

关系及相互影响,在谱图论中,可以将一个图用其对应的拉普拉斯矩阵来表示.通过分析拉普拉斯矩阵及其特征值就可以得到图结构的性质.图的拉普拉斯矩阵定义为 $L=D-A$,其规范化形式为 $L=I_N - D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \in \mathbb{R}^{N \times N}$,其中 A 为邻接矩阵, I_N 为单位阵,度矩阵 $D \in \mathbb{R}^{N \times N}$ 是节点度数组成的对角矩阵, $D_{ii} = \sum_j A_{ij}$.对拉普拉斯矩阵进行特征值分解 $L=UAU^T$,其中 $A=diag([\lambda_0, \dots, \lambda_{N-1}]) \in \mathbb{R}^{N \times N}$ 是 L 的特征值组成的对角矩阵, U 是傅里叶基^[12].以 t 时刻流量数据为例,图信号为 $x = x_t^f \in \mathbb{R}^N$,对图信号进行傅里叶变换可表示为 $\hat{x} = U^T x$.根据拉普拉斯矩阵的性质可知 U 是正交矩阵,因此得到逆傅里叶变换 $x = U\hat{x}$.图卷积是利用定义在傅里叶域中对角化的线性算子来等价代替经典卷积算子^[13]实现的卷积操作,用卷积核 g_θ 对图 G 进行卷积操作^[9]:

$$g_\theta \times_G x = g_\theta(L)x = g_\theta(UAU^T)x = U g_\theta(A)U^T x.$$

由于对图信号进行卷积操作再做傅里叶变换等于对这些信号进行傅里叶变换后的乘积^[14],上式可以理解对 g_θ 和 x 分别做傅里叶变换到谱域,然后对二者的变换结果进行乘法操作,再做傅里叶逆变换得到卷积操作的结果.

将图变换到谱域实现图上的卷积操作即为图卷积,但当图的规模较大时,直接对拉普拉斯矩阵进行特征值分解代价昂贵,因此,本文采用切比雪夫多项式近似展开求解:

$$g_\theta \times_G x = g_\theta(L)x \approx \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L})x.$$

$\theta_k \in \mathbb{R}^K$ 为切比雪夫多项式系数, $\tilde{L} = \frac{2}{\lambda_{\max}} L - I_N$, λ_{\max} 表示拉普拉斯矩阵的最大特征值.切比雪夫多项式的递归定义: $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$,其中, $T_0(x) = 1, T_1(x) = x$.用切比雪夫多项式近似展开求解,相当于对图中的每个节点采用卷积核提取了以该节点为中心的周围 $0 \sim K-1$ 阶邻居的信息^[9].图卷积模块使用线性修正单元(ReLU)为激活函数,即 $ReLU(g_\theta \times_G x)$.

以 $K=2$ 为例,对每个节点提取其 $0 \sim 1$ 阶邻居信息,对拉普拉斯矩阵特征值进行缩放使 $\lambda_{\max}=2$,上述图卷积操作表示为

$$g_\theta \times_G x \approx \theta_0 T_0(\tilde{L})x + \theta_1 T_1(\tilde{L})x = \theta_0 x + \theta_1 (L - I_N)x = \theta_0 x + \theta_1 (-D^{-\frac{1}{2}}AD^{-\frac{1}{2}})x.$$

令 $\theta_0 = -\theta_1 = \theta \in \mathbb{R}$ 以减少参数,则 $g_\theta \times_G x \approx (I_N + D^{-\frac{1}{2}}AD^{-\frac{1}{2}})x\theta$,图上所有节点共享卷积核权重 θ .同时,为避免数值不稳定、梯度爆炸或梯度消失^[15],我们令 $\tilde{A} = A + I_N, \tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$,得到: $g_\theta \times_G x = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}x\theta$.拓展到多维数据,第 r 层卷积的输入数据是 $\mathcal{X}_h^{(r-1)} = (\mathbf{X}^{1,(r-1)}, \mathbf{X}^{2,(r-1)}, \dots, \mathbf{X}^{C_{r-1},(r-1)}) \in \mathbb{R}^{C_{r-1} \times N \times T_{r-1}}$,其中: $r \in \{1, \dots, l\}$ (l 是时空卷积层数); C_{r-1} 是第 r 层网络的输入数据的通道数,当 $r=1$ 时, $C_0 = F; T_{r-1}$ 是输入数据时间维长度,当 $r=1$ 时,近期组件 $T_0 = T_h$ (日周期组件 $T_0 = T_d$ 、周周期组件 $T_0 = T_w$).对 $\mathcal{X}_h^{(r-1)}$ 进行图卷积操作,表示为 $g_\theta \times_G \mathcal{X}_h^{(r-1)} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}\mathcal{X}_h^{(r-1)}\Theta$,其中, $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_{C_r}) \in \mathbb{R}^{1 \times C_{r-1} \times C_r}$ 是卷积核参数.

以 $C_{r-1}=1$ 为例,图 6 展示了图卷积过程中对空间节点的 $0 \sim 1$ 阶邻居信息进行提取的过程.图 6(a)为空间路网结构,图 6(b)为根据图 6(a)计算得到邻接矩阵 A 和度矩阵 $D, \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ 及输入数据 $\mathbf{X}^{1,(r-1)}$ 的矩阵表示如图 6(c)所示,由此计算出用卷积核 g_θ 对数据 $\mathbf{X}^{1,(r-1)}$ 进行图卷积操作的结果 $g_\theta \times_G \mathbf{X}^{1,(r-1)}$ 如图 6(d)所示,其中,节点 F 在 $t=2$ 时刻的值由原来的 $x_{6,2}^1$ 变为 $\Theta_1(0.35x_{4,2}^1 + 0.50x_{6,2}^1)$,即,输入数据被其 $0 \sim 1$ 阶邻居的信息更新.其他特征及卷积核同理,对整个输入数据 $\mathcal{X}_h^{(r-1)}$ 进行一次图卷积操作,得到 $g_\theta \times_G \mathcal{X}_h^{(r-1)} \in \mathbb{R}^{C_r \times N \times T_{r-1}}$,每个节点被该节点的 $0 \sim K-1$ 阶邻居的信息更新.

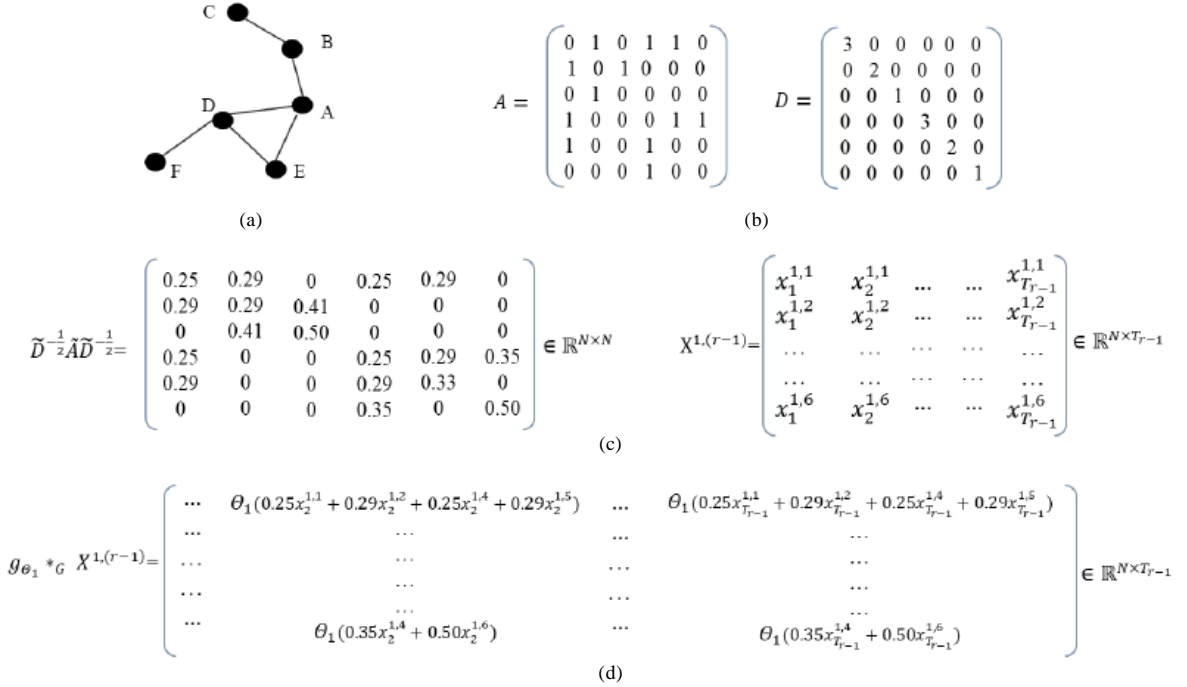


Fig.6 An example of graph convolutions and its matrix representation

图 6 图卷积示例及其矩阵表示

3.2 时间维卷积

通过空间维图卷积操作对输入数据的空间特征进行建模之后,再用标准 2 维卷积捕获时间维特征.我们用线性修正单元激活函数,以提取近期特征为例,得到 $\mathcal{X}_h^{(r)} = \text{ReLU}(\Psi \times (\text{ReLU}(g_{\theta} \times_G \mathcal{X}_h^{(r-1)}))) \in \mathbb{R}^{C_r \times N \times T_r}$ (Ψ 为时间维卷积核的参数),具体卷积过程如图 5 所示.

经过一层时间维卷积之后,节点的信息被该节点相邻时间片信息更新,而节点及其相邻时间片信息在经过图卷积操作后已包含其相邻节点同时刻的信息.因此,通过一层时空卷积操作之后,就会捕获到数据的时间维和空间维特征以及时空相关性.我们使用多层时空卷积,以提取时空维上更远的信息,再通过全连接操作使时空卷积的结果与预测目标维数一致,全连接模块同样使用线性修正单元作为激活函数.

3.3 多组件融合

在本节中,我们将讨论如何融合近期、日周期和周周期组件的输出.以对整个路网周五上午 8 点的流量进行预测为例,有些地方早晚高峰周期规律明显,其日周期、周周期组件的预测结果更为重要;而有些地方不存在明显的交通周期规律,日周期、周周期组件对预测目标起到的帮助较小.可见,不同节点受不同组件的影响程度是不一样的,在融合不同组件的输出结果时,其影响程度也应该不同.因此,应该从历史数据中进行学习.融合后的最终预测结果表示为

$$\hat{Y} = W_h \odot \hat{Y}_h + W_d \odot \hat{Y}_d + W_w \odot \hat{Y}_w,$$

其中, \odot 是矩阵对应元素相乘的哈达马积; W_h, W_d, W_w 是学习参数,反映了近期、日周期、周周期这 3 种时间维特性对预测目标的影响程度.

4 实验与结果分析

为了检验本文提出模型的性能,我们在两个真实世界的数据集上进行了对比实验.本节将对数据集和实验

设置进行说明,并对实验结果进行详细的对比分析.

4.1 数据集介绍

我们使用美国加利福尼亚州的真实高速公路数据集 PeMSD4 和 PeMSD8 来验证我们的模型. PeMS 是由 Caltrans Performance Measurement System 实时采集的高速公路交通数据,该系统拥有超过 39 000 个传感器站,部署在加利福尼亚州高速公路系统的主要大都市区^[16].数据集是基于 30s/次的频率采样得到的原始数据汇总成的以 5m 为时间间隔的样本.数据集包含带时间戳的车流量、平均车速、平均车道占用率这 3 个维度的特征及采集这些信息的检测器的地理位置信息.

- PeMSD4: San Francisco Bay 区域的数据,共包含 29 条路上的 3 848 个检测器,我们选取时间范围为 2018 年 1 月~2 月的数据进行实验,其中,前 50 天作为训练集,后 9 天作为测试集;
- PeMSD8: San Bernardino 区域的数据,共包含 8 条路上的 1 979 个检测器,我们选取时间范围为 2016 年 7 月~8 月的数据进行实验,其中,前 50 天作为训练集,后 12 天作为测试集.

4.2 数据预处理

我们对道路上的检测器进行筛选,去掉那些距离过近的检测器,保证检测器节点间距离大于 3.5 英里.样本以 5m 为时间间隔,因此,道路上每个节点每天包含 288 个数据点.用线性插值法填充缺失值.另外,对数据进行 0 均值标准化(zero-mean)操作 $x' = x - \text{mean}(x)$,即让数据的平均值为 0.

4.3 实验参数设置

我们基于 Pytorch 框架实现了 MCSTGCN 模型,图卷积使用 32 个相同大小的卷积核.时间维卷积同样使用 32 个相同大小的卷积核,这些卷积核沿空间轴维度为 1,沿时间轴维度为 3,通过控制步长调整时间维长度.整个模型训练时间大约为 20s/轮.3 个组件的输入数据长度会对实验结果产生影响,我们通过实验选定 $T_h=3, T_d=1$ 和 $T_w=1$ 为最佳的组合.均方误差是反映估计量与被估计量之间差异程度的一种度量,因此,本文采用该度量指标作为损失函数.

4.4 基准方法

我们将本文提出的模型与以下 7 种已有的时间序列预测方法进行比较.

- HA: 历史均值法,其预测值为近期历史流量状况的平均值.在本文中,我们使用最近 12 个时间片的平均值来预测下一个时间片的值;
- ARIMA^[2]: 自回归积分滑动平均法,是时间序列分析中的一种经典方法;
- LSTM^[17]: 长短期记忆网络,一种特殊的 RNN 模型.LSTM 单元由细胞、输入门、输出门和遗忘门组成;
- GRU^[18]: 门控循环单元网络^[18],一种特殊的 RNN;
- STGCN^[7]: 一种基于空间方法定义的时空图卷积模型,采用了多种划分策略将中心节点的邻居划分到不同子集,实现卷积核的参数共享;
- GCGRU^[11]: 一种将门控循环单元网络与图卷积网络结合的方法;
- GLU-STGCN^[12]: 一种带有门控机制的图卷积网络,在交通数据预测问题上取得了很好的效果.

本文采用均方误差(RMSE)和平均绝对误差(MAE)作为评价指标,具体计算公式下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2},$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i|.$$

4.5 实验结果及分析

我们将 MCSTGCN 模型在数据集 PeMSD4 和 PeMSD8 上与前述 7 种基准方法进行了比较,表 2 展示了对未来 1h 内的流量进行预测的结果.

Table 2 Performance comparison of different approaches on the PeMSD4 and PeMSD8 dataset**表 2** 不同方法在数据集 PeMSD4 和 PeMSD8 上的性能比较

Model	PeMSD4		PeMSD8	
	RMSE	MAE	RMSE	MAE
HA	54.14	36.76	44.03	29.52
ARIMA	68.13	32.11	43.30	24.04
LSTM	45.82	29.45	36.96	23.18
GRU	45.11	28.65	35.95	22.20
STGCN	38.29	25.15	27.87	18.88
GCGRU	37.25	25.00	30.24	20.69
GLU-STGCN	38.41	27.28	30.78	20.99
MCSTGCN(ours)	35.64	22.73	26.47	17.47

从表 2 可以看出,我们的模型在两种评价指标中均达到最佳性能.我们还可以观察到,传统的时间序列预测方法由于建模能力有限,预测结果并不理想.基于深度学习的方法获得了比传统方法更好的预测结果,而包含时空图卷积机制的模型由于考虑了时空依赖性,预测结果要优于一般的深度学习模型.

此外,我们还比较了各种方法随着预测时长的增长性能变化情况.我们以 5m 为间隔,让预测时长从 5m 增长到 1h,实验结果如图 7 所示.

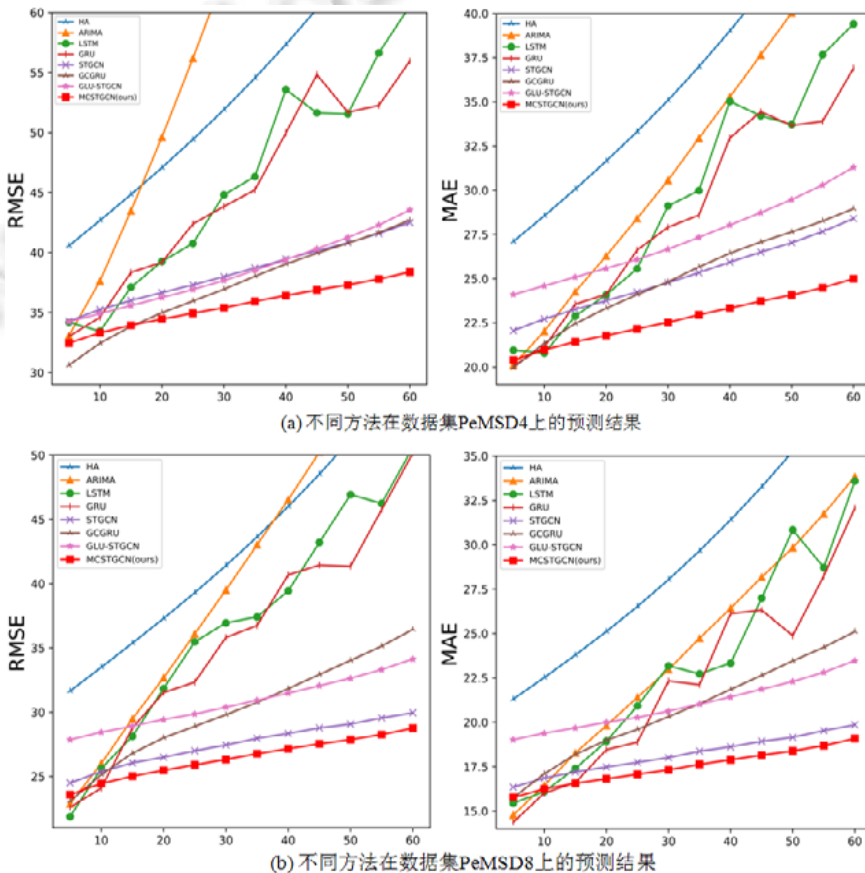
**Fig.7** Performance changes of different methods along with the forecasting duration

图 7 不同方法随着预测时长的性能变化情况

从图 7 可以看出,从整体上,随着预测时长的逐步增大,预测难度越来越大,误差整体呈上升趋势.我们的 MCSTGCN 模型在短期预测中就取得了较优的的预测结果,这表明了多组件与时空图卷积结合的策略能充分

挖掘数据的时空模式;而随着预测时长的增大,我们的模型预测误差比其他方法增长更加缓慢,这是由于该模型显式建模了时间维度上的多种周期特性,因而在中长期预测中显示更加明显的优势。

最后,为了观察每个组件的输入数据的长度变化对预测结果的影响,我们在 PeMSD8 数据集上将各组件的输入数据长度分别设置为 $T_h \in \{1,2,3\}$, $T_d \in \{1,2,3\}$, $T_w \in \{1,2,3\}$ 进行了实验对比,其结果见表 3。

Table 3 MAE with different component input lengths on dataset PeMSD8

表 3 不同组件输入数据长度在数据集 PeMSD8 上的 MAE

	$T_d=1, T_w=1$	$T_d=1, T_w=2$	$T_d=1, T_w=3$	$T_d=2, T_w=1$	$T_d=2, T_w=2$	$T_d=2, T_w=3$	$T_d=3, T_w=1$	$T_d=3, T_w=2$	$T_d=3, T_w=3$
$T_h=1$	19.26	19.27	19.25	18.10	19.25	19.26	19.25	19.25	19.28
$T_h=2$	18.36	18.47	18.71	18.18	18.01	18.71	18.70	18.70	18.70
$T_h=3$	17.48	17.79	17.47	17.58	17.67	17.47	17.75	17.83	17.48

可以看出, T_h 的变化对预测结果影响较大,而 T_d 和 T_w 影响相对较小,我们选定 $T_h=3, T_d=1$ 和 $T_w=1$ 为最佳的组合。

5 总结

本文提出一种新颖的多组件时空图卷积网络 MCSTGCN,该模型结合图卷积和标准卷积构造时空卷积块来同时捕获交通数据的时空特性.在真实的高速公路流量数据集上的实验表明,本文提出的模型的预测效果优于其他已有的交通数据预测方法,验证了该模型在捕获时空特征及时空相关性方面具有优势.事实上,除了高速公路流量预测任务外,我们提出的模型也适用于处理其他基于图结构表示的时空交通数据.未来,我们将进一步通过注意力机制等策略来优化模型的网络结构,进一步提升模型的预测能力。

References:

- [1] Zhang JP, Wang FY, Wang KF, Lin WH, Xu X, Chen Ch. Data-driven intelligent transportation systems: A survey. *IEEE Trans. on Intelligent Transportation Systems*, 2011,12(4):1624–1639.
- [2] Williams BM, Hoel LA. Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. *Journal of Transportation Engineering*, 2003,129(6):664–672.
- [3] Van Lint H, Van Hinsbergen C. Short-term traffic and travel time prediction models. *Transportation Research E-circular*, 2012, 22(1):22–41.
- [4] Jeong YS, Byon YJ, Castro-Neto MM, Easa SM. Supervised weighting-online learning algorithm for short-term traffic flow prediction. *IEEE Trans. on Intelligent Transportation Systems*, 2013,14(4):1700–1707.
- [5] Zhang J, Zheng Y, Qi D, Li RY, Yi XW, Li TR. Predicting citywide crowd flows using deep spatio-temporal residual networks. *Artificial Intelligence*, 2018,259:147–166.
- [6] Niepert M, Ahmed M, Kutzkov K. Learning convolutional neural networks for graphs. In: *Proc. of the ICML*. 2016. 2014–2023.
- [7] Li C, Cui Z, Zheng W, Xu C, Yang J. Spatio-temporal graph convolution for skeleton based action recognition. In: *Proc. of the AAAI*. 2018.
- [8] Bruna J, Zaremba W, Szlam A, Lecun Y. Spectral networks and locally connected networks on graphs. In: *Proc. of the ICLR*. 2014.
- [9] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. In: *Proc. of the NIPS*. 2016. 3844–3852.
- [10] Seo Y, Defferrard M, Vandergheynst P, *et al.* Structured sequence modeling with graph convolutional recurrent networks. *arXiv preprint arXiv:1612.07659*, 2016.
- [11] Li Y, Yu R, Shahabi C, *et al.* Diffusion convolutional recurrent neural network: data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926v1*, 2017.
- [12] Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In: *Proc. of the IJCAI*. 2018.
- [13] Henaff M, Bruna J, LeCun Y. Deep convolutional networks on graph-structured data. *arXiv preprint arXiv:1506.05163*, 2015.

- [14] Simonovsky M, Komodakis N. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In: Proc. of the CVPR. 2017. 29–38.
- [15] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: Proc. of the ICLR. 2017.
- [16] Chen C, Petty K, Skabardonis A, Varaiya P, Jia ZF. Freeway performance measurement system: Mining loop detector data. Journal of the Transportation Research Board, 2001,1748:96–102.
- [17] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997,9(8):1735–1780.
- [18] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv: Neural and Evolutionary Computing, 2014.



冯宁(1993—),女,天津人,硕士,CCF 学生会员,主要研究领域为时空数据挖掘,图数据挖掘.



朱琪超(1994—),女,硕士,主要研究领域为时空数据挖掘.



郭晟楠(1992—),女,博士,CCF 学生会员,主要研究领域为时空数据挖掘,深度学习.



万怀宇(1981—),男,博士,副教授,博士生导师,CCF 专业会员,主要研究领域为社交网络挖掘,交通数据挖掘.



宋超(1995—),男,硕士,主要研究领域为时空数据挖掘,图数据挖掘.

www.jos.org.cn