

基于深度学习的城市轨道交通短时客流量预测

龙小强¹, 李捷², 陈彦如^{3†}

1. 广州市交通运输研究所与广州市公共交通研究中心, 广州 510627;
2. 北京东方科技集团股份有限公司 CIO 组织数字化应用中心, 北京 100016;
3. 西南交通大学 经济管理学院, 成都 610031)

摘要: 我国城市轨道交通已进入快速发展期, 准确预测城轨交通短时客流量, 对于城轨运营安全、运营效率及运营成本具有重要意义. 城轨交通短时客流量由于具有强随机性、周期性、相关性及非线性的特征, 浅层模型的预测精度并不理想. 对此, 基于深度信念网络 (DBN) 和支持向量回归机 (SVM), 提出城轨交通短时客流深层预测模型 (DBN-P/GSVM), 同时基于遗传算法 (GA) 和粒子群算法 (PSO) 实现 SVM 的参数寻优. 最后, 对成都地铁火车北站客流量预测进行实例分析. 结果表明, DBN-P/GSVM 深度预测模型在均方误差、均方根误差、绝对误差均值及绝对百分比误差均值等方面均优于浅层模型——GA-SVM 模型、PSO-SVM 模型和 BP 神经网络模型, 以及深层模型长短期记忆网络 (LSTM) 与 LSTM-Softmax.

关键词: 城轨交通短时客流量; 深度信念网络; 支持向量机; 遗传算法; 粒子群算法; 长短期记忆网络
中图分类号: TP273 **文献标志码:** A

Metro short-term traffic flow prediction with deep learning

LONG Xiao-qiang¹, LI Jie², CHEN Yan-ru^{3†}

1. Guangzhou Transport Research Institute & Guangzhou Public Transport Research Center, Guangzhou 510627, China;
2. CIO-Digital Application Center, BOE Technology Group Co., Ltd, Beijing 100016, China;
3. School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: At present, China's urban rail transit is developing rapidly. Short-term traffic flow prediction plays an important role on the metro safety, efficiency and cost. Many existing approaches with shallow architecture failed to provide favorable results, because short-term traffic flow are highly random, cyclical, correlative and non-linear. Therefore, we propose a prediction model with deep architecture — DBN-P/GSVM based on the deep believe network (DBN) and support vector machine (SVM). The parameters of the SVM are obtained based on the genetic algorithm (GA) and the particle swarm optimization (PSO). Abundant experiments are conducted on the Chengdu Metro North Railway Station. The results show that the proposed DBN-P/GSVM model performs better than such shallow architecture models as the GA-SVM, PSO-SVM and back propagation neural network (BPNN) and such deep architecture models as the long short-term memory (LSTM) and LSTM-Softmax in terms of mean squared error, root mean square error, mean absolute error and mean absolute percentage error.

Keywords: metro short-term traffic flow; deep belief network; support vector machine; genetic algorithm; particle swarm optimization; long short-term memory

0 引言

近年来, 我国的城市轨道交通发展十分迅猛. 截至 2017 年末, 我国内地共计 34 个城市开通城市轨道交通并投入运营, 开通线路 165 条, 运营线路长度达到 5 033 km^[1]. 然而, 随着城市轨道交通网络的日益完善及客流量的不断增加, 客流过度饱和现象经常发生, 尤其是工作日的早晚高峰时段, 由此带来了巨大的安全隐患. 因此, 如何准确预测城轨交通短时客流

量, 并据此制定科学的客运计划和组织方案, 对于维护轨道交通的运营安全、提高运营效率、降低运营成本都具有重要价值.

现阶段城轨交通客流预测方法主要分为以下几种: 1) 线性预测模型, 如时间序列预测模型^[2]、卡尔曼 (Kalman) 滤波模型^[3-4]等; 2) 非线性预测模型, 如小波预测模型^[5]、混沌预测模型^[6-7]及非参数预测模型^[8-10]等; 3) 仿真预测模型, 如元胞自动机预测法^[11]

收稿日期: 2018-10-13; 修回日期: 2019-01-07.

基金项目: 国家自然科学基金项目 (51578465, 71771190).

责任编辑: 魏秀琨.

†通讯作者. E-mail: chenyanru@swjtu.cn.

和交通模拟预测法^[12]等;4)浅层机器学习预测模型,如支持向量机^[13-15]、神经网络^[16-20]等;5)组合预测模型^[21-22]。以上研究成果对于城轨交通短时客流预测具有重要的意义,但也存在局限性,如:线性预测模型在处理具有强随机性和非线性特征的城市轨道交通短时客流时,无法完全获取短时客流的内在特征;线性映射很难充分体现短时客流数据的发展趋势;非线性预测模型能够描述城轨交通短时客流量数据的非线性特征,但面对海量的小粒度短时客流数据,其预测精度有待进一步提高;仿真预测模型通常建模成本较高,且模型计算效率很难满足时效性要求;浅层的机器学习预测模型在处理大数据时容易出现过学或者欠学问题;组合模型中组合方式及组合内容对最终模型的预测效果影响较大,预测精度不稳定。

由于深度学习模型可以较为准确地描述输入层与输出层的复杂关联^[23-24],人们尝试使用深度学习预测模型解决交通流预测问题^[23-26]。在深度学习模型中,深度信任网络(DBN)被认为最为有效^[23]。因此,本文尝试将DBN应用到城市轨道交通短时客流预测中。由于DBN是一种无监督学习模型,而对于城轨交通短时客流预测,需要有监督学习来实现模型的预知能力,以进行客流量预测。作为一种有监督学习的模型——支持向量机(SVM),能够较为精确地预测具有非线性、非静态的时间序列^[14]。因此,本文基于DBN和SVM,提出城轨交通短时客流预测深层模型DBN-P/GSVM。该模型结合DBN的无监督学习和SVM的有监督式学习来对样本数据进行训练,通过底层DBN对数据进行重构与特征提取,为顶层SVM提供有效的数据表达;而顶层的SVM对底层DBN的权值进行微调,使得整体模型更加契合数据变化规律,从而提高城轨交通短时客流量预测精度。

1 DBN-P/GSVM模型的构建

本文所构建的城市轨道短时客流量预测模型DBN-P/GSVM由底层和顶层两部分构成:底层为DBN模型,将根据城市轨道交通短时客流量数据,在无监督式学习模式下逐层进行权值训练,保证每一层的重构误差最小;顶层为SVM,其将DBN顶层输出数据作为输入,对城市轨道交通短时客流数据进行预测,并将预测值与真实值的误差回传到DBN的各个层级之间,通过各个层级的权值调整,使得预测值与真实值的差距最小,从而实现有监督学习。其框架如图1所示。

1.1 底层DBN模型

DBN模型是2006年由Hinton等^[27]提出的一种深度学习模型。它通过在输入观测数据与数据标签

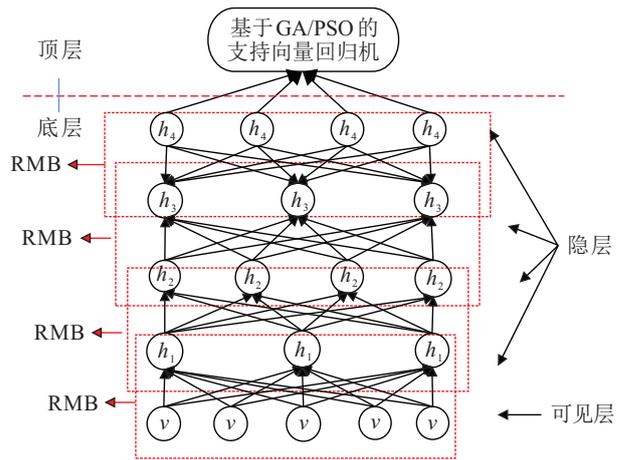


图1 DBN-P/GSVM模型结构

之间形成一个联合分布,对数据标签与观测值之间进行双向评估。DBN由若干限制玻尔兹曼机(RBM)堆叠形成。第1个RBM的可见层即为DBN模型的输入层。通过计算获取第1个RBM隐层的特征表达,并将其转换为第2个RBM的可见层。以此类推,完成DBN模型的整个训练过程。该过程能够对原始输入数据不断地进行特征提取,每一层的输出数据都是输入数据的另一种特征表达。

1.1.1 底层DBN模型输入数据集构造

DBN的输入数据为城轨交通短时客流量。通常客流量统计时间间隔采用5 min、10 min、15 min或30 min。为了更精确地预测短时客流量,本文采用的城轨交通客流量统计时间间隔为5 min。影响客流量的因素主要包括纵向随机性的影响和横向周期性的影响,即某时段客流量会受到同一天内该时段前面若干时段客流的影响以及若干天前该时段客流的影响。假设需要预测的城轨交通客流量为 x_{ij} , i 表示天数, j 表示时段。 $(x_{i-n,j}, \dots, x_{i-2,j}, x_{i-1,j})$ 表示第 i 天的前 n 天第 j 时段的城轨交通客流量,体现城轨交通客流量数据受到的周期性影响; $(x_{i,j-m}, \dots, x_{i,j-2}, x_{i,j-1})$ 表示第 i 天内第 j 时段的前 m 个时段的城轨交通客流量,体现城轨交通客流量数据受到的随机性影响。则DBN-G/PSVM模型预测使用的数据集 X 可以表示如下:

$$X = \begin{bmatrix} x_{i-n,j-m} & \dots & x_{i-2,j-m} & x_{i-1,j-m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{i-n,j-2} & \dots & x_{i-2,j-2} & x_{i-1,j-2} \\ x_{i-n,j-1} & \dots & x_{i-2,j-1} & x_{i-1,j-1} \\ x_{i-n,j} & \dots & x_{i-2,j} & x_{i-1,j} \end{bmatrix} \rightarrow$$

$$\left[\begin{array}{cccc} x_{i,j-m-m} & \cdots & x_{i,j-m-1} & x_{i,j-m} \\ \vdots & & \vdots & \vdots \\ \leftarrow x_{i,j-m-2} & \cdots & x_{i,j-3} & x_{i,j-2} \\ x_{i,j-m-1} & \cdots & x_{i,j-2} & x_{i,j-1} \\ x_{i,j-m} & \cdots & x_{i,j-1} & x_{i,j} \end{array} \right]$$

其中:最后一列为待预测数据,其余列作为输入数据传至DBN模型的输入层.

1.1.2 底层DBN模型的权值训练

DBN模型的权值训练分为两个阶段:第1阶段为无监督学习模式下的逐层权值训练;第2阶段为有监督学习模式下的整体权值微调.

1) 第1阶段权值训练.

在此阶段,将整个DBN模型的底层网络结构进行分解,得到多个单一的RBM模型.在无监督学习模式下逐层产生权值,并且保证每层的权值对该层特征向量的映射最优.

在单个RBM中,可见层设为 $\mathbf{v} = (v_1, v_2, \dots, v_i)$, 隐层设为 $\mathbf{h} = (h_1, h_2, \dots, h_j)$. 比较常见的RBM为Bernoulli-Bernoulli RBM. 因本文中城轨交通短时客流量为实数,故选取 Gaussian-Bernoulli RBM^[23]. 于是,单个RBM的能量函数以及显层和隐层节点对应的条件概率为

$$E(\mathbf{v}, \mathbf{h}; \theta) = - \sum_i^I \sum_j^J \omega_{ij} \frac{v_i}{\sigma_i} h_j - \sum_i^I \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_j^J b_j h_j, \tag{1}$$

$$p(h_j | \mathbf{v}; \theta) = \text{sigm} \left(\sum_i^I \omega_{ij} v_i + b_j \right), \tag{2}$$

$$p(v_i | \mathbf{h}; \theta) = N \left(a_i + \sigma_i \sum_j^J h_j \omega_{ij}, \sigma_i^2 \right). \tag{3}$$

其中: $\theta = \{\omega_{ij}, a_i, b_j\}$, ω_{ij} 为显层节点 v_i 与隐层节点 h_j 之间的关联权值, a_i 为显层节点偏移量, b_j 为隐层节点偏移量, I 为显层结构的节点数, J 为隐层结构的节点数, σ_i 为 v_i 的标准差, $\text{sigm}(x)$ 是 sigmoid 函数, $N(\mu, \sigma^2)$ 是均值为 μ 、方差为 σ^2 的高斯分布.

由式(1)可得,可见层节点与隐层节点的联合概率分布为

$$P_\theta(\mathbf{v}, \mathbf{h}) = \frac{1}{Z(\theta)} \exp(-E(\mathbf{v}, \mathbf{h}; \theta)), \tag{4}$$

$$Z(\theta) = \sum_{\mathbf{v}} \sum_{\mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h}; \theta)), \tag{5}$$

其中 $Z(\theta)$ 为归一化因子. 根据式(4)可得训练数据 \mathbf{v} 的概率分布 $P_\theta(\mathbf{v})$ (即 $P_\theta(\mathbf{v}, \mathbf{h})$ 的边缘概率分布)为

$$P_\theta(\mathbf{v}) = \frac{1}{Z(\theta)} \sum_{\mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h}; \theta)). \tag{6}$$

通过对式(6)构造对数似然函数,使得输入样本在所求分布下概率最大,从而求得最优参数 θ .

2) 第2阶段权值训练.

第1阶段的权值训练目标是每层的权值对该层特征向量的映射最优. 第2阶段的权值训练目标是整个模型最终预测结果与真实值之间的误差最小. 第1阶段权值训练突出单层最优,第2阶段权值训练突出整体最优.

将城轨交通短时客流量数据 X 传至DBN的可见层 \mathbf{v} , 通过第1阶段训练好的DBN得到输出向量 \mathbf{v}^* , 将 \mathbf{v}^* 作为顶层SVM的输入层, 预测出城轨交通短时客流量 \mathbf{v}^{**} , 对比 \mathbf{v}^{**} 与实际客流量, 将两者差值回传到DBN底层网络结构中, 微调整个DBN的权值.

在DBN-P/GSVM模型中体现了有监督和无监督两种学习方式,使得底层DBN可以有效提取出城轨交通短时客流量的内在特征,为顶层SVM预测提供了更加优质的数据准备;而顶层的SVM则微调底层DBN的权值,实现整体模型误差最小.

1.2 DBN-P/GSVM模型顶层支持向量回归机设计

1.2.1 支持向量机回归机原理

假设训练样本集 $T = \{(x_i, y_i), i = 1, 2, \dots, N\}$, $x_i \in R^n, y_i \in R$, 其中 N 为训练的样本容量. SVM通过获取一个非线性映射,将输入数据映射到高维特征空间中,并在该空间构建回归方程 $f(x)$. $f(x)$ 可以定义为

$$f(x) = \mathbf{w}^T \varphi(x) + b. \tag{7}$$

其中: \mathbf{w} 为权向量, b 为偏置量. 引入松弛变量 ξ_i, ξ_i^* , 则优化问题为

$$\min \frac{1}{2} \mathbf{w}^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*). \tag{8}$$

$$\text{s.t. } y_i - (\mathbf{w}^T \varphi(x) + b) \leq \varepsilon + \xi_i, i = 1, 2, \dots, N; \tag{9}$$

$$(\mathbf{w}^T \varphi(x) + b) - y_i \leq \varepsilon + \xi_i^*, i = 1, 2, \dots, N; \tag{10}$$

$$\xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, N.$$

其中: C 为惩罚参数, ξ_i, ξ_i^* 为非负松弛变量. 通过引入拉格朗日乘子,将式(8)变为其对偶形式

$$\max \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N a_i a_j \varphi(x_i) \varphi(x_j) y_i y_j; \tag{11}$$

$$\text{s.t. } \sum_{i=1}^N a_i y_i = 0, 0 \leq a_i \leq C, i = 1, 2, \dots, N. \tag{12}$$

因为二次规划原问题与对偶问题的解相同,所以最优解为 $a^* = \{a_1^*, a_1^*, \dots, a_N^*\}^T$, 从而计算 w^* 和 b^* , 最终获得回归方程

$$f(x) = \sum_{i=1}^N \sum_{j=1}^N a_i^* y_i \varphi(x_i) \varphi(x_j) + b^*. \quad (13)$$

为了避免维数灾难,通常采用核函数 $K(x_i, x_j)$ 替代式(13)中的 $\varphi(x_i)\varphi(x_j)$. 比较常见的核函数为高斯径向基核函数(RBF),其函数形式为

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{\sigma^2}\right), \quad (14)$$

其中 σ 为带宽,是高斯核函数的参数.

采用 SVM 进行回归预测时,需要提前确定惩罚参数 C 与所选核函数的参数 μ (若选择 RBF 核函数,则参数 μ 即为 σ). 传统的 SVM 参数通过交叉验证法获取,容易陷入局部收敛. 本文尝试使用遗传算法(GA)和粒子群算法(PSO)对 SVM 的参数进行优化.

1.2.2 GA-SVM 的设计

遗传算法是一种基于遗传学理论的启发式算法,通过优胜劣汰的竞争机制在解空间内迭代寻优,并通过遗传操作保证解空间的多样性,使求解过程不易陷入局部最优. GA-SVM 的设计主要包含 3 个方面:编码方式设计、适应度函数设计和遗传算子设计.

1) 编码方式.

由于只需要 (C, μ) 进行优化,优化参数较少,本文采取二进制编码方式,即将参数组合 (C, μ) 表示成一条由随机二进制数组成的染色体,其中每一位二进制数表示一个基因位,多个染色体形成种群. 根据各参数的取值范围确定染色体的长度,即

$$C_{\text{length}} = \log_2[(C_{\text{max}} - C_{\text{min}}) \times 100], \quad (15)$$

$$\mu_{\text{length}} = \log_2[(\mu_{\text{max}} - \mu_{\text{min}}) \times 100]. \quad (16)$$

其中: C_{length} 表示参数 C 的二进制编码长度, μ_{length} 表示参数 μ 的二进制编码长度, C_{max} 、 C_{min} 、 μ_{max} 、 μ_{min} 分别为参数 C 和 μ 的最大值与最小值. 染色体具体结构如表 1 所示,其中各二进制数随机产生.

表 1 参数的二进制编码

惩罚参数 C				核函数参数 μ			
1_1	0_2	\dots	$1_{C_{\text{length}}}$	0_1	0_2	\dots	$1_{\mu_{\text{length}}}$

2) 适应度函数.

适应度函数通过评价每一组参数 (C, μ) 的适应度来引导迭代过程向适应度高(解质量好)的方向进行. 本文借助 K 折交叉验证的方法,并采用样本均方误差(MSE)来评估解的质量. MSE 的计算公式如下:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (f(x_i) - y_i^*)^2. \quad (17)$$

其中: N 为训练样本数量, $f(x_i)$ 为城轨交通短时客流量预测值, y_i^* 为城轨交通短时客流量实际值. 相应的适应度函数设计如下:

$$\text{Fit}(C, \sigma) = \begin{cases} \frac{1}{\text{MSE}}, & \text{MSE} \neq 0; \\ \text{Fit}_{\text{best}} + 1, & \text{MSE} = 0. \end{cases} \quad (18)$$

其中 Fit_{best} 表示 GA 当前找到最优的参数组合 (C, μ) 的适应度值. 当 MSE 值较小时,表示 SVM 采用该组参数预测出的城轨交通短时客流量与实际值差距较小,则这组 (C, μ) 的适应度较高.

3) 遗传算子.

GA 通过选择、交叉和变异 3 个遗传操作使算法向最优解不断逼近,通过保持种群多样性来丰富解空间,从而防止算法陷入局部最优的困境. 在选择方面,本文通过轮盘赌法来选取用于交叉变异的父代;在交叉方面,任意选择两个基因位,将两个父代在这两个基因位之间的基因段进行互换,从而生成两个子代;在变异方面,采用单点变异的方式,即随机选取一位变异位,对其进行 $0 \rightarrow 1$ 或者 $1 \rightarrow 0$ 的操作.

运用 GA 算法优化 SVM 参数组合 (C, μ) 的具体实现步骤如下:

Step 1: 确定 GA 及 SVM 运行参数(交叉率、变异率、最大迭代次数、不敏感损失参数和种群规模等);

Step 2: 确定待优化参数 (C, μ) 的取值范围 $[C_{\text{min}}, C_{\text{max}}]$ 和 $[\mu_{\text{min}}, \mu_{\text{max}}]$;

Step 3: 建立种群 $N \times (C_{\text{length}} + \mu_{\text{length}})$ 矩阵,并随机初始化种群;

Step 4: 计算每一组参数 (C, μ) 的适应度;

Step 5: 进行选择、交叉、变异操作,并以此更新种群;

Step 6: 判断算法是否达到终止要求,如果是,则进入 Step 7,否则重复 Step 4 和 Step 5;

Step 7: 将最优参数组合 (C, μ) 作为 SVM 运行参数进行预测.

1.2.3 PSO-SVM 模型的设计与实现

粒子群算法(PSO)是一种利用粒子来模拟生物行为的启发式算法. 在 PSO 算法中,搜索空间中的每个粒子都是待优化问题的一个解,通过粒子的速度及位置的变化在解空间中寻找最优解. 相较于 GA, PSO 的操作更加简单,它的设计包含两个方面:粒子速度及位置更新函数的设计和适应度函数的设计.

1) 粒子速度及位置更新函数.

由于待优化参数为两个—— C 和 μ ,本文设计二维粒子,即粒子在二维的空间进行搜索.粒子速度及位置更新函数如下:

$$v_{id}^{t+1} = \omega v_{id}^t + c_1 r_1 (p_{id}^t - x_{id}^t) + c_2 r_2 (g_d^t - x_{id}^t), \quad (19)$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1}. \quad (20)$$

其中: x_{id}^t 为粒子 i 第 d 维搜索空间第 t 代的位置, v_{id}^t 为粒子 i 第 d 维搜索空间第 t 代的速度, p_{id}^t 为粒子 i 第 d 维搜索空间中前 t 代去过的最优位置, g_d^t 为所有粒子第 d 维搜索空间中前 t 代去过的最优位置, c_1 和 c_2 为局部学习率和全局学习率, r_1 、 r_2 为随机影响因子, ω 为惯性权重, t 为迭代次数.

通过全局最优位置与当前粒子位置的差距来引导粒子向最优位置飞行,同时加入扰动随机影响因子以减小算法陷入局部最优的可能.

2) 适应度函数设计.

PSO中适应度函数的作用与GA中适应度函数的作用相同.为了保证两种算法的可比性,PSO算法的适应度函数设计仍然沿用式(17)和(18)的形式.

3) PSO-SVM模型的实现过程.

Step 1: PSO算法运行参数设置,其中包括种群规模、学习因子、惯性权重和最大迭代次数等;

Step2: 根据 (C, μ) 的取值范围确定两个维度的搜索范围 $[x_d^{\min}, x_d^{\max}]$ 和 $[v_d^{\min}, v_d^{\max}]$;

Step 3: 初始化各粒子的位置和速度;

Step 4: 根据式(17)和(18)计算粒子的适应度,并将各粒子的初始位置设为该粒子的最优位置,寻找所有粒子的最优位置,将其设置成粒子群最优位置;

Step 5: 根据式(19)和(20)更新各粒子的速度及位置,计算更新后粒子的适应度及 p_{id}^t 和 g_d^t ;

Step 6: 判断是否满足终止条件,若满足则算法终止并输出参数 (C, μ) ,否则重复Step 5;

Step 7: 将输出的 (C, μ) 输入到SVM模型中进行预测.

2 实例分析

在成都地铁运营网络中,成都火车北站地铁站是非常重要的交通枢纽,它集火车站、二环高架BRT站和地铁站3种交通方式于一体,客流庞大,并且客流的到达分布由于受到其他交通方式的影响,随机特征十分明显.因此,本文以成都火车北站地铁站为实验对象,收集了2016年3月间成都地铁所有站点的运营数据,运用所构建的DBN-P/GSVM模型进行短时客流预测.将成都火车北站地铁站2016年3月份的运

营数据进行分析后,选择其中数据最为完善的两周作为实验时段.为了验证工作日与非工作日乘客出行行为是否存在较大差异,本文随机选择一周对工作日进站客流量与非工作日进站客流量进行统计.同时,为了体现轨道交通客流的变化特征,对客流量统计时间间隔分别采用5 min、10 min、15 min和30 min对客流量进行统计,统计结果如图2和图3所示.

由图2和图3可知,不同统计时间间隔的工作日数据与非工作日数据都呈现了非线性的变化趋势.就工作日而言,地铁进站客流存在明显的早高峰现象:8点~9点的进站客流量是一天当中的峰值;19点之后地铁客流量逐渐减少,且无明显的晚高峰情况.

此外,周一至周五的客流变化规律存在一定的相似性,表明地铁客流量工作日数据存在一定的周期性.同时,非工作日的地铁进站客流量数据也表现出一定的周期性,但周日地铁客流量高峰集中在12点~17点之间,没有明显的峰值.

对于两类数据集,随着统计时间间隔的增加,进站客流量的变化曲线变得逐渐平滑,说明在短时间间隔的情况下,地铁客流量数据存在较强的随机性.此外,本文对工作日与非工作日的数据集进行了相关分析,结果如表2和表3所示.

表2 工作日数据的相关分析

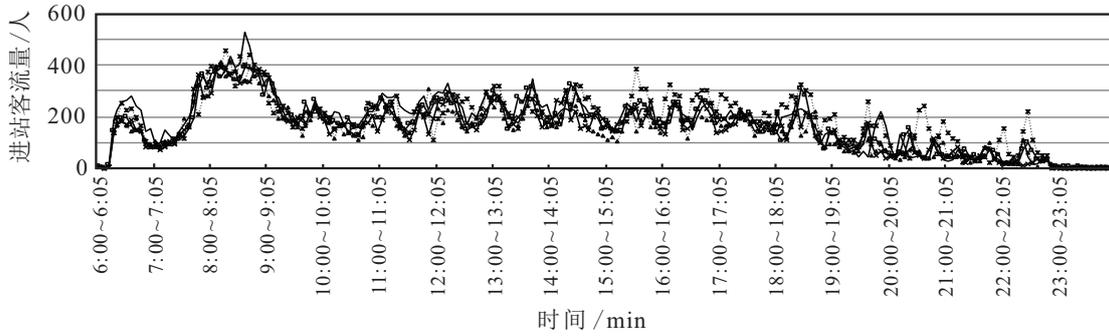
指标	周一	周二	周三	周四	周五	
周一	Pearson相关性	1	0.924	0.926	0.905	0.845
	显著性(双尾)		0	0	0	0
周二	Pearson相关性	0.924	1	0.918	0.930	0.860
	显著性(双尾)	0		0	0	0
周三	Pearson相关性	0.926	0.918	1	0.923	0.851
	显著性(双尾)	0	0		0	0
周四	Pearson相关性	0.905	0.930	0.923	1	0.846
	显著性(双尾)	0	0	0		0
周五	Pearson相关性	0.845	0.860	0.851	0.846	1
	显著性(双尾)	0	0	0	0	

表3 非工作日数据的相关分析

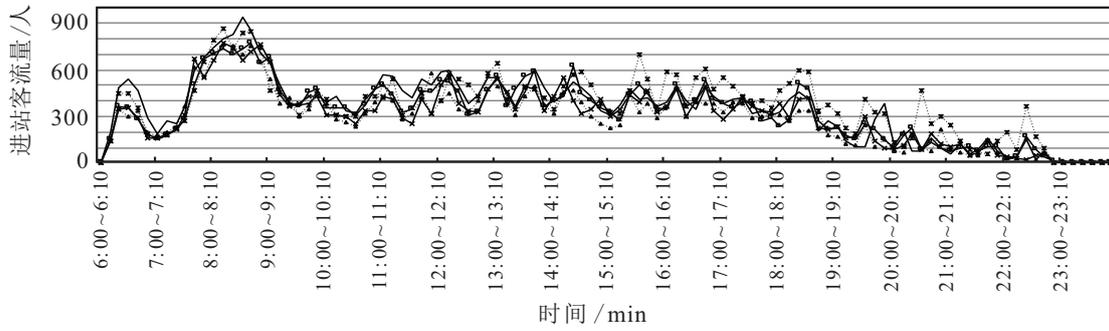
指标		周六	周日
周六	Pearson 相关性	1	0.809
	显著性(双尾)		0
周日	Pearson 相关性	0.809	1
	显著性(双尾)	0	

由表2和表3可知,工作日数据之间与非工作日数据之间的相关性通过显著性检验,体现了其关联性.

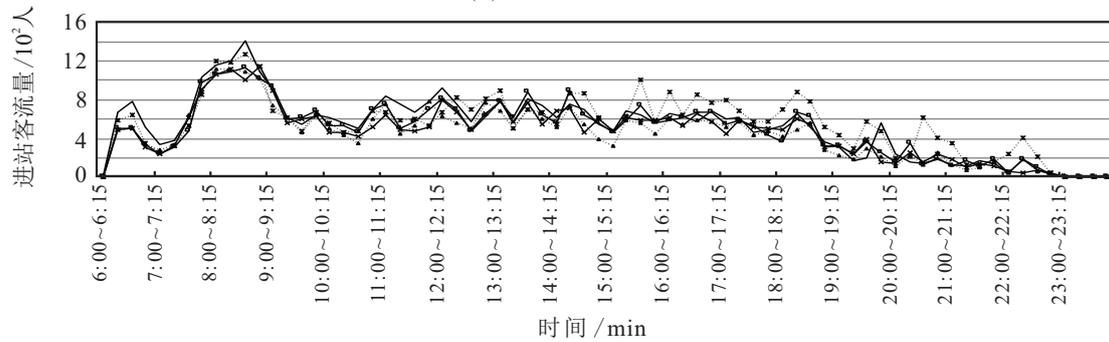
此外由上述图表可知,非工作日与工作日乘客出行行为及客流量的变化趋势存在较大差异,因此,本实验将数据分为工作日数据集和非工作日数据集两个部分,分别对以5 min为间隔的短时客流进行预测.



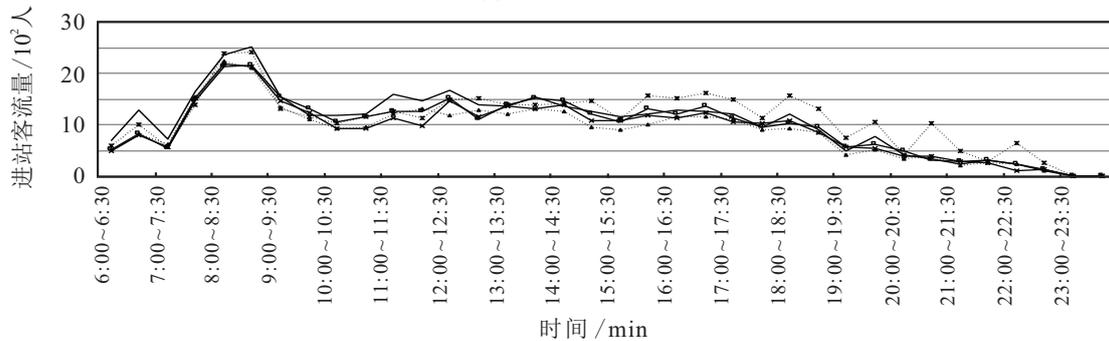
(a) 间隔时间为5min



(b) 间隔时间为10min



(c) 间隔时间为15min



(d) 间隔时间为30min



图2 工作日进站客流量

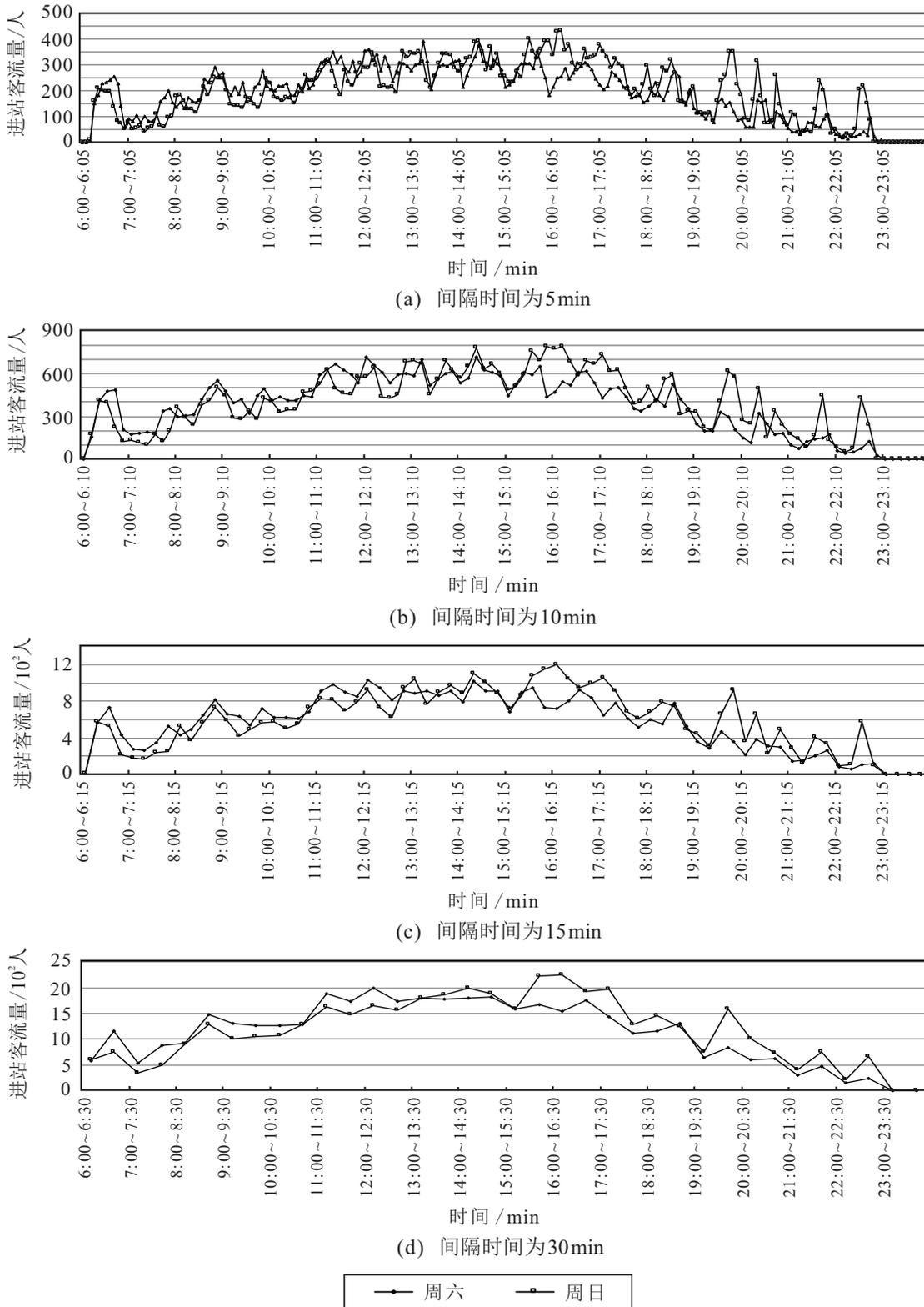


图3 非工作日进站客流量

2.1 数据输入

综合考虑数据规模及预测效率,后续实验中采用 $n = 3, m = 2$ 的输入数据集,即某时段客流量受到前3天该时段客流以及同一天该时段的前两个时段客流的影响.于是,DBN-P/GSVM模型预测使用的样本数据集为 $(x_{i-3,j}, x_{i-2,j}, x_{i-1,j}, x_{i,j-2}, x_{i,j-1}, x_{i,j})$,其中最后一个为待预测数据,前5个为输入数据,将

传输至底层DBN的可见层 v .

2.2 支持向量回归机核函数选择

在支持向量机中,核函数将样本数据从低维非线性特征空间映射到高维线性特征空间.不同的核函数将样本映射到不同的空间,从而导致支持向量机的预测结果存在差异.为了测试样本数据适合的核函数类型,本实验分别选择最常用的4种核函数——线

性核、多项式核、RBF核和Sigmoid核,对其映射性能进行对比.其中,支持向量机的不敏感损失参数为0.01.运用4种核函数的模型预测相对误差如表4所示.

表4 核函数的预测相对误差 %

时间	核函数			
	线性核函数	多项式核函数	RBF核函数	Sigmoid核函数
周一	25.82	27.78	24.54	46.90
周二	20.15	13.72	15.74	146.93
周三	25.95	37.67	22.56	125.85
周四	14.07	22.33	12.47	78.51
周五	19.83	16.19	19.39	82.94
平均误差	21.16	23.54	18.94	96.23
非工作日	34.24	27.37	29.89	31.98

由表4可知,对于工作日数据集而言,运用RBF核函数的模型预测性能最好,其平均预测误差为18.94%.而运用Sigmoid核函数的模型预测效果最差,其平均误差高达96.23%.因此,在后续实验中工作日数据将采用RBF核函数.

对于非工作日,运用多项式核函数的模型预测效果最好,其预测误差为27.37%;而运用线性核函数的模型预测效果最差,其预测误差为34.24%.因此,在后续实验中非工作日数据将采用多项式核函数.

2.3 优化算法选择

SVM参数寻优的过程中,不同的优化算法所获得的参数也是不同的,而参数直接影响SVM的预测性能,进而影响整体预测模型的精度.本实验分别采用GA、PSO对SVM的参数进行寻优,并对工作日与非工作日的预测结果进行对比.算法参数:最大迭代次数为100,种群规模均为20,交叉率为0.4,变异率为0.01,两个待优化参数的寻优区间均为[0.01,1000];局部学习率和全局学习率均为1.5.实验结果如表5所示.

表5 优化算法性能对比 %

时间	相对误差	
	GA	PSO
周一	24.60	24.54
周二	13.91	15.74
周三	25.82	22.56
周四	14.86	12.47
周五	15.64	19.39
平均误差	18.97	18.94
非工作日	27.37	27.48

由表5可知:对于工作日,PSO略优于GA;对于非工作日,GA略优于PSO.因此,在后续实验中,工作日将采用PSO进行SVM参数寻优,而非工作日将采用GA进行SVM参数寻优.

2.4 DBN模型深度选择

DBN模型的深度指隐层层数.深度不同,模型的预测性能会有所差异.研究表明,DBN深度的增加并

不意味着其识别性能的增加,反而还会导致训练时间增加^[28].因此,为了选择合适的模型深度,本文综合考虑样本规模以及类似研究的模型深度,分别选择深度为2、3、4、5进行测试.模型的结构参数设置为:各层节点数为10,激活函数为Sigmoid函数,学习速率为1,动量为0.5.不同深度的模型实验结果如表6所示.

表6 DBN模型深度选择对比 %

时间	相对误差			
	2层	3层	4层	5层
周一	16.94	18.06	10.38	13.21
周二	19.22	27.83	23.10	24.08
周三	24.42	27.30	28.39	28.74
周四	25.70	24.38	12.68	24.61
周五	32.26	18.82	21.68	31.10
平均误差	23.71	23.28	19.28	24.35
非工作日	24.87	24.51	19.74	17.02

由表6可知,对于大多数数据,模型的预测误差没有随着深度的增加而减少.如周二的的数据,随着模型深度的增加,模型预测的相对误差反而在增加.同时对于不同的数据集,不同深度的模型预测误差有所不同.对于工作日的的数据而言,平均预测误差在19.28%~24.35%之间.其中深度为4的模型平均预测误差最小,表明对于城轨交通短时客流量工作日数据,4个隐层的DBN模型的数据特征提取能力更好.因此在后续实例中,工作日数据集将采用4个隐层的网络结构.对于非工作日的的数据而言,整体预测误差在17.02%~24.87%之间,其中深度为5层的DBN模型预测误差最小.因此在后续实例中,非工作日数据将采用5层隐层的网络结构.

2.5 DBN模型隐层节点个数选择

隐层节点个数对DBN模型预测精度会产生影响.考虑到输入数据维度较低,本实验将节点数分别设计为5、10、20、30、40,对不同节点数的网络结构预测性能进行对比.工作日数据集中采用4层隐层的网络结构,非工作日数据集采用5层隐层的网络结构.实验结果如表7所示.

表7 DBN模型隐层节点数选择 %

时间	相对误差				
	5个节点	10个节点	20个节点	30个节点	40个节点
周一	19.36	10.38	25.52	12.89	21.26
周二	23.98	23.10	39.87	17.15	39.58
周三	31.95	28.39	54.13	40.12	57.93
周四	23.74	12.86	44.49	31.52	58.11
周五	24.12	22.99	25.25	22.94	48.30
平均误差	24.63	19.54	37.85	24.92	45.04
非工作日	26.21	17.89	31.07	42.71	43.75

由表7可知,隐层网络节点个数会严重影响模型的预测精度.无论是工作日数据还是非工作日数据,

节点数太多或太少都无法得到理想的结果. 工作日数据的平均误差在 19%~46% 之间, 非工作日数据的平均误差在 17%~44% 之间. 两组数据中隐层节点数为 10 时误差最低. 当隐层节点数为 40 时, 无论是工作日数据, 还是非工作日数据, 模型的预测精度都最低. 原因可能是过多的节点导致更加复杂的数据映射, 从而模型对训练样本拟合程度过高, 出现过学现象.

2.6 DBN-P/GSVM模型算法性能研究

根据前面实验结果, 对于工作日数据, 由 4 层 10 节点的 DBN 及 RBF 核函数的 PSO-SVM 组成的深度模型预测效果最佳; 对于非工作日数据, 由 5 层 10 节点的 DBN 及多项式核函数 GA-SVM 组成的深度模型预测效果最佳.

为验证所提出的 DBN-P/GSVM 模型的有效性, 本文同时采用浅层预测模型 GA-SVM(用 GA 进行参数寻优的 SVM)、PSO-SVM(用 PSO 进行参数寻优的 SVM) 和浅层神经网络模型(BPNN) 分别对同一数据

集进行预测.

以 5 min 间隔的统计数据集为例, 对于工作日, 一天从早上 6:00 至晚上 24:00, 总计 18 个小时, 每小时统计 12 组数据, 共 216 组数据, 剔除一天当中城轨交通客流量比较稀疏的数据, 留存 190 组数据. 其中每组数据包含 5 个输入数据和 1 个输出数据, 输入数据为前 3 天的同一时段的客流量以及同一天当前时段的前两个时段的客流量, 输出数据为当前时段客流量, 由此形成了输入数据与输出数据的一一对应关系. 为了提高模型的鲁棒性, 降低模型过学的可能性, 本文对各输入输出数据对的顺序进行随机排列. 将前 170 组数据作为训练样本, 后 20 组数据作为测试样本. 因为在非工作日数据集中存在数据缺失, 所以在非工作日的数据集实验中提取有效样本 120 组, 并对各数据对的顺序进行随机排列, 以前 100 组作为训练样本, 后 20 组作为测试样本. 预测结果与原始数据对比如图 4 所示.

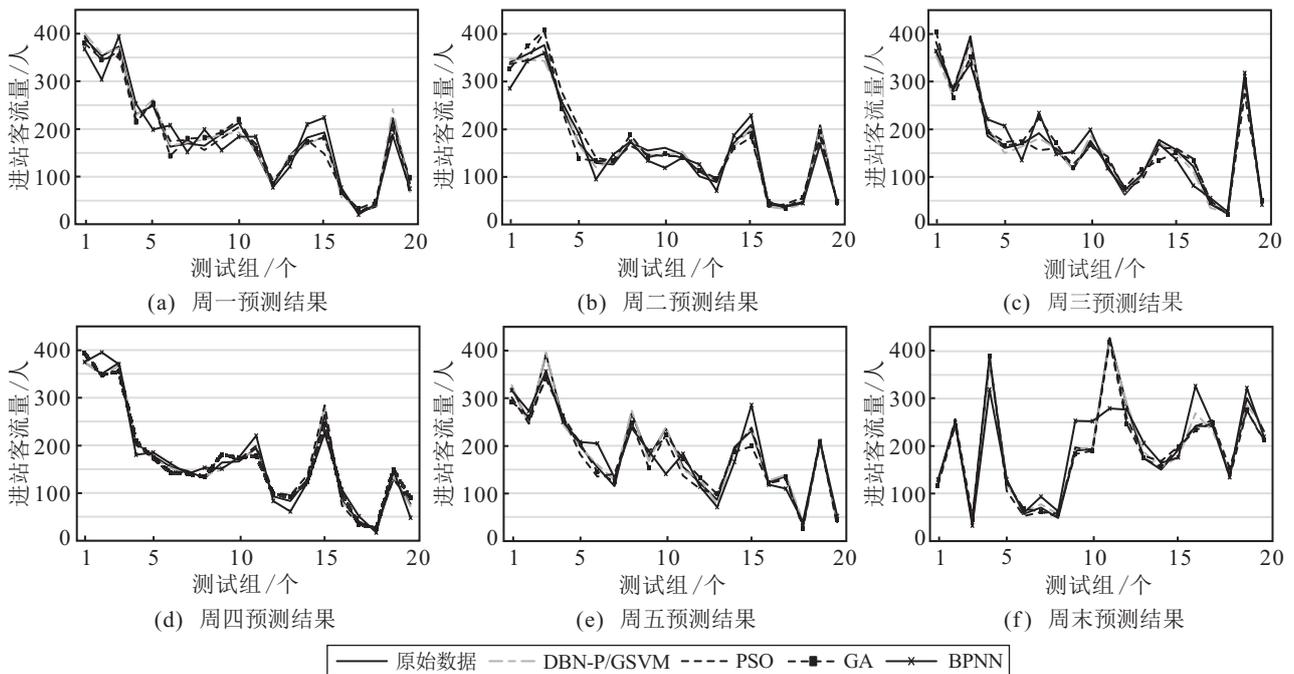


图 4 预测结果与原始数据对比

此外, 本文采用以下 4 种指标更为精确地评价不同算法的预测性能: 1) 均方误差 (MSE); 2) 均方根误差 (RMSE); 3) 绝对误差均值 (MAE); 4) 绝对百分比误差均值 (MAPE). 评价结果如表 8~表 11 所示.

由表 8~表 11 可知, 在城轨交通短时客流预测中, DBN-P/GSVM 模型的 4 个评价指标——MSE、RMSE、MAE 和 MAPE 均优于其他 3 种算法. 与 DBN-P/GSVM 模型相比, PSO-SVM、GA-SVM 和 BPNN 模型都属于浅层机器学习模型. 这类模型由于数据特征提取不如深层模型充分, 数据中的噪声通常会掩

表 8 DBN-P/GSVM 的预测误差

时间	DBN-P/GSVM			
	MSE	RMSE	MAE	MAPE / %
周一	102.90	10.14	7.90	6.70
周二	125.75	11.21	9.15	7.05
周三	123.05	11.09	9.05	6.51
周四	67.55	8.22	6.85	5.75
周五	47.45	6.89	6.25	5.13
周末	96.40	9.82	7.80	5.14
平均	93.85	9.56	7.83	6.05

表9 PSO-SVM的预测误差

时间	PSO-SVM			
	MSE	RMSE	MAE	MAPE/%
周一	205.85	14.35	11.15	8.44
周二	242.75	15.58	12.65	9.20
周三	241.65	15.55	12.65	8.83
周四	104.25	10.21	8.75	7.54
周五	287.85	16.97	13.65	8.56
周末	176.40	13.28	10.70	7.53
平均	209.79	14.32	11.59	8.35

表10 GA-SVM的预测误差

时间	GA-SVM			
	MSE	RMSE	MAE	MAPE/%
周一	137.60	11.73	10.30	8.04
周二	228.60	15.12	12.70	8.89
周三	320.35	17.90	13.05	8.28
周四	135.20	11.63	10.00	7.56
周五	341.55	18.48	14.05	8.19
周末	136.40	11.68	10.55	7.33
平均	216.62	14.42	11.78	8.05

表11 BPNN的预测误差

时间	BPNN			
	MSE	RMSE	MAE	MAPE/%
周一	839.30	28.97	25.50	14.87
周二	559.20	23.65	18.90	12.37
周三	717.45	26.79	21.15	14.60
周四	405.10	20.13	16.30	13.47
周五	1 032.00	32.12	23.40	14.50
周末	2 152.45	46.39	28.85	14.57
平均	950.92	29.68	22.35	14.06

盖数据的内在规律,尤其在处理大量数据时更为明显,从而影响算法预测性能. GA-SVM模型与PSO-SVM预测原理相同,只是选择的优化参数不同,所以在预测精度和算法性能上差别较小. 4种模型中, BPNN模型的算法性能最差. 原因在于BPNN模型在处理大量数据时,数据维度越高,算法越难以收敛. 此外, BPNN模型在使用BP算法时,可能出现梯度减小甚至消失,从而导致远离输出层的隐层无法调整权值和偏置量,最终难以获取最优的模型参数组合. 而本文构建的深度学习模型DBN-P/GSVM能够在无监督式学习模式下不断重构输入数据,充分提取数据特征,展示数据的内在规律,获得更为有效的数据表达. 同时通过有监督式学习进行微调,从整体上使模型更加契合数据变化规律. 两种学习方式的配合使得模型在处理城轨交通短时客流数据时其鲁棒性大大提升,降低了模型出现过学习或欠学习的可能,提升了预测精度.

此外,由于在深度网络中,循环神经网络LSTM常被用于时间序列信号分析,而轨道交通短时客流量本质上为时间序列数据,因此,本文基于python语言构建LSTM模型,对轨道交通短时客流量进行预测,并与所构建的DBN-P/GSVM模型的预测效果进行对比. 同时考虑到Softmax在深度网络中应用广泛,并尝试在LSTM模型中间层使用Softmax函数作为激活函数,设计LSTM-Softmax模型,用于预测轨道交通短时客流量,并与DBN-P/GSVM模型的预测效果进行对比. 对比结果如图5所示.

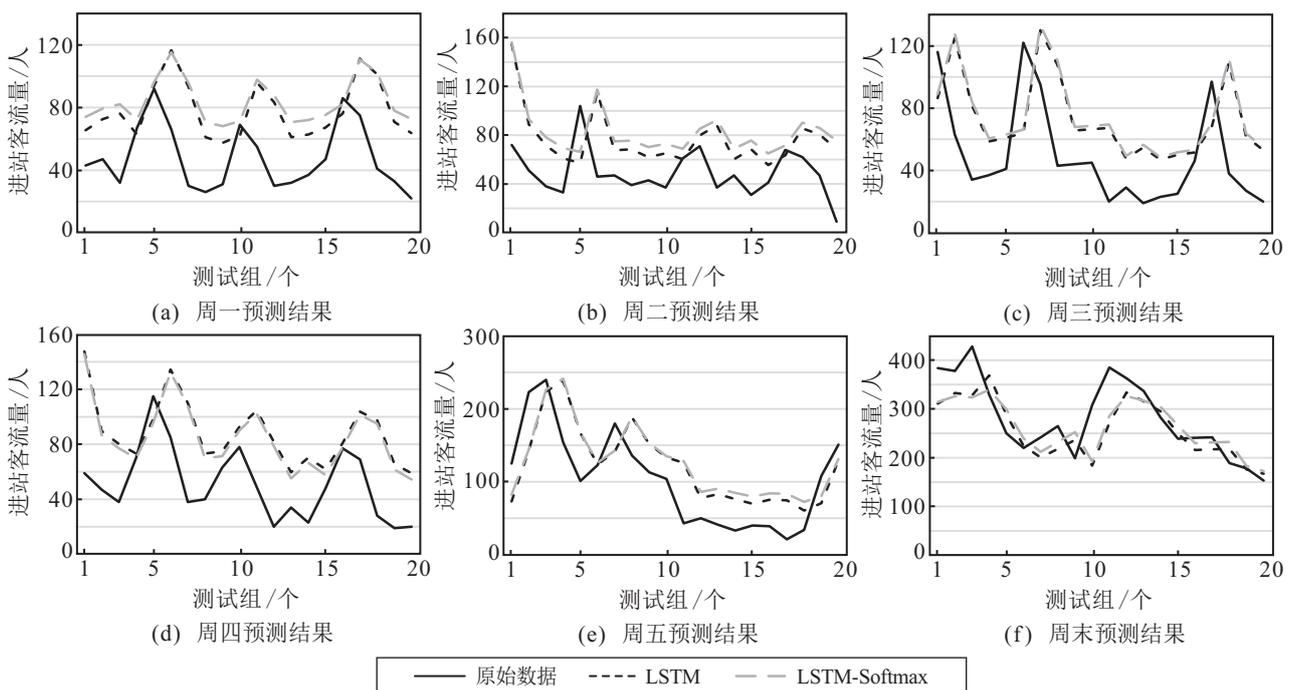


图5 LSTM与LSTM-Softmax的预测结果

LSTM与LSTM-Softmax两个模型的预测误差如表12和表13所示。

表12 LSTM的预测误差

时间	LSTM			
	MSE	RMSE	MAE	MAPE/%
周一	1 315.81	36.27	31.81	84.11
周二	1 470.17	38.34	32.16	90.33
周三	1 541.31	39.26	35.27	85.62
周四	3 027.12	45.02	38.63	109.31
周五	2 316.78	48.13	42.96	52.50
周末	2 899.88	53.85	41.71	5.64
平均	2 095.178	43.478 33	37.09	71.25

表13 LSTM-Softmax的预测误差

时间	LSTM-Softmax			
	MSE	RMSE	MAE	MAPE/%
周一	1 650.13	40.62	36.05	100.17
周二	1 787.00	42.27	37.32	107.85
周三	1 658.22	40.72	36.76	91.39
周四	1 823.84	42.71	35.81	100.87
周五	2 547.03	50.47	45.47	63.47
周末	2 744.06	52.38	41.57	2.64
平均	2 035.047	44.861 67	38.83	77.73

由表12和表13可知,本文构建的DBN-P/GSVM模型的预测效果优于LSTM和LSTM-Softmax模型。

3 结论

本文针对城轨交通短时客流量设计了DBN-P/GSVM深层预测模型。该模型由底层的DBN和顶层的SVM构成。考虑到城轨交通短时客流量的特点,底层DBN由多个Gaussian-Bernoulli RBM堆叠而成。本文同时设计了GA和PSO两种算法实现对顶层SVM的参数优化。研究表明:对于工作日城轨交通短时客流数据,由底层为4层10节点的DBN及顶层为RBF核函数的PSO-SVM组成的深度模型预测效果最佳;对于非工作日城轨交通短时客流数据,由底层为5层10节点的DBN及顶层为多项式核函数GA-SVM组成的深度模型预测效果最佳。

为了验证所建DBN-P/GSVM模型用于城轨交通短时客流量的预测效果,本文同时设计了PSO-SVM、GA-SVM和BPNN三种浅层模型以及LSTM和LSTM-Softmax两个深层模型。研究结果表明,所设计的DBN-P/GSVM模型在MSE、RMSE、MAE和MAPE四个方面均优于其他3种浅层模型和两个深层模型,同时验证了DBN-P/GSVM深度预测模型在处理城市轨道交通短时客流时的有效性。

对于未来研究,可进一步扩大样本容量,对实验中数据集的划分及模型的过学问题进行深入研究。

参考文献(References)

- [1] 中国城市轨道交通协会.城市轨道交通2017年度统计和分析报告[N].中国城市轨道交通协会信息,2018-04-08.
(China Association of Metros. 2017 Annual statistics and analysis report of metros[N]. China Association of Metros Information, 2018-04-08.)
- [2] 韩超,宋苏,王成红.基于ARIMA模型的短时交通流实时自适应预测[J].系统仿真学报,2004,16(7):1530-1532.
(Han C, Song S, Wang C H. A real-time short-term traffic flow adaptive forecasting method based on ARIMA model[J]. J of System Simulation, 2004, 16(7): 1530-1532.)
- [3] Guo J, Huang W, Williams B M. Adaptive Kalman filter approach for atochastic short-term traffic flow rate prediction and uncertainty quantification[J]. Transportation Research, Part C: Emerging Technologies, 2014, 43(1): 50-64.
- [4] 熊杰,关伟,孙宇星.基于Kalman滤波的地铁换乘客流预测[J].北京交通大学学报,2013,37(3):112-116.
(Xiong J, Guan W, Sun Y X. Metro transfer passenger forecasting based on Kalman filter[J]. J of Beijing Jiaotong University, 2013, 37(3): 112-116.)
- [5] Zhao S Z, Ni T H, Wang Y, et al. A new approach to the prediction of passenger flow in a transit system[J]. Computers & Mathematics with Applications, 2011, 61(8): 1968-1974.
- [6] 郭敏,蓝金辉,肖翔,等.基于混沌理论对北京二环路进行短时交通流量预测的研究[J].交通运输系统工程与信息,2010,10(2):106-111.
(Guo M, Lan J H, Xiao X, et al. Forecasting short-time traffic flow for Beijing 2nd ring road using chaos theory[J]. J of Transportation Systems Engineering and Information Technology, 2010, 10(2): 106-111.)
- [7] 张玉梅,曲仕茹,温凯歌.基于混沌和RBF神经网络的短时交通流量预测[J].系统工程,2007,25(11):26-30.
(Zhang Y M, Qu S R, Wen K G. A short-term traffic flow forecasting method based on chaos and RBF neural network[J]. Systems Engineering, 2007, 25(11): 26-30.)
- [8] Clark S. Traffic prediction using multivariate nonparametric regression[J]. J of Transportation Engineering, 2003, 129(2): 161-168.
- [9] Davis G A, Nihan N L. Nonparametric regression and short-term freeway traffic forecasting[J]. J of Transportation Engineering, 1991, 117(2): 178-188.
- [10] Smith B L, Williams B M, Oswald R K. Comparison of parametric and nonparametric models for traffic

- flow forecasting[J]. *Transportation Research, Part C: Emerging Technologies*, 2002, 10(4): 303-321.
- [11] Chrobok R, Wahle J, Schreckenber M. Traffic forecast using simulations of large scale networks[C]. *Proc of the 2001 IEEE Intelligent Transportation Systems*. Oakland: IEEE, 2001: 434-439.
- [12] 马云龙, 王坚, 任子晖. 基于微观仿真的快速路短时交通流预测研究[J]. *系统仿真学报*, 2009, 21(14): 4501-4503.
(Ma Y L, Wang J, Ren Z H. Research on short-term traffic flow prediction based on microcosmic simulation [J]. *J of System Simulation*, 2009, 21(14): 4501-4503.)
- [13] Castro-Neto M, Jeong Y S, Jeong M K, et al. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions[J]. *Int J of Expert Systems with Applications*, 2009, 36(3): 6164-6173.
- [14] Sun Y, Leng B, Guan W. A novel wavelet-SVM short-time passenger flow prediction in beijing subway system[J]. *Neurocomputing*, 2015, 166(C): 109-121.
- [15] Castro-Neto M, Jeong Y, Jeong M K, et al. AADT prediction using support vector regression with data-dependent parameters[J]. *Expert Systems with Applications*, 2009, 36(2): 2979-2986.
- [16] Jiang X, Adeli H. Dynamic wavelet neural network Model for traffic flow forecasting[J]. *J of Transportation Engineering*, 2005, 131(10): 771-779.
- [17] Wei Y, Chen M C. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks[J]. *Transportation Research, Part C: Emerging Technologies*, 2012, 21(1): 148-162.
- [18] Dia H. An object-oriented neural network approach to short-term traffic forecasting[J]. *European J of Operational Research*, 2001, 131(2): 253-261.
- [19] Tsai T H, Lee C K, Wei C H. Neural network based temporal feature models for short-term railway passenger demand forecasting[J]. *Expert Systems with Applications*, 2009, 36(2): 3728-3736.
- [20] 李松, 刘力军, 翟曼. 改进粒子群算法优化BP神经网络的短时交通流预测[J]. *系统工程理论与实践*, 2012, 32(9): 2045-2049.
(Li S, Liu L J, Zhai M. Prediction for short-term traffic flow based on modified PSO optimized BP neural network[J]. *Systems Engineering—Theory & Practice*, 2012, 32(9): 2045-2049.)
- [21] Zhang Y, Zhang Y, Haghani A. A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model[J]. *Transportation Research, Part C: Emerging Technologies*, 2014, 43(1): 65-78.
- [22] Wang J, Deng W, Guo Y. New Bayesian combination method for short-term traffic flow forecasting[J]. *Transportation Research, Part C: Emerging Technologies*, 2014, 43(1): 79-94.
- [23] Huang W, Song G, Hong H, et al. Deep architecture for traffic flow prediction: Deep belief networks with multitask learning[J]. *IEEE Trans on Intelligent Transportation Systems*, 2014, 15(5): 2191-2201.
- [24] 李捷. 基于深度学习的地铁短时客流量预测研究[D]. 成都: 西南交通大学经济管理学院, 2018.
(Li J. Study on metro short-term passenger flow forecasting based on deep learning[D]. Chengdu: School of Economics and Management, Southwest Jiaotong University, 2018.)
- [25] Nicholas G Polson, Vadim O Sokolov. Deep learning for short-term traffic flow prediction[J]. *Transportation Research, Part C: Emerging Technologies*, 2017, 79: 1-17.
- [26] Lv Y, Duan Y, Kang W, et al. Traffic flow prediction with big data: A deep learning approach[J]. *IEEE Trans on Intelligent Transportation Systems*, 2015, 16(2): 865-873.
- [27] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief Nets[J]. *Neural Computation*, 2006, 18(7): 1527-1554.
- [28] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786): 504-507.

作者简介

龙小强(1976—), 男, 高级工程师, 博士生, 从事交通规划与管理的研究, E-mail: 171450427@qq.com;

李捷(1991—), 男, 工程师, 硕士生, 从事机器学习的研究, E-mail: 373039420@qq.com;

陈彦如(1974—), 女, 副教授, 博士生导师, 从事物流与交通规划管理等研究, E-mail: chenyanru@swjtu.cn.

(责任编辑: 李君玲)