

一种基于时空注意力克里金的边坡形变数据插值方法

黎嵘繁¹ 钟婷¹ 吴劲¹ 周帆¹ 匡平¹

¹ 电子科技大学信息与软件工程学院 成都市 610054
(rongfanli1998@gmail.com)

摘要 山体滑坡每年对人们的生命财产安全造成重大损失,是常见的地质灾害之一。为了对山体滑坡进行防控,需要广泛地监测山体表面的沉降过程,但是由于恶劣气候和监测成本等难以克服的困难,山体沉降数据的收集呈现了局部数据不完整、数据采样不均衡和监测点动态变化等特点,使得山体滑坡的防控研究受到了阻碍,给数据的采集和分析工作提出了新的挑战。现存方法从空间角度对缺失进行补充,但忽略了时间维度的依赖关系。为了解决上述的问题,对不完整的INSAR数据填充进行了研究,利用时空掩码矩阵对时空依赖关系进行建模,利用多头注意力对多层次的空间关系进行综合学习,并在克里金法(Kriging)的基础上提出了新的使用时空注意力的克里金插值法,实现了对复杂时空特征的深层理解。在真实数据集上的数据恢复实验中,验证了该算法可以有效地学习复杂的时空特征,并在三种不同的数据缺失情景下都取得了比较现存的插值算法更好的表现。

关键词: 时空数据挖掘; 时空注意力; 克里金; 山体滑坡; 插值法

中图法分类号 TP183

Spatio-Temporal Attention-based Kriging for Land Deformation Data Interpolation

Rongfan Li¹, Ting Zhong¹, Jin Wu¹, Fan Zhou¹ and Ping Kuang¹

¹ School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

Abstract Landslides are among the most common geological hazards, causing significant damage to life and property every year. It is necessary to monitor the land surface extensively to prevent and control landslides. However, insurmountable difficulties such as severe climate and high monitoring cost impede the collection of land surface data, resulting in incomplete local data, unbalanced data sampling and dynamic changes of monitoring points. Existing algorithms tried to handle incomplete data from spatial perspective, which, however, ignored temporal dependencies that is important for data interpolation. To solve these problems, a novel Kriging-based model is proposed, which models spatio-temporal dependencies by spatio-temporal masks and learns hierarchical spatial relationships via multi-head attention network. Interpolation experiments on real-world INSAR datasets show that the proposed model is capable to learn sophisticated spatial and temporal features effectively, and achieves better performance than the state-of-the-art methods in three different data interpolation scenarios.

Keywords Spatio-temporal data mining, Spatio-temporal attention, Kriging, Landslide, Interpolation

1 引言

山体滑坡是世界上最常见的地质灾害之一,

是指岩土在重力作用下发生的沿着一定方向的整体性移动的一种自然现象,其诱因包括降雪、地

到稿日期: 返修日期:

基金项目: 国家自然科学基金资助项目(62072077); 国家重点研发计划(2019YFB1406202); 四川省科技计划项目专项资金(2020YFG0234)

This work was supported by National Natural Science Foundation of China(62072077), National Key R&D Program of China(2019YFB1406202), and Sichuan Science and Technology Program(2020YFG0234)

通信作者: 周帆(fan.zhou@uestc.edu.cn)

震等自然因素，以及矿业、水库修建等人为因素。山体滑坡会导致交通阻塞，各种建筑破坏，并严重威胁人们的生命财产安全。在中国，山体滑坡每年造成数千人死亡，直接或间接造成百亿级别的经济损失。因此，对山体滑坡的成因进行探究，以及对其进行防控和对敏感地区持续性地监控受到了学界和工业界的广泛关注^{[1]-[4]}。

目前，已经有很多针对滑坡敏感地区的滑坡预测研究和案例，但是都非常依赖于相关领域的专家知识，利用人的丰富经验来对现场的各种地质指标和拓扑结构进行分析判断，最终得出一个比较精确的专业结论^{[5]-[6]}。但是这种依赖专业知识的方法主要依靠地质学、生态学、统计学和力学等领域知识积累以及对山体滑坡研究经验。为了对山体滑坡进行自动的监控和预测，各种机器学习方法都被应用到了这个领域^{[7]-[8]}。得益于地理信息系统（Geographic Information System, GIS）和无线传感器网络的发展，各种机器学习的方法能获得足够的信息对山体滑坡做出预测，主要的方法包括但不限于贝叶斯模型、逻辑回归、决策树、随机森林和支持向量机等模型^{[9]-[12]}。

最近几年，深度学习（Deep Learning）发展迅速，并且也有一些研究提出了将深度学习应用到山体滑坡的地质特征捕捉上。主流的方案有两种，一种是将山体表面的照片送入卷积神经网络（Convolutional Neural Networks, CNNs），利用卷积操作学习目标的空间特性^{[3][13]}，但是这种方案有两个明显的问题：首先卷积操作依赖图片，而图像的整体像素往往对结果产生重大影响；其次图像是 2D 的，无法使模型学习到对山体滑坡很重要的倾斜度信息，这种特征需要 3D 数据才能衡量。所以另一种方案使用了图神经网络（Graph Neural networks,

GNN），可以使用任意维度的地理位置信息来表示不同的监测点，通过学习邻接矩阵等图的特征获取目标的空间结构，并结合各种时间感知模块对山体表面沉降等时空（Spatio-temporal）数据做出预测^{[14]-[16]}，然而大部分图神经网络研究都是基于静态图的，无法适应于动态变化的山体监测数据。在动态图的问题领域，已经有了一些有效的方法，但是却对时空性质的图问题研究较少，我们在相关工作中将做进一步介绍^[17-21]。此外，干涉合成孔径雷达技术（Interferometric Synthetic Aperture Radar, InSAR）被广泛应用于地理相关研究，可以提供的高精度地表变化点云（Point-cloud）数据，结合 GNN 可以实现对山体表面更加丰富特征的提取，相关的研究已经取得了较为优异的成果^{[22][23]}。

总结以往的研究，山体滑坡预测有以下几点困难：首先是对时空特征进行建模，使得模型能从时间和空间两个维度同时对特征进行捕捉；其次，在实际情况中，由于传感器失效、雷雨天气干扰、云层遮挡等问题，对目标区域的数据采集往往有缺失，并且这种缺失有局部聚集出现的特点；由于经费不足等条件限制，并不能对整个区域进行足够均匀和充分地监测，导致部分区域监测比较稀疏而部分区域监测比较密集，最终使得模型的学习存在不平衡的现象；最后，监测区域并不是长期一成不变的，会因为地形的变化等原因扩大或缩小范围，因此对应的图（Graph）结构往往是动态的，包括但不限于各种节点的新增和减少，以及节点之间的联系的增加或降低。

为了解决山体滑坡预测中的存在的数据不完整、数据采样不均匀和监测点动态变化等困难，对监测数据进行增强，以辅助其他的山体滑坡预测方法，本文结合了地理特征插值问题代表性算法普通克里金法（Ordinary Kriging, OK）^[24]提

出了一种新的图神经网络结构, 本文的贡献如下:

(1) 我们参考现有的研究^[20]改造了基本的图注意力网络结构, 通过控制时空掩码矩阵的形态和多头注意力的形态, 获得了对时空特征的建模能力, 并且利用基于可达性 (Accessibility) 的子图采样方法, 使得模型能适应于动态图。

(2) 将注意力网络结合克里金 (Kriging) 插值法, 使得模型能够对某监测点的缺失值进行自动地补充, 在较为稀疏的观测区域构造虚拟监测点, 从而获得更全面的图信息和地质特征。

(3) 我们从真实的地理位置上采集了一系列高精度 InSAR 点云数据, 并进行了相关实验, 在和现存的插值法对比中验证了模型的有效性。

2 相关工作

2.1 归纳式图神经网络

现有的对时空数据建模的图神经网络无法适应于动态变化的图结构, 因为图的动态变化会导致谱域中图的全局性变化, 使得图的拉普拉斯矩阵过时, 故而需要重新计算图的空间结构关系^{[14]-[16][25][26]}。这一类无法适应图的动态变化的模型被称为直推式学习 (Transductive), 反之则称为归纳式学习 (Inductive)。Hamilton 等^[17]在 2017 年第一次提出了基于图卷积神经网络 (Graph Convolution Network, GCN)^[25]的归纳式神经网络, 通过子图和 mini-batch 等技巧可以实现归纳式的学习。与之类似的研究^[18]用图聚类方法使得完全图划分成了数个子图, 或者提出了一种对于 GCN 无偏的图采样^[19], 两者都通过对子图的训练来构建更大的动态图。另外一大类归纳式的图网络是基于注意力机制的模型。自从 Vaswani 等^[20]提出基于自注意力 (self-attention) 机制的 Transformer 框架以来, 注

意力网络受到了大量的关注, 在此基础上有研究^[21]提出了一种归纳式的图注意力网络, 通过控制注意力的范围和节点的邻居数量使得模型能适应于动态变化的图。不局限于自然语言处理和图像领域, 注意力机制也被广泛应用到了图网络中, 在人的动作检测和视频处理等时空性质的图问题中具有优秀的表现^{[27][28]}。

2.2 普通克里金法和其他插值算法

克里金法基于这样的一种朴素的直觉: 相近的节点特征也相似, 因此可以根据目标位置周围的信息对目标位置进行特征填补。首先定义两点间特征的相似度为半方差, 普通克里金法认为两点的物理距离和半方差之间有一定的函数关系, 假设特征是均匀的前提下, 通过拟合物理距离和半方差的函数可以得到目标的无偏估计量^[29]。

普通克里金法中有许多关于节点特征的强假设, 并且其物理距离和半方差的函数计算并不能适应各种复杂的情景。为了得到缺失点属性的更好的估计, 很多研究提出了基于同样思想的克里金变体形式, 如泛克里金 (Universal Kriging)、协同克里金 (Co-Kriging) 等等^[24]。

另一方面, 部分研究结合了神经网络和传统的克里金法, 从多种角度增强了传统方法。Tapoglou 等^[30]利用克里金法对空间进行感知, 并利用神经网络对时间进行感知, 从而得到一个对时空数据感知的模型。区别于普通克里金法用协方差矩阵来求解权重系数, Franchi 等^[31]利用深层神经网络来生成权重系数, 从而实现了图片的超分辨率。Wu 等^[32]利用扩散图卷积技术来增强对空间复杂特征捕捉能力, 并扩散到时间维度, 从而完成了一个单纯依赖图神经网络的克里金法实现。

除了经典的克里金法以及其变体，还有一些传统的算法被应用在插值问题中。k 近邻法 (k-Nearest Neighbors, kNN) 作为传统的机器学习算法也常出现在插值算法中，通过找到最相邻的 k 个点，然后直接取他们的平均值作为目标位置的补充。核概率矩阵分解 (Kernelized Probabilistic Matrix Factorization, KPMF) 通过拉普拉斯核将图的核心信息包含进了矩阵分解技术中，可以将空间的特征聚集到目标位置上进行插值^[33]。

但是现有的方法并没有将特征的时空属性同时进行捕捉，没有建立起时空一体的感知体系，也不能适应于动态图的学习。因此我们提出了一种利用了注意力神经网络的克里金法变体，该方法能针对动态变化的，有任意缺失值的图进行时空特征的相互依赖关系的建模，从而实现时空插值，完善原始的数据信息，为后续的数据挖掘提供更完整的信息。实验表明，我们提出的这种方法相比较多种传统方法有显著的优越性。

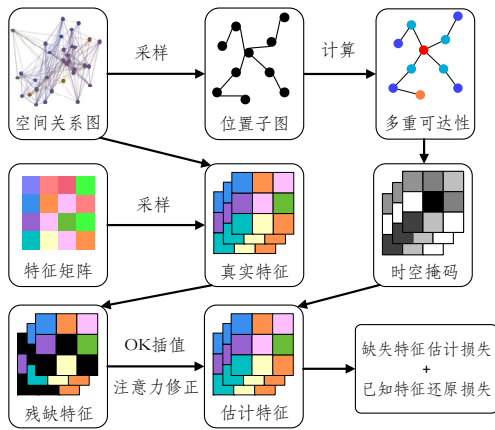


图1 STAK 模型框架

Fig. 1 Architecture of STAK

3 时空注意力-克里金插值法

基于普通克里金法，本文提出了基于时空注意力的克里金插值法 (Spatio-temporal Attention-based Kriging, STAK)，模型框架如图 1 所示。模型主要分成三个部分：首先是在

点云数据上构建了一个图，然后定义了不同层级的邻接关系；其次，构造时空掩码矩阵，使得模型能捕捉时空信息；最后，结合普通克里金法和多头注意力 (Multi-Head Attention) 网络，使用图采样的方案赋予模型归纳式学习的能力，并对缺失数据进行插值和修正。

3.1 问题描述和图的建立

为了更好地描述地质插值问题，不失一般性地，用点云数据定义 N 个地理位置向量 \mathbf{z} 作为山体滑坡的监测点唯一的标识，这个位置向量一般是 3 维，组成矩阵 $\mathbf{Z} \in \mathbb{R}^{N \times 3}$ 。每个位置上有一组对应位置的山体表面沉降记录，监测持续时间 T ，记特征向量为 \mathbf{q} ，组成特征矩阵 $\hat{\mathbf{Q}} \in \mathbb{R}^{N \times T}$ 。为了进行实验，我们将数据集 $\hat{\mathbf{Q}}$ 划分成了两部分，一部分标记为缺失记为 \mathbf{Q}_{miss} ，在后文没有额外说明的情况下，其余部分表示没有缺失的特征记为 \mathbf{Q} ，并且有 $\hat{\mathbf{Q}} = \mathbf{Q} + \mathbf{Q}_{miss}$ 。假设 \mathbf{z}_i 处有不完整的特征向量 \mathbf{q}_i ，记邻居节点为 $\mathcal{N}(\mathbf{z}_i)$ ，在有足够多邻居节点的情况下，可以用邻居节点 $\mathbf{z}_j \in \mathcal{N}(\mathbf{z}_i)$ 对应的特征对 \mathbf{q}_i 进行补充。地质插值问题可以一般性地描述为：根据已知特征 \mathbf{Q} 和位置信息 \mathbf{Z} ，对残缺特征 \mathbf{Q}_{miss} 和全体特征 $\hat{\mathbf{Q}}$ 进行估计。

我们提出利用多头注意力网络来表示这种不同距离的复杂空间关系和时间关系。为了使用图神经网络对图结构强大的解析能力，首先需要定义一个图结构。图中的每个点并不需要和其他全部点相连接，即保持一定的稀疏性，这是因为认为每个点只应该和最近的节点产生互动，而不需要也不能和远处的节点有联系。因此，先构造一个全连接的邻接矩阵 \mathbf{A} ，然后定义一个阈值 ε ，高于这个距离阈值的边都切断，在阈值内的认为是 1 跳 (1-hop) 的邻居，即如果 $\|\mathbf{z}_i - \mathbf{z}_j\|_2 \leq \varepsilon$ 则 $\mathbf{A}_{(i,j)} = 1$ ，表示图中节点 \mathbf{z}_i 和 \mathbf{z}_j

之间有边 e_{ij} 。这里和后文的 $\|\cdot\|_2$ 运算都表示 L2 范数 (Norm)，也称为欧几里得范数 (Euclidean norm)，对于任意一个向量 \mathbf{x} 有 $\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \dots + x_n^2}$ 。完成裁剪边的操作后，就得到了一个图 $\mathcal{G}(\mathbf{Z}, \mathbf{E})$ 。

构建好基本的图后，可以定义多重可达性，可达矩阵用 \mathbf{A}^h 表示，每个点能和 h 跳的邻居产生联系。在这个多重可达性的规范下，我们使用不同的注意力头对不同可达程度的邻居进行学习，从而使得注意力网络可以捕捉更加丰富层次的空间依赖关系。需要注意的是，在算力满足的情况下，同样的可达性矩阵可以被尽可能多的注意力头学习以保证更加稳定的效果，即可以使用 5 个注意力头对 \mathbf{A}^1 的关系进行学习，而使用 3 个注意力头对 \mathbf{A}^2 的关系进行学习，这样的设置可以更加稳定地学习更近邻的邻居节点特征，并且同时保留对更大范围的特征的感知能力。这里使用的可达矩阵数量视监控区域的节点密集程度和特征复杂程度而定。

3.2 构建时空掩码矩阵

注意力应用在自然语言处理领域，使用的位置编码和掩码矩阵只有一个维度^[20]，在我们的情景下，位置编码是天然存在的，但是掩码矩阵是有两种维度，分别是时间和空间。我们设计一个时空掩码矩阵，使得注意力网络可以对不同时空位置的节点分配不同的权重，就可以实现对时空特征的捕捉。在时间维度上，某一时间节点的特征只依赖于前一节点的特征，而不依赖后一时间节点；在空间维度上，某个空间节点的特征更多地依赖其邻居节点，而不是远处的节点。时间上的掩码矩阵很容易建立，空间依赖特性可以被多重可达性矩阵定义，并且在不同注意力头上只使用一个可达性矩阵。如果没有额外说明，后续模型描述都建立在一个注意力头上。

对于空间位置和时间位置分别在 (\mathbf{z}_i, t_i) 和 (\mathbf{z}_j, t_j) 的两个节点，构建对应位置的时空掩码如下：

$$\mathbf{M}_{(\mathbf{z}_i, t_i, \mathbf{z}_j, t_j)} = 1 \quad \text{if } \{\mathbf{z}_j \in \mathcal{N}(\mathbf{z}_i)\} \wedge \{t_j < t_i\} \quad (1)$$

其中， $\mathbf{M} \in \mathbb{R}^{NT \times NT}$ ，如果 $\mathbf{A}_{(i,j)}^h = 1$ ，则认为 \mathbf{z}_j 是 \mathbf{z}_i 的邻居，即 $\mathbf{z}_j \in \mathcal{N}(\mathbf{z}_i)$ 。在时空掩码矩阵中标记为 1 的部分节点特征是会被感知的，其他的都会被忽略。在掩码矩阵的影响下，时空特征可以同时被神经网络捕捉，使得特征的依赖关系更加全面。在不同的注意力头中如果使用不同的可达性矩阵，将会得到不同的时空掩码矩阵，所以模型具有对不同层级的时空特征进行建模的能力。

由于数据中有缺失值，即使有插值法，我们也无法认为这种非真实数据是可靠的。所以对于缺失值，我们需要对其掩码进行进一步的修正处理。定义两种类型的缺失值：一种是因为特殊情况导致观测失效的缺失记录；另一种是因为成本等原因导致没有对某些位置进行观测的缺失记录。对于后一种情况，直接定义一个需要补充的位置向量 \mathbf{z} 即可将问题统一为某个观测点上的数值缺失。这些缺失的特征将首先被普通克里金法进行填充，然后用注意力网络修正，具体的步骤将在后文描述。为了表示这种缺失，掩码矩阵在其他点对缺失点的依赖位置标记为 $\delta \in [0,1]$ ，使其他的点在后续的修正中减少对缺失点特征的依赖，其计算方式如下：

$$\mathbf{M}_{(\mathbf{z}_i, t_i, \mathbf{z}_j, t_j)} = \delta, \quad \text{if } \mathbf{Q}_{(j, t_j)} \text{ is not complete} \quad (2)$$

值得注意的是， δ 并不是一个超参数，而是可以被学习的一个参数，是依赖于普通克里金法对最终填充效果的贡献度的，一般可以设定一个固定初始值 $\delta = 0.5$ ，最终学习出来的 δ 越接近 1 说明普通克里金法的效果越好。通过定义如上的时空掩码矩阵，多头注意力模型不仅能获得对时

空特征的综合分析能力，也能实现对局部和更大范围的空间特征的多重空间敏感性。

3.3 基于多头时空注意力的空间插值法

前文分别定义了多重可达性矩阵的使用和时空掩码矩阵，下面介绍模型的主体部分：基于多头注意力的时空插值法。

为了实现归纳式的学习，参考现存的研究^[17]，模型并不直接从 $\mathcal{G}(\mathbf{Z}, \mathbf{E})$ 上直接学习，而是学习一个子图 \mathcal{G}_{sub} 上的特征。由于时空掩码矩阵的局部特性，我们的模型从设计上就无法从全局角度来理解图的空间结构，这一特点恰好和图注意力网络相契合。

我们设计了一个基于可达性的子图的截取方法。利用前文定义的时空掩码矩阵，可以对图的局部进行注意力的分配，并在这个子图上进行训练。定义需要使用的最大可达性 H ，利用 \mathbf{A}^H 对 $\mathcal{G}(\mathbf{Z}, \mathbf{E})$ 进行节点的裁剪。由于空缺值有聚集性出现的特性，所以后文都讨论一组需要填充的位置 \mathbf{Z}_{miss} 和对应的子图 \mathcal{G}_{sub} 。对一组需要填充的位置 \mathbf{Z}_{miss} ，有：

$$\mathbf{z}_j \in \mathcal{G}_{sub} \quad \text{if } \{\mathbf{z}_i \in \mathbf{Z}_{miss}\} \wedge \{\mathbf{A}_{(i,j)}^H = 1\} \quad (3)$$

构造全部子图后，统计其中子图包含的最多节点数量，记为 N' ，作为图注意力网络的统一输入尺寸，对节点数少于 N' 的子图，额外补充一部分节点但是不改变可达性，使其不可达即可。相对应的，每个子图都有一组对应的规模为 $N' \times N'$ 的子邻接矩阵 $\{\mathbf{A}_{sub}^h\}_{h=1}^H$ ，用于构建不同的时空掩码矩阵。

对 \mathbf{Z}_{miss} 使用普通克里金法进行初次插值，目的是使得自注意力神经网络能获取缺失位置的基准信息，并在这个基础上利用周围区域的特征注意对基准值进行修正，实现更好的插值效果。

记 $\hat{\mathbf{q}}_i$ 为目标点的估计值，根据普通克里金法的特征属性 \mathbf{q} 均匀假设，有期望 $\mathbb{E}[\mathbf{q}] = c$ ，且有方

差 $\text{Var}[\mathbf{q}] = \sigma^2$ ，因此可以推导下面的结论：

$$\begin{aligned} \mathbb{E}[\hat{\mathbf{q}}_i - \mathbf{q}_i] &= \mathbb{E}[\sum_{j=1}^n \lambda_j \mathbf{q}_j - \mathbf{q}_i] \\ &= c \sum_{j=1}^n \lambda_j - c = 0 \end{aligned} \quad (4)$$

因此有约束条件 $\sum_{j=1}^n \lambda_j = 1$ ，然后分析 $\text{Var}[\hat{\mathbf{q}}_i - \mathbf{q}_i]$ 有：

$$\begin{aligned} \text{Var}[\hat{\mathbf{q}}_i - \mathbf{q}_i] &= \text{Var}[\sum_{j=1}^n \lambda_j \mathbf{q}_j - \mathbf{q}_i] \\ &= \text{Var}[\sum_{j=1}^n \lambda_j \mathbf{q}_j] - 2\text{Cov}[\hat{\mathbf{q}}_i, \mathbf{q}_i] + C_{ii} \quad (5) \\ &= \sum_{k=1}^n \sum_{j=1}^n \lambda_k \lambda_j C_{kj} - 2 \sum_{j=1}^n \lambda_j C_{ij} + C_{ii} \end{aligned}$$

其中 $C_{ij} = \text{Cov}[\mathbf{q}_i, \mathbf{q}_j] = \text{Cov}[\mathbf{q}_i - c, \mathbf{q}_j - c]$ 。但是这些协方差比较难以计算，所以再假设 C_{ij} 和两点之间的距离 $\|\mathbf{z}_i - \mathbf{z}_j\|_2$ 有一定的函数关系，通过最小二乘法等方法拟合这种函数关系，就可以快速通过任意两点的三维空间的距离求出两点之间的 C_{ij} 。在 $\sum_{j=1}^n \lambda_j = 1$ 约束下，利用拉格朗日乘数法求得使得目标式最小的权重系数组合 $\sum_{j=1}^n \lambda_j$ 即可完成对目标点的特征估计，对缺失值做出基准估计得到 $\bar{\mathbf{Q}}$ 。

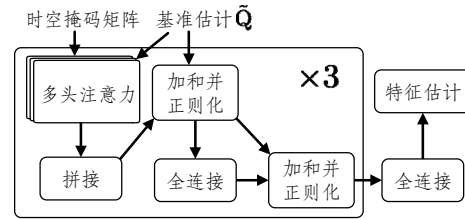


图2 时空多头注意力

Fig. 2 Spatio-temporal multi-head attention

最后构造多头注意力模型，利用时空掩码矩阵对前面得到的基准估计进行修正，参考图 2，其中每个注意力头的计算方式如下。

$$\text{Head}(\bar{\mathbf{Q}}, \mathbf{W}, \mathbf{V}, h) = \text{softmax}\left(\frac{\bar{\mathbf{Q}}\mathbf{W}^T}{\sqrt{NT}} \odot \mathbf{M}^h\right)\mathbf{V} \quad (6)$$

其中 \odot 是矩阵的哈达玛积 (Hadamard product)，即对应位置元素相乘， $\bar{\mathbf{Q}}$ 、 \mathbf{W} 和 \mathbf{V}

分别是查询 (queries)、键 (keys) 和值 (values) 矩阵, 并且都由 $\tilde{\mathbf{Q}}$ 进行初始化, h 表示这个注意力头所使用的可达性阶数。最后我们将全部头拼接到一起并做出对基准估计的修正:

$$\text{MultiHead}(\tilde{\mathbf{Q}}, \mathbf{W}, \mathbf{V}, \mathbf{h}) = \text{concat}(\text{Head}_1, \text{Head}_2, \dots) \phi \quad (7)$$

ϕ 是用于将注意力头融合到一起的可训练的全连接 (fully connected) 网络的参数, 并且可以控制最终的输出 $\hat{\mathbf{Q}}$ 的形态。这个结构重复 3 次便可以输出 $\hat{\mathbf{Q}}$, 包含了全部点的特征估计。计算损失需要对已知特征和未知特征的还原准确度进行权衡, 完全不考虑对已知特征的还原能力会导致模型泛化能力降低, 但同等地考虑两种损失会导致对目标位置的关注度下降, 所以定义损失函数如下:

$$Loss = \sum_{graphs} (\|\hat{\mathbf{Q}}_{miss} - \mathbf{Q}_{miss}\|_2^2 + \gamma \|\hat{\mathbf{Q}} - \mathbf{Q}\|_2^2) \quad (8)$$

其中 γ 是调节两种损失的超参数。在一个区域中 \mathcal{G}_{sub} 中, 缺失部分 \mathbf{Z}_{miss} 占少数, 所以可以近似理解公式 (8) 的第二部分是没有缺失特征的点的恢复损失。算法的整体描述如算法 1 所示。

算法 1 时空注意力克里金插值法

输入: 地理位置 \mathbf{Z} , 特征矩阵 \mathbf{Q} , 缺失标记 \mathbf{Z}_{miss} 和 \mathbf{Q}_{miss} , 裁剪阈值 ε , 最大可达性 H , 损失调节参数 γ ;

输出: 训练参数 ϕ ;

1. for $i \leftarrow 1$ to N do;

a) for $j \leftarrow 1$ to N do;

if $(\|\mathbf{z}_i - \mathbf{z}_j\|_2 \leq \varepsilon) \mathbf{A}_{(i,j)} = 1$;

2. 计算 $\{\mathbf{A}^h\}_{h=1}^H$, 基于 \mathbf{A}^H 随机采样子图

$\{\mathbf{A}_{sub}^h\}_{h=1}^H$;

3. for $h \leftarrow 1$ to H do

a) 根据公式 1 和 2, 计算 \mathbf{M} ;

b) 根据克里金法^[24]计算基准 $\tilde{\mathbf{Q}}$;

c) 用 $\tilde{\mathbf{Q}}$ 初始化 $\bar{\mathbf{Q}}$ 、 \mathbf{W} 和 \mathbf{V} , 计算

$\text{Head}(\bar{\mathbf{Q}}, \mathbf{W}, \mathbf{V}, h)$;

d) 计算 $\hat{\mathbf{Q}} = \text{MultiHead}(\bar{\mathbf{Q}}, \mathbf{W}, \mathbf{V}, \mathbf{h})$;

4. 计算 $Loss$, 并利用梯度下降法优化参数 ϕ ;

5. 重复 2-4 步直到收敛。

4 实验

4.1 数据集构造

我们的数据集是取自于大渡河上的猴子岩水电站附近的山坡, 位于四川甘孜藏族自治州丹巴县和康定市交界处。数据集中包含了从 2019 年 1 月 9 日到 8 月 31 日, 利用 InSAR 技术获取的目标区域长达 8 个月, 每次间隔一周的地表沉降点云数据, 其中大渡河西岸有 4569 个监测点, 东岸有 2164 个监测点, 每次的沉降数据从 -29.06mm 到 30.05mm 不等。对水电站等重要设施附近的山体进行监测, 从而对山体滑坡进行防控, 对保护人们的生命财产安全有重要意义。在图 3 中, 画出了数据集的点云视图 (a) 和某时间点的西岸特征视图 (b-d)。点云数据直接用归一化后的经度、维度和海拔高度数据作为 RGB 进行着色。从特征视图中可以看出沉降特征在空间维度上存在连续性, 即深色和深色的点有聚集性。从图 3 中 (b)、(c) 和 (d) 的比较中可以发现, 在时间维度上特征也有连续性, 3 (b) 中深色的部分在 3 (c) 中依然保持了一定的深色, 而在时间较远的 3 (d) 中特征的尺度和分布情况都完全不相同。这种时间和空间的双重连续性观察也正是 STAK 将时空同时进行分析的动机。

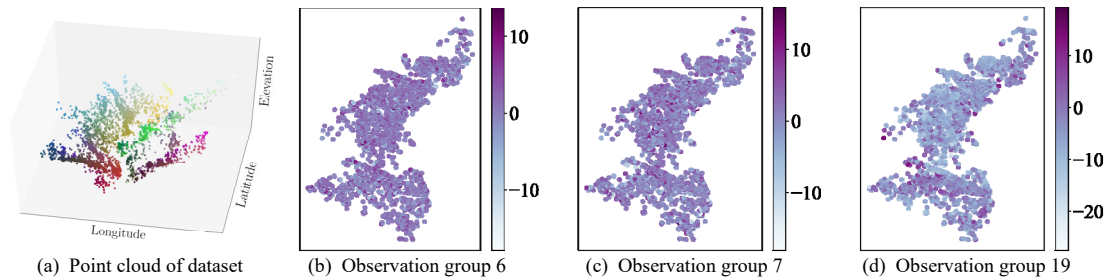


图3 监测区域点云视图和特征视图

Fig. 3 Point cloud of the monitored area and features of west side

为了检验 STAK 对缺失值的恢复效果，我们从完整的数据集中标记了一部分特征为缺失值，然后利用这些值进行对比。我们设计了三种不同的缺失值构造方案对插值算法进行检验：（1）针对数据不完整，我们假设缺失位置在局部集中分布，每个残缺位置的特征空缺量小于或等于 5；（2）针对数据采样不均匀，我们假设部分监测位置特征全部空缺，并在局部集中分布，其余位置全部完好；（3）针对复合问题，我们有上述两种情况的综合数据。三种情况的缺失值数量相同，占总量的 10%。通过对真实情况的三种模拟，我们认为可以反映 STAK 的插值性能，并且运用到现实的插值问题中。需要注意的是，虽然左右两岸的数据存在不连续，但是我们分别在两岸进行子图采样，避免了对两岸分别讨论的问题。为了防止在同样的子图上过拟合，我们将不同的缺失子图组合随机分割成了两份，70%作为训练集，剩下的 30%作为测试集。我们取十次实验的平均值作为文中展示的结果。

4.2 实验设置

为了体现 STAK 模型特点和性能，我们用几种广泛使用的插值法进行了对比，分别是：

（1）普通克里金法（OK），在地质插值问题上经受检验的算法；（2）KPMF 算法^[33]，传统的图处理算法，利用矩阵分解技术将图的特征进行聚集然后进行插值；（3）kNN，通过找到目标位置

周围在欧几里得空间中最近的 k 个节点，然后对

到稿日期：返修日期：
基金项目：国家自然科学基金资助项目（62072077）；国家重点研发计划（2019YFB1406202）；四川省科技计划项目专项资金（2020YFG0234）

This work was supported by National Natural Science Foundation of China(62072077), National Key R&D Program of China (2019YFB1406202), and Sichuan Science and Technology Program (2020YFG0234)

通信作者：周帆 (fan.zhou@uestc.edu.cn)

他们的特征做平均化处理来获得预测。

为了学习动态图结构，STAK 需要从子图上进行学习，为了保持公平，对其他的对比算法都使用了类似的子图，其大小都被设定为 $N' = 500$ ，人工挑选了东西岸合计共 100 个子图。STAK 使用了最多 3 阶的可达性矩阵，即 A^3 。对 1, 2, 3 阶的可达性矩阵分别使用了 5, 4, 3 个注意力头进行学习。在多头注意力的内部使用了三层时空注意力层，见图 2。从点云数据中构造图的时候，令 $\varepsilon = 200$ 。时空掩码矩阵的初始化 $\delta = 0.5$ ，未缺失值重建损失权重 $\gamma = 0.5$ 。最后，初始学习率设置为 0.002，每 100 轮训练将学习率调整为之前的 0.8，损失 20 轮不下降便停止，取整个训练中最优损失对应的参数用于测试。对于 kNN 算法，我们使用了 $K = 100$ 进行计算。

4.3 性能比较和结果分析

我们使用了均方根误差（Root Mean Square Error, RMSE），平均绝对误差（Mean Absolute Error, MAE），精确度（Accuracy, ACC）和决定系数（Coefficient of Determination, R^2 ）来衡量插值性能的高低，RMSE 和 MAE 越小，ACC 和 R^2 越接近 1 的算法，其插值性能越好。与训练模型的时候需要关注算法的泛化能力不同，这里只需要考虑对缺失值的恢复重建能力。

表 1 不同模型的插值性能对比

Table 1 Performance of different algorithms

Algorithms	RMSE	MAE	ACC	R ²
KPMF	0.125	0.093	0.854	0.243
kNN	0.109	0.071	0.931	0.586
OK	0.096	0.065	0.947	0.636
STAK (1)	0.072	0.050	0.978	0.763
STAK (2)	0.091	0.063	0.952	0.662
STAK (3)	0.087	0.060	0.957	0.689

表 1 中给出了 4 种不同类型的模型插值性能对比, 其中最优结果已经加粗。总体来说, KPMF 的表现最差, kNN 比较 OK 略差, STAK 总体上优于其他模型。在我们的实验中, 除了 STAK, 其他的算法在不同的情景中表现非常接近, 这是因为其他算法总是只考虑空间特征, 导致三种情景在空间角度都有相似的问题结构, 所以没有在表格中体现三种不同情形。STAK 在情景 1 中的效果明显优于情景 2, 在情景 3 中表现比较平衡。我们分析这是因为在情景 1 和 3 中, STAK 可以利用历史数据进行估计, 使得特征的估计具有了时间上的连续性, 这一点从直觉上是成立的, 因为地表的沉降总是连续的, 不会发生突变。针对三种缺失情景的比较证明了同时从时间和空间角度对特征进行估计的策略有效性, 此观点会在后续的实例分析中得到进一步的验证。

STAK 和 OK 的对比可以视为时空注意力的修正效果消融实验, 从表中可以看出, STAK 的性能全面好于 OK, 并且时空注意力的提升效果在情景 1 中好于情景 2 和 3, 我们认为原因是在情景 1 中, STAK 可以依赖缺失位置的历史数据进行时空依赖的估计, 而在情景 2 中, STAK 和 OK 一样只能通过空间特征进行预测。我们可以得出结论, 时空注意力可以有效提高 OK 算法的效果, 并且时间维度的感知可以显著提高插值算法

的效果。此外, 掩码矩阵中的 δ 参数在三种情景下的最终的学习结果都是 0.7 左右, 说明了 STAK 能对 OK 的效果进行一定程度的拔高。

计算精确度时, 我们设置当绝对误差小于 1 的时候认为该预测是正确的。为了对不同方法的精确度进行更直观的描述, 我们在缺失情景 3 中将各个算法预测的总体误差做了一个统计如图 4 所示。在图中可以看出, STAK 在 0.3 误差级对其他方法明显的优势。在高精度预测中 kNN 表现很差, 我们认为这是因为 kNN 通过计算平均值进行预测, 严重依赖于估计位置的周围点的均匀性, 其平均值容易受到某邻居特征极端变化影响导致估计精度快速下降, 尤其是在比较陡峭的区域和采样不均匀的区域。OK 用高阶线性模拟一定程度上克服了这个, 但是我们认为线性解的灵活度不足, 故而用 STAK 通过注意力网络对其进行修正, 因此对比 OK 有进一步明显的提升。

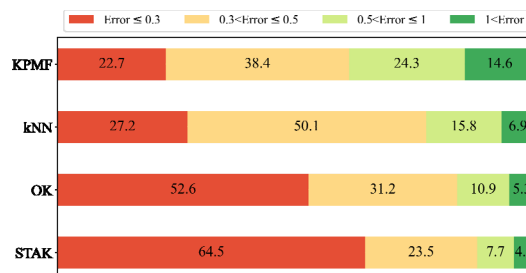


图 4 估计误差分布

Fig. 4 Discrete distribution of estimated error

到稿日期: 返修日期:

基金项目: 国家自然科学基金资助项目 (62072077); 国家重点研发计划 (2019YFB1406202); 四川省科技计划项目专项资金 (2020YFG0234)

This work was supported by National Natural Science Foundation of China (62072077), National Key R&D Program of China (2019YFB1406202), and Sichuan Science and Technology Program (2020YFG0234)

通信作者: 周帆 (fan.zhou@uestc.edu.cn)

不同的特征缺失率对不同算法的效果有非常直接的影响，以精确度为评价标准，随着缺失率从 5% 上升到 10% 和 20%，全部算法的插值效果都有不同程度的下降，见图 5。其中 STAK 下降的幅度最小，对比之下 kNN 和 KPMF 下降幅度非常大，说明了 STAK 对缺失程度的敏感性有最好的表现，可以适应于不同的缺失程度的插值问题，并且始终优于其他的方法。

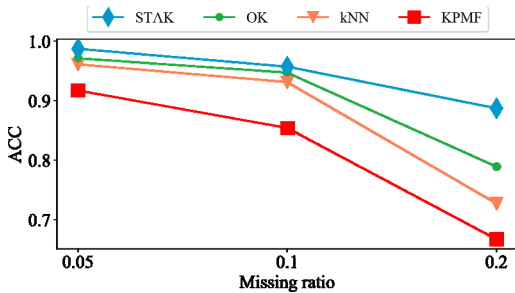


图 5 缺失率对性能的影响

Fig. 5 Performance influence of missing ratio

最后，从时间和空间两个角度对各个算法的插值性能进行可视化的展示。

首先是时间维度，随机选择两个形变变量不同的监测点的 10 个历史特征数据，分两次每次 5 个缺失处理，然后进行缺失数据的估计并整理到了一起。为了避免线条杂乱这里去掉了 KPMF 的插值结果。从图 6 可以看出，STAK 最贴近真实特征的走向，总体的偏离程度最小。此外，其他算法只能依靠邻居节点的特征保持时间维度的连贯，而 STAK 直接对不同时间的特征进行感知，我们认为这种能力使得 STAK 能更好地保持时间维度的平滑性，更加符合边坡形变插值问题的渐变属性。

然后我们选择了某时刻的 6 个彼此相邻的位置从空间角度对插值效果进行可视化，见图 7。可以看出，KPMF 整体最差，STAK 和 OK 始终保持了比较好的特征估计，同时 STAK 基本上保持了对 OK 的优越性。kNN 呈现出五个位置的特征都非常接近的特性，这也是因为 kNN 同等考虑周

围大量邻居特征，但是在 3 和 4 这种相对特征变化较大的位置，kNN 便难以从邻居特征中得到对目标特征的较好估计。与之相反的，OK 和 STAK 并不同等地对邻居节点进行参考，所以能得到更精确的结果。

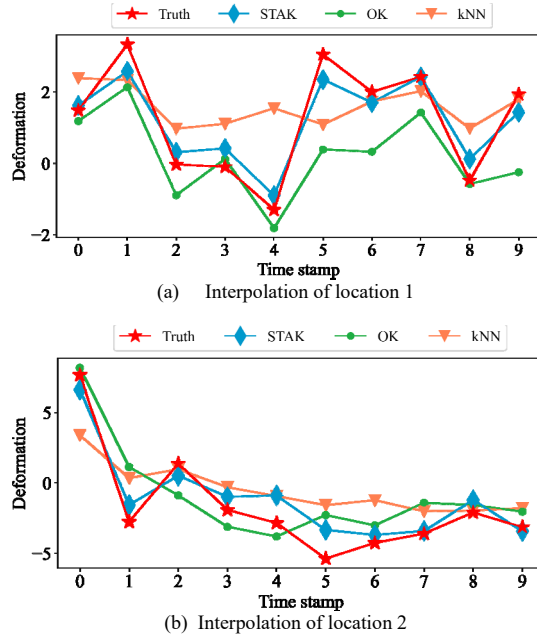


图 6 时间维度插值效果可视化

Fig. 6 Visualization of interpolation in the time dimension

结束语

本文提出了一种基于时空注意力的克里金插值法，利用时空掩码矩阵可以有效捕捉时空特征，利用多头注意力机制可以对不同层次的空间特征进行学习，最后得到插值结果。通过在真实地面数据集上和其他的几种插值算法进行对比，从整体性的性能表现和案例分析两方面验证了本文算法的有效性，证明了在边坡形变插值问题中同时对时间和空间进行学习分析的有效性和必要性，可以应用此思路在山体滑坡的形变插值问题中。本框架具有一定的普适性，也适用于任何的具有时空属性的图网络插值问题如城市车流量预测，可以解决各种传感器失效或经济等原因引起的特征不完整不均匀问题。在未来的工作中，不局限于利用神经网络强大的特征学习能力，我们考虑挖掘更多的对特征估计和插值有价值的深层

特征, 进一步地研究地质数据的内在规律, 并显式地进行分析和学习, 从而增强模型的可解释性和插值效果。此外, 真实情况下的特征缺失比我们设计的实验往往更加复杂, 例如缺失值会大量

集中出现或者均匀分布在每个监测点上, 我们将研究如何设计更加贴近现实的缺失算法以辅助插值算法的实用性进一步提升。

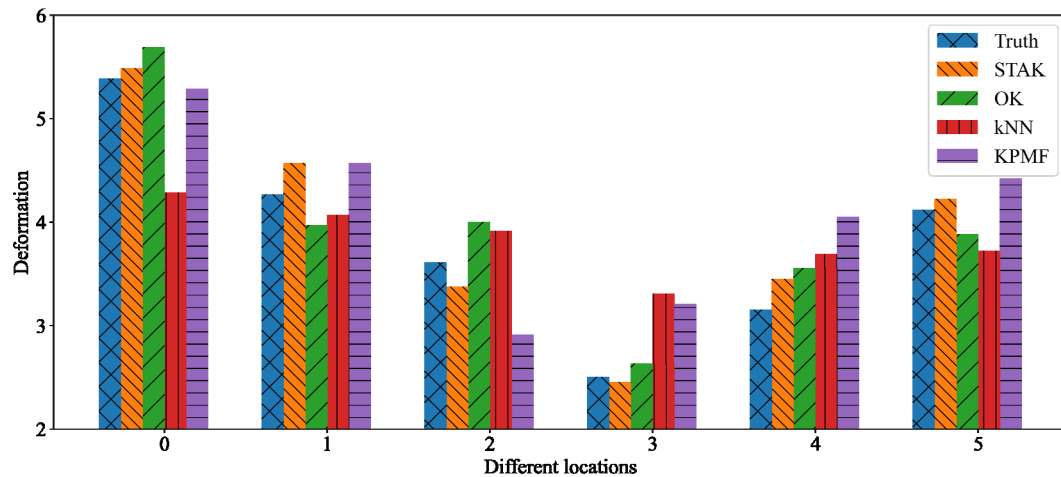


图7 空间维度插值效果可视化

Fig. 7 Visualization of interpolation in the spatial dimension

参考文献

- [1] BOZZANO F, CIPRIANI I, MAZZANTI P, et al. Displacement patterns of a landslide affected by human activities: insights from ground-based InSAR monitoring[J]. *Natural hazards*, 2011, 59(3): 1377-1396.
- [2] GAO W, DAI S, CHEN X. Landslide prediction based on a combination intelligent method using the GM and ENN: two cases of landslides in the Three Gorges Reservoir, China[J]. *Landslides*, 2020, 17(1): 111-126.
- [3] HAJIMORADLOU A, ROBERTI G, POOLE D. Predicting Landslides Using Locally Aligned Convolutional Neural Networks[J]. *arXiv preprint arXiv:1911.04651*, 2019.
- [4] HUANG R Q. LARGE-SCALE LANDSLIDES AND THEIR SLIDING MECHANISMS IN CHINA SINCE THE 20TH CENTURY [J]. *Chinese Journal of Rock Mechanics and Engineering*, 2007(03):433-454. (in Chinese)
- 黄润秋. 20 世纪以来中国的大型滑坡及其发生机制[J]. *岩石力学与工程学报*, 2007(03):433-454.
- [5] ZHU A X, WANG R, QIAO J, et al. An expert knowledge-based approach to landslide susceptibility mapping using GIS and fuzzy logic[J]. *Geomorphology*, 2014, 214: 128-138.
- [6] VAKHSHOORI V, ZARE M. Landslide susceptibility mapping by comparing weight of [键入文字]
- evidence, fuzzy logic, and frequency ratio methods[J]. *Geomatics, Natural Hazards and Risk*, 2016, 7(5): 1731-1752.
- [7] ZHOU J, LU P, YANG Y. Reservoir landslides and its hazard effects for the hydropower station: a case study[C]//Workshop on World Landslide Forum. Springer, Cham, 2017: 699-706.
- [8] GAN B R, YANG X G, ZHOU J W. GIS-based remote sensing analysis of the spatial-temporal evolution of landslides in a hydropower reservoir in southwest China[J]. *Geomatics, Natural Hazards and Risk*, 2019, 10(1): 2291-2312.
- [9] CHEN W, XIE X, PENG J, et al. GIS-based landslide susceptibility modelling: a comparative assessment of kernel logistic regression, Naive-Bayes tree, and alternating decision tree models[J]. *Geomatics, Natural Hazards and Risk*, 2017, 8(2): 950-973.
- [10] KALANTAR B, PRADHAN B, NAGHIBI S A, et al. Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN)[J]. *Geomatics, Natural Hazards and Risk*, 2018, 9(1): 49-69.
- [11] HONG H, POURGHASEMI H R, POURTAGHI Z S. Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models[J]. *Geomorphology*, 2016, 259: 105-118.

- [12] HONG H, PRADHAN B, JEBUR M N, et al. Spatial prediction of landslide hazard at the Luxi area (China) using support vector machines[J]. *Environmental Earth Sciences*, 2016, 75(1): 40.
- [13] LEI T, ZHANG Y, LV Z, et al. Landslide inventory mapping from bitemporal images using deep convolutional neural networks[J]. *IEEE Geoscience and Remote Sensing Letters*, 2019, 16(6): 982-986.
- [14] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[C]//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018: 3634-3640.
- [15] LI Y, YU R, SHAHABI C, et al. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting[C]//*International Conference on Learning Representations*, 2018: 1-16.
- [16] WANG X, MA Y, WANG Y, et al. Traffic flow prediction via spatial temporal graph neural network[C]//*Proceedings of The Web Conference*, 2020: 1082-1092.
- [17] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[J]. *arXiv preprint arXiv:1706.02216*, 2017.
- [18] CHIANG W L, LIU X, SI S, et al. Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks[C]//*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019: 257-266.
- [19] ZENG H, ZHOU H, SRIVASTAVA A, et al. GraphSAINT: Graph Sampling Based Inductive Learning Method[C]//*International Conference on Learning Representations*, 2019:1-19.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All you Need[C]//*Conference on Neural Information Processing Systems*, 2017:6000-6010.
- [21] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. *arXiv preprint arXiv:1710.10903*, 2017.
- [22] DONG J, ZHANG L, LIAO M, et al. Improved correction of seasonal tropospheric delay in InSAR observations for landslide deformation monitoring[J]. *Remote Sensing of Environment*, 2019, 233: 111370.
- [23] CARLÀ T, INTRIERI E, RASPINI F, et al. Perspectives on the prediction of catastrophic slope failures from satellite InSAR[J]. *Scientific reports*, 2019, 9(1): 1-9.
- [24] Hooten M B. *Statistical Analysis of Environmental Space-Time Processes*. Nhu D. Le and James V. Zidek[J]. *Journal of the American Statistical Association*, 2007, 102: 1477-1477.
- [25] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. *arXiv preprint arXiv:1609.02907*, 2016.
- [26] WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32 (1):4-24.
- [27] SONG S, LAN C, XING J, et al. An end-to-end spatio-temporal attention model for human action recognition from skeleton data[C]//*Proceedings of the AAAI conference on artificial intelligence*, 2017, 31(1).
- [28] ZHAO Z, YANG Q, CAI D, et al. Video Question Answering via Hierarchical Spatio-Temporal Attention Networks[C]//*International Joint Conference on Artificial Intelligence*, 2017: 3518-3524.
- [29] CRESSIE N. The origins of kriging[J]. *Mathematical geology*, 1990, 22(3): 239-252.
- [30] TAPOGLOU E, KARATZAS G P, TRICHAKIS I C, et al. A spatio-temporal hybrid neural network-Kriging model for groundwater level simulation[J]. *Journal of hydrology*, 2014, 519: 3193-3203.
- [31] FRANCHI G, YAO A, KOLB A. Supervised deep Kriging for single-image super-resolution[C]//*German Conference on Pattern Recognition*. Springer, Cham, 2018: 638-649.
- [32] WU Y, ZHUANG D, LABBE A, et al. Inductive graph neural networks for spatiotemporal kriging[J]. *arXiv preprint arXiv:2006.07527*, 2020.
- [33] ZHOU T, SHAN H, BANERJEE A, et al. Kernelized probabilistic matrix factorization: Exploiting graphs and side information[C]//*Proceedings of the 2012 SIAM international Conference on Data mining*. Society for Industrial and Applied Mathematics, 2012: 403-414.
- 黎嵘繁, 1998年生, 硕士在读, 主要研究方向是图神经网络, 数据挖掘。
- 钟婷, 1977年生, 博士, 副教授, 主要研究方向是云计算安全, 机器学习, 数据挖掘。
- 吴劲, 1972年生, 博士, 副教授, 主要研究方向是软件定义网络, 大数据技术, 软件安全技术。
- 周帆, 1981年生, 博士, 副教授, CCF会员, 主要研究方向是机器学习, 时空数据挖掘, 数据挖掘和知识发现。

匡平，1977 年生，博士，副教授，CCF 会员，主要研究方向是人工智能，时空数据处理和数据挖掘。



Rongfan Li , born in 1998, postgraduate, is not a member of CCF. His main research interests include graph neural network and data mining.



Fan Zhou , born in 1981, PhD, associate professor, is a member of CCF. His main research interests include machine learning, spatio-temporal data mining, data mining and knowledge discovery.