

时空序列预测方法综述*

黎 维^a, 陶 蔚^a, 周星宇^b, 潘志松^{a†}

(陆军工程大学 a. 指挥控制工程学院; b. 通信工程学院, 南京 210007)

摘要: 随着数据采集技术的进步, 带有地理位置信息的时空数据迅速增长, 迫切需要探索有效的时空数据建模方法。时空序列预测是时空数据建模的基础方法之一, 它广泛应用于很多领域。目前缺乏对它进行综述的中文文献, 因而对这些方法进行归纳和总结具有重要的研究意义。针对时空序列预测问题进行了研究, 首先回顾了其应用背景和发展历程, 介绍了它的相关定义及特点。然后按其类别介绍了传统的时空序列预测方法、基于传统机器学习的时空序列预测方法和基于深度学习的时空序列预测方法, 并分析了这些方法的应用范围和优缺点。最后对时空序列预测未来的研究方向进行了梳理和展望, 为研究者们进一步深入研究时空序列预测问题奠定了理论基础。

关键词: 机器学习; 深度学习; 时空数据; 时空序列预测

中图分类号: TP181

文献标志码: A

文章编号: 1001-3695(2020)10-001-08

doi: 10.19734/j.issn.1001-3695.2019.05.0184

Survey of spatio-temporal sequence prediction methods

Li Wei^a, Tao Wei^a, Zhou Xingyu^b, Pan Zhisong^{a†}

(a. College of Command & Control Engineering, b. College of Communication Engineering, Army Engineering University of PLA, Nanjing 210007, China)

Abstract: With the advancement of data acquisition technology, spatiotemporal data with geographic location information are growing rapidly, so it is urgent to explore effective spatiotemporal data modeling methods. Spatiotemporal sequence prediction is one of the basic methods in spatiotemporal data modeling and widely used in many fields. At present, there is no Chinese literature on its review, so it is of great significance to summarize these methods. For the spatiotemporal sequence prediction problem, this paper reviewed its application background and development history. Then it introduced some related definitions and characteristics. According to its category, the spatiotemporal sequence prediction methods based on traditional methods, classical machine learning methods and deep learning methods are analyzed respectively as well as their application scope, advantages and disadvantages. Finally, this paper made a review and prospect over the research directions of spatiotemporal sequence prediction, laying a theoretical foundation for researchers to further study this problem.

Key words: machine learning; deep learning; spatiotemporal data; spatiotemporal sequence prediction

0 引言

移动互联网(mobile Internet, MI)^[1]、传感器网络(sensor network, SN)^[2]和全球定位系统(global positioning system, GPS)^[3]等技术的快速普及和广泛应用使得各行业的数据量呈爆发式增长, 其中最典型的就是时空序列数据。时空序列数据是一种具有空间维度的时间序列数据。它广泛存在于多个领域, 如天气预测^[4-6]、经济统计^[7-9]、交通预测^[10-12]、环境监测和粮食产量预测^[13-15]等。准确地提取数据中的时空特征, 并以此为据构建出预测模型是时空序列预测的关键^[16]。

时间序列数据^[17]存在着时间维度上的前后依赖关系, 它通常是在事先指定的相等间隔时间段内, 依据设定的采样频率, 对研究的随机变量进行观测记录从而得到的一组序列数据^[18]。图1展示的是山西省河曲水文站自1952年以来50年降水量时间序列。对时间序列在空间维度进行扩展可以得到时空序列(spatio-temporal sequence, STS), 它通常是多个时间序列的集合^[19]。图2展示的是从2015年11月起36个月内长江三角洲地区价格指数时空序列。

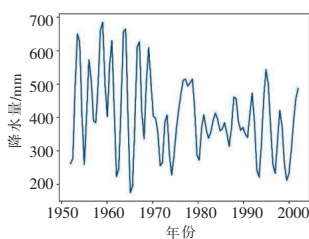


图1 山西省河曲水文站降水量时间序列图

Fig. 1 Time series of precipitation in Hequ hydrological station, Shanxi Province

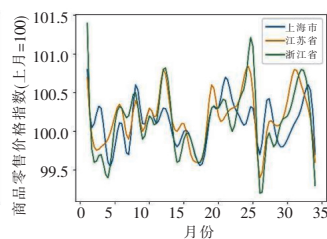


图2 长江三角洲价格指数时空序列图

Fig. 2 Spatio-temporal sequence of Yangtze River delta price index

时空序列较为复杂, 进行时空序列预测时不仅需要考虑时间上的连续性和周期性, 还需要考虑不同区域之间的空间相关性, 并且这些空间相关性还会随着时间发生变化^[20]。因此, 要准确地从复杂的时空数据中挖掘出所需要的时空特征具有较大的难度。传统的时空序列预测方法主要是基于统计学原理, 这些传统方法基本都将时空序列看做是多个时间序列来处理, 很难捕获空间的相关性, 也难以挖掘出其中的非线性时空模式。传统的机器学习方法虽然能有限地捕获时空数据中的非

收稿日期: 2019-05-28; 修回日期: 2019-07-17 基金项目: 国家自然科学基金资助项目

作者简介: 黎维(1995-), 男, 湖北咸宁人, 博士研究生, 主要研究方向为深度学习、时空序列预测; 陶蔚(1991-), 男, 安徽合肥人, 博士研究生, 主要研究方向为凸优化算法在机器学习中的应用; 周星宇(1985-), 男, 湖南邵阳人, 讲师, 博士研究生, 主要研究方向为深度学习; 潘志松(1973-), 男(通信作者), 福建南安人, 教授, 博导, 主要研究方向为模式识别、机器学习和网络安全(hotps@hotmail.com)。

线性关系,但是长期预测上效果不佳,也很难适应大规模的时空数据集。近年来兴起的深度学习方法在时空序列预测问题中取得了比前两者更好的效果。无论是在预测精度还是在针对大规模数据集的处理上,深度学习方法都比前两者有优势。在用深度学习方法进行时空序列预测中,一般的处理方法是将时空序列视为一系列有时间顺序的图片,用卷积网络提取数据中的空间特征,然后将其送入循环神经网络中捕获时序特征,最后得出预测结果^[21]。当前,使用深度学习中的各种方法从多方面对时空数据进行分析,进而构建一个时空序列预测混合模型已成为一种趋势。

总而言之,目前时空序列预测问题中还存在许多问题。一方面时空数据本身存在较多缺失和错误,而现有的预处理方法较为简单,亟需改进;另一方面是现有的预测方法都有一定的局限性,难以全盘考虑时空序列数据各个方面的特征,因而在预测精度上受到了限制。此外,将时空序列预测方法较好地应用于各种实际问题也存在一定困难,易于部署的模型预测精度不理想(如历史平均法),而相对准确的基于深度学习的方法则存在模型较大、实时性难以保证等问题。

尽管针对时空序列预测问题的研究不断推进,但是相关的研究都较为零散^[4,9,10,15],很少有对于该领域各方面整体介绍的中文文献。文献[22]从较为抽象的层次介绍了时空序列预测中的机器学习方法,本文除了介绍了时空序列预测中的机器学习方法,还介绍了近两年较新的利用深度学习进行时空序列预测的方法,并和传统的方法进行了对比分析。

1 时空序列的相关定义及其性质特点

时空序列数据反映了多个区域内多个随机变量随时间不断变化的趋势。除了时间和空间特征,时空数据还有属性特征,因而呈现出多维、动态和时空相关等复杂性质^[16]。本章主要介绍时空序列的相关定义及其特点,并给出一些常用的时空序列数据集。

1.1 时空序列预测

时空序列由多条相关的时间序列组合而成^[19],在进行时空序列预测时通常把要预测的目标空间 A 划分为 p 个子区域,分别记为 A_1, A_2, \dots, A_p ,第 i 个子区域 A_i 中的变量 X 在时刻 t 的观测值可以表示为 $X_{i,t}$,时空序列 STS 可以表示为

$$\text{STS} = \{\{A_1, TS_1\}, \{A_2, TS_2\}, \dots, \{A_p, TS_p\}\} \quad (1)$$

其中: TS_i 为区域 A_i 中变量 X 对应的时间序列 $\{X_{i,1}, X_{i,2}, \dots, X_{i,T}\}$ 。

记 $X_{i,1:T} = [X_{i,1}, X_{i,2}, \dots, X_{i,T}]$,多步时空序列预测是根据历史时空变量 $X_{i,1:T}$ ($i = 1, 2, \dots, p$) 来预测 $T+1$ 时刻到 $T+L$ 时刻地区 A 内所有 p 个子区域内变量 X 的观测值 $X_{i,T+1:T+L}$ ($i = 1, 2, \dots, p$)。建立时空序列预测模型的过程即寻找一个满足 $X_{i,T+1:T+L} = \mathcal{F}(X_{i,1}, X_{i,2}, \dots, X_{i,T})$ ($i = 1, 2, \dots, p$) 的映射关系 \mathcal{F} 的过程。在用时空序列预测对实际问题建模的过程中,通常把目标空间 A 划分为 $M \times N$ 个单位网格以便处理不同子区域间的时空关系,其中 M, N 分别为目标空间的长和宽。

当 $L = 1$ 时,问题为常见的单步时空序列预测问题,单步时空序列预测是多步时空序列预测的基础。当 $L > 1$ 时,问题为多步时空序列预测问题。多步时空序列方法主要分为间接多步预测和直接多步预测^[22]两大类。间接多步时空序列预测是先训练一个单步时空序列预测器,然后不断将产生的样本数据送到预测器中得到多步预测结果。该方法的缺点是在预测过程中上一时刻预测误差会传到下一个时刻,预测误差会不断累加。为了解决误差传递的问题,研究者们提出了直接多步时空序列预测方法,该方法对于每个预测间隔 h ($h = 1, 2, \dots, L$) 分

别训练一个模型,为了降低整个模型大小,这 L 个模型有时候会像间接多步时空序列预测模型一样均为一个相同的单步预测器,只不过 L 个模型的损失函数会被相加后共同优化。

1.2 时空序列数据的性质特点

时空序列是对研究对象在不同区域随时间变化过程的抽象表示。本节主要介绍时空数据的性质和特点,对于时空序列数据的预测起到了较好的指导作用。

时空序列数据比时间序列数据更为复杂,其具体性质特点包括:海量性、多尺度、动态性、高维性、模糊性、时空相关性、非线性等^[16]。

海量特征是时空数据多角度采集和长时间积累的结果,时空序列预测的首要目标就是研究出新的高效策略来克服海量数据带来的困难。目前基于大数据驱动的深度学习方法在处理海量时空数据方面表现比传统时空序列预测方法要好。

动态特征是指时空数据和时间高度相关,它随时间变化且具有一定周期性。例如在出租车需求预测问题^[23]中,出租车需求量在上下班时段和其他时段,在工作日和周末就变化很大,并且呈现出一定的周期性。针对时空序列的周期性,基于频域的傅里叶变换^[24]、小波分析^[25]等方法效果比较好。

高维特征是指时空数据具有时间维、空间维和属性维等多维特征^[26]。如在人流量预测问题^[27]中,交通人流量可能会被许多其他的复杂因素所影响,如天气、事件等,故时空数据一般是高维的。预测维度之外的其他维度可用做额外信息辅助时空序列的预测^[28]。

模糊特征指的是时空数据在如空间位置等方面具有一定的模糊性^[29]。时空相关性体现了研究对象在时间和空间上的关联,可以分为时空自相关性和时空偏相关性。非线性特征广泛存在于实际时空数据中,也是时空序列预测的重点和难点。不完整性是由时空数据部分缺失或错误造成的^[30]。

在研究时空序列预测方法过程中,必须充分考虑到时空数据的这些特点并进行相应的处理,这也是预测工作的重点和难点。结合时空数据的这些特点进行预测,才能针对实际问题给出令人满意的预测结果。

1.3 常用时空序列数据集

一些常用的标准时空数据集如表 1 所示,其中 NYC-bike^[31]、radar echo^[32] 是气象数据集, TaxiBJ^[27]、air quality^[4] 和 PEMS-BAY^[33] 是交通数据集。这些数据集需要经过筛选、去噪、变换等预处理后才能用于时空序列预测。

表 1 标准时空数据集介绍
Tab. 1 Introduction of benchmark spatio-temporal datasets

数据集名称	数据类型	采集地区	区域划分	时间跨度/d	时间片长度/min	应用领域
air quality dataset	气象监测数据	天津	无	1 094	60	天气预测
TaxiBJ dataset	出租车 GPS 数据	北京	32 × 32	510	30	人群预测
radar echo dataset	雷达数据	香港	100 × 100	97	6	天气预测
NYC-bike dataset	自行车轨迹数据	纽约	10 × 20	60	30	交通量预测
PEMS-BAY dataset	传感器数据	加利福尼亚	无	180	15	交通速度预测

2 传统的时空序列预测方法

一般而言,传统的时空序列预测方法都是基于参数模型的,即在确定时空序列预测模型的基础上求解出模型的参数,然后完成预测工作。本章主要介绍几种常见的传统时空序列预测方法。

2.1 基于历史平均的时空序列预测

历史平均法 (historical average, HA) 是一种较为简单的预测方法, 该方法将研究对象在同一位置相同时间间隔内的历史观测值取平均或加权取平均后作为预测结果, 其计算复杂度低且易于部署。但由于缺乏深层次的挖掘, 通常在使用该方法进行时空序列预测时存在预测结果精度不高的问题。

另一种比较典型的平均法是指平滑法 (exponential smoothing, ES), 该方法给不同历史观测值分配不同的权重, 较近期的观测值的权重比较远期的观测值权重大。依据平滑次数不同, 指数平滑法可分为一次指数平滑法、二次指数平滑法和三次指数平滑法。后两者均是在前者的基础上进行再次平滑, 故这里只介绍一次指数平滑法在时空序列预测中的应用。将时空序列按 p 个区域划分为 p 条时间序列, 对于每条时间序列分别用指数平滑法进行预测。一次指数平滑法的定义为

$$X_{t+1} = \alpha v_t + (1 - \alpha) X_t \quad (2)$$

其中: X_t 为 t 时刻一次指数平滑法的预测值; v_t 为 t 时刻的观测值; $\alpha \in [0, 1]$ 为平滑常数。

2.2 基于差分自回归移动平均模型的时空序列预测

Box 等人^[34] 在 20 世纪 60 年代提出了一种比较流行的 Box-Jenkins 方法用于时间序列的分析、预测和控制, 该方法首先确定适当的参数值, 然后通过参数估计得出模型中具体的参数值, 最后检验所拟合模型的适当性并且适当修正该模型。ARIMA^[35] 是一种经典的时序预测模型, 全称是差分自回归移动平均模型 (autoregressive integrated moving average model, ARIMA)。在用 ARIMA 模型对时空序列预测的过程中, 模型对目标空间 A 每个位置对应的时间序列分别进行预测, 最后形成对时空序列的预测结果。然而 ARIMA 只能较好地预测具有线性特征的平稳序列, 无法挖掘出时空序列中的非线性特征。本节将介绍 ARIMA 模型的建模过程。

2.2.1 时间平稳性及其检验

ARIMA 等传统方法对时空序列预测时首先需要对序列进行平稳性检验。时间序列的平稳性定义为

假设某个时间序列是由某一个随机过程生成的, 如果满足下列三个条件则称该随机时间序列是弱平稳的 (weakly stable): a) 期望 $E(X_t) = \mu$ 是常数, 与时间 t 无关; b) 方差 $\text{Var}(X_t) = \sigma^2$ 是常数, 与时间 t 无关; c) 协方差 $\text{Cov}(X_t, X_{t+k}) = \gamma_k$ 是常数, 与时间间隔 k 有关, 与时间 t 无关。

检验时空序列平稳性的方法通常有图形检验法、自相关函数检验法、单位根检验法等。对于非平稳的时空序列需要使用差分方法使其变为平稳的序列。对于时间序列 $\{X_1, X_2, \dots, X_n\}$, 其一阶差分为

$$\nabla X_t = X_t - X_{t-1} \quad (3)$$

二阶差分为

$$\nabla^2 X_t = \nabla(\nabla X_t) = X_t - 2X_{t-1} + X_{t-2} \quad (4)$$

如图 3 所示, 曲线 1 基本上不存在趋势, 随机波动不存在某种规律, 是平稳时间序列。常见的平稳时间序列有白噪声序列等。曲线 2 呈现出很明显的上升趋势, 是非平稳时间序列。常见的非平稳时间序列有 GDP、物价指数等。

2.2.2 空间平稳性

若 N 个空间位置相互独立的点 z_1, z_2, \dots, z_N 的值是一组联合分布的随机变量, 即点 z_1, z_2, \dots, z_N 为联合概率分布函数 $p(z_1, z_2, \dots, z_N)$ 的某一特定结果, 则空间平稳性可定义为

若空间过程 z_i 是平稳的, 则对空间内 N 个位置中的任意点 i 和点 $i+h$, 都有

$$p(z_i | z_1, z_2, \dots, z_N) = p(z_{i+h} | z_1, z_2, \dots, z_N) \quad (5)$$

2.2.3 时空平稳性

若 $z_i(t)$ 为时空变量, 其中 i 为空间位置, t 为时刻且 ($i = 1, 2, \dots, N; t = 1, 2, \dots, T$), $M_{z_i(t)} = E[z_i(t)]$ 为 $z_i(t)$ 的期望, 则结合时间平稳性和空间平稳性可得出:

如果 $M_{z_i(t)}$ 是平稳的, 则对空间内 N 个位置中的任意点 i 和点 $i+h$ 以及时间 t 和 $t+k$, 都有

$$E[z_i(t)] = E[z_{i+h}(t+k)] \quad (6)$$

即时空变量的期望与时刻和空间位置均无关, 且时空变量的方差和协方差均为与时间、空间无关的常量。这样的时空变量形成的时空序列是平稳的。

2.2.4 模型公式

ARIMA(p, d, q) 模型有三个参数, 参数 p 为自回归项 (autoregressive, AR), 是时序数据滞后的阶数, 即模型的最高阶数; 参数 d 为差分阶数, 即时序数据需要进行 d 次差分才能转换为平稳时空序列; 参数 q 为移动平均项 (moving average, MA), 为了使模型的序列变得光滑。整个模型公式为

$$\begin{aligned} X_t &= \sum_{i=1}^p \alpha_i X_{t-i} + \sum_{i=1}^q \beta_i X_{t-i} + \varepsilon_t + c, \quad \alpha_p \neq 0 \\ \text{s. t. } E(\varepsilon_t) &= 0, \text{Var}(\varepsilon_t) = \sigma^2, E(\varepsilon_t \varepsilon_s) = 0, t \neq s \\ E(X_s \varepsilon_t) &= 0, \forall s < t \end{aligned} \quad (7)$$

其中: $\{\alpha_i\}, \{\beta_i\}$ 为模型参数, $\{\varepsilon_t\}$ 均值为 0, 是随机干扰序列, c 为常数, $\alpha_p \neq 0$ 保证了模型的最高阶数为 p , 即只通过前面连续 p 个时刻的历史观测值来预测下一个点的值, $E(X_s \varepsilon_t) = 0, \forall s < t$ 表明当前的随机干扰值和之前的序列无关。

ARIMA 模型的一些特殊形式如下:

- 当 $p=0, d=1, q=0$ 时, ARIMA(0, 1, 0) 为随机游走模型 (random walk), 研究对象每一个时刻的状态是在上一个时刻状态的基础上随机变化, 类似于布朗运动。
- 当 $d=0, q=0$ 时, ARIMA($p, 0, 0$) 为 p 阶自回归模型, 模型由前面 p 个连续时刻研究对象的状态递归推导出当前时刻状态。
- 当 $p=0, d=0$ 时, ARIMA(0, 0, q) 为 q 阶移动平均模型, 模型是历史白噪声的线性组合, 常用于刻画预期之外的影响。

2.3 基于时空自回归差分移动平均模型的时空序列预测

ARIMA 模型仅仅考虑了时空序列的时间维度的特征, 而忽略了空间相关关系, 因而预测准确度难以令人满意。在 ARIMA 的基础上, Martin 等人^[36] 首次提出了时空自回归差分移动平均模型 (spatio-temporal autoregressive integrated moving average model, STARIMA), 他们使用差分方法将不平稳时空序列转换为平稳时空序列以便模型处理, 但 STARIMA 模型无法处理空间非平稳序列。

在求解 STARIMA 模型的过程中, 文献[37, 38] 提出了模型识别、参数估计和模型验证三步迭代建模过程。模型识别主要是根据时空自相关函数来确定模型的三个阶数参数; 参数估计则主要是在确定某个准则的情况下平衡预测误差和参数个数, 最后求解出模型的各个参数^[39]; 模型验证主要是检验参数估计以及检验残差序列的随机性^[40, 41]。

1981 年, Hooper 和 Hewings 将时间序列模型和时空序列模型进行了对比分析, 进一步研究了 STARIMA 模型的各个参数。文献[42]在交通流预测问题中使用了 STARIMA 模型, 得到了较好的结果, 但仍只能解决时空平稳的序列。文献[43]阐明了在不知道概率密度函数的情况下, STARIMA 模型的参数不易获得, 因而该文献采用了遗传算法估计 STARIMA 模型参数, 在波士顿东部犯罪逮捕事件预测中证明了该算法高效可行。STARIMA 的一般形式为

$$z_i(t) = \sum_{k=0}^p \sum_{h=1}^m \varphi_{kh} w^{(h)} z_i(t-k) - \varepsilon_i(t) + \sum_{k=0}^q \sum_{h=1}^n \theta_{kh} w^{(h)} \varepsilon_i(t-k) \quad (8)$$

其中: $z_i (i=1, 2, \dots, N)$ 代表空间变量 $z_i(t) (t=1, 2, \dots, T; i=1, 2, \dots, N)$ 表示时间序列, T 指时间序列的长度, N 是空间单元的个数, k 为时间延迟, h 为空间延迟, p 为自相关阶数, q 为移动平均阶数, μ^h 为空间权重矩阵, $\varepsilon_i(t)$ 为随机误差。

2.4 基于向量自回归模型的时空序列预测

时空序列比时间序列更为复杂, 目前研究的一般是一元时空序列。Sims 等人提出了向量自回归模型 (vector autoregressive model, VAR) [44-47], 该模型是由单变量自回归模型 (autoregressive model, AR) 拓展而来, 其中单变量变为了由多元时间序列变量组成的向量。将时空序列目标空间的每一个区域看做多元时间序列的一个维度, 则时空序列可以看成是多元时间序列预测的一种特例, 因此可以使用 VAR 模型进行预测。一般的 VAR(p) 模型的公式为

$$Y_t = c + \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \dots + \Pi_p Y_{t-p} + \mu_t \quad (9)$$

其中: Y_t 是 $N \times 1$ 阶时间序列列向量; c 为 $N \times 1$ 阶常数; $\Pi_1, \Pi_2, \dots, \Pi_p$ 均为 $N \times N$ 阶参数矩阵; $\mu_t \sim \text{IID}(0, \Omega)$ (IID: 独立同分布) 是 $N \times 1$ 阶随机误差列向量。

VAR 模型具有以下特点: a) VAR 模型变量的值除了依赖它自己的滞后值, 还依赖于其他变量的滞后值; b) 该模型需要估计的参数较多, 当样本容量较小时多数参数估计的精度较差, 故模型需要大量样本。

3 基于传统机器学习的时空序列预测方法

时空序列数据预测工作本质上与机器学习方法中的回归分析之间存在着紧密的联系。决策树、支持向量机和隐马尔可夫模型等传统机器学习方法在时空序列预测方面均取得了不错的效果。

3.1 基于决策树的时空序列预测

决策树 [48-50] (decision tree, DT) 是一种树结构形式的预测分析模型, 能够基于输入特征较好地捕获近乎随机的非线性特征。树中的每个节点表示某个对象, 分叉路径代表属性值, 叶节点代表类别。利用决策树来进行时空序列预测工作, 主要是针对给定的离散时空序列数据学习出决策树的结构, 这部分工作是构建整个决策树预测模型的基础, 其目的就是快速地挖掘出数据特征。Hal 等人 [51] 提出了一种集成了搜索和学习来解决复杂结构化预测问题的算法 SEARN (integrating search and learning), 对时空序列预测问题也是通过将其分解后再进行预测, 分解时着重于控制更广泛范围依赖产生的级联错误。Kim 等人 [52] 设计了一种能够学习出时空序列预测器的决策树框架, 通过滑动窗口分解, 将时空序列预测问题分解成许多固定长度的预测问题, 然后分别进行预测。在视觉语音合成问题 [53] 中, 该方法首先对时空序列预测问题使用滑动窗口分解, 之后使用决策树这种非参数基预测器分别进行预测, 最后得到预测结果。

3.2 基于时空支持向量回归的时空序列预测

Vapnik 等人在统计学习理论的基础上提出了支持向量机 (support vector machine, SVM) [54-56], 这种有监督学习方法。该方法通过核函数将原始的非线性数据映射到高维空间内使其变得线性可分, 从而将低维问题转换为易于处理的高维问题。因此支持向量机能够较好地解决非线性、高维度等问题, 具有较好的泛化性能。实验证明, SVM 在小样本上效果特别好。

王佳璆等人 [16] 首次将支持向量应用于时空序列预测中, 在支持向量机核函数的基础上提出了时空核函数 (spatio-temporal kernel function), 并使用时空核函数设计了时空支持向量回归模型 (spatio-temporal support vector regression, STSVR)。支持向量回归 (support vector regression, SVR) [57] 模型的模型

函数也是一个线性函数, 和线性回归不同的是 SVR 在线性函数两侧设置了一个间隔带, 对所有在间隔带内的样本均不计算其损失, 模型最后是通过最小化间隔带的宽度和总损失来最优化模型。因为将多维问题等价地转换为多个一维问题还未得到明确的证明, 且 SVR 训练比较耗时, 多个 SVR 进行训练时参数调试工作量太大, 故单输出的 SVR 算法不能解决多输出的时空序列预测问题, 所以先需要将 SVR 回归算法从单输出扩展到多输出 SVR 算法 [58]。在构建时空支持向量回归模型的过程中, 为了解决时空序列数据的时间非平稳性和空间非平稳性, 文献 [16] 将折扣最小平方方法 (discounted least squares, DLS) 应用于支持向量机中, 该方法对距离较近的训练数据分配较大的权重, 距离较远的训练数据分配较小的权重, 最后使用 STSVR 完成了对时空非平稳序列的建模。在测试过程中, 使用 STSVR 模型在东莞市 GDP 产值预测和我国 137 个气象站气温预测上均取得了比 ARIMA 和 STARIMA 更好的拟合效果。

3.3 基于隐马尔可夫模型的时空序列预测

隐马尔可夫模型 (hidden Markov model, HMM) [59, 60] 是由马尔可夫模型发展而来的一种统计模型, 它在马尔可夫模型的基础上增加了多个隐状态, 因而能够较好地起到预测作用。该模型在语音识别、词性标注、机器翻译等领域得到了广泛应用。模型的难点是从可观测状态的参数来确定隐含状态的参数, 最后进行分析得到模型。一个隐马尔可夫模型 λ 可以用一个三元组表示 [61], 即:

$$\lambda = (A, B, \Pi) \quad (10)$$

其中: A 为状态转移概率矩阵; B 为观测概率矩阵; Π 为初始状态概率向量。 Π 和 A 决定了隐藏的马尔可夫链状态序列, B 决定了如何从状态序列生成观测序列。通过前向概率、后向概率或者是两者结合的方法可以计算出某个观测序列的概率。

在传统的时空序列预测方法中, ARIMA 模型和 STARIMA 模型只适合对平稳的时空序列进行预测, 而难以处理空间不平稳的时空序列。STSVR 模型虽然预测准确度不错, 但是对历史样本依赖程度较大, 且比较复杂。针对这个问题, 柳姣姣等人 [62] 提出了一种基于时空密度聚类的隐马尔可夫模型时空序列预测方法。该方法主要分为三步: 第一步使用了一种基于转折点的 PLR 方法 (piecewise linear representation, CP-PLR) 识别并提出序列中的噪声干扰, 得到时空序列的分段线性表示; 第二步使用 ST-DBSCAN (spatio-temporal density-based spatial clustering of applications with noise) 方法对分段后的时空序列数据集进行聚类获得几个隐含状态 (ST-DBSCAN 是 Birant 等人 [63] 在经典的密度聚类算法 DBSCAN [64] 基础上进一步考虑了时间维度, 发明的一种基于密度的时空聚类方法); 最后一步使用隐马尔可夫模型将时空序列预测问题转换为状态预测问题, 将得到的预测最优的状态序列作为时空序列预测结果。在时空序列预测中使用隐马尔可夫模型, 分析出数据在时间和空间上存在相关性的同时还可以用分段方法有效地去除数据噪声, 增加预测的准确度。文献 [62] 在药品冷藏库中的时空序列温度数据预测问题上验证了该模型的有效性。

4 基于深度学习的时空序列预测方法

近年来随着基于数据驱动的深度学习的兴起, 各领域海量的数据得到了更有效的利用。深度学习方法在语音识别 [64-67]、计算机视觉 [68-70] 和自然语言处理 [71, 72] 等领域都取得了颠覆性的成果。在时空序列预测领域中, 基于模型驱动的传统方法只能处理平稳的时空序列数据, 在对非平稳时空序列建模时需要先将其转换为平稳时空序列或者从中提取出非线性非平稳的时空趋势, 这使得建模过程较为复杂。同时, 传

统方法并不能捕获时空序列中隐藏的非线性特征,因而不适合处理非线性时空序列。基于深度学习的方法能够更有效地利用数据,充分挖掘出隐藏在时空序列数据中的线性和非线性特征。本章主要介绍人工神经网络、卷积网络和残差网络等深度学习在时空序列中的应用。2008 年,王佳璆等人^[16]从静态和动态神经元原理出发,提出了时空神经元模型,并在此基础上建立了基于深度学习的时空序列预测模型时空神经网络(spatio-temporal artificial neural network, STANN),该模型是一种多输入多输出(multiple-input multiple-output, MIMO)的动态神经网络模型。先验知识可以提高学习精度^[73],通过增加空间邻接性等先验知识,STANN 不再只依赖训练数据,该模型基本单元时空神经元 $ST_n(m, p)$ 定义为

$$\hat{z}_i(t) = \sum_{j=1}^n iw_{ji}^{(m)} \cdot z_j(t-1) + lw^{(p)} \cdot \hat{z}_i(t-1) + b \quad (11)$$

其中: i 为目标空间单元; j 为与 i 有空间相关性的邻域单元; z_1, z_2, \dots, z_j 为输入变量; \hat{z}_i 为输出值; $iw_{ji}^{(m)}$ 代表 m 阶延迟邻域的空间连接权重; $lw^{(p)}$ 代表 p 阶时间延迟的反馈权重; b 表示阈值。一个 STANN 网络结构拓扑图如图 4 所示。

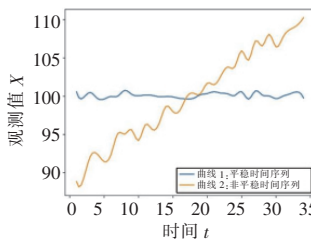


图 3 平稳时间序列和不平稳时间序列对比

Fig. 3 Comparison between stationary time series and non-stationary time series

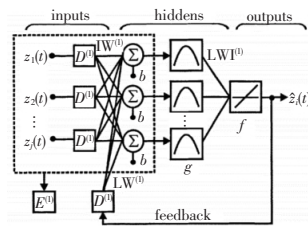


图 4 时空神经网络 STANN 结构拓扑图(引自文献[16])

Fig. 4 STANN structure topology of spatio-temporal neural network

图 4 所示的模型中,有 j 个输入和 i 个输出; $\hat{z}_1, \hat{z}_2, \dots, \hat{z}_i$ 为要预测的空间单元; z_1, z_2, \dots, z_j 为与 i 有空间相关性的邻域空间单元; D 是 1 阶时间延迟算子; E 是 1 阶空间延迟算子。模型中 p 期时间延迟操作是通过增加时间延迟算子 D 的个数实现的, m 期空间延迟操作则是通过增加输入节点的数目来实现。由此可以得到 STANN 模型的表达式为

$$z(t) = \begin{bmatrix} z_1(t) \\ z_2(t) \\ \vdots \\ z_i(t) \end{bmatrix}$$

$$\begin{bmatrix} f_1(\sum_{j=1}^n lw_{ji}^{(1)} \cdot g_j(\sum_{j=1}^n iw_{ji}^{(1)} \cdot z_j(t-1) + lw_1^{(1)} \cdot \hat{z}_i(t-1) + b_j)) + b_1 \\ f_2(\sum_{j=1}^n lw_{ji}^{(1)} \cdot g_j(\sum_{j=1}^n iw_{ji}^{(1)} \cdot z_j(t-1) + lw_2^{(1)} \cdot \hat{z}_i(t-1) + b_j)) + b_2 \\ \vdots \\ f_N(\sum_{j=1}^n lw_{ji}^{(1)} \cdot g_j(\sum_{j=1}^n iw_{ji}^{(1)} \cdot z_j(t-1) + lw_N^{(1)} \cdot \hat{z}_i(t-1) + b_j)) + b_N \end{bmatrix} \quad (12)$$

其中: f_1, f_2, \dots, f_N 为输出层传递函数; g_j 为隐含层传递函数; 其他参数和式(11)中定义相似。式(12)中每一个维度的输出都使用了加权回归,故 STANN 模型有解释局部空间变异的能力。STANN 这种动态神经网络模型能够通过学习自动调整适应空间权重,在时空序列预测上相比于 STARIMA 模型精度和泛化性能有了较大的提升。

Shi 等人^[32]在 2015 年提出了一种卷积长短期记忆模型 ConvLSTM,该模型对全连接长短期记忆模型(fully connected long short-term memory, FC-LSTM)进行拓展,将其中 input-to-state 和 state-to-state 部分由前馈式计算替换成卷积的形式,有效地解决了 LSTM^[74]在预测时空数据时出现的冗余问题。Zhang 等人^[75]在 2016 年提出了一种基于深度网络的时空序列数据预

测模型 DeepST 来解决交通流预测问题。DeepST 模型中,取三个不同大小的时间间隔从小到大将时空序列数据等间隔分为时间邻近(temporal closeness)、周期性(period)、季节性(season)三个不同的子序列,对每个子序列进行卷积操作后再融合,融合后再进行卷积操作,然后将元数据通过多个全连接层提取出的特征融合到一起作为最终的预测结果。DeepST 模型能够同时提取出时间和空间特征,相比 ARIMA、STARIMA、CNN 等模型预测准确度有了较大的提升。在 DeepST 基础上, Zhang 等人^[27]在 2017 年提出了一个基于时空残差网络的模型 ST-ResNet 来预测一个城市中各个区域某个时刻的人群流入量和流出量。ST-ResNet 模型将文献[75]中的 DeepST 中的卷积模块替换为残差卷积模块,对这三个子序列特征每个都设计一个残差卷积模块,这三个模块结构相同。模型最后动态地将这三个模块的输出进行加权后聚合,然后用全连接层提取出外部因素包含的特征并与聚合结果拼接起来形成最终结果。ST-ResNet 较好地利用了外部因素中的特征,比 DeepST 预测效果更好。

文献[76]提出了一种多视角时空神经网络模型(deep multi-view spatial-temporal network, DMVST-Net)来解决出租车需求预测这种时空序列预测问题。以前的方法主要是基于时间序列预测的,忽视了数据的空间特征;将时空序列视为一系列图像数据使用 CNN^[77]进行卷积后得到预测结果,虽然效果较好,但是没能同时考虑时间特征和空间特征。DMVST-Net 则充分考虑到这些情况,该模型首先将时空序列视为一系列图像,然后分别用卷积网络和全连接网络结合提取出空间特征,再使用 LSTM 进行预测进一步提取出时间特征,最后得到时空特征。对于用语义网络表示不同区域间的内在关系,进行加权和嵌入^[78]并映射到向量空间中,再通过全连接层提取出特征。将得到的特征和之前提取出的时空特征拼接起来通过全连接层后得到预测结果,最终把预测结果映射到真实数据中得到真实的预测值。通过对比实验,DMVST-Net 模型在出行预测问题上达到了当前最优效果。

深度学习中的其他技术,如注意力机制^[79,80]、迁移学习^[81~83]、图神经网络^[84,85]、多任务学习^[86,87]等在时空序列预测问题中也取得了较好的效果。Yao 等人^[31]针对城市交通量预测问题提出了一种 STDN 模型,该模型在用局部 CNN 提取数据空间特征后送入 LSTM 中后将注意力机制应用于 LSTM 中使得提取时序特征更有选择性。Wang 等人^[88]提出了 RegionTrans 这个跨城市迁移模型,通过建立区域级表示的映射关系,将在一个数据充足的城市中学到的时空序列模型迁移到数据匮乏的城市中,该模型解决了部分地区数据不平衡的问题。在此基础上, Yao 等人^[89]提出了 MetaST 模型,将文献[88]中迁移学习的源城市从一个拓展到多个,使得迁移的模型更加鲁棒和有效。Wang 等人^[90]提出了一个基于图卷积神经网络的框架 DST-GCNN 来进行交通量预测,该框架分为两个部分:第一部分用于提取用图表示的交通流特征;第二部分预测图结构的变化并将预测得到的图结构应用于交通流预测。该模型解决了交通流之间的图关系随时间变化的问题,在 METR-LA 和 TaxiBJ 这两个实际数据集上都取得了很好的效果。计算机视觉领域内各种技术的发展使得时空序列预测方法也得到了相应的进步。Niu 等人^[91]将图像分割问题中的 U-Net^[92]网络引入交通速度预测问题中,并和 LSTM 组合成 L-U-Net 网络,较好地预测了市内的交通状况。Du 等人^[93]提出了 DST-ICRL 模型,该模型将多通道流量表示和不规则卷积网络结合来学习交通数据中的空间特征,然后利用具有权重采样策略的 LSTM 模块来学习时间相关性,有效解决了交通人流量预测问题。

5 进一步研究方向

时空序列预测方法虽然有了较大的进步,但仍然存在很多问题。在传统时空序列预测方法中,模型的选取仍然依赖于特定数据集和个人经验,没有一种指导性的思路,并且这些传统方法难以捕获数据中动态的时空关系。在时空支持向量回归和深度学习等基于数据驱动模型中,模型的泛化能力是从历史样本中学出的,因而对训练数据的依赖程度较大。而时空序列数据往往较为复杂,经常受到各种因素影响,原始数据中存在较多的缺失和错误,这也影响了时空序列预测的精度^[94]。数据的处理和模型的设计与选取是提高时空序列预测精度的关键,因此时空数据预处理特别是数据去噪,尤为重要;对于某些数据较少的地区,直接进行时空序列建模得到的预测结果精度不高,需要将其他数据充足的地区学到的知识迁移过来。对于某些时空序列预测问题(如局部地区降雨的预测^[95]、风速^[20]等),预测的实时性很重要,因而需要建立一个能够实时进行更新的模型。基于深度学习的模型一般可解释性不足,因此对基于深度学习的时空序列预测模型进行可视化,然后理解它们是学习时空数据特征的过程也很重要。最后,本文进一步研究方向可总结如下:

a) 时空数据去噪和迁移学习。时空序列数据受各种可见和不可见因素的影响,具有高维、动态、多尺度和模糊等性质,因而数据中通常存在较多的错误样本,这极大地影响了各种模型的预测准确性。目前对于时空序列数据去除噪声部分的相关研究工作还比较少,可以考虑设计一定的算法对模型的训练样本进行筛选和修正,使得训练出来的模型更加精准。例如利

用滑动时间窗口的方法从原始数据中等间隔(如一天、一周等)地抽取数据组成新的时空序列^[74-75],使得时空数据的周期性变得更易于抽取。对于数据缺乏的城市,利用迁移学习可以从一个或多个有充足数据的城市中学到的时空序列预测模型迁移过去,从而解决数据匮乏的问题。

b) 时空序列在线预测和模型简化。本文主要讨论的是基于批量学习的时空序列预测方法,而实际应用中一般对时空序列预测的实时性比较高。在后续的工作中,可以考虑将在线学习和时空序列预测方法结合起来,使得模型既能够实时快速地更新模型,又能够在线快速反馈预测结果。此外,时空序列预测模型往往比较复杂,模型参数较多^[96],在保持预测精确度满足要求的前提下对模型进行压缩和简化也是一个值得研究的方向。

c) 时空序列模型的可解释性。基于深度学习的时空序列预测方法中,在提取其他影响因素特征时,除了直接使用卷积、循环神经网络、全连接提取特征和使用语义网络分析外,是否还有其他的方法进行提取,这些提取方法背后是否有对应的逻辑解释,这都是值得考虑的问题。通过对模型进行可视化,深入了解模型提取时空序列数据特征的过程,进而可以更好地解释该模型。

6 结束语

本文首先介绍了时空序列的定义,阐明了时空序列数据的特点。随后分三个方面分别介绍了时空序列预测方法,包括传统的时空序列预测方法、基于传统机器学习的时空序列预测方法以及基于深度学习的时空序列预测方法。文中提及的各种时空序列预测方法总结如表2所示。

表2 时空序列预测方法总结
Tab.2 Summary of spatiotemporal sequence prediction methods

类别	方法名称	特点
传统方法	HA ^[97]	简单,易于部署,精度低
	ES ^[98]	依据时间间隔的远近给数据分配不同权重,进行滑动平均
	ARIMA ^[35]	经典时序预测模型,将时空序列视为多条不相关的时间序列,未考虑空间相关性
	STARIMA ^[36]	ARIMA模型在空间上的扩展,进一步考虑了数据的时空相关性
	VAR ^[44]	自回归模型的推广,用于估计联合内生变量的动态关系,将时空数据的多个地点的观测值视为一个向量
传统机器学习方法	DT ^[52]	使用滑动窗口分解,将问题分解为固定长度的子问题用决策树进行预测
	STSVR ^[16]	在支持向量回归的基础上,依据数据距离不同分配不同权重
	HMM ^[62]	使用隐马尔可夫模型去除数据噪声,将时空序列预测转换为状态预测
	STANN ^[16]	通过定义时空算子,构建出首个时空神经网络
	ConvLSTM ^[32]	将卷积操作融入LSTM中,利用LSTM挖掘区域间的动态联系
深度学习方法	DeepST ^[75]	取不同时间间隔(一小时、一天、一周)的数据共同建模,挖掘出数据中的周期性
	ST-ResNet ^[27]	在DeepST的基础上,加入了残差网络使得
	DMVST ^[76]	从时间、空间、语义三个视角建模
	STDN ^[31]	在LSTM中加入了注意力机制,对时间间隔不同的数据分配不同的注意力
	RegionTrans ^[86-88]	使用了迁移学习,把在数据充足的城市学到的模型迁移到数据不足的区域
	MetaST ^[89]	使用了多源迁移学习,把多个模型从多个源城市迁移到一个目标城市
	DST-GCNN ^[90]	使用了图卷积网络,有效地学习到了时空数据间的动态变化
	L-U-Net ^[91]	将U-Net和LSTM相结合
	DST-ICRL ^[93]	将多通道流量表示、不规则卷积网络和带权重采样的LSTM结合

时空序列预测方法在股票预测、交通流预测和气候预测等方面均取得了很好的效果。随着时空序列数据规模的急剧增长,如何处理大量的数据并进行去噪,如何将时空序列预测方法和在线学习相结合来处理时空大数据,如何将复杂的时空序列预测模型进行简化,这些都是本文下一步要考虑的问题。

参考文献:

- [1] 罗军舟,吴文甲,杨明.移动互联网:终端、网络与服务[J].计算机学报,2011,34(11):2029-2051. (Luo Junzhou,Wu Wenjia,Yang Ming. Mobile Internet: terminals, networks and services [J]. Chinese Journal of Computers, 2011, 34(11): 2029-2051.)
- [2] 李志刚,周兴社.传感器网络[J].计算机应用研究,2004,21(12):9-12. (Li Zhigang,Zhou Xingshe. Sensor network [J]. Appli-

cation Research of Computers, 2004, 21(12): 9-12.)

- [3] Hofmann-Wellenhof B, Lichtenegger H, Collins J. Global positioning system: theory and practice [M]. Berlin: Springer Science & Business Media, 2012.
- [4] Olaiya F, Adeyemo A B. Application of data mining techniques in weather prediction and climate change studies [J]. International Journal of Information Engineering and Electronic Business, 2012, 4(1): 51-59.
- [5] Li Yaguang, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting [EB/OL]. (2017-07-06). [2018-02-22]. https://arxiv.org/abs/1707.01926v3.
- [6] Radzuan N F M, Othman Z, Bakar A A. Uncertain time series in weather prediction [J]. Procedia Technology, 2013, 11: 557-564.

- [7] Lin Yupei, Yang Yiwen. Stock markets forecasting based on fuzzy time series model [C]//Proc of IEEE International Conference on Intelligent Computing and Intelligent Systems. Piscataway, NJ: IEEE Press, 2009: 782-786.
- [8] Kim K. Financial time series forecasting using support vector machines [J]. *Neurocomputing* 2003 55(1-2): 307-319.
- [9] Li C, Hu J W. A new ARIMA-based neuro-fuzzy approach and swarm intelligence for time series forecasting [J]. *Engineering Applications of Artificial Intelligence* 2012 25(2): 295-308.
- [10] 王艳辉, 贾利民, 王卓, 等. 基于时空序列的铁路客流量数据挖掘预测 [J]. *中国铁道科学* 2005 26(4): 130-135. (Wang Yanhui, Jia Limin, Wang Zhuo, et al. Data mining and prediction of railway passenger volume based on spatiotemporal sequence [J]. *China Railway Science* 2005 26(4): 130-135.)
- [11] Chen Yuanfang, Shu Lei, Wang Lei. Traffic flow prediction with big data: a deep learning based time series model [C]//Proc of IEEE Conference on Computer Communications Workshops. Piscataway, NJ: IEEE Press 2017: 1010-1011.
- [12] Polson N G, Sokolov V O. Deep learning for short-term traffic flow prediction [J]. *Transportation Research Part C: Emerging Technologies* 2017 79(2): 1-17.
- [13] Cao Liying, San Xiaohui, Zhao Yueling, et al. The application of the spatio-temporal data mining algorithm in maize yield prediction [J]. *Mathematical and Computer Modelling* 2013 58(3-4): 507-513.
- [14] Das M, Ghosh S K. A probabilistic approach for weather forecast using spatio-temporal inter-relationships among climate variables [C]//Proc of the 9th International Conference on Industrial and Information Systems. Piscataway, NJ: IEEE Press 2014: 1-6.
- [15] 郑姗姗. 城市降雨积水时空序列分析与短时预测的初步研究 [D]. 北京: 中国科学院大学 2014. (Zheng Shanshan. Spatial-temporal sequence analysis of urban rainfall and water accumulation and preliminary study of short-term prediction [D]. Beijing: University of Chinese Academy of Sciences 2014.)
- [16] 王佳璐, 邓敏, 程涛. 时空序列数据分析和建模 [M]. 北京: 科学出版社 2012. (Wang Jiaqu, Deng Min, Cheng Tao. Spatio-temporal sequence data analysis and modeling [M]. Beijing: Science Press, 2012.)
- [17] Kendall M, Wold H. A study in the analysis of stationary time series [J]. *Journal of the Royal Statistical Society Series A (General)* 1954 117(4): 484.
- [18] 原继东, 王志海. 时间序列的表示与分类算法综述 [J]. *计算机科学* 2015 42(3): 1-7. (Yuan Jidong, Wang Zhihai. Review of time series representation and classification techniques [J]. *Computer Science* 2015 42(3): 1-7.)
- [19] Cliff A D, Ord J K. Space-time modelling with an application to regional forecasting [J]. *Transactions of the Institute of British Geographers* 1975(64): 119-128.
- [20] Zhu Qiaomu, Chen Jinfu, Zhu Lin, et al. Wind speed prediction with spatio-temporal correlation: a deep learning approach [J]. *Energies* 2018 11(4): 705.
- [21] Zhu Qiaomu, Chen Jinfu, Shi Dongyuan, et al. Learning temporal and spatial correlations jointly: a unified framework for wind speed prediction [J]. *IEEE Trans on Sustainable Energy* 2019(99): 1.
- [22] Shi Xingjian, Yeung D-Y. Machine learning for spatiotemporal sequence forecasting: a survey [EB/OL]. 2018-08-21. <https://arxiv.org/abs/1808.06865v1>.
- [23] Wang Dong, Cao Wei, Li Jian, et al. DeepSD: supply-demand prediction for online car-hailing services using deep neural networks [C]//Proc of the 33rd IEEE International Conference on Data Engineering. Piscataway, NJ: IEEE Press 2017: 243-254.
- [24] Bracewell R N, Bracewell R N. The Fourier transform and its applications [M]. New York: McGraw-Hill 1986.
- [25] Percival D B, Walden A T. Wavelet methods for time series analysis [M]. Cambridge: Cambridge University Press 2006.
- [26] 王家耀. 时空大数据及其在智慧城市中的应用 [J]. *卫星应用*, 2017(3): 10-17. (Wang Jiayao. Spatio-temporal data and its application in smart cities [J]. *Satellite Applications* 2017(3): 10-17.)
- [27] Zhang Junbo, Zheng Yu, Qi Dekang. Deep spatio-temporal residual networks for citywide crowd flows prediction [C]//Proc of the 31st AAAI Conference on Artificial Intelligence. Cambridge, MA: AAAI Press 2017.
- [28] Liang Xiaoyuan, Wang Guiling, Min M R, et al. A deep spatio-temporal fuzzy neural network for passenger demand prediction [C]//Proc of SIAM International Conference on Data Mining. 2019: 100-108.
- [29] 裴韬, 周成虎, 骆剑承, 等. 空间数据知识发现研究进展评述 [J]. *中国图象图形学报: A 辑* 2001 6(9): 854-860. (Pei Tao, Zhou Chenghu, Luo Jiancheng, et al. Progress in knowledge discovery of spatial data [J]. *Journal of Image and Graphics: Issue A* 2001 6(9): 854-860.)
- [30] 范明, 孟小峰. 数据挖掘概念与技术 [M]. 北京: 机械工业出版社, 2010. (Fan Ming, Meng Xiaofeng. Data mining concept and technology [M]. Beijing: Machinery Industry Press 2010.)
- [31] Yao Huaxiu, Tang Xianfeng, Wei Hua, et al. Revisiting spatial-temporal similarity: a deep learning framework for traffic prediction [C]//Proc of AAAI Conference on Artificial Intelligence. 2019.
- [32] Shi Xingjian, Chen Zhourong, Wang Hao, et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting [C]//Advances in Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc. 2015: 802-810.
- [33] Li Yaguang, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network: data-driven traffic forecasting [EB/OL]. (2017-07-06). [2018-02-22]. <https://arxiv.org/abs/1707.01926>.
- [34] Box G E P, Jenkins G M, Reinsel G C, et al. Time series analysis: forecasting and control [M]. Hoboken: Wiley 2015.
- [35] Dickey D. Time series theory and methods [J]. *Technometrics*, 1989 31(1): article No. 121.
- [36] Martin R L, Oeppen J E. The identification of regional forecasting models using space: time correlation functions [J]. *Transactions of the Institute of British Geographers* 1975(66): 95-118.
- [37] Pfeifer P E, Deutch S J. A three-stage iterative procedure for space-time modeling [J]. *Technometrics* 1980 22(1): 35-47.
- [38] Deutsch P S J. A STARIMA model-building procedure with application to description and regional forecasting [J]. *Transactions of the Institute of British Geographers (New Series)* 1980 5(3): 330-349.
- [39] Anselin L. Spatial econometrics: methods and models [M]. Berlin: Springer Science & Business Media 2013.
- [40] Cressie N. Statistics for spatial data [J]. *Terra Nova* 1992 4(5): 613-617.
- [41] Anselin L, Bera A K. Spatial dependence in linear regression models with an introduction to spatial econometrics [M]//Ullah A, Giles D. Handbook of Applied Economic Statistics. 1998: 237-289.
- [42] Kamarianakis Y, Prastacos P. Space-time modeling of traffic flow [J]. *Computers & Geosciences* 2005 31(2): 119-133.
- [43] Halim S, Bisoño I N, Sunyoto D, et al. Parameter estimation of space-time model using genetic algorithm [C]//Proc of IEEE International Conference on Industrial Engineering and Engineering Management. Piscataway, NJ: IEEE Press 2009: 1371-1375.
- [44] Christiano L J, Christopher A. Sims and vector autoregressions [J]. *Scandinavian Journal of Economics* 2012 114(4): 1082-1104.
- [45] Sims C A. Money, income, and causality [J]. *American Economic Review* 1972 62(4): 540-552.
- [46] Sims C A. Macroeconomics and Reality [J]. *Econometrica* 1980 48(1): 1-48.
- [47] Sims C A. A comparison of interwar and postwar cycles: monetarism reconsidered [J]. *American Economic Review* 1980 70(2): 250-257.
- [48] Quinlan J R. Induction on decision tree [J]. *Machine Learning*, 1986 1(1): 81-106.
- [49] Rokach L, Maimon O. Decision trees [M]//Data Mining and Knowledge Discovery Handbook. Boston: Springer 2005: 165-192.
- [50] Quinlan J R. Induction of decision trees [J]. *Machine Learning*, 1986 1(1): 81-106.
- [51] Hal D, Langford J, Marcu D. Search-based structured prediction [J]. *Machine Learning* 2009 75(3): 297-325.
- [52] Kim T, Yue Yisong, Taylor S, et al. A decision tree framework for spatiotemporal sequence prediction [C]//Proc of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press 2015: 577-586.
- [53] Taylor S L, Mahler M, Theobald B J, et al. Dynamic units of visual speech [C]//Proc of ACM SIGGRAPH/Eurographics Symposium on Computer Animation. New York: ACM Press 2012: 275-284.
- [54] Osuna E, Freund R, Girosi F. Support vector machines: training and applications, AIM-1602CBCL-144 [R]. 1997.

- [55] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32-42. (Zhang Xuegong. Introduction to statistical learning theory and support vector machines [J]. *Acta Automatica Sinica*, 2000 26(1): 32-42.)
- [56] Vapnik V. The nature of statistical learning theory [M]. Berlin: Springer Science & Business Media, 2013.
- [57] Drucker H, Burges C J C, Kaufman L, et al. Support vector regression machines[C]//Advances in Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 1997: 155-161.
- [58] 赵耀红, 钟萍, 王来生. 一种多输出支持向量机的增量学习算法[J]. 计算机应用与软件, 2010, 27(6): 14-16. (Zhao Yaohong, Zhong Ping, Wang Laisheng. An incremental learning algorithm for multi-output support vector machines [J]. *Computer Applications and Software* 2010 27(6): 14-16.)
- [59] Baum L E, Petrie T. Statistical inference for probabilistic functions of finite state Markov chains[J]. *The Annals of Mathematical Statistics*, 1966 37(6): 1554-1563.
- [60] Baum L E, Petrie T, Soules G, et al. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains[J]. *The Annals of Mathematical Statistics*, 1970 41(1): 164-171.
- [61] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012. (Li Hang. Statistical learning method [M]. Beijing: Tsinghua University Press, 2012.)
- [62] 柳姣姣, 禹素萍, 吴波, 等. 基于隐马尔可夫模型的时空序列预测方法[J]. 微型机与应用, 2016, 35(1): 74-76. (Liu Jiaojiao, Yu Suping, Wu Bo, et al. Spatio-temporal series prediction method based on hidden Markov model [J]. *Microcomputer Its Applications*, 2016 35(1): 74-76.)
- [63] Birant D, Kut A. ST-DBSCAN: an algorithm for clustering spatial-temporal data[J]. *Data & Knowledge Engineering* 2008 60(1): 208-221.
- [64] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//Proc of the 2nd International Conference on Knowledge Discovery and Data Mining. [S. l.]: AAAI Press, 1996: 226-231.
- [65] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks[C]//Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE Press, 2013: 6645-6649.
- [66] Sak H, Senior A, Beaufays F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling [EB/OL]. (2014-02-05). <http://arxiv.org/abs/1402.1128v1>.
- [67] Xiong W, Droppo J, Huang Xuedong, et al. Achieving human parity in conversational speech recognition [EB/OL]. (2016-10-14). [2017-02-17]. <https://arxiv.org/abs/1610.05256>.
- [68] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2012: 1097-1105.
- [69] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Trans on Pattern Analysis & Machine Intelligence* 2015 39(6): 1137-1149.
- [70] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. *IEEE Trans on Pattern Analysis & Machine Intelligence* 2014, 37(9): 1904-1916.
- [71] Bengio Y, Vincent P, Janvin C. A neural probabilistic language model [J]. *Journal of Machine Learning Research* 2003 3(6): 1137-1155.
- [72] Collobert R, Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning [C]//Proc of the 25th International Conference on Machine Learning. New York: ACM Press, 2008: 160-167.
- [73] Jozefowicz R, Vinyals O, Schuster M, et al. Exploring the limits of language modeling [EB/OL]. (2016-02-07). [2016-02-11]. <https://arxiv.org/abs/1602.02410>.
- [74] Hochreiter S, Schmidhuber J. Long short-term memory [J]. *Neural Computation* 1997 9(8): 1735-1780.
- [75] Zhang Junbo, Zheng Yu, Qi Dekang, et al. DNN-based prediction model for spatio-temporal data[C]//Proc of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM Press, 2016: 92.
- [76] Yao Huaxiu, Wu Fei, Ke Jintao, et al. Deep multi-view spatial-temporal network for taxi demand prediction [C]//Proc of International Conference on Artificial Intelligence. 2018: 2588-2595.
- [77] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [78] Mikolov T, Chen Kai, Corrado G, et al. Efficient estimation of word representations in vector space[C]//Proc of International Conference on Learning Representations. 2013: 1-12.
- [79] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//Advances in Neural Information Processing Systems. Piscataway, NJ: Curran Associates Inc., 2017: 5998-6008.
- [80] Mnih V, Heess N, Graves A. Recurrent models of visual attention [C]//Advances in Neural Information Processing Systems. Piscataway, NJ: Curran Associates Inc., 2014: 2204-2212.
- [81] Pan S J, Yang Qiang. A survey on transfer learning [J]. *IEEE Trans on Knowledge and Data Engineering* 2009 22(10): 1345-1359.
- [82] Long Mingsheng, Cao Yue, Wang Jianmin, et al. Learning transferable features with deep adaptation networks [EB/OL]. (2015-02-10). [2015-05-27]. <https://arxiv.org/abs/1502.02791>.
- [83] Yao Yi, Doretto G. Boosting for transfer learning with multiple sources [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2010: 1855-1862.
- [84] Battaglia P W, Hamrick J B, Bapst V, et al. Relational inductive biases, deep learning, and graph networks [EB/OL]. (2018-06-04). [2018-10-17]. <http://arxiv.org/abs/1806.01261>.
- [85] Zhou Jie, Cui Ganqu, Zhang Zhengyan, et al. Graph neural networks: a review of methods and applications [EB/OL]. (2018-12-20). [2019-07-10]. <https://arxiv.org/abs/1812.08434>.
- [86] Ruder S. An overview of multi-task learning in deep neural networks [EB/OL]. (2017-06-15). <https://arxiv.org/abs/1706.05098>.
- [87] Zhang Junbo, Zheng Yu, Sun Junkai, et al. Flow prediction in spatio-temporal networks based on multitask deep learning [J]. *IEEE Trans on Knowledge and Data Engineering* 2019 32(3): 468-478.
- [88] Wang Leye, Geng Xu, Ma Xiaojuan, et al. Cross-city transfer learning for deep spatio-temporal prediction [EB/OL]. (2018-02-01). [2018-05-19]. <https://arxiv.org/abs/1802.00386>.
- [89] Yao Huaxiu, Liu Yiding, Wei Ying, et al. Learning from multiple cities: a meta-learning approach for spatial-temporal prediction [EB/OL]. (2019-01-24). [2019-02-25]. <https://arxiv.org/abs/1901.08518>.
- [90] Wang Menglin, Lai Baisheng, Jin Zhongming, et al. Dynamic spatio-temporal graph-based CNN for traffic prediction [EB/OL]. (2018-12-05). [2020-03-05]. <https://arxiv.org/abs/1812.02019>.
- [91] Niu Kun, Zhang Huiyang, Zhou Tong, et al. A novel spatio-temporal model for city-scale traffic speed prediction [J]. *IEEE Access* 2019, 7: 30050-30057.
- [92] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C]//Proc of International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2015: 234-241.
- [93] Du Bowen, Peng Hao, Wang Senzhang, et al. Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction [J]. *IEEE Trans on Intelligent Transportation Systems* 2019(99): 1-14.
- [94] Deng Shaojiang, Jia Shuyuan, Chen Jing. Exploring spatial-temporal relations via deep convolutional neural networks for traffic flow prediction with incomplete data [J]. *Applied Soft Computing* 2019, 78: 712-721.
- [95] Shi Xingjian, Gao Zhihan, Lausen L, et al. Deep learning for precipitation nowcasting: a benchmark and a new model [C]//Advances in Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2017: 5617-5627.
- [96] Atluri G, Karpatne A, Kumar V. Spatio-temporal data mining: a survey of problems and methods [J]. *ACM Computing Surveys* 2018 51(4): 83.
- [97] Campbell J Y, Thompson S B. Predicting excess stock returns out of sample: can anything beat the historical average? [J]. *The Review of Financial Studies* 2007 21(4): 1509-1531.
- [98] Gardner Jr E S. Exponential smoothing: the state of the art [J]. *Journal of Forecasting* 1985 4(1): 1-28.