

基于动态稀疏注意力的地铁客流预测模型^{*}马茜¹ 梁奕^{2,3} 段毅^{2,3} 曾尚琦^{2,3}

(1. 西咸新区轨道交通发展有限公司, 710086, 西安; 2. 南瑞集团(国网电力科学研究院)有限公司, 210003, 南京; 3. 国电南瑞科技股份有限公司, 211106, 南京//第一作者, 高级工程师)

摘要 地铁客流预测是随时间演变的多维时间序列数据, 不同序列之间存在复杂的动态相互依赖关系。为挖掘多种监测指标之间存在的内在复杂关系, 提出动态稀疏注意力模型: 利用全局变量注意力自动选择相关驱动序列, 增强模型预测的判别性; 根据局部紧密相关和全局稀疏相关的先验知识, 对历史时间步和相关变量分别卷积和稀疏卷积, 提取局部时间和局部变量特征; 设计了稀疏注意力对相关时间步加权和变量加权, 提高预测表现。结果表明, 与其他常用客流预测模型相比, 动态稀疏注意力模型能高度准确地预测客流。

关键词 地铁; 客流预测; 动态稀疏注意力模型

中图分类号 U293.13

DOI:10.16037/j.1007-869x.2022.04.005

Metro Passenger Flow Prediction Model Based on Dynamic Sparse Attention

MA Qian, LIANG Yi, DUAN Yi, ZENG Shangqi

Abstract Metro passenger flow prediction is multi-dimensional time series data that changes as time passes, and complex dynamic interdependent relationship exists among various series. To explore the internal complex relationship among various monitoring indicators, dynamic sparse attention model (DSANNs) is proposed, which uses global variable attention to automatically select relevant driver sequence to enhance the discriminability of model prediction. According to the prior knowledge of local close correlation and global sparse correlation, the convolution and sparse convolution of historical time steps and related variables are conducted to extract the characteristics of local time and local variables, and the weight of related time steps and variables by sparse attention is designed to improve prediction performance. The experiment results show that, compared with other commonly used passenger flow prediction model, DSANNs can carry out the prediction with high level of accuracy.

Key words metro; passenger flow prediction; DSANNs

First-author's address Xixian New District Rail Transit Development Co., Ltd., 710086, Xi'an, China

地铁客流的精准预测可帮助优化行车间隔和制定合理的运行方案。然而, 影响客流的变量在时间和空间上存在不均衡性, 表现为: 时间序列的高峰平峰, 日、周、月客流以及空间存在位置性和方向性, 包括车站、线路、线网瞬时和历史数据(部分受气候条件的影响)。其动态复杂度显著提升。

目前针对客流的时间序列预测, 主要基于历史客流进行预测建模。随着深度学习表征能力的增强, 深度学习技术在时空数据预测领域得到广泛应用。RNN(递归神经网络)可捕获客流的时序依赖关系^[1]。CNN(卷积神经网络)结构侧重于空间特征, 却丢失了时间维度信息^[2]。ConvLSTM(卷积长记忆网络)将RNN和CNN的架构结合起来, 将全连接层替换为卷积层^[3]来捕获时空相关性; 然而, ConvLSTM在状态转换中使用卷积, 相当于把循环连接结构模型变为时空恒定结构模型。总之, 客流预测仍存在复杂非线性建模、动态关联及计算复杂度等亟待解决的问题。

在周一早高峰时段, 进站客流特征大致相同, 体现了短期模式和局部数据紧密相关性; 在其余工作日的早高峰及晚高峰时段, 进站客流特征大致相同, 体现了长期模式和全局数据稀疏相关性。对此本文提出DSANNs(动态稀疏注意力)模型, 利用全局变量注意力自动选择相关驱动序列, 增强模型预测的判别性; 根据局部数据紧密相关性和全局数据稀疏相关性的先验知识, 对历史时间步和相关变量分别卷积和稀疏卷积, 提取局部时间和变量特征, 并设计稀疏注意力对相关时间步加权和变量加权, 进而提高客流预测的准确性。

* 国家重点研发计划“云计算和大数据”重点专项(2017YFB1001800); 江苏省工业和信息化厅重点质量攻关项目(2019-305)

1 客流预测的影响因素

现有的地铁客流预测研究通常基于历史数据进行,其预测结果完全依赖于历史数据的准确性,并未挖掘多种监测指标之间存在的内在复杂关系。客流预测主要受以下因素影响:

1) 长短周期时间依赖的关联。进出站客流的多维时间序列蕴含一种局部紧密和全局稀疏相关的内在关联,混合着长短周期时间依赖关系。变量的变动可分为趋势性、周期性、随机性及循环性变动。客流预测需过滤随机性变动,仅反映时间序列中的趋势性变动和周期性变动,综合捕获时间序列数据中长短周期时间依赖关联。

2) 变量的动态相关性。各维度变量之间的相关性是动态的,即动态相关性。车站的进出站客流、线网拓扑、天气状况和空气质量等变量间均存在动态相关性,而且进、出站客流间的相关性强,客流与温度的相关性弱。可见,测点各维度变量的时间序列数据对于客流预测的影响并不一致。测点监测每项指标的时间相关性也是动态的:非高峰时段的客流数据差异不大,相关性较强;在高峰时段,测点的监测数据可能存在较大差异,相关性较弱。此外,测点监测数据也可能存在周期性变化规律,因此需要选取合适的数据统计时间间隔。

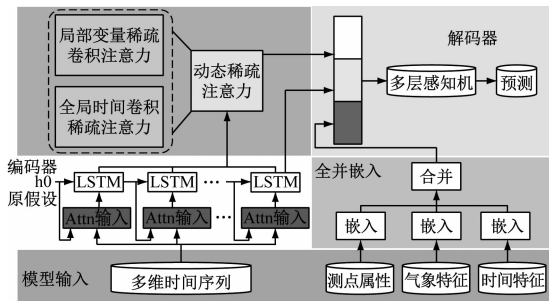
3) 数据受到外部因素的影响。各变量的时间序列数据对目标预测序列数据的驱动作用不仅随时间的变化而变化,还会受外部因素的影响。恶劣环境因素及特殊活动等事件因素,都会导致时间序列数据出现突变、陡降甚至是拐点。

2 DSANNs 模型

为了捕获变量多维时间序列中长短周期时间依赖关系和动态性,针对客流预测的影响因素,本文提出 DSANNs 模型。如图 1 所示,DSANNs 模型包括全局变量注意力机制的编码器、局部变量稀疏卷积注意力机制和全局时间卷积稀疏注意力的解码器。

1) 全局变量注意力机制自适应地选择相关驱动序列,增强预测的判别性,捕获动态全局变量相关性。

2) 局部紧密、全局稀疏的全局时间卷积稀疏注意力机制对变量多维时间序列中相关时间步蕴含的信息加权,提取各时间步下的内在局部特征。稀



注: LSTM 为长短时记忆网络;Attn 输入指注意力机制输入。

图1 动态稀疏注意力模型框架

Fig. 1 Dynamic sparse attention model framework

疏注意力可减少关联,降低模型复杂度,防止过拟合。

3) 局部变量稀疏卷积注意力会自动选择相关变量进行加权。经过稀疏卷积核提取的局部特征可以理解为每个变量在固定的时间窗口形成的局部内在关系。通过局部变量注意力,将所有横跨固定窗口的时间序列数据信息整合。

4) 将相关的外部时间特征、气象特征、目标测点的相关属性等特征因素合并为 1 个低维向量嵌入到模型中。

2.1 全局变量注意力

在预测过程中,时间序列同输入序列数据和目标序列数据的相关程度并非一致。基于每个测点输入序列数据得到的对一时刻预测数据是随时间动态变化的。全局变量注意力机制设计如图 2 所示。

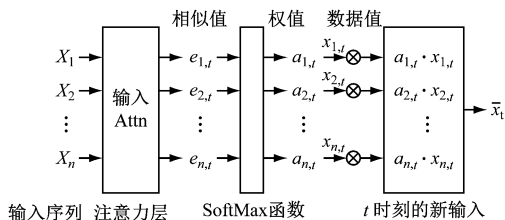


图2 全局变量注意力机制

Fig. 2 Mechanism of global variable input attention

全局变量注意力机制能自动判定输入序列每个时间步下各序列对目标序列预测的重要程度,突出与当前预测任务最相关的输入变量,弱化与任务预测无关的输入变量。

编码器将输入序列编码成机器翻译中的特征表示^[4]。第 k 输入序列 $X_k = (x_{k,1}, x_{k,2}, \dots, x_{k,T}) \in \mathbb{R}_T, x_{k,t} \in \mathbb{R}_{n \times T}$ 表示在驱动序列 k 时间戳 t 下的数据值, n 为驱动预测序列的数量, T 为历史时间序列

的长度。该编码器用来学习从 $x_t (x_t \in \mathbb{R}_T)$ 到 \mathbf{h}_t 的 1 个映射函数,可表示为

$$\mathbf{h}_t = f_1(\mathbf{h}_{t-1}, x_t), \mathbf{h}_t \in \mathbb{R}_m \quad (1)$$

式中:

\mathbf{h}_t ——编码器在时间步为 t 的隐藏向量,蕴含 t 时刻前的历史信息, m 为该隐藏向量大小;

f_1 ——非线性激活函数。

DSANNs 使用 LSTM 单元作为非线性激活函数 f_1 去捕获原始时间序列中的长期依赖关系。将其与历史时间步隐藏层状态 \mathbf{h}_{t-1} 和 LSTM 中的细胞状态 S_{t-1} 进行匹配,得到该时间步下驱动预测时间序列的各变量重要性,故有:

$$e_{k,t} = v_e \tanh(W_e [\mathbf{h}_{t-1}, S_{t-1}] + U_e x_k)$$

$$a_{k,t} = \frac{\exp(e_{k,t})}{\sum_{j=1}^n \exp(e_{j,t})} \quad (2)$$

式中:

v_e, W_e, U_e ——均为调优的超参数,且 $v_e \in \mathbb{R}, W_e \in \mathbb{R}_{T \times 2m}, U_e \in \mathbb{R}_{T \times T}$;

$e_{k,t}$ ——在时间步 t 下衡量第 k 输入序列的相似性值;

$a_{k,t}$ ——通过 SoftMax 函数得到的注意力权值,且所有 $a_{k,t}$ 归一化和为 1。

输入注意力机制通过前馈神经网络和递归循环神经网络部分一同训练,得到驱动预测序列随着时间演变的权重,并在每一个时间步自动选取最重要的特征进行预测,同时调整时间步 t 下隐藏层状态向量。具体为:

$$\bar{\mathbf{x}}_t = (a_{1,t}x_{1,t}, a_{2,t}x_{2,t}, \dots, a_{n,t}x_{n,t}) \quad (3)$$

2.2 动态稀疏注意力

每个时间序列在固定窗口大小为 ω 的数据对目标预测序列预测的重要性不同,往往蕴含着一种局部紧密相关和全局稀疏相关的内在关系,而密集全局关联任务是很少的。因此,基于局部紧密相关和全局稀疏相关的内在关系,DSANNs 中设计了 2 种注意力机制(如图 3 所示),分别对变量和时间进行卷积操作和注意力权值分配。

2.2.1 局部变量稀疏卷积注意力

DSANNs 模型利用卷积核卷积编码后的时间序列,增强学习到时间序列的局部变量内在关系。设计卷积核 $C_j \in \mathbb{R}_{T_0} (C_{j,T-ns}, \dots, C_{j,T-2s}, C_{j,T-s}, \dots, C_{j,T-1}, C_{j,T}) \neq 0, T-ns \geq 1, s$ 为局部紧密全局稀疏程度。

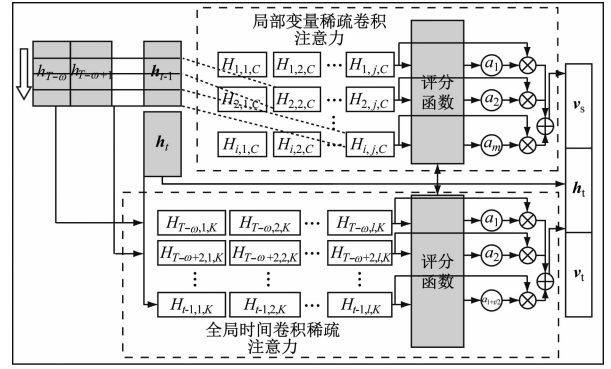


图 3 DSANNs 模型的注意力机制

Fig. 3 Mechanism of DSANNs attention

卷积核卷积每个变量,捕获其局部紧密相关和全局稀疏相关的内在特征。全局变量注意力机制的隐藏层向量表示 $H \in \mathbb{R}_{m \times T}$ 经过卷积操作,最终的变量维度为 $H_c \in \mathbb{R}_{m \times k}$ 。其中, k 代表 k 个具有不同内在关联关系的类似卷积核, m 代表每个时间步编码后变量的维度。RELU 是激活函数 $\text{RELU}(x) = \max(0, x)$, 故有:

$$H_{i,j,C} = \text{RELU}\left(\sum_{p=1}^{\omega} H_{i,(T-\omega+p)} * C_{j,T-\omega+p} + b_{jC}\right) \quad (4)$$

式中:

$H_{i,j,C}$ ——卷积核对隐藏层向量提取结果;

$C_{j,T-\omega+p}$ ——第 j 卷积核;

$H_{i,(T-\omega+p)}$ ——隐藏层第 i 行向量;

b_{jC} ——偏差的参数。

局部变量稀疏卷积注意力选择相关变量进行加权,将时间序列数据整合成 1 个固定尺度的向量 \mathbf{v}_s :

$$a_i = \text{sigmoid}(H_{i,C} \mathbf{h}_t)$$

$$\mathbf{v}_s = \sum_{i=1}^m a_i H_{i,C} (\mathbf{v}_s \in \mathbb{R}_k) \quad (5)$$

式中:

$H_{i,C}$ ——卷积后第 i 行隐藏层行向量表示;

a_i ——sigmoid 函数计算注意力权值。

由式(5),通过权重对 H_c 重新分配得到隐藏层向量 \mathbf{v}_s 。

2.2.2 全局时间卷积稀疏注意力

时间序列数据通常混合长期和短期模式,并且时间序列中历史时间步蕴含的信息对目标预测的重要性不同,往往蕴含着一种局部紧密相关和全局稀疏相关的内在关系。DSANNs 根据此先验知识,利用卷积神经网络提取各变量序列中的短期局部

变量,构建稀疏注意力捕获时间序列中的局部紧密相关和全局稀疏相关的内在关系。

全局时间卷积稀疏注意力自动选择相关时间步加权。卷积核提取所有变量在每一个时间步的内在局部特征,通过全局时间稀疏注意力,将所有横跨固定窗口的时间序列数据整合成一个固定尺度的低纬度向量 \boldsymbol{v}_t 。 \boldsymbol{v}_t 为固定尺度的上下文向量,具备时空特征。则有:

$$a_j = \frac{\exp(f(\boldsymbol{H}_{j,K}, \boldsymbol{h}_t))}{\sum_{b=t-3}^{t-1} \exp(f(\boldsymbol{H}_{b,K}, \boldsymbol{h}_t)) + \sum_{i=1}^{t-5} \exp(f(\boldsymbol{H}_{i,K}, \boldsymbol{h}_t))}$$

$$\boldsymbol{v}_t = \sum_{j=t-3}^{t-1} a_j \boldsymbol{H}_{j,K} + \sum_{i=1}^{t-5} a_i \boldsymbol{H}_{i,K}$$

$$(i = 1, 3, \cdots, t - 5, \quad \boldsymbol{v}_t \in \mathbb{R}_l) \quad (6)$$

式中:

$\boldsymbol{H}_{j,K}$ ——卷积后形成的第 j 时间步隐藏层列向量表示;

a_j ——自动选取时间步计算注意力权值。

提取各时间步下的内在局部变量特征后,通过局部紧密相关和全局稀疏相关的注意力机制,对单测点多维时间序列中相关时间步蕴含的信息加权,捕获时间序列短期和长期依赖关系。

2.3 模型预测

DSANNs 模型包括:局部变量稀疏卷积注意力提取的、相关变量加权的、隐藏层向量表示 \boldsymbol{v}_s ,全局时间卷积稀疏注意力提取的、相关时间步加权的、隐藏层向量表示 \boldsymbol{v}_t ,全局变量注意机制的隐藏层向量 \boldsymbol{h}_t 。通过多层感知机,得到最终预测结果:

$$\gamma_t = W_h \boldsymbol{h}_t + W_s \boldsymbol{v}_s + W_t \boldsymbol{v}_t$$

$$\widehat{y_{t-1+\Delta}} = W_\gamma \gamma_t \quad (7)$$

式中:

W_h, W_s, W_t, W_γ ——分别为各隐藏层的权重参数, $W_h \in \mathbb{R}_{m \times m}$, $W_s \in \mathbb{R}_{m \times k}$, $W_t \in \mathbb{R}_{m \times l}$, $W_\gamma \in \mathbb{R}_{(n \times \Delta) \times m}$

$\widehat{y_{t-1+\Delta}}$ ——预测数据表示, $\widehat{y_{t-1+\Delta}} \in \mathbb{R}_{n \times \Delta}$

3 试验与分析

为了测试不同变量多维时间序列预测方法的性能,采用基于 Python 语言和 Keras 深度学习框架开发的 DSANNs,以某城地铁站点的月度客流量为例,进行试验。按照 10 min 聚合 1 次,则聚合数据总计 4 464 条。

3.1 试验配置

1) 参数设置。通过窗口大小进行网格搜索

“ $\omega = \{3, 5, 10, 15, 25, 50, 72, 100\}$ ”,确定时间窗口 ω ;对隐藏层进行网格搜索“ $m = \{16, 32, 64, 128, 256\}$ ”,确定隐藏层 m ;为了确定不同注意力卷积核个数带来预测表现的变化,设置 $k = l$,对卷积核个数进行网格搜索“ $k = l = \{16, 32, 64, 128\}$ ”,最终参数选择验证集的最佳性能参数。

2) 评价指标。为了衡量各种时间序列预测方法的有效性,考虑了均方误差 E_{MS} 、平均绝对误差 E_{MA} 及平均绝对百分误差 E_{MAP} 等 3 种不同的评价指标。

$$E_{MAP} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_{t,i} - \hat{y}_{t,i}}{y_{t,i}} \right| \times 100\% \quad (8)$$

其中, y_t 为 t 时间步的目标, \hat{y}_t 为 t 时刻的预测值。

3.2 结果分析

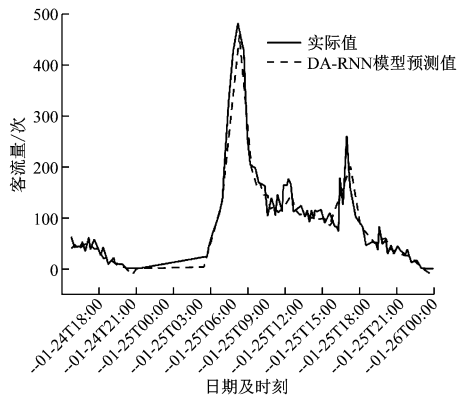
在客流多维时间序列数据集上将 DSANNs 模型同 ARIMA(差分整合移动平均自回归)模型^[6]、LSTM 模型^[7]、ConvLSTM 模型^[3]、TCN(时间卷积网络)模型^[8]、DA-RNN(注意力机制-递归神经网络)模型^[9]进行比较。各模型的时间序列预测效果如表 1 所示。

表 1 各模型对单测点多维时间序列的预测结果评价指标
Tab.1 Evaluation indicators of prediction result of single-point multi-dimensional time series in each model

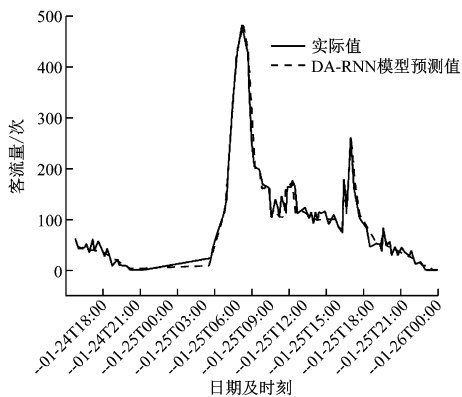
模型	E_{MA}	$E_{MAP}/\%$	E_{MS}
ARIMA	0.126 5	18.433 2	0.072 3
LSTM	0.027 9	7.952 3	0.001 6
ConvLSTM	0.025 3	3.364 5	0.001 5
TCN	0.029 1	7.425 2	0.002 1
DA-RNN	0.022 6	2.156 8	0.001 2
DSANNs	0.019 2	1.814 5	0.000 8

从表 1 可以看出:基于实测数据,采用 DSANNs 模型得到的 E_{MA} 、 E_{MAP} 和 E_{MS} 最佳,与采用 DA-RNN 模型的结果相比分别降低了 21.07%、18.47%、29%;与 LSTM 模型相比,ARIMA 模型的预测精度较差,这是因为 ARIMA 模型只考虑目标序列,忽略了其他相关维度的驱动序列;ConvLSTM 模型、TCN 模型结合了时间和变量的相关模型,预测表现要优于单纯捕获时间依赖的 LSTM 模型;DA-RNN 比采用卷积和长短记忆网络的 ConvLSTM 模型、TCN 模型的性能更好,因为它关注了时间和变量的动态性。综合来看,DSANNs 模型表现最优。这主要因为 DSANNs 模型通过局部紧密全局稀疏的时间注意机制以及相关变量的注意力机制,捕获了时间动态性和变量的动态内在关系,提高了预测准确性。图 4 展示了对比基线最佳的 DA-RNN 模型和

DSANNs 模型在地铁客流量数据集的预测结果。



a) DA-RNN 的地铁客流预测及对比



b) DSANNs 模型的地铁客流预测及对比

图4 DSANNs 模型与 DA-RNN 模型的地铁客流量预测值同实际值对比

Fig. 4 Comparison between DSANNs and DA-RNN metro passenger flow prediction value and actual value

与 DA-RNN 模型相比,DSANNs 模型在捕获时间动态性和变量内在关系方面表现更好,预测性能也更佳。经分析,DSANNs 模型通过全局稀疏局部紧密的时间注意力机制的权重分配,减少关联性后,加大了全局稀疏的权重分配,使得模型整体的预测效果良好。

4 结语

客流预测需要预测地铁车站的潜在流量,表现为随时间演变的多维时间序列数据,受到外界复杂因素的影响而发生动态变化。通过分析客流的时空特征,提出 DSANNs 模型。DSANNs 模型自动提取动态时间和变量维度的内在关系。其中,全局变量注意力机制自适应地选择相关驱动序列,增强了模型预测的判别性;稀疏注意力对相关时间步加

权,注意力机制对相关变量进行加权,不仅能够从全局和局部捕获动态内在时间和变量关联,提高模型预测准确性,也能减少计算的复杂性。相比于 ConvLSTM 模型、TCN 模型、DA-RNN 模型等,DSANNs 模型具有明显优势。未来可将时空数据预测模型和时序图进一步结合,提高图神经网络的预测准确性。

参考文献

- [1] 杨斌,钟金英. 卷积神经网络的研究进展综述[J]. 南华大学学报(自然科学版), 2016(3):66.
YANG Bin, ZHONG Jinying Review of convolutional neural network[J]. Journal of University of South China (Science and Technology), 2016(3):66.
- [2] SHI X, CHEN Z, WANG H, et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting [EB/OL]. (2015-09-19) [2021-12-20]. <https://arxiv.org/abs/1506.04214v2>. arXiv:1506.04214[cs.CV].
- [3] QIN Y, SONG D, CHEN H, et al. A dual stage attention based recurrent neural network for time series prediction [C] // IJCAI-2017. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne: IJCAI, 2017:2627.
- [4] LIANG Y, JIANG Z, ZHENG Y. Inferring traffic cascading patterns [C] // SIGSPATIAL'17. Proceedings of the 25 th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. Los Angeles, CA. New York: Association for Computing Machinery, 2017:1-10.
- [5] KINGMA D P, BA J L. Adam: a method for stochastic optimization [C] // 3rd International Conference on Learning Presentations. Conference Track Proceedings. San Diego: ICLR, 2017.
- [6] CHEN S, WANG X X, HARRIS C J. NARX-based nonlinear system identification using orthogonal least squares basis hunting [J]. IEEE Transactions on Control Systems Technology, 2007, 16(1): 78.
- [7] DEY R, SALEM F M. Gate-variants of gated recurrent unit (GRU) neural networks [C] // IEEE 60 th International Midwest Symposium on Circuits and Systems (MWSCAS). Boston, MA. New Jersey: IEEE, 2017:1597-1600.
- [8] KOZMA R, ILIN R. Optimization via efficient learning in CNNs: cognitively-motivated temporal discount functions in SRNNs [C] // 12 th International Workshop on Cellular Nanoscale Networks and Their Applications (CNNA 2010). Berkeley, CA. New Jersey: IEEE, 2010:1-6.
- [9] LAI G, CHANG W C, YANG Y, et al. Modeling long- and short-term temporal patterns with deep neural networks [C] // SIGIR'18. The 41 st International ACM SIGIR Conference on Research & Development in Information Retrieval. Ann Arbor, MI. New York: Association for Computing Machinery, 2018:95-104.

(收稿日期:2021-08-19)