导致了分类精度的极大变化,从而显著影响了系统的稳定性^[16]。利用核超限学习机(kernel-Extreme Learning Machine, KELM),采用核函数代替 ELM 隐含层的计算方法,解决了 ELM 中输入层和隐含层参数的随机分布所导致的问题,在牺牲一定训练速度的情况下,获得了与相应数据集更高的相关性,以及更高的稳定性^[17]。

本文以南宁市的交通监测数据为基础,将核超限学习机应用于交通预测领域,创新性地提出了核超限学习机群组(Kernel-Extreme Learning Machine Group, KELM-Group)的学习框架。核超限学习机群组(KELM-Group)是由多个 KELM 子模型组成,每个 KELM 子模型只负责某一类样本的学习,在该类样本上达到了全局最优;在预测时,样本会根据特征进入不同的子模型训练。KELM-Group 每一类样本均能达到全局最优,可以获得比 ELM 更高的预测准确率。

2 超限学习机群组算法

2.1 超限学习机(ELM)

超限学习机是一种基于广义逆矩阵理论提出的新型单隐层前向型网络^[15]。与传统的神经网络相比,可以通过一步计算来解决网络的输出权重,极大地提高了网络的泛化能力和学习速度,并具有较强的非线性拟合能力,大大减少了计算量和搜索空间。

设有 N 个不重复的输入样本 (x_i, t_i) , 其中 $x_i = [x_{i1}, x_{i2}, \dots, x_{im}]^T \in R^n$ 是一个 n 维的输入, $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$ 输入 x_i 对应的 m 维输出。有 \tilde{N} 个隐层节点, 激励函数为 $g(x)$ 的标准单隐层前馈神经网络在数学上可以记为:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(x_j) = \sum_{i=1}^{\tilde{N}} \beta_i g_i(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{O}_j, j=1, \dots, N \quad (1)$$

其中, $\mathbf{w}_i = [\omega_{i1}, \omega_{i2}, \dots, \omega_{im}]^T$ 是第 i 个隐层节点的输入权重; $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ 是第 i 个隐层节点的输出权重; b_i 是第 i 个隐层节点的偏置; $\mathbf{w}_i \cdot \mathbf{x}_j$ 是 \mathbf{w}_i 和 \mathbf{x}_j 的内积。

具有 \tilde{N} 个隐层节点、激励函数为 $g(x)$ 的标准单隐层前馈神经网络如果能够完全拟合 N 个输入样本的输出, 即神经网络的误差 $\sum_{i=1}^{\tilde{N}} \|\mathbf{O}_j - t_j\| = 0$, 则存在 β_i, \mathbf{w}_i 和 b_i 使得:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = t_j, j=1, \dots, N \quad (2)$$

N 个样本的情况下, 可以被记为:

$$\mathbf{H}\beta = \mathbf{T} \quad (3)$$

其中, \mathbf{H} 为隐层节点的输出矩阵。

$$\mathbf{H} = \mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}, \mathbf{x}_1, \dots, \mathbf{x}_{\tilde{N}}) = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \dots & g(\mathbf{w}_{\tilde{N}} \cdot \mathbf{x}_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \dots & g(\mathbf{w}_{\tilde{N}} \cdot \mathbf{x}_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}} \quad (4)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m}, \mathbf{T} = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}$$

根据 ELM 理论, 如果激励函数无限可微, 则输入层的权重和隐层偏置可以随机赋值。固定输入权重 \mathbf{w}_i 和隐层偏置 b_i 后, 训练该单隐层前馈神经网络就是找到线性系统 $\mathbf{H}\beta = \mathbf{T}$ 的一个满足最小二乘的 $\hat{\beta}$ 。

$$\hat{\beta} = \min_{\beta} \|\mathbf{H}\beta - \mathbf{T}\| \quad (5)$$

如果隐层节点的数量 \tilde{N} 和输入的不重复的训练样本数 N 相等, 即 $\tilde{N} = N$ 时, 矩阵 \mathbf{H} 正定可逆, 则易知 $\hat{\beta} = \mathbf{H}^{-1} \mathbf{T}$ 。此时, 该单隐层神经网络的输出误差为 0。

然而实际在大多数情况下, 隐层节点数都远远小于训练样本数 N , 即 $\tilde{N} \ll N$ 。此时, \mathbf{H} 不是正定矩阵, 不存在 $\mathbf{w}_i, b_i, \beta_i$ ($i = 1, \dots, \tilde{N}$) 使得 $\mathbf{H}\beta = \mathbf{T}$ 。此时可以转向求使损失函数 $\|\mathbf{H}\beta - \mathbf{T}\|$ 最小的 β 。

$$\hat{\beta} = \min_{\beta} \|\mathbf{H}\beta - \mathbf{T}\| \quad (6)$$

根据极小范数解准则(即同时满足 $\min \|\mathbf{H}\beta - \mathbf{T}\|$ 和 $\min \|\beta\|$), 式(6)存在最小二乘解:

$$\hat{\beta} = \mathbf{H}^+ \mathbf{T} \quad (7)$$

其中, \mathbf{H}^+ 是隐层矩阵 \mathbf{H} 的 Moore-Penrose 增广逆矩阵。

为了获得更好的泛化性能和鲁棒性, 可以加入正则项。

$$\hat{\beta} = \left(\frac{1}{C} + \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{T} \quad (8)$$

如果隐层节点的输出函数 $h(x)$ 是未知的, Huang 等^[16]将核函数引入 ELM, 提出了 KELM 算法。ELM 的随机矩阵 $\mathbf{H}^T \mathbf{H}$ 由核矩阵代替, 可以写成如下形式:

$$\chi_{\text{ELM}} = \mathbf{H} \mathbf{H}^T \quad (9)$$

$$\chi_{\text{ELM}, ij} = h(x_i) \cdot h(y_j) = K(x_i, y_j)$$

其中, $K(x_i, y_j)$ 是核函数。

KELM 的输出公式可以写成以下形式:

$$y = F_{ELM}(x) = h(x)\beta = h(x)H^T \left(\frac{I}{C} + HH^T \right)^{-1} Y$$

$$= \begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_n) \end{bmatrix} \left(\frac{I}{C} + \chi_{ELM} \right)^{-1} Y \quad (10)$$

根据核函数理论^[18],核函数 $K(\mu, \nu)$ 包括 RBF 核函数、线性核函数和多项式核函数等,通常设定 RBF 核为:

$$K(\mu, \nu) = \exp[-(\mu - \nu)^2 / \gamma] \quad (11)$$

在 KELM 算法中,不需要具体给出隐层节点的特征映射函数 $h(x)$ 的具体形式,而仅需要知道核函数 $K(\mu, \nu)$ 的特定形式就可以获得输出函数的值。同时,由于内核函数直接使用内积形式,在输出函数值求解时不必设置隐层节点数,从而不必设置隐层的初始权值和偏移量。

2.2 核超限学习机群组算法

对于交通拥堵指数而言,“路段周围是否有学校”“路段周围是否有医院”这类特征对道路拥挤的规律性有很大的影响,应作为特征输入模型。然而,这一特征是很难提取和难以找到的。解决该问题的办法是为每一条道路提供一种模型,这样道路环境相关的很多信息都被隐含在了模型之中,无须进行复杂的、基于经验和实验的特征挖掘。

神经网络系统具有学习复杂非线性系统的能力,非常适用于交通预测,但是神经网络的结构复杂、运算速度慢,为每个路段提供一个模型显然是难以训练的。基于以上两个问题,本文提出了一种新型的 KELM 学习框架——核超限学习机群组(KELM-Group)。核超限学习机群组能够有效解决多模型训练的问题。

如图 1 所示,多个子模型具有相同的结构,其输入节点数、隐层节点数、输出节点数相同;同时,KELM 算法要求只需要知道核函数 $K(\mu, \nu)$ 的具体形式即可,输入 X 和隐层之间的网络结构可以共用。

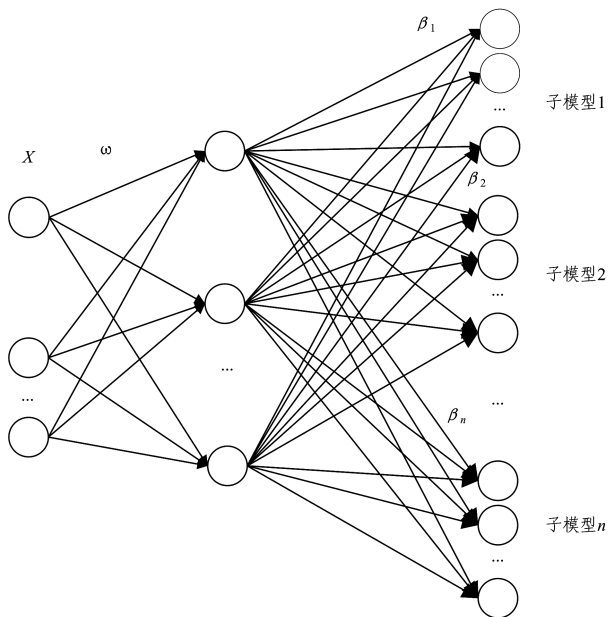


图 1 多个 KELM 可以共享同一个随机权重 ω

Fig. 1 Multiple KELMs sharing the same random weight ω

特征是机器学习的输入,其按照取值类型可以分为两种:

1)数值型特征,如年龄、收入等具有相对大小的特征。

2)枚举型特征,如道路等级、性别、城市等,不同取值之间不存在相对大小关系。

在实际应用中,特征输入复杂多样,既包括数值型特征,又包括枚举型特征,特征取值范围、类型、维度不同,无法直接进行学习,因此需要对特征进行预处理和标准化后再进行学习。一般地,数值型特征需要进行离散化,将其离散化为多个区间,数值型特征和枚举型特征均统一为枚举特征。

枚举型特征是一系列标签,需要对其进行进一步二值化处理后再送入模型。所谓二值特征主要是 0/1 特征,即特征只取两种值:0 或者 1。枚举值转为二值特征的方法是:将枚举特征映射为多个特征,每个特征对应一个特定枚举值。为了区别,下文将枚举特征称为簇特征,将由枚举特征转化而来的多个二值特征称为单特征。

设一个输入样本由 p 个簇特征组成,即 $x = (x_1, x_2, x_3, \dots, x_i, \dots, x_p)$ 。特征簇 x_i 可以转为 q_i 个二值特征, $x_i = (\alpha_{i,1}, \alpha_{i,2}, \alpha_{i,3}, \dots, \alpha_{i,j}, \dots, \alpha_{i,q_i})$, 则:

$$x = (x_1, x_2, x_3, \dots, x_i, \dots, x_p)$$

$$= ((\alpha_{1,1}, \alpha_{1,2}, \dots, \alpha_{1,j}, \dots, \alpha_{1,q_1}), \dots, (\alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,j}, \dots, \alpha_{i,q_i}), \dots, (\alpha_{p,1}, \alpha_{p,2}, \dots, \alpha_{p,j}, \dots, \alpha_{p,q_p})) \quad (12)$$

x 特征的总数为 $\sum_{i=1}^p q_i$ 。

如图 1 所示,如果使用单模型,输入特征数为 $\sum_{i=1}^p q_i$,需要的隐层节点数大,网络结构复杂。如果以多模型的方式,用簇特征 x_i 划分原始模型,则需要 q_i 个子模型,每个子模型的输入特征数为 $(\sum_{i=1}^p q_i - q_i)$ 。多个子模型可以共用同一个输入权重 ω ,形成一个 KELM 群组,每个节点是一个子模型。对于任意一个确定的输入样本,根据簇特征 x_i 的具体取值,进入不同的子模型。相对于原始模型,子模型特征数变少,隐层节点数也会相应减少,模型复杂度降低,每个子模型的训练时间大大缩短,KELM-Group 训练的时间总和可能比单模型短。特别是在 x_i 单特征数比较大时,加速效果明显。

一个复杂的 KELM 网络可以转化为 KELM 群组,将复杂的任务分而治之,提高处理速度。同时这种转化将输入样本分成了许多块,不同的输入进入不同的 KELM 子模型训练,互相之间互不影响,非常适合并行处理。与此同时,在采用多模型的训练方法时,模型本身隐含了很多信息,减少了特征处理的工作量。

3 基于核超限学习机群组算法的短时拥堵预测

3.1 特征提取

对于机器学习应用来说,特征提取是一个非常重要的部分,它将直接影响预测结果。显然,诸如道路上是否有交通灯、道路所在区域、附近是否有商业区或学校等因素对道路的交通状况的影响更大。但是在提出的系统中,这些特征很难自动获得,因此不被选取作为系统的输入特征。在使用特征选择工具进行大量测试之后,我们选择了以下 5 个特征。

1)当前时间:离散值,即 06:05,08:05, ..., 21:55。总共有 191 个。

2)道路逻辑区域:离散值,1,2,3, ..., 50,共 50 个。

3) 上次交通拥挤指数:连续值,从 0 到 100 不等。

4) 道路类型:分类、快速路、主干道、支路。

5) 相邻道路数:连续且为正整数。

物理区域在一定程度上描述了道路的环境,在道路拥挤预测中起着重要的作用。例如,学校附近的道路除了孩子们早晨上学、下午放学,可能总是状况良好。同时,交通条件具有较强的时空周期性,应考虑区域特征。但物理区域信息不是由 TRMS 自动提供的,对变化非常敏感,特别是在系统运行时有一条新的道路修建的情况下。此外,人们标记的区域可能不准确。本文试图用逻辑区域代替物理区域来描述道路的环境信息。

我们将 $x_{i,j}$ ($0 \leq x_{i,j} < 100$) 定义为在 j 时刻道路 i 的交通拥挤指数(从 6:00 到 22:00,每 5 min 评估一次,共 191 个指数)。对于特定的道路 i ,一天的拥挤指数向量为 $[x_{i,0}, x_{i,1}, x_{i,2}, \dots, x_{i,190}]^T$ 。因此,所有道路的矢量可表示为如下形式:

$$\mathbf{x}_0 = [x_{0,0}, x_{0,1}, x_{0,2}, \dots, x_{0,190}]^T$$

$$\mathbf{x}_1 = [x_{1,0}, x_{1,1}, x_{1,2}, \dots, x_{1,190}]^T$$

...

$$\mathbf{x}_n = [x_{n,0}, x_{n,1}, x_{n,2}, \dots, x_{n,190}]^T$$

利用 K 均值算法可以对上述向量进行聚类,从而将道路聚类为 K 聚类,本文称其为逻辑区域。同一群组道路一天内的交通状况相似,不同群组道路的交通状况的差异较大。道路群组或所谓的道路逻辑区域可以从逻辑上区分道路。本文对 k 值进行了调试,设 $k=50$ 。

3.2 交通拥堵指数

交通拥堵的具体定义目前尚无统一标准,对交通拥堵进行评价的目的是提出一种计算指标,以对当前时刻道路的拥堵情况进行量化描述。中国南宁市已经建立了完整的道路交通运行监测系统,通过长期调研,结合南宁市的道路特点,本文提出了交通拥堵指数评价指标。交通拥堵指数的取值范围为 $[0, 100]$,当车速为 0 时,表示道路极度拥堵,此时拥堵值为 100;当车速无限接近于无穷大时,道路趋近于无限畅通,道路拥堵值为 0。根据市民的出行习惯,参考南宁市《城市道路交通运行评价指标体系》,可将交通拥堵状态分为 3 个层次,如表 1 所列。

表 1 不同道路按照平均速度和拥堵值标准划分拥堵状态层次

Table 1 Different roads divided into congestion levels according to average speed and congestion value criteria

	畅通	基本畅通	拥堵
高速路、快速路	(65, ∞)	(35, 65]	(20, 35]
主干道	(40, ∞)	(30, 40]	(15, 20]
次干道、支路	(35, ∞)	(25, 35]	(10, 15]

交通拥堵指数 $f(x)$ 和车速 $x(x \geq 0)$ 的关系如下:

$$f(x) = 100 - \left(\frac{1}{1 + e^{-\partial x}} - \frac{1}{2} \right) \cdot 200 \quad (13)$$

其中,参数 ∂ 用于不同道路等级下的微调。高速路、快速路、主干道、次干道、支路等不同的道路等级,在相同的速度下具有不同的拥堵体验。车速相同时,道路等级越高,拥堵指数越大。参数 ∂ 就是用来反映道路等级对交通拥堵指数的影响。

从图 2 可以看出,在相同的速度下,道路等级越高,拥堵指数就越大。

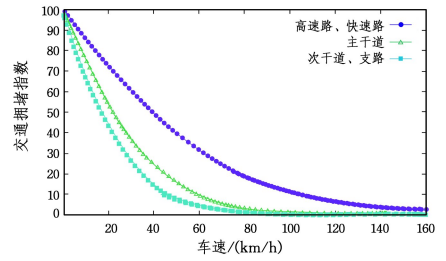


图 2 不同等级道路的交通拥堵指数

Fig. 2 Traffic congestion index of different grades of roads

目前南宁市已经有 8000 多辆浮动车(装载了 GPS 的行驶车辆,这些车辆定时向指挥中心报告车速、角度等信息),每 5 min 可以覆盖南宁市 70% 的市内道路。根据浮动车的实时路况报告,通过一些复杂的算法将浮动车匹配到周围的道路上,可以计算出路段的平均车速,进而计算出道路拥堵指数。

本文的目的在于利用核超限学习机群组学习结构,对未来 0~30 min 的交通拥堵指数进行分钟级别的预测。南宁市每天积累的监测数据、交通拥堵指数评价、路网、市政等信息为交通拥堵指数的预测提供了坚实的数据支持。

利用 KELM-Group 算法,为每个路段生成一个子模型。这些子模型具有相同的网络结构,多个子模型可以共享一个输入层节点,形成 KELM 模型群组, KELM 群组内的子模型串行训练,多个 KELM 群组可以并行训练。作为整体模型的一个节点,每个子模型只实现某个特定路段的预测,功能有限;所有子模型聚集在一起形成了一个巨大的复杂网络,能够预测整个路网。本文中,共有 18328 个路段需要预测,因此 KELM-Group 的子模型数为 18328。

4 实验

4.1 实验数据与环境

本节主要对基于时序特征的 KELM-Group 模型在南宁市交通拥堵数据集上的性能进行实验,将 KELM-Group 算法和 ELM 算法以及流行的其他机器学习算法如 LR, SVM, GBDT 进行比较。除特殊说明外,所有的实验均在 Ubuntu12.04 的机器上运行,实验机器的物理内存为 32 GB,处理器为 Duo i7 2.50 GHz。KELM-Group 使用 Python 实现,运行在 Python2.7 环境下。

拥堵评价系统每天产生约 2400000 条训练数据,其中有 1515446 个不重复的训练样本。本文实验均使用 1515446 个不重复的样本对模型进行训练和验证。所有的实验均进行了交叉验证(Cross validation),全部的训练集被平均随机分成 4 份较小的子集。每次取其中的 3 份作为训练集,另外一份作为测试集,参与的实验数据如表 2 所列。

表 2 基于 KELM-Group 的拥堵预测实验的训练和测试数据

Table 2 Training and test data of congestion prediction experiment

based on KELM-Group

训练样本	测试样本	特征数
1136 585	378 861	282

实验中,将测试样本的预测结果与测试样本的实际拥堵值进行比较,如果实际拥堵值属于预测结果的拥挤层级,则可以认为预测结果准确,否则认为预测不准确,可以

统计拥堵预测模型的准确率。

4.2 KELM-Group 模型调优实验

基于 KELM-Group 预测模型需要调优两个参数:惩罚系数 C_0 和核参数 γ 。通过 2.1 节对 KELM 理论的阐述可以看出,使用 RBF 核函数的 KELM 包含的两个重要参数分别是惩罚系数 C_0 和核参数 γ ,图 3 给出了两个参数与测试准确率的关系。为了获得良好的泛化性能,需要选择适当的惩罚系数 C_0 和核参数 γ ,它们都在集合中进行选择。最佳的性能通常是在惩罚系数 C_0 非常窄的范围内实现的,这意味着 KELM-Group 对惩罚系数 C_0 非常敏感。然而,当正确选择 C_0 时,性能将对内核参数 γ 不敏感。

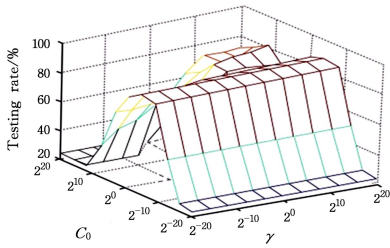


图 3 惩罚系数 C_0 和核参数 γ 的关系

Fig. 3 Relationship of penalty coefficient C_0 and kernel parameter γ

4.3 KELM-Group 和其他算法的比较

本文将 KELM-Group 学习算法与目前比较常用的多种神经网络和其他类型的机器学习算法在交通拥堵指数预测训练数据集上进行了比较。比较对象包括逻辑回归(Logistic Regression)、岭回归(Ridge Regression)、支持向量机(Support Vector Machine for Regression)和 GBDT 回归(Gradient Boosting Decision Tree Regression)等。这些算法均在 Python Sklearn 1.4 机器学习库中实现,其中 GBDT 中的深度为 9,LR 引入了 L1 正则,其他参数均保持默认。本节所有的实验均进行交叉验证,实验主要得到这些机器学习算法在交通数据集上的训练时间和预测准确度。实验结果如表 3 所列。

表 3 不同机器学习算法的比较

Table 3 Comparison of different machine learning algorithms

算法	训练时间/s	准确度/%
KELM-Group(80 进程)	3.28	92.03
GBDT	117.16	92.10
Ridge	22.23	90.49
Lasso	19.34	86.79
LR	0.69	74.15
bp 网络	271.36	63.54

由表 3 可以看出,KELM-Group 和 GBDT 的准确度都很高,准确度均在 92% 以上,GBDT 的准确度略微比 KELM-Group 高。但是,从训练时间上来看,KELM-Group 的优势明显。KELM-Group 算法具有非常优秀的拟合复杂非线性系统的能力,泛化性能好,学习速度快。

Lasso 算法和 LR 算法都是常用的线性模型,LR 算法可以看作无隐层的神经网络。由于计算速度快,这些线性模型被广泛应用于诸如统计分析、计算广告和搜索排名等大型数据场景中。在该数据集下,两者的训练速度都非常快,仅需十几秒,但其预测精度与 KELM-Group 有很大的差异。在相同的训练时间下,KELM-Group 的准确度明显高于 Lasso 和 LR 的准确度。从 KELM-Group 与线性模型的对比分析可以看

出,KELM-Group 算法既具有线性模型的训练速度,又不失神经网络的逼近能力。

值得说明的是,在此数据集上也对 Adaboost-SVM 进行了实验,但是实验运行了 2080 min(大约 35 h)后依然没有结果,只好终止实验。为方便比较,在子数据集上进行了实验,保留了 1/4 的训练样本,结果如表 4 所列。

表 4 KELM-Group 与 Adaboost-SVR 的对比

Table 4 Comparison of KELM-Group and Adaboost-SVR

算法	训练时间/s	准确度/%
KELM-Group(80 进程)	10.49	87.16
Adaboost-SVR	688.85	84.27

结果表明,相比 Adaboost-SVM,KELM-Group 不仅有更高的预测准确度,而且训练速度快 60 多倍。Adaboost-SVM 并不适合在大数据集上应用。

4.4 KELM-Group 和 ELM 的对比实验

本节对单个 ELM 模型和 KELM 群组模型进行了对比。ELM 的隐层节点数均为 1000。所有的实验均进行了交叉验证(Cross validation),全部的训练集被平均随机分成 4 份较小的子集,每次取其中的 3 份作为训练集,另外 1 份作为测试集。

表 5 ELM 与 KELM-Group 的对比

Table 5 Comparison of ELM and KELM-Group

算法	训练时间/s	准确度/%
ELM	2.32	64.28
Cluster-ELM(单进程)	3.28	92.99
Cluster-ELM(20 进程)	0.184	93.12
Cluster-ELM(40 进程)	0.102	93.04

从实验结果来看,KELM-Group 算法比 ELM 的准确度高出很多。大量的不同路段的时间序堆积在一起,呈现出了无规律的状态,单一的模型无法正确学习其中的规律,训练获得的模型没有实用价值。如果想用单一模型提高预测准确度,就必须人工凭经验挖掘其他特征,与时序特征一起作为输入特征进行训练。但这样势必会引起特征维数增加,导致灾难,延迟训练时间。

使用 KELM-Group 算法后,多份小样本分别被送入不同的 KELM 模型,输入样本数少,单个 KELM 的计算复杂度降低,共享不同输入权重的子模型可以同时训练,总体的运行时间反而比单一模型更快。从本节的实验数据可以发现,在某些情况下,采用多模型的 KELM-Group 要比单一模型的计算速度快(很容易并行化),同时还能获得更好的泛化能力。

4.5 预测结果评估

预测结果的准确性不仅需要定量分析其准确度,还需要结合实际情况来检验其预测结果是否符合实际情况以及是否满足市民的直观感受。为此,本文将预测结果与南宁市的视频监控相结合,对其进行了定性评价。

图 4 是预测系统在 2017 年 3 月 5 日 15 时 50 分产生的东葛路区域 10 分钟的预测结果与真实拥堵情况的对比。图 4(a)是东葛路与滨湖北路的交叉路口的 10 分钟预测可视化结果,图 4(b)是 10 分钟后的真实路口,可以看出,车流由西向东缓行,而预测结果可视化结果显示为黄色;其他方向通行畅通,对应的预测结果显示为绿色,预测结果和真实吻合;图 4(c)是东葛路与茅桥路路口的 10 分钟预测可视化结果。西

段左车道与东段右车道行车缓慢,有拥堵现象,这与10分钟前的预测结果相吻合。

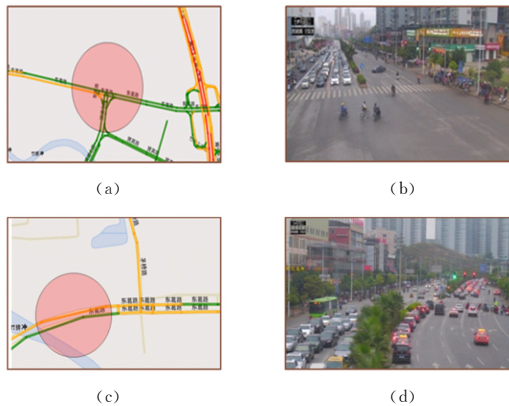


图4 东葛路区域10分钟的预测结果与10分钟后的真实路况(电子版为彩色)

Fig. 4 10-minute forecast results of Dongge Road and real road condition in 10 minutes

可以看出,该预测系统的预测准确度高,与实际情况较为符合,可靠性高,具有很强的实用价值。事实上,该系统已经投入使用一年,为南宁市交通管理部门提供了很多的决策支持。

结束语 本文对超限学习机算法进行了阐述,受到多模型的启发,提出了核超限学习机群组学习框架。核超限学习机群组是由多个KELM子模型组成,每个KELM子模型只负责某一类样本的学习。KELM群组每一类样本均能达到全局最优,可以获得比ELM更高的预测准确率。同时,KELM子模型之间相互独立,容易并行化,适合大规模计算。实验结果表明,基于核超限学习机群组的预测方法在短时预测时准确度高,并且预测的结果与实际情况较为符合,可靠性高,具有很强的实用价值。与仅使用默认值的其他算法相比,文中所提算法有较好的效果。本文未展示其他算法调参之后的对比结果,未来准备进一步对其他算法进行调参,将其与所提算法进行对比,以证明其是否仍然具有更好的效果,使实验结果更有说服力。

参考文献

[1] XIE Y L. Analyzing the application of floating vehicle data in urban intelligent transportation [J]. China Science and Technology Expo, 2015, 41(41): 260. (in Chinese)
谢彦良. 分析浮动车数据在城市智能交通中的应用[J]. 中国科技博览, 2015, 41(41): 260.

[2] SUN Z, FOX G C. Traffic Flow Forecasting Based on Combination of Multidimensional Scaling and SVM [J]. International Journal of Intelligent Transportation Systems Research, 2014, 12(1): 20-25.

[3] YU G, ZHANG C. Switching ARIMA model based forecasting for traffic flow [C] // Proc. IEEE Int. Conf. Acoust. Speech Signal Process. 2004, 2: 429-432.

[4] ZHANG Y L, YE Z R. Short-Term Traffic Flow Forecasting Using Fuzzy Logic System Methods [J]. Journal of Intelligent Transportation Systems Technology Planning & Operations, 2008, 12(3): 102-112.

[5] XUE J, SHI Z. Short-Time Traffic Flow Prediction Based on Chaos Time Series Theory [J]. Journal of Transportation Systems Engineering & Information Technology, 2008, 8(5): 68-72.

[6] CETIN M, COMERT G. Short-Term Traffic Flow Prediction with Regime-Switching Models [J]. Transportation Research Record Journal of the Transportation Research Board, 2006, 1965(1): 23-31.

[7] ABDULHAI B, PORWAL H, RECKER W. Short-Term Traffic Flow Prediction Using Neuro-Genetic Algorithms [J]. ITS Journal, 2002, 7(1): 3-41.

[8] ZHAO L, WANG F Y. Short-term traffic flow prediction based on ratio-median lengths of intervals two-factors high-order fuzzy time series [C] // IEEE International Conference on Vehicular Electronics and Safety (ICVES 2007). IEEE, 2007: 1-7.

[9] MUSSONE L. A review of feedforward neural networks in transportation research [J]. E & I Elektrotechnik Und Informationstechnik, 1999, 116(6): 360-365.

[10] JIAO G T, XU J M, MA Y H. Research on traffic volume prediction method based on QPSO-RBF [J]. Traffic information and Safety, 2008, 26(4): 128-131. (in Chinese)
焦光庭, 徐建闽, 马永红. 基于 QPSO-RBF 的交通量预测方法研究 [J]. 交通信息与安全, 2008, 26(4): 128-131.

[11] ZHENG W, LEE D, SHI Q, et al. Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach [J]. Journal of Transportation Engineering-asce, 2006, 132(2): 114-121.

[12] XU Q H, DING Z K, BI X Y. A real-time traffic flow prediction method based on dynamic recurrent neural network [J]. Journal of Huaihai Institute of Technology (Natural Science Edition), 2004, 12(4): 14-17. (in Chinese)
徐启华, 丁兆奎, 毕训银. 一种基于动态递归神经网络的交通流量实时预测方法 [J]. 淮海工学院学报(自然科学版), 2004, 12(4): 14-17.

[13] YANG Q F, ZHANG B, GAO P. Short-term traffic volume prediction method based on improved dynamic recurrent neural network [J]. Journal of Jilin University (Engineering Edition), 2012, 4(4): 887-891. (in Chinese)
杨庆芳, 张彪, 高鹏. 基于改进动态递归神经网络的交通量短时预测方法 [J]. 吉林大学学报(工学版), 2012, 4(4): 887-891.

[14] YUN S Y, NAMKOONG S, SHIN S W, et al. Application of a recurrent neural network to traffic volume forecasting [C] // Intelligent Transportation: Realizing the Future. Abstracts of the Third World Congress on Intelligent Transport Systems, 1996.

[15] HUANG G B, ZHU Q Y, SIEW C K. Extreme Learning Machine: Theory and Applications [J]. Neurocomputing, 2006, 70(1): 489-501.

[16] HUANG G B, ZHOU H, DING X, et al. Extreme learning machine for regression and multiclass classification [J]. IEEE Trans. Syst. Man Cybern. Part B, 2012, 42(2): 513-529.

[17] PAL M, MAXWELL A E, WARNER T A. Kernel-based extreme learning machine for remote-sensing image classification [J]. Remote Sensing Letters, 2013, 4(9): 853-862.

[18] KIVINEN J, SMOLA A J, WILLIAMSON R C. Learning with Kernels [M]. MIT Press, 2002.