

# 基于数字孪生技术的元宇宙 空气污染物浓度推断模型

彭一非<sup>1</sup>, 袁贞<sup>1</sup>, 张旭龙<sup>2</sup>, 姜桂林<sup>3</sup>, 刘逾江<sup>4</sup>

1. 湖南财信数字科技有限公司, 湖南 长沙 410035;
2. 平安科技(深圳)有限公司, 广东 深圳 518063;
3. 湖南财信金融控股集团有限公司, 湖南 长沙 410035;
4. 墨尔本大学, 澳大利亚 墨尔本 3010

## 摘要

空气污染与人们的健康和经济社会的发展息息相关。然而, 监测站点分布稀疏, 无法提供细粒度的空气污染物浓度。此外, 现有的空气污染物浓度推断方法缺乏实时处理相关数据的能力, 具有滞后性。为了解决上述问题, 提出了一种基于数字孪生技术的元宇宙空气污染物浓度推断模型。该模型将现实数据映射到元宇宙空间中, 并构建数据仓库, 通过构建空气污染物特征库实现对空气污染物浓度的实时精确推断。实验结果表明, 该模型能提高空气污染物浓度推断的准确性和有效性。

## 关键词

空气污染; 数字孪生; 元宇宙

中图分类号: TP399

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2023005

## *Metaverse air pollutant concentration inference model based on digital twin technology*

PENG Yifei<sup>1</sup>, YUAN Zhen<sup>1</sup>, ZHANG Xulong<sup>2</sup>, JIANG Guilin<sup>3</sup>, LIU Yujiang<sup>4</sup>

1. Hunan Chasing Digital Technology Co., Ltd., Changsha 410035, China
2. Ping An Technology (Shenzhen) Co., Ltd., Shenzhen 518063, China
3. Hunan Chasing Financial Holdings Co., Ltd., Changsha 410035, China
4. The University of Melbourne, Melbourne 3010, Australia

## Abstract

Air pollution is closely related to people's health and economic and social development. However, monitoring sites are sparsely distributed and cannot provide fine-grained air pollutant concentrations. In addition, the existing air pollutant concentration inference methods lack the ability to process relevant data in real time, so they have a hysteresis. To solve the above problems, a metaverse air pollutant concentration inference model based on digital twin technology was proposed. The model mapped the real data into the metaverse space, and built a data warehouse to achieve real-time accurate inference

of air pollutant concentrations through the construction of an air pollutant feature library. The experimental results show that the model can improve the accuracy and validity of air pollutant concentration inference.

### Key words

air pollution, digital twin, metaverse

## 0 引言

随着工业文明的推进,煤炭、石油等化石燃料的大量使用给生产生活提供了源源不断的动力,同时也给人们带来了空气污染问题。空气污染严重危害人们的健康,研究<sup>[1]</sup>表明,长期暴露于受污染的空气中的人更容易患呼吸系统疾病和心血管疾病,即使空气污染水平很低。还有研究<sup>[2-3]</sup>发现,在典型空气污染地区,当大气污染物(如PM<sub>2.5</sub>、PM<sub>10</sub>、NO<sub>2</sub>、O<sub>3</sub>、CO等)排放量增加时,因卒中、房颤和高血压入院的风险也随之增加。空气污染不仅危害居民健康,也制约着我国经济的可持续发展。我国的《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》提出:“深入打好污染防治攻坚战,建立健全环境治理体系,推进精准、科学、依法、系统治污,协同推进减污降碳,不断改善空气、水环境质量,有效管控土壤污染风险。”与此同时,城市空气质量作为经济社会发展约束性指标之一,已经被明确列入《国家新型城镇化规划(2021—2035年)》<sup>[4]</sup>。因此,为了提高人们的生活质量,降低医疗成本,并给政府部门的监管提供理论依据,需要实现对空气污染物浓度的监测和推断。

国内外对空气污染的预报方法主要有3种:潜势预报、统计预报和数值预报<sup>[5]</sup>。使用上述方法需要翔实的气象同步观测数据,在实际应用中往往无法具有良好的时效性。研究人员曾提出细粒度空气污染物数

据监测方法,该方法能提高空气污染物数据推断的准确率和F1值<sup>[6]</sup>。然而,其仍无法满足获取细粒度空气污染物数据的时效性和有效性。

元宇宙的出现为实时细粒度空气污染物数据监测提供了着力点。相对于传统数字建模仿真,元宇宙中的数字孪生技术是动态、实时、双向、全生命周期的。因此,可以基于数字孪生技术在元宇宙平台建立一个空气污染监测模型来监控空气污染物浓度。

本文希望通过数字孪生技术构建接近真实场景的元宇宙空间。本文在该虚拟场景中采用机器学习方法,提出了一种空气污染物浓度推断模型来推断实时细粒度空气污染物数据,以满足居民需求,并达到元宇宙指导现实生活的目的。

## 1 相关研究与问题分析

本文围绕提出的以下两个科学问题进行相关工作的分析和论证。针对大城市中站点分布稀疏问题,本文对前人研究成果进行了总结,同时分析了其不足之处;针对无法快速获取实时细粒度空气污染物数据问题,本文通过引入元宇宙背景及介绍与数字孪生相关的技术来阐述在元宇宙虚拟场景中进行细粒度空气污染物数据推断的可行性,并分析与现实场景相比,元宇宙虚拟场景的优越性。接下来本文将介绍前人对细粒度空气污染物浓度推断问题的研究分析以及在元宇宙中部署空气污染物浓度推断模型的优势和可行性。

### 1.1 细粒度空气污染物浓度推断问题研究

针对空气质量推断问题,一些研究人员提出了统计方法,利用来自站点的有限空气质量数据来推断空气质量。Jutzeler A等人<sup>[7]</sup>采用一种基于区域的高斯过程模型估计城市空气污染扩散,该模型适用于暴露评估和异常检测。Xu Y N等人<sup>[8]</sup>使用基于张量分解的方法推断完整的空气质量值,这种算法可以有效减小噪声。此外,还有一些研究人员采用数据驱动的思想来解决这一问题。如Zheng Y等人<sup>[9]</sup>提出了一种基于协同训练框架的半监督学习方法,该框架包括基于人工神经网络(artificial neural network, ANN)的空间分类器和基于条件随机场(conditional random field, CRF)的时间分类器,通过协同训练好的模型来推断空气质量。相比统计模型,数据驱动模型可以学习到更多与空气质量有关的因素,能适用于更广泛的城市环境。

关于数据的特征构建,为了提高空气质量推断的准确性,研究人员一直在研究来自多个来源的数据的融合,他们采用了许多其他类型的城市数据,如兴趣点(points of interests, POI)和气象数据。Zheng Y等人<sup>[9]</sup>利用有限数量的现有监测站报告的(历史和实时)空气质量数据和在城市中观察到的各种数据集(如气象、交通流、人员流动、道路网络结构和POI),推断整个城市的实时和细粒度空气质量信息;Yu H M等人<sup>[10]</sup>建议使用城市监测站点的稀疏分布特征对大气污染物浓度进行推断;Wei J等人<sup>[11]</sup>引入卫星遥感数据,利用数据的时空特征,采用随机森林方法推断细粒度空气污染浓度。

对于某一种特定的污染物(如空气污染物数据)的推断,研究人员也尝试引

入气溶胶数据。气溶胶光学深度(aerosol optical depth, AOD)是气溶胶消光系数从地面到大气层顶的积分,表示无云大气铅直气柱中气溶胶散射造成的消光程度<sup>[12]</sup>。它们分布在从地球表面到大气顶部的空气柱中。有研究表明<sup>[13]</sup>,利用相对湿度和气溶胶标高等气象条件进行校正后的AOD与空气污染物具有高相关性。要想获得AOD值,就必须通过卫星上的传感器进行测量,因此,遥感作为在远离目标和非接触目标物体条件下探测目标地物的技术,将完成测量任务。然而,研究人员引用的数据覆盖范围过大,空间分辨率过低,无法完成细粒度空气污染物数据推断。实际上利用气溶胶数据能将空间分辨率提高到 $1 \text{ km}^2$ ,但研究人员并未采纳。

### 1.2 元宇宙场景问题

本节分析实现空气污染物浓度推断模型的场景问题。通过介绍元宇宙的概念和历史发展,详细阐述在元宇宙中部署空气污染物浓度推断模型的原因和优势。

元宇宙(metaverse)一词首次出现是在1992年一部名叫《雪崩》的科幻小说中,小说描述了一个平行于现实世界的平行宇宙。2021年3月,游戏公司Roblox在官方招股说明书上提出要打造早期元宇宙的想法;同年10月,脸书(Facebook)改名为Meta,并宣布将彻底转型为一家元宇宙公司。元宇宙不是网络游戏,也不仅仅是一个虚拟世界,它是人们依靠先进的计算机通信技术打造的一个具备新型社会体系的虚实相融的数字生活空间,是将多种新技术整合而形成的<sup>[14]</sup>。元宇宙关键技术主要包括:区块链技术、交互技术、5G网络技术、云计算技术、人工智能技术、数字孪生技术、物联网技术<sup>[15]</sup>。区块链技术为搭建经济系统提供支撑,实现元宇宙中的价值交

换<sup>[16]</sup>；交互技术为用户构建综合感官，提供元宇宙中的沉浸式体验；5G网络技术高速率、低延时、大连接的服务，为元宇宙提供实时流畅的体验；大容量、高效敏捷的云计算技术，为元宇宙提供强大的云端服务；人工智能技术实现元宇宙中的“智慧大脑”；数字孪生技术和物联网技术打造元宇宙中的虚实融合的综合环境。

在元宇宙世界构建空气污染物浓度系统有三大优势。优势一：能有效降低成本。一方面，可以搭建一个动态的数据仓库，让数据的获取变得容易，有效降低数据收集的成本；另一方面，元宇宙是一个自动运行的平台，可以减少人工干预，降低运营成本。优势二：这个系统是稳定的。在元宇宙世界中可以构建稳态特征的知识图谱，计算出的特征库权重是稳定不变的，因此这个系统具有很高的鲁棒性。优势三：能实现最接近现实世界的模拟场景。元宇宙世界是现实世界的映射，在元宇宙平台实现的场景模拟演练是最贴近现实的。

为了生成现实世界的镜像，需要借助数字孪生技术。数字孪生技术充分利用各种传感器大数据、数字模型，构建一个现实世界在虚拟空间中的整体反映的仿真空间<sup>[17]</sup>。由于通信技术的蓬勃发展，现已具备对物理仿真建模技术进行升级的条件，通过数字孪生技术来构建元宇宙虚拟场景。相比于传统场景仿真建模，数字孪生技术构建的虚拟场景是实时动态的，数字孪生体会反映本体的实时动态；数字孪生也是双向的，数字孪生体的数据可以反向传送给本体，用户可以根据数字孪生体的反馈采取相应措施；数字孪生是全生命周期的，它贯穿于产品的整个周期，包括设计、开发、制造、服务以及维护。数字孪生体借助物联网拥有了精准的感知能力，借助云计算拥有了强大的算力，借助先进通信网络拥有了实时响应能力。虽然数字孪

生和元宇宙在概念和技术上有诸多相似之处，但是元宇宙比数字孪生的持久性更强，是完全去中心化的，并且会一直与现实世界保持协同进化<sup>[15]</sup>。

以数字孪生技术为基础，可以构建元宇宙虚拟场景<sup>[18]</sup>。相比于现实生活中的空气污染物浓度监测系统，在元宇宙场景下，数据来源更加丰富，能融合更多维度的数据特征。由于元宇宙是全数字的，而且元宇宙是一个实时在线的系统，通过这些关联数据计算出的空气污染物浓度是实时准确的。此外，可以引入外部数据成果，这样就可以将计算模型解耦，优化系统性能。

在元宇宙场景中，用户借助虚拟现实（virtual reality, VR）眼镜可以根据人眼所看位置的图像，在元宇宙空间中重建虚拟图像，通过物联网快速采集到相应位置的有关计算数据，借助元宇宙虚拟计算推断细粒度空气污染物浓度数据，反馈给用户，以指导用户采取相关措施。

## 2 模型描述

根据上述问题分析，本文通过虚拟现实交互规则，设计了一个基于数字孪生技术的元宇宙空气污染物浓度推断模型BSLInf。该模型不仅能够获取多维度、多种类现实场景准确数据，其数据还具有高精度性，并最终快速反馈细粒度场景下的实时空气污染物浓度值。基于数字孪生技术的元宇宙空气污染物浓度推断模型框架示意图如图1所示，模型包含4个模块，自下而上依次为数据采集模块、孪生模型仿真模块、特征库构建模块和场景应用模块。各模块功能简要描述如下。

(1) 数据采集模块主要负责采集多维度、少样本的长时序空间数据，包括空气监测站通过传感器获得的各类污染物浓度

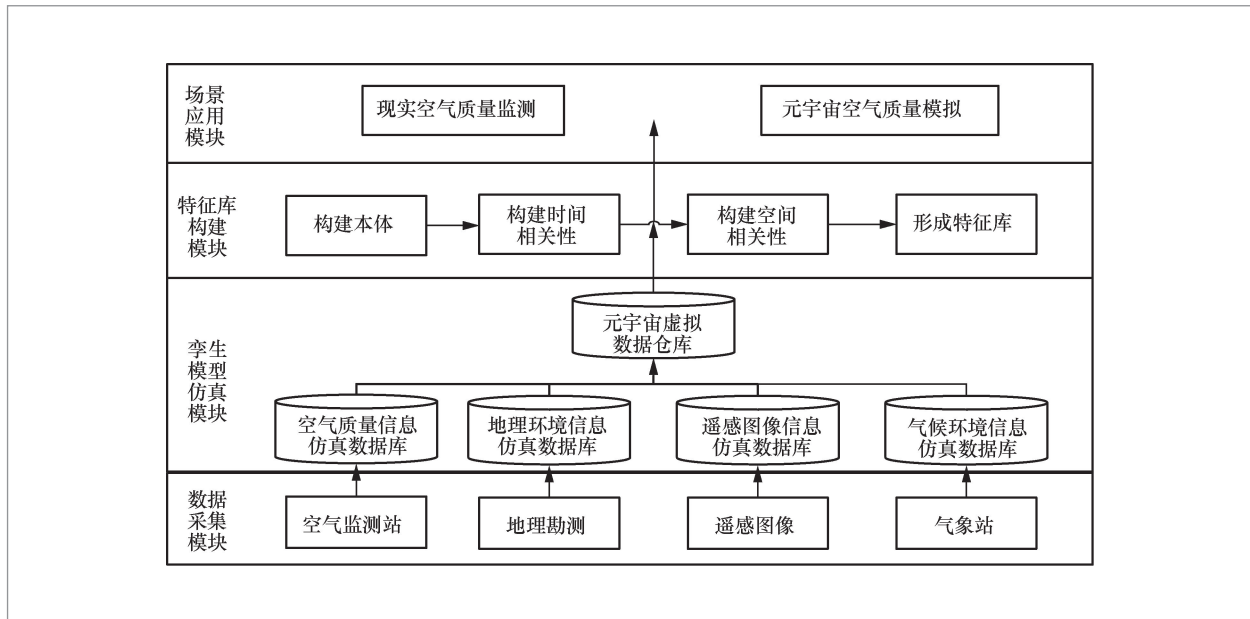


图1 基于数字孪生技术的元宇宙空气污染物浓度推断模型框架示意图

数据、地理勘测得到的地理环境数据、卫星采集的遥感图像数据和气象站获取的气象数据。

(2) 孪生模型仿真模块采用数字孪生建模方法，在数据管理平台、数据共享交换平台、地理信息系统 (geographic information system, GIS) 平台和云计算平台的支撑下，对数据采集模块获取的物理现实数据进行实时3D仿真建模，以形成空气质量信息仿真数据库、地理环境信息仿真数据库、遥感图像信息仿真数据库和气候环境信息仿真数据库，并最终在元宇宙平台搭建元宇宙虚拟数据仓库。

(3) 特征库构建模块旨在采用联机分析方法提取对空气污染物浓度影响大的特征数据，并最终构建空气污染物浓度特征库。其中，构建特征库的方法为：通过本体语言构建空气污染物浓度本体分类体系；采用机器学习模型迭代选取特征数据；在时间和空间两个维度上采用度量方法对选取出的特征数据进行评估；最终形成与空气污

物浓度相关的知识图谱，即特征库。

(4) 场景应用模块将场景数据与元宇宙虚拟数据匹配，结合特征库，通过空气污染物浓度推断模型实时显示所在区域的空气污染物浓度。

## 2.1 数据采集模块

为了搭建元宇宙虚拟数据仓库，首先需要对物理数据进行采集，该数据能为有效构建数据仓库提供支撑，是数字孪生模型实体数据的来源和数字孪生决策的作用对象，也是数字孪生模型构建的首要前提。在采集物理数据的过程中，本文收集了空气污染物浓度、地理环境信息、卫星遥感和气象站测量设备和传感器等工具获取的数据，并在城市中的任意地点随机获取多维度实时数据。接下来详细介绍物理数据的采集过程。

首先，本文获取国控空气质量监测站点选取的空气污染物数据，监测数据指标

为:  $\text{SO}_2$ 、 $\text{CO}$ 、 $\text{PM}_{2.5}$ 、 $\text{NO}_2$ 、 $\text{O}_3$ 、 $\text{PM}_{10}$ 的浓度, 时间单位为1 h。然后, 为了有效提升后续模块的准确性和鲁棒性, 本文从不同角度考虑各因素与空气污染物浓度的相关性以采集多维度数据, 并根据数据来源将所有影响因子分为3类: 地理因素、卫星遥感因素和气象因素, 即气溶胶、温度、湿度、风速、气压和植被覆盖率。本文从美国航空航天局(National Aeronautics and Space Administration, NASA)采集了公开数据集MCD19A2, 从欧洲中期天气预报中心(European Centre for Medium-Range Weather Forecasts, ECMWF)采集了公开数据集ERA-Land。其中, MCD19A2是陆地气溶胶光学深度网格2级产品, 其通过分辨率成像光谱仪(moderate-resolution imaging spectroradiometer, MODIS)中的Terra卫星和Aqua卫星组合并实施多角度大气校正算法(multi-angle implementation of atmospheric correction algorithm, MAIAC), 提供用于地表大气属性和视图几何形状的双向反射系数(bidirectional reflectance factor, BRF)或表面反射率, 其投影方式为正弦投影。MCD19A2产品能提供空间分辨率单位为 $1 \text{ km}^2$ 的气溶胶粒径, 时间单位为1天, 每天能提供4~6组数据; ERA5-Land是一个再分析数据集, 它通过重放ECMWF ERA5气候并利用物理定律将模型数据与来自世界各地的观测结果结合, 最终生成完整且可靠的数据集。ERA5-Land产品的数据内容包括温度、湿度、风速、气压和植被覆盖率, 其时间单位为1 h。

为了将各项多维度数据与空气污染物数据匹配, 本文对各项数据自带的经纬度进行匹配, 具体步骤如下。

步骤1: 对MCD19A2每日提供的4~6组数据进行加权平均。

步骤2: 将ERA5-Land数据的时间调

整为东八时, 并同样进行以日为单位的加权平均。

步骤3: 分别计算MCD19A2数据经纬度、ERA5-Land数据经纬度和国控空气质量监测站点数据经纬度的曼哈顿距离, 取最小值并形成节点-属性表结构。

本文采集的数据分布稀疏, 无须采集所有地区的空气污染物浓度和其他输入数据, 该采集方法不仅减少了人力成本和计算量, 还有效降低了各模块的过拟合风险。经过上述匹配操作后, 本文根据各站点的空间分布和时间不变性构建空气污染物节点时空特征拓扑结构图。现实世界样本拓扑结构如图2所示, 圆形为国控空气质量监测站点覆盖地区, 三角形为监测站点覆盖范围以外地区, 其空气污染物数据未知。本文希望利用细粒度输入数据来构建拓扑结构, 达到通过覆盖地区得到未覆盖地区在一定时间范围内的空气污染物数据的目的。

## 2.2 孪生模型仿真模块

成功采集物理数据后, 本文希望通过数字孪生技术将数据与环境映射到元宇宙的虚拟世界中, 并构建元宇宙虚拟数据仓库。通过在数据仓库中模拟仿真研究空气污染物与其他影响因子的权重关系, 指导现实世界应对紧急空气污染事件。

与传统仿真模型不同, 本文提出的孪生模型仿真模块通过不同地区的空气污染物浓度数据与输入数据对比, 使用循环一致性对抗生成方法, 分别设计仿真域生成器和现实域生成器。其中, 前者以现实域数据为输入, 以仿真域数据为输出; 后者以前者的输出为输入, 以现实域数据为输出, 两者形成循环生成结构。通过最小化现实域和循环生成的仿真域之间的距离, 学习不同空气污染物浓度数据和输入数据机理的域间迁移函数, 实现仿真数据库的高质量生成。

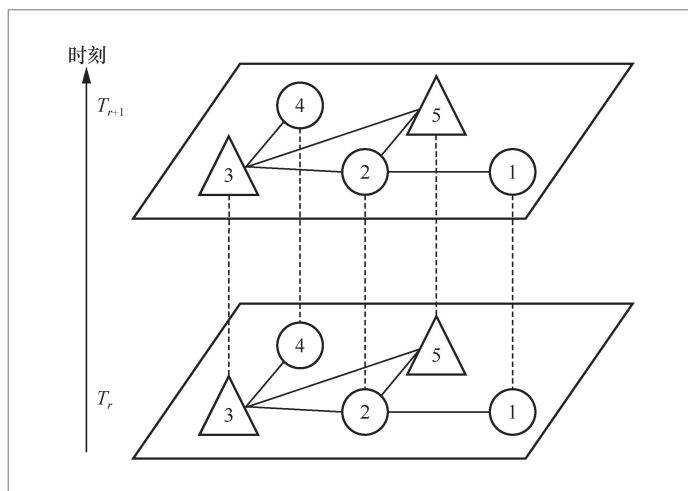


图2 现实世界样本拓扑结构

为了生成相同工况下的多维度稀疏仿真数据,整合生成孪生数据仓库,该模型基于三维重建技术,先将空气污染物浓度数据和输入数据映射到隐变量空间,形成场景在三维空间中不同气候和不同地理环境下的隐式表征,在约束关键隐式表征不变的前提下,使用渐进式生成器合成新的孪生特征数据,并用扩散模型实现新工况下的虚拟场景数据。生成的虚拟场景数据中可保留原现实场景的数据逻辑规则,但

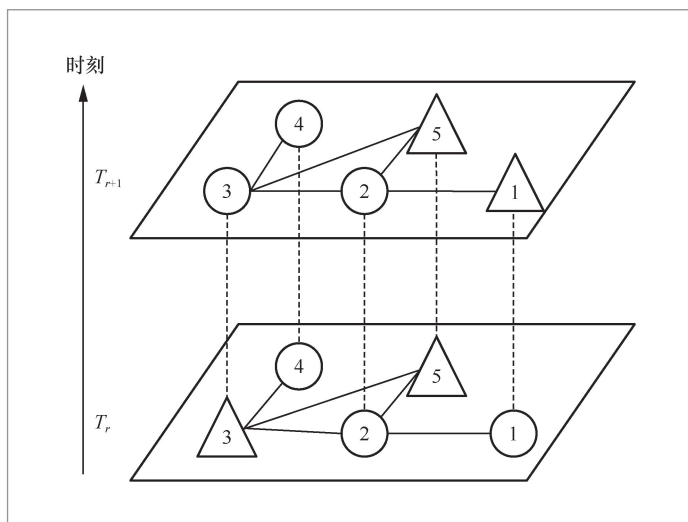


图3 元宇宙样本拓扑结构

具有不同影响系数,进而实现相同工况下的多维度数据生成。

为了迁移生成元宇宙场景下的虚拟数据仓库,模型采用神经辐射场方法,以神经辐射场输出的多维度稀疏仿真数据为输入,生成跨工况的连续数据。本文设计了一种探索-开发编码器,构建跨工况迁移生成模型,将空气污染物浓度数据和输入数据的高层语义变量映射到生成模型中的非规范化嵌入空间,通过初始规范化嵌入空间,允许模型在规范化嵌入空间中中等概率地生成多种工况的仿真数据样本,通过梯度更新约束模型,在开发阶段逐渐提升跨工况仿真数据仓库生成的概率,通过多尺度空间特征和语义保存变量约束渐进融合解码器,对跨工况的孪生数据仓库生成模型进行收敛,完成对孪生数据仓库的工况迁移,实现高质量的元宇宙虚拟场景数据生成。

将现实世界中采集的物理数据映射到元宇宙虚拟仓库中后,数据分布的规则将会更丰富,并有所变化,这是因为在元宇宙虚拟场景中,人们可在不同的时间段任意定义稀疏的采样点,即空气污染监测站点的位置,这些采样点的位置和数量既可固定,也可随着时间的推移非线性变化。元宇宙样本拓扑结构如图3所示,其中,圆形为空气污染物采样点,三角形的空气污染物数据未知。从图3中可知,采样点的位置不仅可不变化(如2号节点),还能随机调整(如1号节点、3号节点)。上述采样步骤不仅使元宇宙虚拟场景能使用鲁棒性更强的算法模型,而且将元宇宙与现实生活紧密联系在一起,使元宇宙具有更明显的引领作用。

### 2.3 特征库构建模块

该模块能够从元宇宙多维度稀疏虚拟数据仓库中训练学习各影响因子与空气污染物浓度的权重关系,从而构建空气污染

物浓度特征库。为了完成此目标,本文首先通过本体语言构建空气污染物浓度分类体系,然后采用机器学习方法学习各分类特征之间的权重关系,最后根据权重关系建立空气污染物浓度特征库。

在空气污染物数据推断问题中,现有的研究模型通过邻接矩阵初始化等方法创建空气污染物节点之间的相关系数矩阵,并将该矩阵用于描述所有时间戳范围内的节点空间相关性。然而,上述方法忽略了一点,其构建的静态相关系数矩阵在每一个时刻均相同,而各节点之间的空气污染物相关性会随着时间的推移非线性改变,因此,上述模型无法准确推断空气污染物数据。为了解决上述问题,本文从互联网中爬取和解析与空气污染物相关的描述数据,以及相关知识分类方法。本文采用SPARQL查询模型,该查询模型提供

了一种连接和查询来自各种来源的数据的方法,以便轻松搜索与信息图关联的本体描述框架数据。自动分析并分类数据后,本文通过Protege软件成功构建本体分类体系,空气污染物本体分类体系如图4所示。

构建本体分类体系后,本文提出了基于平衡交互长短期记忆(long-short term memory, LSTM)的空气污染物浓度推断模型BSLInf。BSLInf模型设计了交互池化模块,动态更新相邻节点的隐藏状态信息;创建了BMLoss损失函数,用于估计不同影响因子的空气污染物浓度影响权重。实验证明,与基准模型相比,BSLInf模型在特征库构建方面有更好的准确性和精确性。BSLInf模型包含时间LSTM和交互池化两大模块,其框架如图5所示。

对于时间LSTM模块,本文将整个研

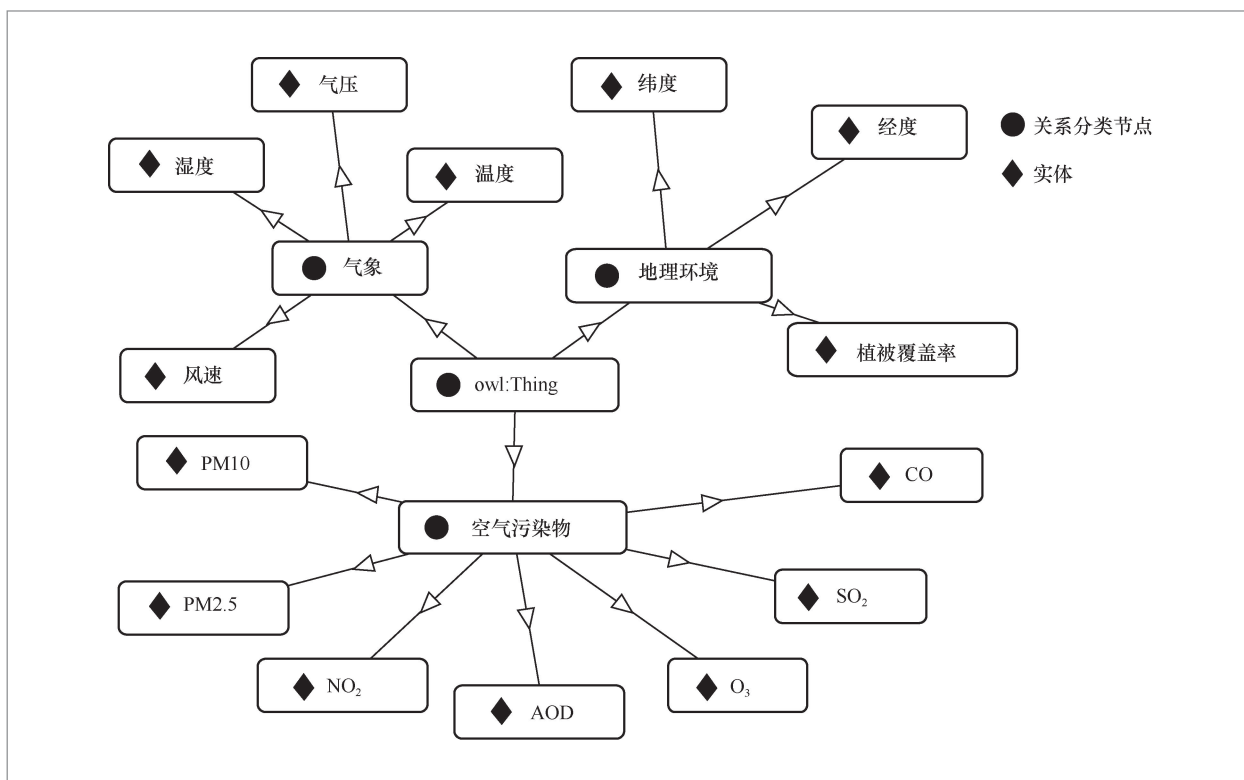


图4 空气污染物本体分类体系

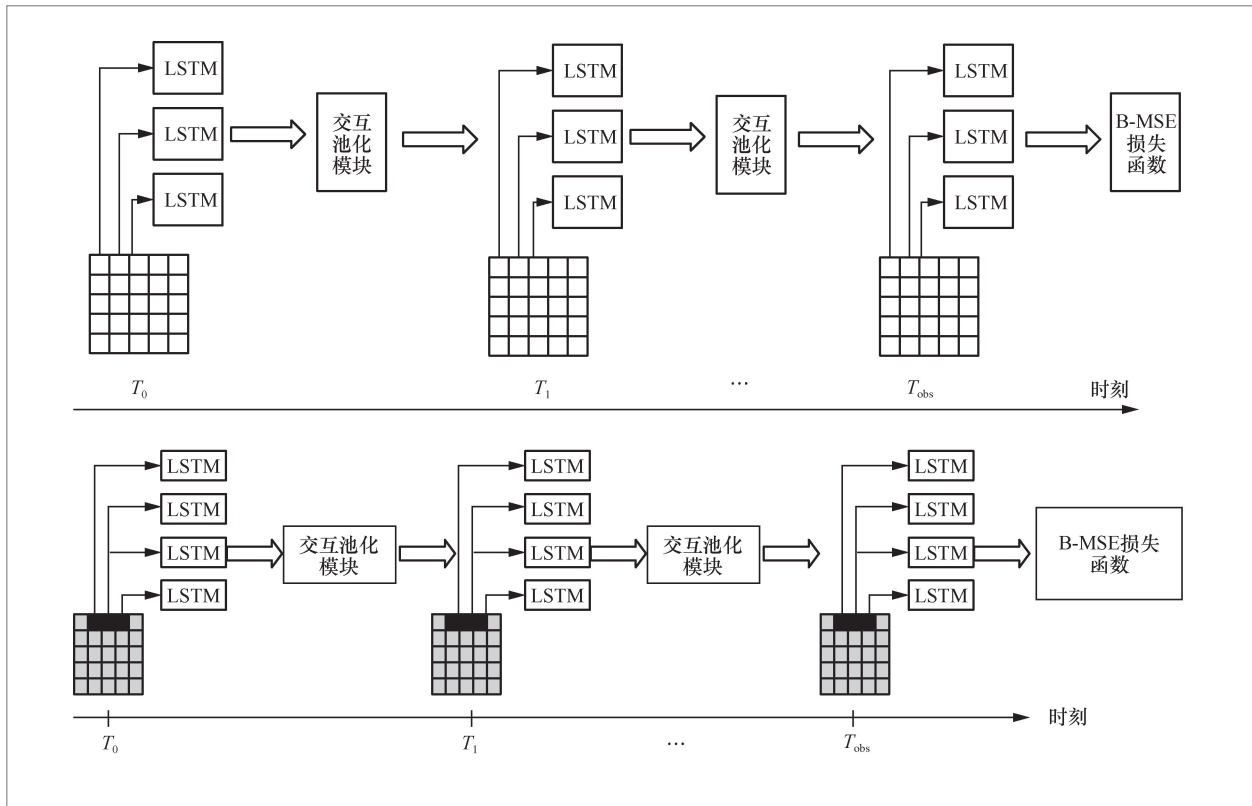


图5 BSLinf 模型框架

究区域划分为相同的网格,其大小为 $R \times R$ ,并对小网格中的每个站点分别建模,通过交互池化层考虑小网格之间的空间相关性。在模型训练中,本文利用BMLoss损失来探究周围区域不同空气污染物浓度级别的影响。对于交互池化模块,由于受到污染物排放和扩散的影响,目标节点空气污染物浓度变化趋势与邻接节点密切相关。为了考虑空气污染物浓度节点之间的局部空间相关性,本文将相邻的节点视为“邻居”。交互池化模块整合邻居节点的LSTM隐藏状态信息,为目标节点生成当前交互张量。

由于不同污染等级下PM2.5污染事件的比例高度不平衡,较高浓度的PM2.5污染事件发生的可能性较小,但对人体健康和日常出行的影响较大。因此为了平衡给定时刻节点周围区域的影响,本文提出了BMLoss损失函数,在计算训练损失时根

据空气污染物浓度为每个节点赋予不同的权重。BMLoss损失函数如下:

$$\text{BMLoss} = \frac{1}{N} \sum_{i=T_{\text{obs}}+1}^{T_{\text{obs}}+T_{\text{pred}}} \sum_{j=1}^N w'_j (w'_j - \tilde{w}'_j)^2 \quad (1)$$

其中, $T_{\text{obs}}$ 表示训练数据的时间戳长度, $T_{\text{pred}}$ 表示预测数据的时间戳长度。

通过BMLoss损失函数度量后,各影响因子(特征)之间的权重系数将保持稳定,即特征库构建完成。

### 3 场景应用和技术验证

#### 3.1 场景验证模块

成功构建特征库后,本文提出的推断模型已具备通过空气污染物浓度历史数据

和输入数据映射计算实时空气污染物浓度数据的能力。用户只需在该位置戴上空气污染物浓度监测设备(如VR眼镜),就能够通过实时图片和位置信息获取相关数据,并最终测量计算得到该地区的空气污染物浓度。

### 3.2 技术验证

为了验证BSLInf模型在不同时间和空间维度上的性能,本文将其与最近提出的两种推断模型进行了比较,简要说明如下。

- SLInf模型:一种数据驱动模型,整合多个监测数据,推断特定空气污染物监测站的浓度数据。

- LSTMInf模型:一种混合模型,结合了LSTM和集成学习模型AdaBoost来推断未来空气污染物浓度的动态变化。

为了保证实验结果的可信度,本文采集的数据均来自公开数据集。如第2.1节所述,本文获取了国控空气质量监测站点的空气污染物数据,数据来源于中国环境监测总站;本文获取了分辨率为1 km<sup>2</sup>的气溶胶数据MCD19A2,数据来源于NASA;本文获取了气象数据集ERA-Land,数据来源于ECMWF。

本文将10%的实验数据作为测试集,共64天的数据。为了全面度量BSLInf模型和其他模型的推断性能,本文选择准确率和F1值进行比较。准确率可以衡量模型推断的正确程度,F1值可以同时考虑模型的精确率和召回率,计算式如下:

$$Accu = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$F1 = \frac{2 \cdot po \cdot re}{po+re} \quad (3)$$

其中,TR表示真阳性,TN表示真阴性,

FP表示假阳性,FN表示假阴性;po表示精确率,  $po = \frac{TP}{TP+FP}$  ; re表示召回率,

$$re = \frac{TP}{TP+FN}。$$

本文提出的BSLInf模型与SLInf、LSTMInf模型在不同推断天数下的准确率和F1值比较结果如图6和图7所示。

从图6和图7中可以看到,本文提出的

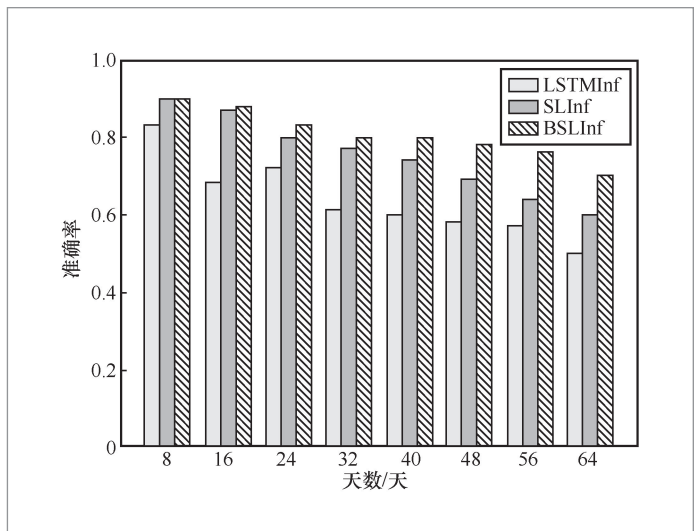


图6 模型准确率

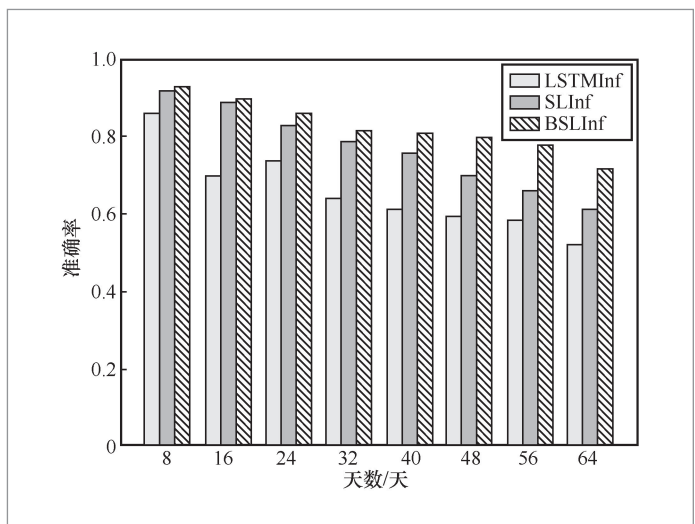


图7 模型F1值

BSLInf模型在总体准确率和F1值上要优于另外两种模型。当预测天数为8天时,BSLInf模型的准确率与另外两种模型相差不大;随着时间的推移,当预测天数达到64天时,BSLInf模型的准确率和F1值已明显高于另外两种模型(比SLPInf模型高出近10个百分点),这说明BSLInf模型在推断未来长时间的空气污染物浓度上有明显提升效果。

本文提出的模型取得了良好性能的原因是:监测站点空气污染物受当地污染源排放、气候、地理环境等因素的影响,加入交互池化层后,模型的推断精度有所提高。交互池化层通过考虑相邻站点的相互作用来捕捉局部区域的空间相关性,从而更准确地推断空气污染物浓度的变化。

## 4 结束语

本文提出了一种基于数字孪生技术的元宇宙空气污染物浓度推断模型BSLInf,成功解决了空气污染物浓度细粒度和实时性推断问题。本文通过技术验证证明,该模型在推断准确率上优于基准模型。然而,该模型迁移能力较差、准确率仍需提升。未来,笔者将尝试通过设计对抗域泛化方法解决上述问题。

## 参考文献:

- [1] YAZDI M D, WANG Y, DI Q, et al. Long-term association of air pollution and hospital admissions among medicare participants using a doubly robust additive model[J]. *Circulation*, 2021, 143(16): 1584–1596.
- [2] EBENSTEIN A, FAN M Y, GREENSTONE M, et al. New evidence on the impact of sustained exposure to air pollution on

life expectancy from China's Huai River Policy[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(39): 10384–10389.

- [3] SONG J, LU M X, LU J G, et al. Acute effect of ambient air pollution on hospitalization in patients with hypertension: a time-series study in Shijiazhuang, China[J]. *Ecotoxicology and Environmental Safety*, 2019, 170: 286–292.
- [4] 寇江泽. 持续攻坚让蓝天白云常驻[N]. *人民日报*, 2021-08-07.  
KOU J Z. Continue in tackling difficulties, and make blue sky and white clouds permanent[N]. *The People's Daily*, 2021-08-07.
- [5] 马元婧. 基于深度学习的大气环境监测系统关键技术研究[D]. 北京: 中国科学院大学, 2021: 7.  
MA Y J. Research on key technologies of atmospheric environment monitoring system based on deep learning[D]. Beijing: University of Chinese Academy of Sciences, 2021: 7.
- [6] KANG Y, CHEN J, CAO Y, et al. A higher-order graph convolutional network for location recommendation of an air-quality-monitoring station[J]. *Remote Sensing*, 2021, 13(8): 1600.
- [7] JUTZELER A, LI J, FALTINGS B. A region-based model for estimating urban air pollution[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2014, 28(1): 424–430.
- [8] XU Y N, ZHU Y M, SHEN Y Y, et al. Fine-grained air quality inference with remote sensing data and ubiquitous urban data[J]. *ACM Transactions on Knowledge Discovery from Data*, 2019, 13(5): 1–27.
- [9] ZHENG Y, LIU F R, HSIEH H P. U-Air: when urban air quality inference meets big data[C]//*Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York: ACM Press, 2013: 1436–1444.
- [10] YU H M, LI Q Y, WANG R, et al. A deep calibration method for low-cost

- air monitoring sensors with multilevel sequence modeling[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(9): 7167-7179.
- [11] WEI J, HUANG W, LI Z Q, et al. Estimating 1-km-resolution PM<sub>2.5</sub> concentrations across China using the space-time random forest approach[J]. Remote Sensing of Environment, 2019, 231: 111221.
- [12] 杨颖川, 葛宝珠, 郝赛宇, 等. 基于能见度及 AOD 数据的北京市 PM<sub>2.5</sub> 浓度的反演[J]. 气候与环境研究, 2020, 25(5): 521-530.  
YANG Y C, GE B Z, HAO S Y, et al. Inversion of PM<sub>2.5</sub> concentration in Beijing based on visibility and AOD data[J]. Climatic and Environmental Research, 2020, 25(5): 521-530.
- [13] 王静, 杨复沫, 王鼎益, 等. 北京市 MODIS 气溶胶光学厚度和 PM<sub>2.5</sub> 质量浓度的特征及其相关性[J]. 中国科学院研究生院学报, 2010, 27(1): 10-16.  
WANG J, YANG F M, WANG D Y, et al. Characteristics and relationship of aerosol optical thickness and PM<sub>2.5</sub> concentration over Beijing[J]. Journal of the Graduate School of the Chinese Academy of Sciences, 2010, 27(1): 10-16.
- [14] 王文喜, 周芳, 万月亮, 等. 元宇宙技术综述[J]. 工程科学学报, 2022, 44(4): 744-756.  
WANG W X, ZHOU F, WAN Y L, et al. A survey of metaverse technology[J]. Chinese Journal of Engineering, 2022, 44(4): 744-756.
- [15] 郑诚慧. 元宇宙关键技术及与数字孪生的异同[J]. 网络安全技术与应用, 2022(9): 124-126.  
ZHENG C H. Key technologies of the metauniverse and their similarities and differences with digital twin[J]. Network Security Technology & Application, 2022(9): 124-126.
- [16] 孟永辉. 当数字货币不再是终局, 区块链从理想进入现实[J]. 大数据时代, 2020(10): 6-13.  
MENG Y H. Blockchain moves from ideal to reality as digital currencies are on the right track[J]. Big Data Time, 2020(10): 6-13.
- [17] 程祖国, 罗敏. 环境偶双极的数字孪生及其应用[J]. 科技创新与应用, 2020(6): 170-171.  
CHENG Z G, LUO M. Digital twin of environmental dipole and its application[J]. Technology Innovation and Application, 2020(6): 170-171.
- [18] 郎为民, 田尚保, 李宇鸽, 等. 数字孪生技术架构研究[J]. 电信快报, 2022(8): 1-6.  
LANG W M, TIAN S B, LI Y G, et al. Research on the technical architecture of digital twin[J]. Telecommunications Information, 2022(8): 1-6.

## 作者简介



彭一非 (1995- ), 男, 湖南财信数字科技有限公司数据开发员, 主要研究方向为大数据分析、云计算等。



袁贞 (1997- ), 女, 湖南财信数字科技有限公司数据开发工程师, 主要研究方向为人工智能、大数据等。



张旭龙 (1988- ), 男, 博士, 平安科技 (深圳) 有限公司高级算法研究员, 主要研究方向为语音合成、语音转换、音乐信息检索以及机器学习和深度学习方法在人工智能领域的应用。



姜桂林 (1981- ), 男, 博士, 湖南财信金融控股集团有限公司首席信息官, 主要研究方向为深度学习。



刘逾江 (1999- ), 男, 墨尔本大学工程管理硕士, 蒙纳士大学荣誉学士, 中南大学学士, 曾在中金资本、IDG资本、平安集团科技会、财信证券实习, 主要研究方向为机器学习。

收稿日期: 2022-10-08

通信作者: 姜桂林, [jianguilin@hnchasing.com](mailto:jianguilin@hnchasing.com)