

参考综述：Graph Neural Network for Traffic Forecasting: A Survey (2021 年 清华大学)

1. 问题引入+算法发展

交通流的预测任务可以看做是一个 seq2seq 的任务，给定图上前 t1 个的交通数据来预测接下来 t2 个交通数据。

即：学习一个函数 $h(\cdot)$ 将 T 映射到 T' 上。

$$[X^{(t-T'+1)}, \dots, X^{(t)}; \mathcal{G}] \xrightarrow{h(\cdot)} [X^{(t+1)}, \dots, X^{(t+T)}]$$

传统算法：

模型分类	具体算法	缺点
传统时间序列分析 (统计模型)	历史平均(HA)、 自回归综合移动平均(ARIMA) 向量自回归(VAR)	这些方法需要数据满足一定的假设。 难以处理不稳定、非线性数据
传统机器学习	KNN、SVM	虽然能够处理复杂数据，但是 (1) 难以同步考虑高维交通数据的时空相关性 (2) 依赖于特征工程，需要较多专家经验
传统深度学习	CNN, LSTM	虽然可以提取时空特征，但是只能处理结构化的 2D 或者 3D 等规则数据
基于图的深度学习	GCN, ChebNet, GGCN	没有同时考虑动态时空相关性

要解决的问题：Spatial-Temporal GNNs

交通流的（各个时间点间的）时间动态性和（交通网络图各个节点间的）空间依赖性。

2. 图构建

- 图定义: $G = (V, E, A)$
 - 其中 V 为节点集, E 为边集, A 为邻接矩阵。
 - A 中的元素 a_{ij} 表示节点 i 和 j 之间的“边权值”。
- 对于单个时间步 t , G 的节点特征矩阵 ' $X_t \in \mathbb{R}^{N \times d}$ ' 包含特定的交通状态
 - 其中 N 为节点数, d 为特征数。
- 基于图的交通预测定义: 找到一个函数 f , 生成 $y = f(X; G)$
 - 其中 y 为待预测的交通状态, $X = \{x_1, x_2, \dots, x_T\}$ 为图 G 上定义的历史流量状态, T 为历史窗口大小中的时间步长数。
- 考虑外部因素 ε : 例如天气和节假日, $y = f(X, \varepsilon; G)$

3. 经典模型: DCRNN、STGCN、Graph

WaveNet

- **DCRNN** (ICLR 2018)

原文: DIFFUSION CONVOLUTIONAL RECURRENT NEURAL NETWORK: DATA-DRIVEN TRAFFIC FORECASTING

参考链接: [DCRNN 理解笔记](#)

通过 DCNN 结合正反向的流量卷积获取空间依赖特征, 通过

DCGRU 部分获取时间依赖特征。

1. DC (Diffusion Convolution) : 扩散卷积

利用双向的随机游走来获取图上的空间关系（随机游走：类比物理中的布朗运动），**从一个顶点开始遍历，以概率 $1-\alpha$ 走到顶点的邻居节点，以 α 的概率跳跃到图中的任意一个顶点。**

■ 扩散（随机游走的计算方式）：

$$\mathcal{P} = \sum_{k=0}^{\infty} \alpha(1-\alpha)^k (D_O^{-1}W)^k$$

α 是表示跳转的任意一个顶点的概率， $1-\alpha$ 表示跳转到邻居节点，而后面的 $D_O^{-1}W$ 表示状态转移矩阵，即从当前节点跳转到每一个邻居节点的概率， K 则表示扩散步骤。

■ 扩散卷积：

$$X_{:,p} \star_{\mathcal{G}} f_{\theta} = \sum_{k=0}^{K-1} \left(\theta_{k,1} (D_O^{-1}W)^k + \theta_{k,2} (D_I^{-1}W^{\top})^k \right) X_{:,p} \quad \text{for } p \in \{1, \dots, P\}$$

对出度的图进行卷积操作 + 对入度的图进行卷积操作（考虑了两个方向的流量卷积）， $D_O^{-1}W$ 是上面提到的随机游走计算出来的状态转移矩阵。使用这个卷积运算操作就可以构建出一个扩散卷积层。

参考链接：[DCRNN 算法讲解视频](#)

$$X_{:,p} * G f_0 = \sum_{k=0}^{K-1} \theta_{k,1} (D_0^{-1} W)^k X_{:,p} + \sum_{k=0}^{K-1} \theta_{k,2} (D_1^{-1} W^T)^k X_{:,p}$$

对输出的图进行卷积操作 对输入的图进行卷积操作

D_{output} 输出矩阵

D_{input} 输入矩阵

$X_{:,p}$ 所有的第 p 个特征

$*G$ 对图 G 做图卷积

f_0 卷积核

θ 核里的参数

$W \cdot A$
邻接矩阵 特征矩阵

$w^2 \cdot A, w^3 \cdot A$

$$D_0 = \text{diag}(W1), \quad W * 1 = \begin{bmatrix} \text{---} \\ \text{---} \\ \text{---} \end{bmatrix}_{N \times N} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{N \times 1} = \begin{bmatrix} \bullet \\ \bullet \\ \bullet \end{bmatrix}_{N \times 1}$$

该节点的出度

$$\begin{bmatrix} \bullet & & \\ & \ddots & \\ & & \bullet \end{bmatrix}_{N \times N}$$

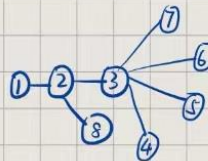
$$\sum_{k=0}^{K-1} \theta_{k,1} (D_0^{-1} W)^k X_{:,p}$$

K : 打散程度

$k=0$: 不打散, 直接取

If $k=1$

$$\sum_{k=0}^{k-1} \theta_{k,1} (D_0^{-1} W)^k X_{:,p} = \underbrace{\theta_{0,1} X_{:,p}}_{\text{自身}} + \underbrace{\theta_{1,1} (D_0^{-1} W) X_{:,p}}_{\text{1跳节点卷积}}$$



$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1/2 & 0 & \dots & 0 \\ 0 & 0 & 1/4 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1/4 \end{bmatrix}_{D_0^{-1} \cdot N \times N} \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 1 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 0 \end{bmatrix}_{W: N \times N} \begin{bmatrix} 0.2 \\ 0.3 \\ 0.4 \\ \vdots \\ 0.4 \end{bmatrix}_{X_{:,p}}$$

$D_0^{-1} \rightarrow W \cdot N \times N$ 对 W 归一化 (按行)

WA 一跳节点有关的
信息进行聚合

$W^2 A$ 二跳

$W^3 A$ 三跳
聚合到 1 里

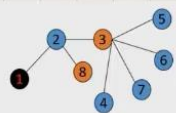
Diffusion Convolution

If $k=2$

$$\sum_{k=0}^{k-1} \theta_{k,1} (D_0^{-1} W)^k X_{:,p} = \underbrace{\theta_{0,1} X_{:,p}}_{\text{自身}} + \underbrace{\theta_{1,1} (D_0^{-1} W) X_{:,p}}_{\text{1跳}} + \underbrace{\theta_{2,1} (D_0^{-1} W)^2 X_{:,p}}_{\text{2跳}}$$

$$\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1/2 & \dots & 1/2 \\ 1/4 & 1/4 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1/4 & 1/4 & 1/4 & \dots & 0 \end{bmatrix}_{W': N \times N} \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1/2 & \dots & 1/2 \\ 1/4 & 1/4 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1/4 & 1/4 & 1/4 & \dots & 0 \end{bmatrix}_{W'': N \times N} \begin{bmatrix} 0.2 \\ 0.3 \\ 0.4 \\ \vdots \\ 0.4 \end{bmatrix}_{X_{:,p}}$$

$W'' \leftarrow \text{聚合了 3, 8 节点的信息}$



2. RNN 模块: DCGRU 扩散卷积门控单元

- GRU 的变体: 将 GRU 中的矩阵乘法替换为扩散卷积

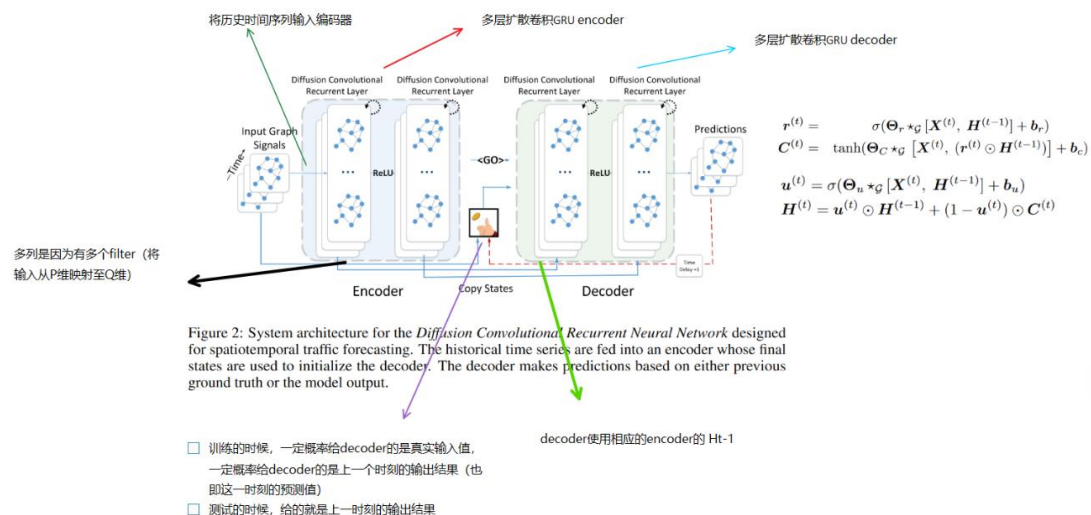
$$\begin{aligned} r^{(t)} &= \sigma(\Theta_r \star_G [X^{(t)}, H^{(t-1)}] + b_r) & u^{(t)} &= \sigma(\Theta_u \star_G [X^{(t)}, H^{(t-1)}] + b_u) \\ C^{(t)} &= \tanh(\Theta_C \star_G [X^{(t)}, (r^{(t)} \odot H^{(t-1)})] + b_c) & H^{(t)} &= u^{(t)} \odot H^{(t-1)} + (1 - u^{(t)}) \odot C^{(t)} \end{aligned}$$

其中 $X(t), H(t)$ 表示时间 t 的输入和输出， $r(t)$ 和 $u(t)$ 分别是时间 t 的重置门和更新门。

- 采用 seq2seq 架构：编码器和解码器都是带有 DCGRU 的循环神经网络

3. 模型结构

参考链接：[DCRNN 模型图详解](#)



● STGCN (IGCAI 2018)

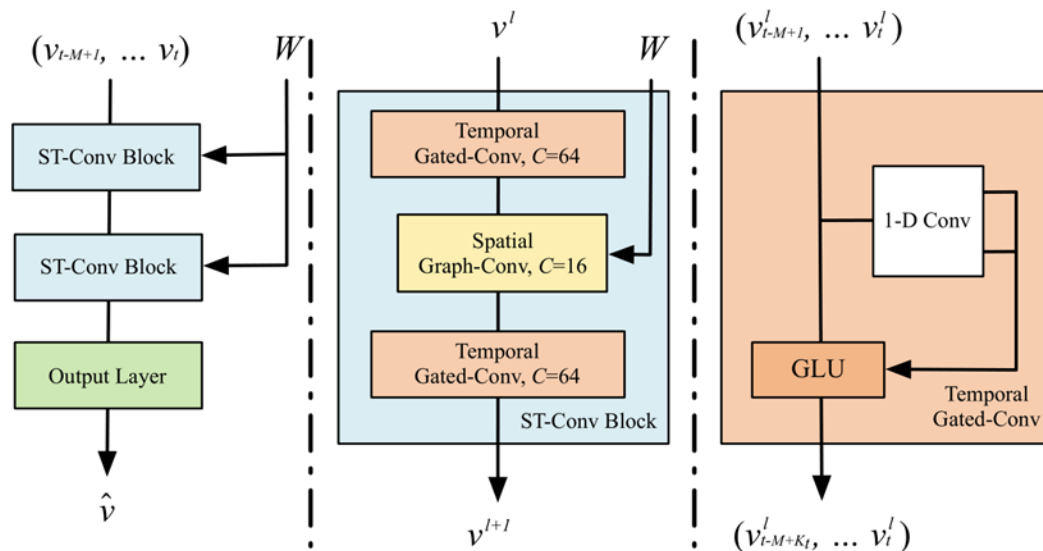
1. STGCN (北大, IGCAI 2018)

原文：Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic

参考链接：[STGCN 详解](#)

在交通研究中首次应用纯卷积结构同时提取图结构时间序列的时

空特征，即建立了完全卷积的结构，并不是直接应用传统的卷积以及循环神经网络，这可以让训练速度更快，参数更少。



堆叠多个时空卷积块，每个块串联两个（时间）门控卷积层和一个（空间）图卷积层。

2. ASTGCN (北交大, AAAI 2019)

原文: Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting

参考链接: [ASTGCN 详解](#)

图卷积 + 注意力机制

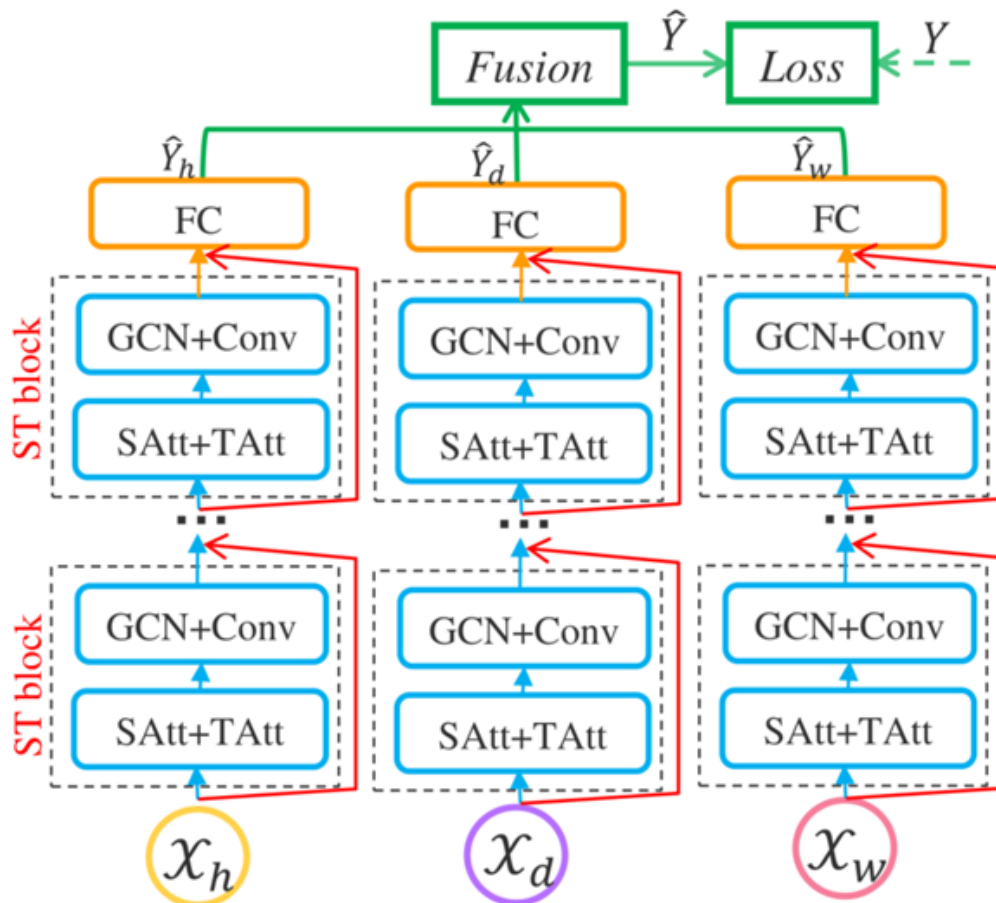
ASTGCN 主要由三个独立的组件组成，对时间序列上三个不同长短的周期模式（当前、日、周）进行建模。

每个组件 (ST BLOCK) 包含: 时空注意机制 + 时空卷积

ST BLOCK=(Satt+Tatt)+(GCN+conv):

(1) 先用注意力机制提取更有效的时间和空间特征,

- (2) 再输入给卷积层,
- (3) 卷积层中,用图卷积提取空间特征,用标准卷积提取时间特征,
- (4) 最后将注意力层和卷积层堆叠, 完成 ST BLOCK 的构建。

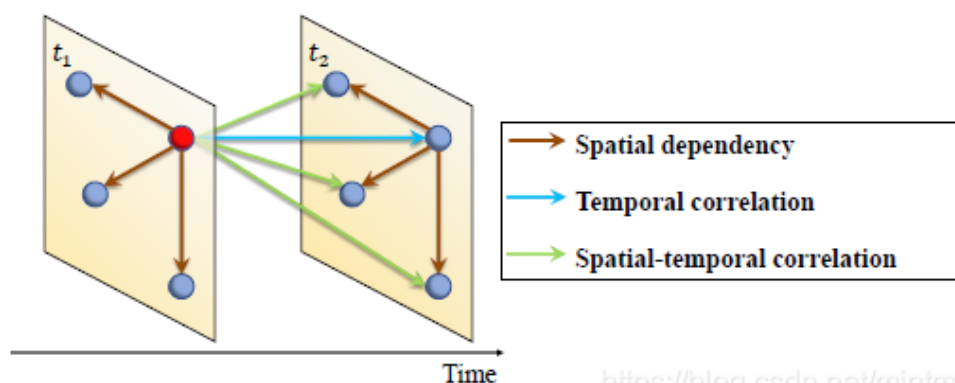


3. STSGCN (北交大, AAAI 2020)

原文: Spatial-Temporal Synchronous Graph Convolutional Networks: A New Framework for Spatial-Temporal Network Data Forecasting

参考链接: [STSGCN 论文笔记](#), [STSGCN 模型结构讲解](#)

■ 问题引入



三类依赖关系

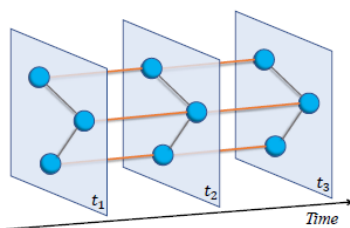
- 在同一时间片中，邻居节点与特定点的依赖关系（棕色箭头）
- 在不同时间片中，特定点对其自身的依赖关系（蓝色箭头）
- 在不同时间片中，邻居节点对特定点的依赖关系（绿色箭头）

两个单独的组件，捕获空间依赖性和时间相关性，再将空间表示形式加到是渐渐磨中，这么做不能很好的体现时空相关性。

■ 时空图卷积模块 (STSGCM)：同步捕获局部时空相关性

局部时空图：连续 3 个时间片组成一个局部时空图，体现节点与其多阶邻居的关系

局部时空图



(a) Localized Spatial-Temporal Graph

局部时空图的邻接矩阵

$A(t_1)$	$A^{t_1 \rightarrow t_2}$	
$A^{t_2 \rightarrow t_1}$	$A(t_2)$	$A^{t_2 \rightarrow t_3}$
	$A^{t_3 \rightarrow t_2}$	$A(t_3)$

在时间片 i 和时间片 j 节点与自身的连接关系

$$A'_{i,j} = \begin{cases} 1, & \text{if } v_i \text{ connects to } v_j \\ 0, & \text{otherwise} \end{cases}$$

在时间片 i 的空间图的邻接矩阵

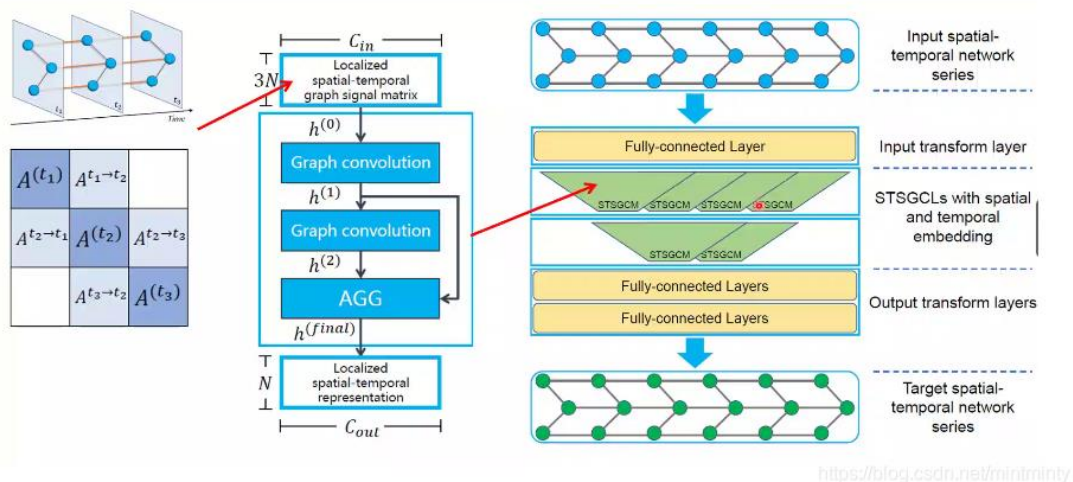
(b) Adjacency matrix of Localized Spatial-Temporal Graph

1	1	0	1	1	1										
1	1	1	1	0		1									
0	1	1	0	1			1								
1	1	0	1	1				1							
1	0	1	1	1					1						
1						1	1	0	1	1	1				
	1					1	1	1	1	0		1			
		1				0	1	1	0	1			1		
			1			1	1	0	1	1				1	
				1		1	0	1	1	1					1
					1						1	1	0	1	1
						1					1	1	1	1	0
							1				0	1	1	0	1
								1			1	1	0	1	1
									1	1	0	1	1	1	1
									1	1	0	1	1	1	1

- 时空图卷积层 (STSGCL) : 捕获长期时空网路数据的异质性 (不均匀性)

在不同时间段部署多个单独的 STSGCM, 再将多个 STSGCL 堆叠以聚合。

- 模型结构



• Graph WaveNet (IJCAI 2019)

原文: Graph WaveNet for Deep Spatial-Temporal Graph Modeling

MLP（多层感知器）。

■ TCN

本模型的 TCN 组件采用了门限机制，包括两个 TCN 块，这两个 TCN 使用了不同的激活函数，TCN-a 使用 tanh，是神经网络训练过程中的最常使用的激活函数（非线性映射），右侧 TCN-b 使用 sigmoid 激活函数，然后将二者的值按元素进行乘积。

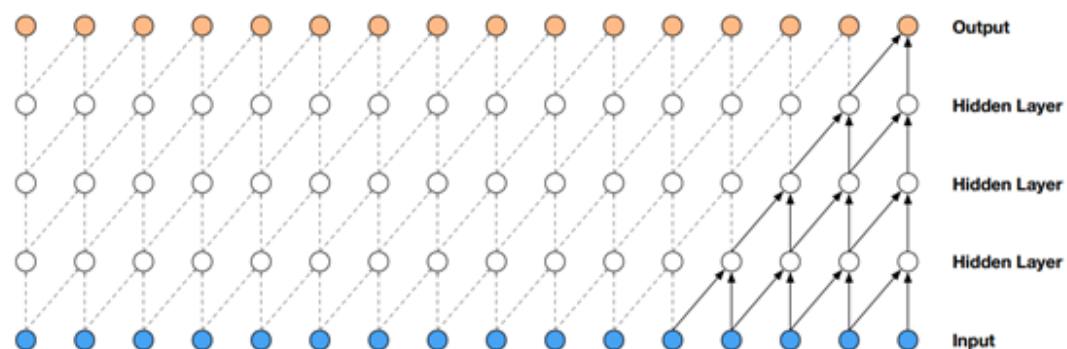
这里的门限机制和 LSTM 中的是一致的，作用就是控制有效信息的流入和无效信息的丢弃。在 Gated-TCN 中，TCN-a 就相当于 input（被控制的向量），TCN-B 就是 gate,最后将两者的值进行乘积得到该组件的输出。

$$h = g(\Theta_1 \star \chi + b) \odot \sigma(\Theta_2 \star \chi + c),$$

■ 扩张因果卷积

在 TCN 中采用 Dilated Casual Conv（扩张因果卷积），将所有的输入数据以时间维度为轴看作一维向量，对所有的输入向量进行 1D 卷积操作。

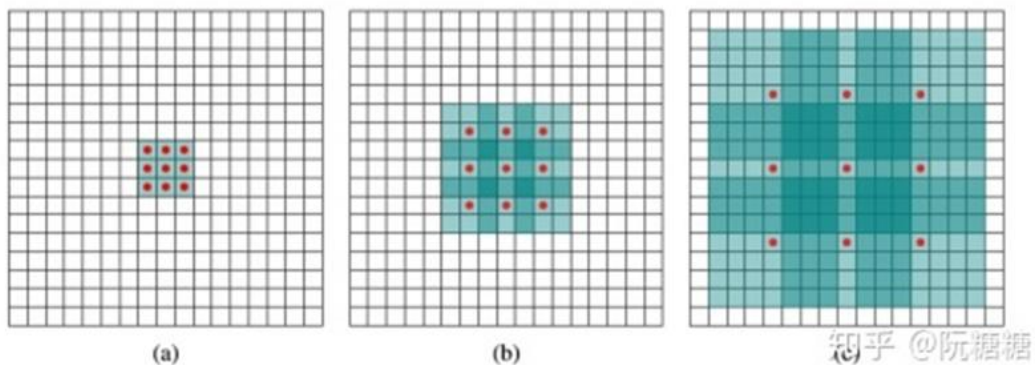
因果卷积：就是某时刻的预测值仅仅和历史时刻的值相关，保证了在预测任务中，只是用历史信息进行预测，不包含未来的信息。



扩张卷积

扩张因果卷积：空洞卷积神经网络（替换传统叠加 CNN+池化层）

空洞卷积神经网络的作用类似卷积神经网络联合 pooling 层。比起 pooling 层，空洞卷积神经网络能够减少信息的损失，并在每一层操作后不减少节点数量；比起叠加 CNN 网络，空洞卷积神经网络能够以指数级别的增长速度扩大视野域。这是不是说明空洞卷积神经网络能够代替传统 CNN+pooling 层呢？在图像领域，传统 CNN 依然有着强势的地位，其原因之一是在图像领域，位置相邻的特征之间的关系是非常重要的，而空洞卷积神经网络不能够很好捕捉到这一信息。但在本文中，用它来解决一维的时序信息是比较合适的。



4. 数据集

- **METR-LA**，洛杉矶高速公路数据集，207 个传感器并收集了 2012 年 3 月 1 日至 2012 年 6 月 30 日 4 个月的数据

备注：数据应用在 **DCRNN** 文章中。

- **PeMS04**，加利福尼亚高速公路数据集，shape 为(307, 16992, 3)，307 个传感器，16992 组数据（5 分钟采集一次，共 59 天），3 维特征（flow, occupy, speed）

- **PeMS08**，加利福尼亚高速公路数据集，shape 为(170, 17856, 3)，170 个节点，连续 62 天

备注：数据应用在 **ASTGCN** 文章中。

- TaxiBJ，北京出租车数据集，郑宇，
"BJ15_M32x32_T30_InOut.h5"，原始数据
shape=(5596,2,32,32)

备注：数据应用在 **ST-ResNet** 文章中。

- NYC-Taxi，纽约出租车数据集，"volume.train.npz"，原始数据
shape=(1920,10,20,2)

备注：数据应用在 **STDN** 文章中。

5. 评估指标

- 均方根误差(RMSE)
- 平均绝对误差(MAE)
- 平均绝对百分比误差(MAPE)

6. 总结和未来工作的启发