

# 2005—2020年城市研究中机器学习方法应用进展与综述

曾文菁, 周 恺\*, 熊益群

(湖南大学建筑与规划学院, 长沙 410000)

**摘要:**机器学习作为实现数据挖掘和问题预测的新方法,近年来在城市研究各领域中获得广泛使用,本研究对相关文献进行阶段性总结。首先,本文从数据的类型与选择及预处理出发,介绍了各类机器学习算法的性质特点和适用性,利用 Citespace 对 2005—2020 年以来机器学习与城市研究的交叉领域、热点、前沿和趋势做出分析。其次,研究重点对近五年相关文献中的有监督机器学习算法应用情况进行研读,从城市交通、城市生态、自然地理学、人文地理学四个主要领域展开综述,并解读城市研究中无监督学习、半监督学习及强化学习方法的初步探索。最后,文章总结了机器学习方法的优势,提出未来应当挖掘各类机器学习方法在城市研究多领域多视角中的应用潜力,把握智能技术方法与城市研究高效结合的前沿趋势。

**关键词:**机器学习;城市研究;数据;有监督学习;研究综述

**DOI:**10.48014/cgsr.20220711001

**引用格式:**曾文菁,周恺,熊益群. 2005—2020 年城市研究中机器学习方法应用进展与综述[J]. 中国地理科学评论, 2023, 1(1): 16-30.

由于城市系统集中了大部分的社会经济资源和各类活动设施,城市动态机制的复杂性使得相关研究很难通过简单定性或传统定量方式来充分模拟预测;因此,城市研究的推进一直依赖于分析工具的不断更新完善。“机器学习”是通过计算机学习数据中的内在规律性信息,以获得新的经验知识和解决问题的途径即算法模型,从而使计算机能够像人一样做出智能化决策的分析方法<sup>[1]</sup>。机器学习是人工智能的核心研究领域和主要发展方向之一<sup>[2]</sup>,该方法具有强大的问题处理能力和发展潜力,近些年来被广泛应用于各个领域。

机器学习以其高度发达的感知能力和优越的并行信息处理能力,在城市研究各个领域的应用中取得了持续的进展,近十余年的研究有必要进行阶段性的总结。本文将简要介绍机器学习的数据和算法,从近十五年机器学习与城市研究的交叉领域

中总结研究热点、前沿和趋势,并分析近五年来机器学习在城市交通、城市生态、自然地理学、人文地理学四大领域的应用进展,以探索机器学习在城市研究中未来发展的更多可能性。

## 1 数据和算法

### 1.1 数据准备

数据是信息的载体<sup>[3]</sup>,好的数据是研究得到有意义结果的基本保障<sup>[4]</sup>。机器学习方法所使用数据的性质、数量和质量在避免偏差、保证可靠性和精确度上对模型的性能有不可忽视的影响,所以数据的认识、获取以及处理对于整个机器学习过程是非常重要的(图 1)。

\* 通讯作者 Corresponding author: 周恺, kai\_zhou@hnu.edu.cn

收稿日期:2022-07-12; 录用日期:2022-08-10; 发表日期:2023-03-28

基金项目:国家自然科学基金项目“城市收缩治理的理论模型、国际比较和关键规划领域研究”(项目号:52078197)

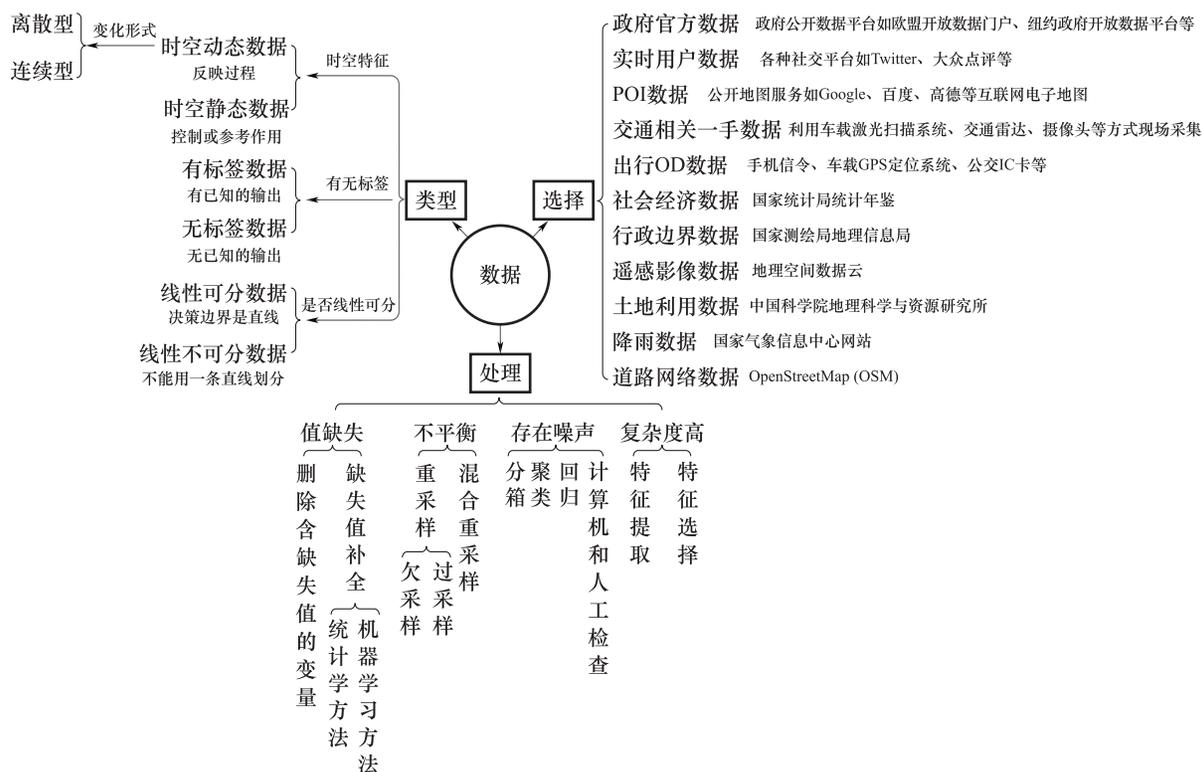


图 1 数据类型、选择及预处理

Fig. 1 Data type, selection and preprocessing

### 1.1.1 数据类型及选择

从时空特征的角度看,数据可以根据变化性强弱划分为时空静态数据和时空动态数据。时空静态数据在一定时空范围内不会变化,在运行过程中主要发挥控制或参考作用,如行政边界数据、地形地貌和土壤数据等;时空动态数据则常随时间或空间改变,是反映过程的数据,如房价数据、气象数据、人口数据、土地利用数据、出行 OD 数据等。时空动态数据又可以按变化形式的不同进一步分为离散型和连续型数据,在变化过程中,离散的数据之间不存在特定的规律联系,在机器学习时通常对其进行分类处理,而连续的数据一般可以通过回归以函数形式表达其变化规律。此外,根据数据标记情况可以分为有标签数据和无标签数据,有标签数据具有已知的输出,而无标签数据没有确定的相应输出。数据还可以分为线性可分数据和线性不可分数据,线性可分数据分类的决策边界是直线的,而线性不可分数据不能用一条直线对其进行划分。最后,数据维度即数据的特征数量也是一个很重要的特性,不同的数据会存

在维度上的差异。因此,各项研究应当根据对数据类型和性质的正确判断选用适合的机器学习算法和建模方式。

整体数据环境的发展,促进了全球数据量的无限扩增,各种数据来源正在逐渐丰富且获取方式更加方便,数据集维护的社区也在不断成熟,这为城市研究创造了更友好的数据条件和支撑。从数据资源类型看,各国政府的开放数据集是最易获取且权威的数据源之一,各个社交平台能获取到很多最新的实时用户数据,现场采集的方式已成为收集一手数据的重要来源,利用移动通信设备等可以收集到实时性强的时空全覆盖出行数据源,等等。城市研究的数据选择,需要在保证数据权威性、真实性和可获得性的基础上,尽量覆盖更丰富的样本类型和保持数据的完整性。根据实际问题和经验,研究需要选取合适的数据数量和数据粒度,并尽可能实现低成本的数据获取,实时性强的数据应注重选择最新数据,保障研究的可靠性。

### 1.1.2 数据处理

数据值缺失是常见数据预处理的问题之一。

当变量的缺失率大,包含的有效值少,且重要性较低时,可以直接将变量删除;否则可以运用统计学方法如均值插补、回归插补等单一插补法以及多重插补法或者机器学习方法来填充缺失值,使信息完整<sup>[5]</sup>。删除变量的方法简单易行,但会导致资源浪费、信息丢失,从而影响数据的客观性和分析结果的正确性<sup>[6]</sup>。插补方法保留了所有信息,但耗时多,而且由于数据缺失机制复杂多样,使得补全处理难度大。因此,针对不同的数据缺失问题还需要进行最佳方法的进一步探索<sup>[7]</sup>。

如因数据收集方法等问题造成了数据比例存在偏差,部分类别的数据较丰富而其他类别中数据数量较少,导致数据不平衡和不具有充分代表性,则可以通过重采样方法(欠采样和过采样)来进行数据集平衡<sup>[8]</sup>。欠采样是指对丰富类的样本进行再抽取,丢弃大量数据仅保留一部分样本以平衡数据;过采样则是重复少数类中的一些样本,增加其基数使数据平衡<sup>[9]</sup>。欠采样随机舍弃样本可能会丢失有用信息,而过采样复制样本将增加过拟合的可能性。因此,为取得良好的数据处理效果,将过采样和欠采样技术相结合进行混合重采样也是不错的解决方案<sup>[10]</sup>。

另外,由于数据源的随机性,数据中难免存在一些错误或偏差,可以用分箱、聚类和回归等方法处理数据噪声。分箱是一种处理数据噪声简单常用的预处理方法,通过数据排序,将相邻数据划分到多个箱子中,再对箱子中的数据按平均值、边界值或中值方法进行平滑处理;聚类方法是将数据分为由相似对象组成的多个类,然后找出并清除那些分布在类之外的值即噪声;回归方法是通过寻找变量之间的关系并用一个函数拟合数据,以此来平滑数据。另外,研究者还可以利用计算机和人工检查结合的方法处理数据噪声问题<sup>[11]</sup>。

针对数据复杂度高的问题,为减轻“维度灾难<sup>①</sup>”,研究通常对数据进行降维。数据降维的方法主要有两种:①“特征提取”是以线性或者非线性的方式使原特征空间发生改变,得到一个维度更低且各维之间更加独立的新特征空间<sup>[12]</sup>;②“特征选择”则是直接在原数据特征的基础上进行筛选,处理后的特征集是原特征集的子集,原特征空间没有改变<sup>[13]</sup>。一般来说,特征提取只适合数值型数据,而

特征选择既适合数值型数据也适合非数值型数据。

## 1.2 机器学习算法

由于数据的特征属性和质量不同、实际研究问题的情况差异,应当结合先验知识选择合适的机器学习算法,算法选择中有关模型效率与结果精度的评估都非常重要。当前常用的机器学习算法包含:有监督学习、无监督学习、半监督学习和强化学习<sup>[14]</sup>。

### 1.2.1 有监督学习

“有监督学习”是通过对有标签数据集的学习得到一个最优模型,再利用这个模型预测新数据的标签。有监督学习主要完成回归和分类任务。其中,回归技术预测的数据对象是连续值,而分类技术预测的数据对象是离散值<sup>[15]</sup>。通常,在有监督机器学习模型构建时,会将数据划分为训练集、验证集和测试集。训练集是由输入和输出结果即标签组成的,用来拟合模型;训练出模型后,利用验证集调整参数,优化模型;得出最优模型后,使用测试集衡量和评价模型性能。当样本数量不足时,为保证模型训练效果,可以不设置验证集。常见的监督学习算法有线性回归(Linear Regression)、逻辑回归(Logistic Regression)、K近邻(KNN)、支持向量机(SVM)、随机森林(RF)等。

### 1.2.2 无监督学习

“无监督学习”的输入数据无标签,因此无确定输出,学习模型可以推断数据内在结构<sup>[16]</sup>。无监督学习的主要任务是根据样本间的相似性对样本集进行聚类,试图使类内差距最小化,类间差距最大化,常见算法如K均值(K-means)聚类等。无监督学习另外一个非常重要的功能就是对数据进行降维处理,在尽量保留相关结构的基础上降低数据的复杂度,使数据之间的潜在关系更加清晰,提高机器学习算法的运行效率,常见的无监督降维方法有主成分分析(PCA)、奇异值分解(SVD)、自编

① 维度灾难:当数据的维度增加到一定程度时,将使学习任务变得相当复杂和困难,导致机器学习性能下降,易产生过拟合和泛化性能差等问题。

码器(Auto-encoder)等算法;除其之外,无监督学习还可以利用孤立森林(IF)、单分类支持向量机(One-class SVM)等算法进行异常的检测,剔除异常点。

### 1.2.3 半监督学习

“半监督学习”的部分数据有标签,其余数据无标签。半监督学习的应用场景包括分类、回归、聚类等<sup>[17]</sup>,可以在有监督算法基础上加入无标记样本,增强有监督学习的效果,也可以在没有监督算法基础上加入有标签的样本,辅助无监督学习。一般而言,半监督学习侧重于前者,即主要是对常用有监督学习算法的改进,一般先用无监督方法学习数据的内在结构,以合理组织数据,再用有监督方法训练模型并实现预测。常用的半监督学习算法有概率生成模型(Generative Model)、半监督支持向量机(S3VM)等。

### 1.2.4 强化学习

“强化学习”的算法逻辑是“主体”根据周围“环境”的“状态”采取“行动”,并根据“收益”情况学习行动方式的过程<sup>[18]</sup>。强化学习系统在行动—收益的环境中获得经验,改进行动方案以适应环境,通过一轮又一轮行动与收益的过程逐渐强化智能<sup>[19]</sup>。常见的强化学习算法有Q-Learning、马尔可夫决策过程(MDP)、策略梯度(PG)、时间差分学习(TD)等。AlphaGo就是强化学习的一个典型应用,此外最前沿的无人驾驶和具有强人工智能的机器人研究,都需要强化学习领域的知识。

## 2 近十五年机器学习与城市研究的结合

### 2.1 交叉研究领域

2005—2020年,机器学习应用于城市研究中的相关文献<sup>①</sup>总体上呈现逐年上升的趋势,2015年以后出现显著增长。机器学习在城市研究中的文献量激增表明,以人工智能与机器学习技术为主导的智慧城市研究框架和建设体系正在形成,智能技术的应用对城市研究的效率与准确性的提升存在积极效果。

从时间上看,机器学习早期在城市中的研究领

域以环境科学和生态学、人文地理学、图像科学为主,早在2005年就已有少量机器学习与环境科学和生态学领域相结合的相关文献,此后逐渐向工程学、计算机科学、公共管理、交通学等方向衍生,后期研究范围更加广泛,涵盖区域和城市规划、气象学、能源等领域,方法也更加多元化。从研究的数量上看,环境科学和生态学的文献量最大,其次是工程学、人文地理学、图像科学、计算机科学等学科。同时,工程学、环境科学和生态学、计算机科学的中介中心度较大,与其他领域具有较密切且重要的联系作用(图2)。

### 2.2 热点、前沿和趋势分析

#### 2.2.1 关键词共引聚类热点分析

由关键词组成的共引聚类关系图可知,不同时期机器学习应用于城市研究的共引文献可根据互相的共引关系归纳为9个有效聚类(图3)。聚类0#“PM<sub>2.5</sub>”、聚类2#“不透水面”、聚类3#“污染物质”和聚类8#“滑坡易发性”获得了大量共被引文献,体现了机器学习在环境科学和生态学领域与工程领域中的广泛应用,证明了机器学习在环境评价、污染分析和预测、灾害探测等实践中的高研究热度。聚类1#“元胞自动机”常与多种机器学习算法结合用于模拟城市增长和土地利用变化,聚类6#“人工神经网络”方法的应用则覆盖了更多元的领域,广泛用于能源需求预测、交通预测等多学科当中。聚类4#“建筑物检测”和聚类7#“土地”反映了机器学习在人文地理学领域中完成了土地覆盖分类、土地利用变化和城市增长模拟等多方面城市研究。另外,聚类5#“城市设计”显示了机器学习在城市空间研究中也具有较高的热度,主要通过研究人们的空间感知、分析城市视觉质量,以支持更科学的城市形态学研究。

① 在Web of Science引文索引数据库中以“machine learning(机器学习)”和“urban studies(城市研究)”为主题词,检索2005—2020年发表的文献(会议录论文除外),共得到832条记录构成本研究的文献数据集。

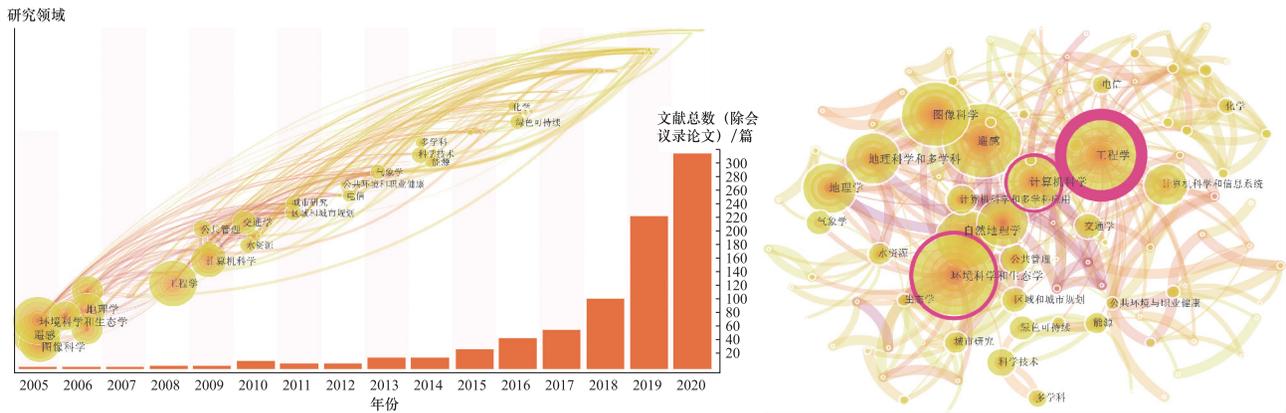


图 2 相关文献数量及交叉研究领域

Fig. 2 The number of relevant literature and intersecting research field

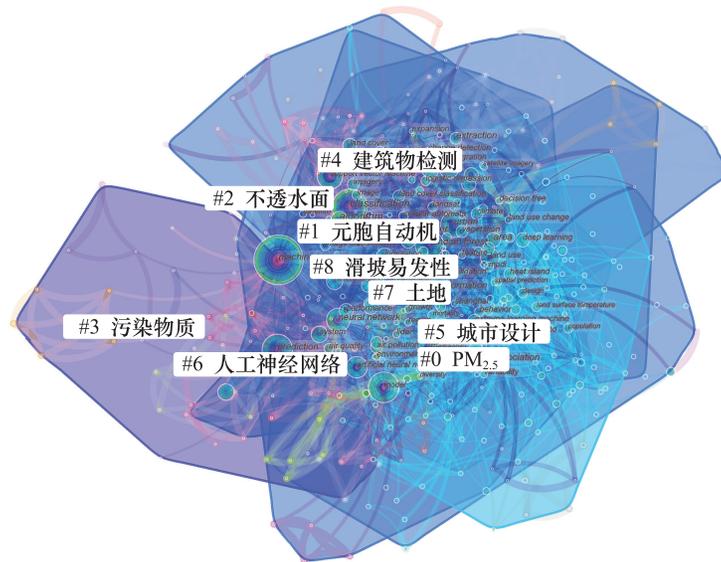


图 3 关键词共引聚类关系

Fig. 3 Keywords co-citation clustering

图 3 中节点的大小反映了关键词的共引频率, 由图可知关键词“机器学习”是近年来共引频率最高的词汇, 2~5 名依次为“分类”“模型”“随机森林”和“预测”。“分类”和“预测”是机器学习在城市研究中的重要功能和用途, 各项研究通过不同机器学习算法实现对城市研究问题的分类预测作用, 这个过程通常借助机器学习“模型”的辅助, 而“随机森林”作为一种包含多个决策树(DT)的学习器, 是目前在城市研究中应用最多的机器学习模型之一, 尤其在 2019 年和 2020 年被引数量大幅度增加。

### 2.2.2 关键词突现的前沿及研究趋势分析

关键词突现反映了引发研究突变关键节点的最早文献及高引用文献情况, 根据关键词突现可以

追溯不同研究前沿的出现时段及彼此之间的衍生关系, 进而更好把握研究前沿并对未来城市研究中机器学习的应用及发展趋势做出合理的判读(图 4)。

2005 年关键词“预测”和“支持向量机(SVM)”的突现及相关文献高被引反映了支持向量机在城市问题预测上发挥了显著的优势作用。2005 年 Lu 等<sup>[20]</sup>考虑到人工神经网络(ANN)局部极小、过拟合训练、泛化性能差等固有缺点, 探讨应用 SVM 预测空气污染物水平的可行性, 通过对支持向量机功能特性的研究, 证明了 SVM 模型在预测不同时间序列的空气质量指数时优于 ANN, 且具有更好的泛化性能。2014 年突现关键词是“地球资源卫星(Landsat)”, 在 2014 年 NASA 更新 Landsat8 数据

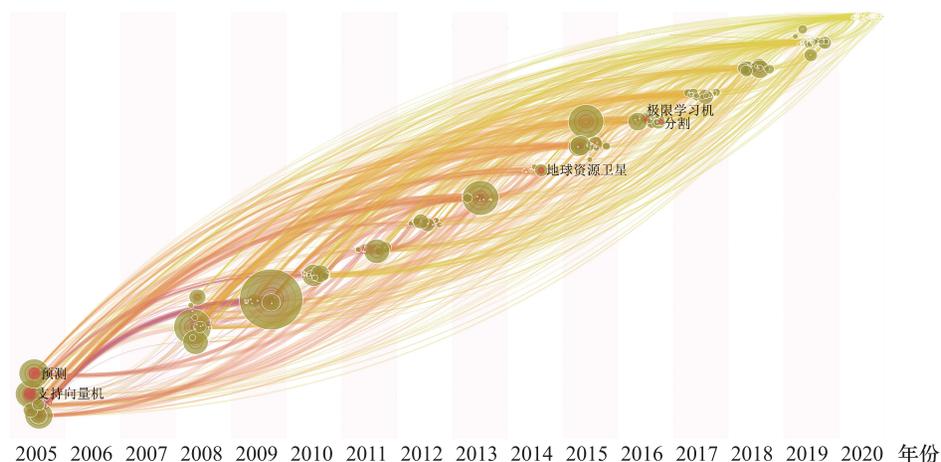


图4 关键词突现

Fig. 4 Keywords burst

后,相关研究开始利用地球资源卫星图像结合机器学习算法解决部分环境资源及生态问题<sup>[21]</sup>。2016年突现关键词“分割”反映了多种机器学习算法在不同城市研究领域中的发挥的分割和分类方面的优势。以 Tan 等<sup>[22]</sup>、Mallinis 等<sup>[23]</sup>、Anandakumar 等<sup>[24]</sup>的研究为代表,体现了随机森林等多种机器学习算法用于分类的良好性能。2016年突现的另一个关键词“极限学习机(ELM)”是2004年由南洋理工大学黄广斌教授提出的一种简单易用、有效的单隐层前馈神经网络(SLFNs)学习算法<sup>[25]</sup>。2016年 Yaseen 等<sup>[26]</sup>的研究表明 ELM 模型在许多统计测量中优于 SVR 和广义回归神经网络(GRNN)模型; Kariminia 等<sup>[27]</sup>、Ivana 等<sup>[28]</sup>还证明了 ELM 在预测人热舒适度时的准确性和效率优于 ANN 和遗传规划等其他方法。

### 3 近五年机器学习在城市研究中的应用

城市研究是一门多领域交叉学科,机器学习方法的引入对于城市系统中的复杂问题具有更好的描述能力,为城市研究提供了一条更加灵活、包容的路径。

#### 3.1 机器学习方法引入的重要性

##### 3.1.1 传统统计学方法在城市研究中存在不足

在当前数据总量丰富,更新速度迅速的时代,传统研究方法在数据挖掘潜力和采集效率方面存

在明显的不足。在许多情境下,传统统计学方法对数据具有一定的要求,通常在对数据做出预设后再明确数学模型来拟合,因此,传统统计模型的优劣易受数据分布的影响。然而,真实数据的分布是难以假设的,在数据庞杂的情况下,不仅需要花费大量时间处理,重复基础工作,还容易因人为主观因素影响最终结果,而且很多复杂的城市问题也很难用比较明确且有限的数学形式来描述。这些因素都使得传统统计分析的效率低,主观性强,精确度低。另外,传统统计学方法处理复杂高维、非线性城市问题的能力有所欠缺,灵活性低,且针对各种不断更新中的城市问题的泛化性能弱。

##### 3.1.2 机器学习方法推动多领域城市研究发展

机器学习的智能特性使得其在数据处理、模型构建等方面均优于传统统计学方法。机器学习方法不仅打破了传统方法在样本大小上的局限,还避免了结论泛化性能普遍较弱的弊端;且不同于传统统计学方法,机器学习方法对数据没有任何假定,其基于算法的模型在获得良好预测效果的同时也便于理解。因此,机器学习在处理城市问题的效率、结果精确度以及适用性上都拥有更大的优势。

#### 3.2 有监督机器学习方法在各领域的应用

从2016—2020年的城市研究相关文献中机器学习方法的具体应用情况来看,有监督算法占大部分。通过收集和归纳整理,近五年来有监督机器学习在城市研究中应用最为广泛的主要领域包含城

市交通、城市生态、自然地理学、人文地理学四大领域。

### 3.2.1 城市交通领域

为最优化配置交通资源,依靠有监督机器学习技术,相关研究实现了向高效率、大规模转变,智能交通迅速发展。深度学习是机器学习的一个分支,具有更复杂的结构和更强的计算能力,因而受到了广泛的关注和应用<sup>[29]</sup>。先进的有监督深度学习算法以其强大的特征学习能力,在辅助城市交通规划中具有良好的潜力。

短时车流量预测是智能交通系统中非常关键的功能。支持向量回归(SVR)、随机森林以及深度学习长短时记忆神经网络(LSTM)等有监督算法可以在保证准确率较高的前提下,使短时车流预测更简便易行且具有较强适应性和鲁棒性,从而提高预测模型的实际应用价值<sup>[30,31]</sup>;此外通过开发定制的机器学习方法还可以有效提高城市交通网络流量预测模型的准确性<sup>[32]</sup>。

除交通流量预测外,有监督机器学习方法还可以应用于更加广泛的的城市交通预测,给智能交通系统提供更全面的决策支持。极限学习机(ELM)作为一种简单易用的单隐层前馈神经网络在检测交通事故时的性能优于其他传统基本模型,能够及时、准确地捕获交通事件,防范事故风险<sup>[33]</sup>;在发生城市轨道交通突发事件后,采用在解决小样本、非线性等复杂问题上具有独特优势的SVR和蒙特卡罗方法,能准确把握乘客路径选择和客流动态分布状况<sup>[34]</sup>;根据交通流量等变量的稳定性,合理选择随机森林、SVR、多层感知器(MLP)和多元线性回归等机器学习模型,可以高效并准确地预测城市交通速度,预防城市交通拥堵<sup>[35]</sup>。另外,不少学者还证明了深度学习在城市交通预测中的优势<sup>[36]</sup>,通过结合丰富多源的移动交通数据可以显著提高大规模交通预测如交通事故安全黑点识别<sup>[37]</sup>、交通流模式分析<sup>[38]</sup>、城市交通客流预测<sup>[39]</sup>等问题的效率和精度。

此外,更多研究还关注在多元交通方式共存的情境下,机器学习在各类交通问题研究中的应用潜力,如共享单车智能预测服务的可行性探究<sup>[40]</sup>、共享单车和出租车之间的最优出行方式选择<sup>[41]</sup>、乘客

出入地铁站的方式选择预测<sup>[42]</sup>、网约车服务与家庭汽车拥有率的关系研究<sup>[43]</sup>等,借助有监督机器学习方法实现了高效率、高精度和强适用性的城市交通预测和分析效果,为城市交通规划提供了科学依据。

### 3.2.2 城市生态领域

机器学习回归算法在城市生态领域实现了广泛运用。城市气温预测和模拟对于极端天气应对和人类的舒适健康等方面有非常重要的意义,在时间维度上,借助SVR算法可以对月平均气温进行预测,其性能优于MLP以及其他方法<sup>[44]</sup>;若将空间维度纳入考虑,利用公开的气象站观测数据和随机森林回归模型,还能绘制表观温度地图<sup>[45]</sup>。城市空气质量以及水体水质同样显著影响着人们的健康和可持续发展,利用空气质量数据,采用SVM、M5P模型树、ANN以及深度学习方法LSTM等能够实现城市空气质量智能预测,监测未来城市空气质量水平<sup>[46,47]</sup>;结合历史水质数据,运用ANN和时间序列分析可以建立水质因子的预测模型<sup>[48]</sup>。

此外,有监督分类和回归算法还以其处理非线性复杂关系时的独特优越性,大量用于城市绿色基础设施现状绘制和规划、土地生态安全格局识别与优化以及城市生态可持续性水平预测当中。如将支持向量机、随机森林、人工神经网络和Naive贝叶斯分类器等机器学习方法应用于卫星图像能完成城市绿色基础设施图的绘制<sup>[49]</sup>;在此基础上,运用人工神经网络(ANN)和自适应的基于网络的模糊推理系统(ANFIS)算法,还可以辅助绿色基础设施规划方案探索<sup>[50]</sup>。此外,运用ANN、决策树(DT)、SVM等进行土地生态安全水平评价以及城市生态系统可持续性水平预测,有助于提出土地生态安全优化策略并科学把控城市未来发展态势<sup>[51,52]</sup>。

随着人们越来越重视城市的可持续性和生活质量,可视街道绿化已经得到越来越多研究的关注,城市绿地的评估重心已经转移到人的感知角度,已有不少学者结合谷歌街景(GSV)图像、兴趣点(PoIs)、移动定位服务(LBS)等数据,采用SVM、深度学习SegNet、PSPNet等机器学习算法辅以空间句法等分析方法和地理信息系统(GIS)等工具,准确提取、测度并评估人本尺度城市绿化<sup>[53-55]</sup>,让

以往难以测度要素的衡量变得具有可操作性,同时还能提高数据处理效率,降低评估的主观性,增强绿化要素提取的自动化和智能化。

### 3.2.3 自然地理学领域

有监督机器学习算法在自然地理领域也取得了一些成就,研究内容主要包括野火、洪水、地震灾害等。

利用高空间分辨率 RapidEye 图像和多时相 Landsat8 数据,采用随机森林分类器,能较为可靠地预测火灾风险<sup>[24]</sup>。另外,考虑到人类对野火发生也具有一定影响作用,Vilar 等<sup>[56]</sup>还运用了机器学习最大熵模型(MaxEnt),将社会经济驱动因素按照时空维度建模,并集成到野火风险评估中。

洪水的发生频率高、造成损失大,具有突发性,且易产生其他灾害的连锁反应,因此洪涝风险的防范在城市规划和管理中至关重要,机器学习为洪涝风险的评估和预测提供了可靠的解决方案。RF、ANN 等多种机器学习算法可以根据洪水风险的重要指标如降雨阈值等,对洪水和非洪水事件进行分类<sup>[57]</sup>,并构建风险评估模型<sup>[58]</sup>;此外,运用最大熵(MaxEnt)和遗传算法规则序列生成(GARP)等机器学习模型还能绘制洪水风险图<sup>[59]</sup>,协助预警系统,为防洪减灾和应急管理提供决策支持。

地震作为一种不可避免的自然灾害,会对城市产生不同程度的生命威胁和财产损失后果,机器学习算法与遥感图像的结合可以对地震损害进行分类<sup>[60]</sup>。

### 3.2.4 人文地理学领域

有监督机器学习分类算法广泛用于土地利用和土地覆盖(LULC)分类及制图中。已有研究利用谷歌街景图像、HJ1B 和 ALOS-PALSAR 等数据结合 SVM、DT、RF、深度卷积神经网络(DCNN)等方法可以帮助完成土地利用分类规则集构建、LULC 分类及制图<sup>[61-64]</sup>。

除了对现状土地利用类型进行分类制图外,机器学习在土地利用变化模拟及预测和城市增长模拟的研究中也起了重要的推进作用,其通过从历史数据中学习全局性的规则来模拟复杂行为,帮助决策者在城市规划周期中实现不同情景的评估。机器学习(ML)模型作为模拟土地利用变化最有效的

模型<sup>[65]</sup>,在获取规则方面,不仅比传统方法具有更高的准确度,还具有更高的灵活性,更擅长处理数据中的非线性关系,以 ANN、SVR、RF、逻辑回归(LR)和多元自适应回归样条(MARS)为例的多种机器学习模型在处理土地利用变化预测和城市增长模拟问题时具有良好的性能<sup>[66]</sup>。元胞自动机(CA)是一种具有模拟复杂系统时空演化过程能力的动力学模型,广泛应用于土地利用变化和城市扩张模拟等研究中。基于多源数据,一些学者在预测城市 LULC 变化和模拟城市扩张的过程中,将 CA 与 SVM、RF、深度学习卷积神经网络(CNN)、循环神经网络和(RNN)、LSTM 和 MLP 等机器学习算法耦合,增强了时空数据挖掘能力,促进了模型整体泛化性能,有效提高了城市扩张推演模拟的精度和效率,改善了城市扩张模拟的性能<sup>[67-70]</sup>。

随着城市不断增长扩张,准确地提取城市区域对城市研究非常重要。与传统的逻辑回归等方法相比,应用机器学习方法能增加此类研究的客观全面性。如结合夜间灯光数据,SVM 分类法能够有效进行大尺度城市土地的提取<sup>[71]</sup>,具有较好的泛化性和较强的鲁棒性,能实现以非常低的成本高效地对城市建成区进行自动识别,为城市规划和城市管理部门提供及时有效的参考信息。此后,深度学习如 CNN 被证明在城市边缘区界定和城市功能区划分时具有突出的优越性<sup>[72]</sup>,对城市区域提取的精确度和稳定性均优于 SVM 模型<sup>[73]</sup>,在描述复杂城市区域方面具有很大潜力<sup>[74]</sup>,多元数据的融合还能增强模型的识别效果。

有监督回归算法如线性回归(LR)和 RF 等还可以用于人口空间分布模拟。对于人口密集区,RF 模型的效果要优于 LR 模型,对于非人口密集区,LR 模型的效果则要优于 RF 模型,根据模型的不同特性进行分区建模完成人口分布格网化模拟,能够有效提高人口数据空间化的精度<sup>[75]</sup>。

另外,一些研究还结合遥感技术、地理信息系统技术等信息地理学工具,开展了城市固废检测<sup>[76]</sup>、城市制图<sup>[77]</sup>、城市修补<sup>[78]</sup>、城市公园选址<sup>[79]</sup>等研究,推动了多源大数据和数字化技术在城市研究中更加深入和前沿的探索。

### 3.3 无监督学习、半监督学习及强化学习的应用优势

从现有研究来看,无监督或半监督学习在数据上具有显著的优势。相比有监督学习过程,无监督或半监督学习极大地降低了数据获取和处理的难度,在模型构建和产生的规则方面也更加自主。无监督及半监督学习方法目前在城市研究中的应用较少,且通常作为整个过程中的预处理方法或辅助手段。各类型已有研究常在研究主体部分展开之前,运用原理简单、可解释性较强、模型收敛速度快、适用于庞大数值型数据的 K-Means 聚类算法和尤其适用于不平衡簇类的层次聚类等,对相关要素进行初步分类(聚类)或区间划分,优化分析思路,为后续研究奠定重要基础<sup>[80-85]</sup>。更多无监督和半监督方法如生成对抗网络(GAN)、t-SNE、PCA 等也能在各项研究进入有监督学习阶段前,预先实现分类和降维等作用,以减少数据清洗和样本集制作的成本与误差,有效降低任务复杂度,显著提高研究效率、可行性和操作性能<sup>[86,87]</sup>。

近年来也已有部分学者将无监督或半监督学习作为主要方法应用到各领域的城市研究当中,解决了交通学领域的城市交通状态分类<sup>[88]</sup>、客流分布模式识别<sup>[89]</sup>以及人文地理学领域的城市形态类型识别等问题<sup>[90]</sup>;此外,K-means 聚类等无监督或半监督学习方法还能量化测度城市尺度下的建筑色彩,协助城市设计色彩分析与风貌引导<sup>[91]</sup>。

强化学习具有较好的决策能力,但欠缺对感知问题的处理能力,而深度学习则具有较强的感知能力,作为机器学习中的两大重要分支,强化学习与深度学习的创新性融合,为复杂城市研究问题的感知决策提供了有益思路。

强化学习 Q learning 与深度神经网络的集成,即深度 Q 网络(DQN),可以在保证强化学习的能力的同时,为高度结构化、高维度的数据提供减弱的特征集,已有学者通过探讨深度强化学习 DQN 在城市中的应用优势,为城市交通研究指明了新的方向,也给智慧城市服务和建设提供了强大的支持<sup>[92]</sup>;由于深度学习 CNN 在图像处理方面具有显著优势,将 CNN 和强化学习结合用于处理图像数据的感知决策任务,还可以优化传统机器学习方法

在大量样本采集以及个性化和效率上的不足<sup>[93]</sup>。

## 4 结论与讨论

相比传统的统计学研究方法,机器学习算法可以显著提升各类城市研究的效率和准确性,促进现代城市研究向智能化、精细化方向发展。同时,深度学习的进一步发展引发了各领域新一轮的城市研究热潮。机器学习方法在城市研究中的应用给规划和决策者提供了量化可视的可靠成果,可以作为未来城市发展方向的有价值依据。

目前大量应用于城市研究中的有监督机器学习算法,通过结合多元数据,已经作为可借鉴范例解决了不少以往使用传统方法存在困难的问题。然而,有监督算法的优化过程不透明且不具有可解释性。相比较而言,无监督算法通常具有更好的解释性,而且在数据获取和处理上更加简单,具有很强的拓展性。作为二者的结合,半监督学习方法在减少标注代价,提高学习机器性能方面也具有十分重要的实际意义。但是不能忽略的是,无监督学习相比有监督学习在结果准确性和有效性上处于弱势。因此,合理选择城市研究中的机器学习算法类型,适当加强无监督学习和半监督学习在各领域中的应用,并发挥强化学习的自动决策优势是一个未来值得探索的有益方向。

机器学习方法的应用已经涵盖了众多领域,解决了各个领域的许多重要问题,但多年来同领域中的研究问题较为类似,较少在研究对象或角度上发生突破性变化。实际上,各个领域的大多数问题均可以变换为机器学习可以解决的分类型、回归和聚类等问题。未来的研究可以通过将新问题进行形式和方式上的转变,从而充分利用机器学习实现更广泛的的城市研究。

**利益冲突:**作者声明无利益冲突。

## 参考文献(References)

- [1] 张润,王永滨. 机器学习及其算法和发展研究[J]. 中国传媒大学学报(自然科学版), 2016, 23(02): 10-18, 24.  
<https://doi.org/10.3969/j.issn.1673-4793.2016.02.002>
- [2] 闫友彪,陈元琰. 机器学习的主要策略综述[J]. 计算机应用研究, 2004, (07): 4-10, 13.

- <https://doi.org/10.3969/j.issn.1001-3695.2004.07.002>
- [3] 陈康,向勇,喻超.大数据时代机器学习的新趋势[J].电信科学,2012,28(12):88-95.  
<https://doi.org/10.3969/j.issn.1000-0801.2012.12.014>
- [4] 韩京宇,徐立臻,董逸生.数据质量研究综述[J].计算机科学,2008,35(2):1-5,12.  
<https://doi.org/10.3969/j.issn.1002-137X.2008.02.001>
- [5] 熊中敏,郭怀宇,吴月欣.缺失数据处理方法研究综述[J].计算机工程与应用,2021,57(14):27-38.  
<https://doi.org/10.3778/j.issn.1002-8331.2101-0187>
- [6] 刘星毅,农国才.几种不同缺失值填充方法的比较[J].南宁师范高等专科学校学报,2007,(03):148-150.  
<https://doi.org/10.3969/j.issn.1674-8891.2007.03.049>
- [7] 陈娟,王献雨,罗玲玲,等.缺失值填补效果:机器学习与统计学习的比较[J].统计与决策,2020,36(17):28-32.  
<https://doi.org/10.13546/j.cnki.tjyc.2020.17.006>
- [8] 王和勇,樊泓坤,姚正安,等.不平衡数据集的分类方法研究[J].计算机应用研究,2008,25(5):1301-1303,1308.  
<https://doi.org/10.3969/j.issn.1001-3695.2008.05.006>
- [9] 杨明,尹军梅,吉根林.不平衡数据分类方法综述[J].南京师范大学学报(工程技术版),2008,(04):7-12.  
<https://doi.org/10.3969/j.issn.1672-1292.2008.04.002>
- [10] 李艳霞,柴毅,胡友强,等.不平衡数据分类方法综述[J].控制与决策,2019,34(04):673-688.  
<https://doi.org/10.13195/j.kzyjc.2018.0865>
- [11] 刘莉,徐玉生,马志新.数据挖掘中数据预处理技术综述[J].甘肃科学学报,2003,(01):117-119.  
<https://doi.org/10.3969/j.issn.1004-0366.2003.01.027>
- [12] 胡洁.高维数据特征降维研究综述[J].计算机应用研究,2008,(09):2601-2606.  
<https://doi.org/10.3969/j.issn.1001-3695.2008.09.009>
- [13] 张丽新.高维数据的特征选择及基于特征选择的集成学习研究[D].北京:清华大学,2004.
- [14] Naqa I E, Murphy M J. What Is machine learning? [M]. //Naqa I E, Li R, Murphy M J. Machine learning in radiation oncology. Cham, Switzerland: Springer International Publishing, 2015: 3-11.  
[https://doi.org/10.1007/978-3-319-18305-3\\_1](https://doi.org/10.1007/978-3-319-18305-3_1)
- [15] 杨剑锋,乔佩蕊,李永梅,等.机器学习分类问题及算法研究综述[J].统计与决策,2019,35(06):36-40.  
<https://doi.org/10.13546/j.cnki.tjyc.2019.06.008>
- [16] 殷瑞刚,魏帅,李晗,等.深度学习中的无监督学习方法综述[J].计算机系统应用,2016,25(08):1-7.  
<https://doi.org/10.15888/j.cnki.csa.005283>
- [17] 梁吉业,高嘉伟,常瑜.半监督学习研究进展[J].山西大学学报(自然科学版),2009,32(04):528-534.  
[https://doi.org/10.13451/j.cnki.shanxi.univ\(nat.sci.\).2009.04.030](https://doi.org/10.13451/j.cnki.shanxi.univ(nat.sci.).2009.04.030)
- [18] 高阳,陈世福,陆鑫.强化学习研究综述[J].自动化学报,2004(01):86-100.  
<https://doi.org/10.16383/j.aas.2004.01.011>
- [19] 黄炳强.强化学习方法及其应用研究[D].上海:上海交通大学,2007.
- [20] Lu W Z, Wang W J. Potential assessment of the “support vector machine” method in forecasting ambient air pollutant trends [J]. Chemosphere, 2005, 59 (5): 693-701.  
<https://doi.org/10.1016/j.chemosphere.2004.10.032>
- [21] Chang N B, Vannah B, Yang Y J. Comparative sensor fusion between hyperspectral and multispectral satellite sensors for monitoring microcystin distribution in lake Erie[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(6): 2426-2442.  
<https://doi.org/10.1109/jstars.2014.2329913>
- [22] Tan K, Jin X, Plaza A, et al. Automatic change detection in high-resolution remote sensing images by using a multiple classifier system and spectral-spatial features [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016, 9(8): 3439-3451.  
<https://doi.org/10.1109/jstars.2016.2541678>
- [23] Mallinis G, Mitsopoulos L, Beltran E, et al. Assessing wildfire risk in cultural heritage properties using high spatial and temporal resolution satellite imagery and spatially explicit fire simulations: The case of Holy Mount Athos, Greece[J]. Forest, 2016, 7(2): 46.  
<https://doi.org/10.3390/f7020046>
- [24] Anandakumar R, Nidamanuri R R, Rishnan R. A super-voxel-based spectro-spatial approach for 3D urban point cloud labeling [J]. International Journal of Remote Sensing, 2016, 37(17): 4172-4200.  
<https://doi.org/10.1080/01431161.2016.1211348>
- [25] 徐睿,梁循,齐金山,等.极限学习机前沿进展与趋势[J].计算机学报,2019,42(07):1640-1670.  
<https://doi.org/10.11897/SP.J.1016.2019.01640>
- [26] Yaseen Z M, Jaafar O, Deo R C, et al. Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq[J]. Journal of Hydrology,

- 2016,542:603-614.  
<https://doi.org/10.1016/j.jhydrol.2016.09.035>
- [27] Kariminia S, Shamsirband S, Motamedi S, et al. A systematic extreme learning machine approach to analyze visitors' thermal comfort at a public urban space[J]. *Renewable and Sustainable Energy Reviews*, 2016, 58: 751-760.  
<https://doi.org/10.1016/j.rser.2015.12.321>
- [28] Ivana B P, Vukadinovi A, Radosavljevi J M, et al. Forecasting of outdoor thermal comfort index in urban open spaces: The Nis fortress case study[J]. *Thermal Science*, 2016, 20(5): 1531-1539.  
<https://doi.org/10.2298/TSCI16S5531B>
- [29] Hao X, Zhang G G, Ma S. Deep learning[J]. *International Journal of Semantic Computing*, 2016, 10(03): 417-439.  
<https://doi.org/10.1142/S1793351X16500045>
- [30] 程政. 城市道路短时车流量预测模型研究[D]. 合肥: 中国科学技术大学, 2016.
- [31] 金玮. 基于周期性分量提取的城市快速路交通流短时预测理论与方法研究[D]. 北京: 北京交通大学, 2017.
- [32] Liu Z Y, Liu Y, Meng Q, et al. A tailored machine learning approach for urban transport network flow estimation[J]. *Transportation Research Part C: Emerging Technologies*, 2019, 108: 130-150.  
<https://doi.org/10.1016/j.trc.2019.09.006>
- [33] Li L C, Qu X, Zhang J, et al. Traffic incident detection based on extreme machine learning[J]. *Journal of Applied Science and Engineering*, 2017, 20(4): 409-416.  
<https://doi.org/10.6180/jase.2017.20.4.01>
- [34] 李春晓. 城市轨道交通突发事件下乘客路径选择行为建模与仿真[D]. 北京: 北京交通大学, 2017.
- [35] Bratsas C, Koupidis K, Salanova J M, et al. A comparison of machine learning methods for the prediction of traffic speed in urban places[J]. *Sustainability*, 2020, 12(1): 142.  
<https://doi.org/10.3390/su12010142>
- [36] Liu Z D, Li Z J, Wu K, et al. Urban traffic prediction from mobility data using deep learning[J]. *IEEE Network*, 2018, 32(4): 40-46.  
<https://doi.org/10.1109/mnet.2018.1700411>
- [37] Fan Z Y, Liu C, Cai D J, et al. Research on black spot identification of safety in urban traffic accidents based on machine learning method[J]. *Safety Science*, 2019, 118: 607-616.  
<https://doi.org/10.1016/j.ssci.2019.05.039>
- [38] Peppas M V, Bell D, Komar T, et al. Urban traffic flow analysis based on deep learning car detection from CCTV image series[J]. *Remote Sensing and Spatial Information Sciences*, 2018, 42(4): 499-506.  
<https://doi.org/10.5194/isprs-archives-xxii-4-499-2018>
- [39] Du B W, Peng H, Wang S Z, et al. Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 21(3): 972-985.  
<https://doi.org/10.1109/tits.2019.2900481>
- [40] Bacciu D, Carta A, Gnesi S, et al. An experience in using machine learning for short-term predictions in smart transportation systems[J]. *Journal of Logical and Algebraic Methods in Programming*, 2017, 87: 52-66.  
<https://doi.org/10.1016/j.jlamp.2016.11.002>
- [41] Zhou X L, Wang M S, Li D Y. Bike-sharing or taxi? Modeling the choices of travel mode in Chicago using machine learning[J]. *Journal of Transport Geography*, 2019, 79: 102479.  
<https://doi.org/10.1016/j.jtrangeo.2019.102479>
- [42] Aqib M, Mehmood R, Alzahrani A, et al. Rapid transit systems: Smarter urban planning using big data, in-memory computing, deep learning, and GPUs[J]. *Sustainability*, 2019, 11(10): 2736.  
<https://doi.org/10.3390/su11102736>
- [43] Sabouria S, Brewer S, Ewing R. Exploring the relationship between ride-sourcing services and vehicle ownership, using both inferential and machine learning approaches[J]. *Landscape and Urban Planning*, 2020, 198: 103797.  
<https://doi.org/10.1016/j.landurbplan.2020.103797>
- [44] Salcedo-Sanz S, Deo R C, Carro-Calvo L, et al. Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms[J]. *Theoretical and Applied Climatology*, 2016, 125: 13-25.  
<https://doi.org/10.1007/s00704-015-1480-4>
- [45] Hung C H, Knudby A, Xu Y M, et al. A comparison of urban heat islands mapped using skin temperature, air temperature, and apparent temperature (Humidex), for the greater Vancouver area[J]. *Science of the Total Environment*, 2016, 544, 15: 929-938.  
<https://doi.org/10.1016/j.scitotenv.2015.12.021>
- [46] Shaban K B, Kadri A, Rezk E. Urban air pollution monitoring system with forecasting models[J]. *IEEE Sen-*

- sors Journal, 2016, 16(8):2598-2606.  
<https://doi.org/10.1109/jsen.2016.2514378>
- [47] Kök I, Şimşek M U, Özdemir S. A deep learning model for air quality prediction in smart cities [A]//2017 IEEE International Conference on Big Data (Big Data) [C]. Boston, MA, USA: IEEE, 2017, 1983-1990.  
<https://doi.org/10.1109/bigdata.2017.8258144>
- [48] Khan Y, Chai S S. Predicting and analyzing water quality using machine learning: A comprehensive model [A]//2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT) [C]. Farmingdale, NY, USA: IEEE, 2016: 1-6.  
<https://doi.org/10.1109/lisat.2016.7494106>
- [49] Kranjcic N, Medak D, Župan R, et al. Machine learning methods for classification of the green infrastructure in city areas[J]. International Journal of Geo-information, 2019, 8(10):463.  
<https://doi.org/10.3390/ijgi8100463>
- [50] Labib S M. Investigation of the likelihood of green infrastructure (GI) enhancement along linear waterways or on derelict sites (DS) using machine learning[J]. Environmental Modelling and Software, 2019, 118: 146-165.  
<https://doi.org/10.1016/j.envsoft.2019.05.006>
- [51] 黄烈佳, 杨鹏. 基于机器学习的武汉城市圈土地生态安全格局识别与优化策略[J]. 生态与农村环境学报, 2020, 36(07):862-869.  
<https://doi.org/10.19741/j.issn.1673-4831.2019.0517>
- [52] Molina-Gómez N I, Rodríguez-Rojas K, Calderón-Rivera D, et al. Using machine learning tools to classify sustainability levels in the development of urban ecosystems[J]. Sustainability, 2020, 12:3326.  
<https://doi.org/10.3390/su12083326>
- [53] Lu Y. The association of urban greenness and walking behavior: Using google street view and deep Learning techniques to estimate residents' exposure to urban greenness[J]. International Journal of Environmental Research and Public Health, 2018, 15:1576.  
<https://doi.org/10.3390/ijerph15081576>
- [54] Ye Y, Richards D, Lu Y, et al. Measuring daily accessed street greenery: A human-scale approach for informing better urban planning practices[J]. Landscape and Urban Planning, 2019, 191:103434.  
<https://doi.org/10.1016/j.landurbplan.2018.08.028>
- [55] Tang Z Y, Ye Y, Jiang Z D, et al. A data-informed analytical approach to human-scale greenway planning: Integrating multi-sourced urban data with machine learning algorithms[J]. Urban Forestry & Urban Greening, 2020, 56:126871.  
<https://doi.org/10.1016/j.ufug.2020.126871>
- [56] Vilar L, Gómez I, Martínez-Vega J, et al. Multitemporal modelling of socio-economic wildfire drivers in central Spain between the 1980s and the 2000s: Comparing generalized linear models to machine learning algorithms[J]. PLoS ONE, 2016, 11(8):e0161344.  
<https://doi.org/10.1371/journal.pone.0161344>
- [57] Ke Q, Tian X, Bricker J, et al. Urban pluvial flooding prediction by machine learning approaches—a case study of Shenzhen city, China[J]. Advances in Water Resources, 2020, 145:103719.  
<https://doi.org/10.1016/j.advwatres.2020.103719>
- [58] Chen J F, Li Q, Wang H M, et al. A machine learning ensemble approach based on random forest and radial basis function neural network for risk evaluation of regional flood disaster: A case study of the Yangtze River Delta, China[J]. International Journal of Environmental Research and Public Health, 2020, 17(1):49.  
<https://doi.org/10.3390/ijerph17010049>
- [59] Eini M, Kaboli H S, Rashidian M, et al. Hazard and vulnerability in urban flood risk mapping: Machine learning techniques and considering the role of urban districts[J]. International Journal of Disaster Risk Reduction, 2020, 50:101687.  
<https://doi.org/10.1016/j.ijdrr.2020.101687>
- [60] Bialas J, Oommen T, Rebbapragada U, et al. Object-based classification of earthquake damage from high-resolution optical imagery using machine learning[J]. Journal of Applied Remote Sensing, 2016, 10(3):036025.  
<https://doi.org/10.1117/1.JRS.10.036025>
- [61] Wang X Y, Guo Y G, He J, et al. Fusion of HJ1B and ALOS PALSAR data for land cover classification using machine learning methods[J]. International Journal of Applied Earth Observation and Geoinformation, 2016, 52:192-203.  
<https://doi.org/10.1016/j.jag.2016.06.014>
- [62] Gu H Y, Li H T, Liu Z Y, et al. A semi-automatic rule set building method for urban land cover classification based on machine learning and human knowledge[J]. Remote Sensing and Spatial Information Sciences,

- 2017,42(2)/W7:729-732.  
<https://doi.org/10.5194/isprs-archives-xlii-2-w7-729-2017>
- [63] Huang B,Zhao B,Song Y M. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery [J]. *Remote Sensing of Environment*, 2018, 214 (1): 73-86.  
<https://doi.org/10.1016/j.rse.2018.04.050>
- [64] Srivastava S,Vargas-Muñoz J E,Tui D. Understanding urban landuse from the above and ground perspectives:A deep learning, multimodal solution[J]. *Remote Sensing of Environment*, 2019, 228: 129-143.  
<https://doi.org/10.1016/j.rse.2019.04.014>
- [65] Aburas M M, Ahamad M S S, Omar N Q. Spatio-temporal simulation and prediction of land-use change using conventional and machine learning models: A review[J]. *Environmental Monitoring and Assessment*, 2019, 191: 205.  
<https://doi.org/10.1007/s10661-019-7330-6>
- [66] Shafizadeh-Moghadam H, Asghari A, Tayyebi A. Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth [J]. *Computers, Environment and Urban Systems*, 2017, 64: 297-308.  
<https://doi.org/10.1016/j.compenurbysys.2017.04.002>
- [67] He J L, Li X, Yao Y, et al. Mining transition rules of cellular automata for simulating urban expansion by using the deep learning techniques [J]. *International Journal of Geographical Information Science*, 2018, 32 (10): 2076-2097.  
<https://doi.org/10.1080/13658816.2018.1480783>
- [68] Mu L, Wang L Z, Wang Y W, et al. Urban land use and land cover change prediction via self-adaptive cellular based deep learning with multisourced data[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019, 12(12): 5233-5247.  
<https://doi.org/10.1109/jstars.2019.2956318>
- [69] 徐朗. 土地生态适宜性约束下的未来城市扩张优化研究[D]. 南京: 南京大学, 2019.
- [70] Xing W R, Qian Y H, Guan X F, et al. A novel cellular automata model integrated with deep learning for dynamic spatio-temporal land use change simulation[J]. *Computers and Geosciences*, 2020, 137: 104430.  
<https://doi.org/10.1016/j.cageo.2020.104430>
- [71] Dou Y Y, Liu Z F, He C Y, et al. Urban land extraction using VIIRS nighttime light data: An evaluation of three popular methods [J]. *Remote Sensing*, 2017, 9 (2): 175.  
<https://doi.org/10.3390/rs9020175>
- [72] 王胜利. 深度学习在城市功能区域划分中的应用研究[D]. 成都: 电子科技大学, 2018.
- [73] 刘星南. 基于深度神经网络的都市边缘区界定研究[D]. 广州: 广州大学, 2020.  
<https://doi.org/10.27040/d.cnki.ggzdu.2020.000056>
- [74] Guo J X, Ren H Z, Zheng Y T, et al. Identify urban area from remote sensing image using deep learning method [C]. *IGARSS 2019- 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019: 7407-7410.  
<https://doi.org/10.1109/igarss.2019.8898874>
- [75] 成方龙, 赵冠伟. 分区策略与机器学习的人口分布精细化模拟[J]. *测绘科学*, 2020, 45(09): 165-173.  
<https://doi.org/10.16251/j.cnki.1009-2307.2020.09.025>
- [76] 肖莎. 基于机器学习的高分辨率遥感影像城市固废检测[D]. 福州: 福州大学, 2018.
- [77] 吕浩博. 基于深度学习的长时间序列城市制图与变化检测研究[D]. 北京: 清华大学, 2018.  
<https://doi.org/10.27266/d.cnki.gqhau.2018.000194>
- [78] 郑屹, 杨俊宴. 基于大规模街景图片人工智能分析的精细化城市修补方法研究[J]. *中国园林*, 2020, 36(08): 73-77.  
<https://doi.org/10.19775/j.cla.2020.08.0073>
- [79] 秦和天. 基于 GIS 和机器学习的未来城市公园选址研究——以常州市为例[D]. 南京: 南京大学, 2020.  
<https://doi.org/10.27235/d.cnki.gnjj.2020.000649>
- [80] Lai Y, Kontokosta C E. Quantifying place: Analyzing the drivers of pedestrian activity in dense urban environments [J]. *Landscape and Urban Planning*, 2018, 180: 166-178.  
<https://doi.org/10.1016/j.landurbplan.2018.08.018>
- [81] Gui R Z, Chen T J, Nie H. In-depth analysis of railway and company evolution of Yangtze River Delta with deep learning[J]. *Complexity*, 2020, 2020: 5192861.  
<https://doi.org/10.1155/2020/5192861>
- [82] 高梦琦. 基于机器学习的城市轨道交通客流需求预测[D]. 北京: 北京交通大学, 2020.  
<https://doi.org/10.26944/d.cnki.gbfju.2020.003623>
- [83] Zhao J H, Fan W, Zhai X H. Identification of land-use characteristics using bicycle sharing data: A deep learning approach [J]. *Journal of Transport Geography*, 2020, 82: 102562.

- <https://doi.org/10.1016/j.jtrangeo.2019.102562>
- [84] 刘镇熙. 基于机器学习算法的中国城市圈层特征测度及其与产业发展的耦合研究[D]. 长沙:湖南大学,2019.  
<https://doi.org/10.27135/d.cnki.ghudu.2019.003889>
- [85] 林豪,江竹,李树彬. 基于机器学习的城市快速路速度—密度关系模型[J]. 西安科技大学学报,2020,40(6):1109-1116.  
<https://doi.org/10.13800/j.cnki.xakjdxsb.2020.0623>
- [86] 廖自然. 基于街景图片机器学习技术的城市建筑风貌分类研究[D]. 南京:东南大学,2019.  
<https://doi.org/10.27014/d.cnki.gdnau.2019.001699>
- [87] Ferreira D L, Nunes B A A, Campos C A V, et al. A deep learning approach for identifying user communities based on geographical preferences and its applications to urban and environmental planning[J]. ACM Transactions on Spatial Algorithms and Systems,2020,6(3):1-24.  
<https://doi.org/10.1145/3380970>
- [88] Cheng Z Y, Wang W, Lu J, et al. Classifying the traffic state of urban expressways: A machine learning approach[J]. Transportation Research Part A: Policy and Practice,2020,137:411-428.  
<https://doi.org/10.1016/j.tra.2018.10.035>
- [89] 张一帆. 基于机器学习的城市轨道交通新线开行下常规公交站点客流预测研究[D]. 北京:北京交通大学,2020.  
<https://doi.org/10.26944/d.cnki.gbfju.2020.002161>
- [90] Rhee J, Llach D C, Krishnamurti R. Context-rich urban analysis using machine learning: A case study in Pittsburgh, PA[A]//Sousa J P, Xavier J P, Henriques G C, Architecture in the Age of the 4th Industrial Revolution- Proceedings of the 37th eCAADe and 23rd SIGRADi Conference[C]. Porto, Portugal: University of Porto,2019:343-352.  
<https://doi.org/10.52842/conf.ecaade.2019.3.343>
- [91] 叶宇,仲腾,钟秀明. 城市尺度下的建筑色彩定量化测度——基于街景数据与机器学习的人本视角分析[J]. 住宅科技,2019,39(05):7-12.  
<https://doi.org/10.13626/j.cnki.hs.2019.05.002>
- [92] Zhao L, Wang J D, Liu J J, et al. Routing for crowd management in smart cities: A deep reinforcement learning perspective[J]. IEEE Communications Magazine,2019,57(4):88-93.  
<https://doi.org/10.1109/mcom.2019.1800603>
- [93] 胡思润,杨晓旭,宋靖华. 基于机器学习的城市生成方法研究[J]. 智能建筑与智慧城市,2019,(11):106-109.  
<https://doi.org/10.13655/j.cnki.ibci.2019.11.036>

# Advances and Review of Machine Learning Applications in Urban Studies from 2005 to 2020

ZENG Wenjing, ZHOU Kai<sup>\*</sup>, XIONG Yiqun

(School of Architecture and Planning, Hunan University, Changsha 410000, China)

**Abstract:** Machine learning, as a new method for data mining and problem prediction, has been widely used in various fields of urban studies in recent years, which requires a periodical summary of relevant literature. Start with data types, selection and preprocessing, this paper introduces the characteristics and applicability of various machine learning algorithms, and analyzes the cross-fields, hot spots, frontiers and trends of machine learning and urban studies from 2005 to 2020 by using Citespace. Second, focusing on the application of supervised machine learning algorithms from relevant literature in the past five years, a review is made from four main aspects including urban traffic, urban ecology, physical geography, human geography, and the tentative explorations of unsupervised learning, semi-supervised learning and reinforcement learning method in urban studies are unscrambled as well. Finally, the advantages of machine learning methods are summarized, and it's proposed that the application potential of various machine learning methods in multiple fields and perspectives of urban research should be explored in the future, and the cutting-edge trend of efficient combination of intelligent technology and methods with urban research should be grasped.

**Keywords:** Machine learning; urban studies; data; supervised learning; research review

**DOI:** 10.48014/cgsr.20220711001

**Citation:** ZENG Wenjing, ZHOU Kai, XIONG Yiqun. Advances and review of machine learning applications in urban studies from 2005 to 2020[J]. Chinese Geography Sciences Review, 2023, 1(1): 16-30.

Copyright © 2023 by author(s) and Science Footprint Press Co., Limited. This article is open accessed under the CC-BY License (<http://creativecommons.org/licenses/by/4.0/>).

