

# 机器学习在国内城市洪涝灾害研究中的 进展与热点分析

杨梦杰<sup>1,2,3,4</sup>, 吕永鹏<sup>1,2,3,4</sup>, 东阳<sup>1,2,3,4</sup>

[1.上海市市政工程设计研究总院(集团)有限公司,上海市 200092; 2.江苏申武先进技术研究院有限公司,江苏 常州 213168;  
3.长三角绿色建筑与韧性城市产业技术创新中心,江苏 常州 213168; 4.上海城市排水系统工程技术研究中心,上海市 200092]

**摘要:**机器学习作为一种新兴的技术方法,近年来在城市洪涝灾害研究中的应用价值不断凸显。利用文献计量可视化工具 CiteSpace 对 1986—2024 年来国内基于机器学习的城市洪涝灾害研究进行梳理与分析,揭示研究领域的整体发展脉络、研究热点及未来趋势。主要结论:(1)机器学习在国内城市洪涝灾害中的研究成果数量经历平稳—升温—波动—快速上升 4 个阶段;(2)研究作者和研究机构呈现出一定程度的聚集;发文章期仅约有一半的比例属于核心期刊,且 CSCD 与 CSSCI 期刊占比不高;(3)机器学习在城市洪涝灾害研究中呈现内容多样化特征,在以往洪水预报研究为主逐步向洪涝灾害风险评估为趋势转变。

**关键词:**机器学习;城市洪涝灾害;CiteSpace;文献计量

**中图分类号:** TV87

**文献标志码:** A

**文章编号:** 1009-7716(2024)10-0127-06

## 0 引言

随着全球气候变化不断加剧以及城镇化快速发展,城市洪涝灾害事件的发生频率、强度、广度也在逐步增加,如近些年的郑州“2021.7.20”、北京“2023.7.31”、深圳“2023.9.7”、广东梅州“2024.6.16”及湖南资兴“2024.7.26”等特大暴雨灾害事件。城市内涝灾害的频发已成为影响城市公共安全的突出问题,也是制约社会经济可持续发展的重要因素<sup>[1-3]</sup>。为缓解洪涝灾害对城市发展的影响,亟需建立准确、高效的市内涝模拟评估模型,开展城市洪涝过程模拟分析与灾害风险评估,从而为城市防洪排涝与防灾减灾等提供决策参考<sup>[4-5]</sup>。

当前国内有关城市洪涝灾害的研究主要集中在洪涝灾害的机理分析<sup>[6]</sup>、洪水灾害风险认知研究<sup>[7]</sup>、洪涝灾害分析模型研究<sup>[8]</sup>、洪涝灾害损失评估研究<sup>[9]</sup>、洪涝灾害遥感监测研究<sup>[10]</sup>等。此外,已有学者尝试将机器学习方法引入到城市洪涝灾害研究中,

已开展了较为丰富的探索,如基于数据驱动的 CNN 方法开展洪水预测分析<sup>[11]</sup>、基于深度学习方法开展道路积水分析<sup>[12]</sup>等。但是,围绕机器学习在国内城市洪涝灾害研究中的应用开展整体进展的系统分析仍相对较少,这使得难以以为该领域研究发展脉络的系统把握提供重要指导。

因此,本文依托文献计量学方法,借助可视化的知识图谱分析手段,绘制机器学习在国内城市洪涝灾害研究应用领域的知识图谱,厘清该领域研究的主要学者与知识网络结构、发展历史与演进脉络、研究热点与未来趋势,旨在为我国城市洪涝灾害研究与防治提供理论参考与启示。

## 1 研究方法与数据来源

CiteSpace 作为科学文献计量分析软件,通过计量建模和图谱绘制,对文献信息的测度和相似性进行可视化分析,可实现特定研究领域的前沿辨识和热点演进趋势的分析<sup>[13-14]</sup>。本文以 CiteSpace 为研究工具,对国内机器学习在城市内涝风险评估研究中应用的发展历史、前沿热点及发展趋势进行可视化分析。

以中国知网 CNKI 数据库为数据来源,以主题为检索项,文献数据的检索公式为:(内涝+洪涝+暴雨+雨洪+洪水+防洪+防汛+排涝)×(机器学习+人

收稿日期:2024-03-09

作者简介:杨梦杰(1995—),男,博士,工程师,从事智慧排水等相关研究工作。

通信作者:东阳(1986—),女,博士,高级工程师,从事排水运行分析、智慧排水、排水模型等研究工作。电子邮箱:dongyang@smedi.com

工智能+大数据+AI+数据驱动+深度学习+神经网络+随机森林+支持向量机)。检索文献的类型为学术期刊,检索时间为1986年1月—2024年9月,通过筛选、剔除与本研究主题不相关的文献,最终得到1320篇中文文献,时间跨度为1986—2024年<sup>[13,15]</sup>。

## 2 结果分析

### 2.1 研究成果数量特征

研究成果的数量变化可反映出机器学习在国内城市洪涝灾害研究中的整体趋势。通过趋势线可以得知(见图1),该领域研究成果数量整体上呈现出上升的态势。从图中可以发现研究成果数量可大致分为四个阶段:(1)1986—1997年,相关研究处于初始阶段,年发文量较为平稳,处于0~10篇之间;(2)1998—2009年,研究成果数量开始升温,呈现出逐步增长趋势;(3)2010—2013年,出现小幅下降,并在20篇左右波动;(4)2014—2024年,表现出飞速增长的趋势,并于2023年达到150篇。随着物联网、大数据、人工智能等新技术的日趋成熟,以及智慧水利、智慧城市建设的现实需求,未来机器学习在城市洪涝灾害中的研究应用将会持续增长。

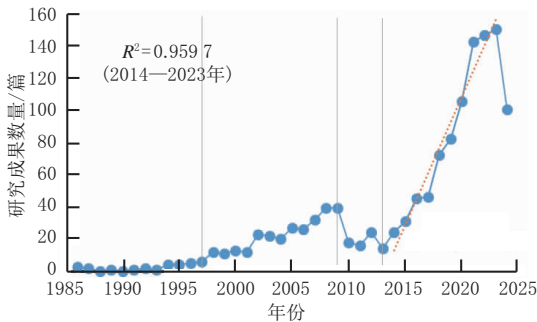


图1 CNKI文献1986—2024年研究成果数量变化

### 2.2 研究作者与机构特征

为进一步了解研究作者与研究机构在机器学习应用于国内城市洪涝灾害研究的动态和实际贡献,将各个研究作者与研究机构的研究成果数量进行了统计。李致家、郭生练、刘业森及李鸿雁均发表了超过10篇的相关论文(见表1)。其中,李致家第一篇有关机器学习在城市洪涝灾害领域应用的文章发表于1997年的《湖泊科学》期刊上,题目为“洪水预报中特征值预报的若干数学方法比较”。该论文并对比分析了统计回归模型、神经网络模型和模糊回归模型在洪水预报中的优劣<sup>[16]</sup>。

为更直观地展示相关机构的发文情况,对TOP20发文量的高频研究机构进行了统计(见表2)。排名前

表1 1986—2024年TOP20的高频研究作者 单位:篇

序号	作者	数量	序号	作者	数量
1	李致家	12	11	徐宗学	7
2	郭生练	11	12	王本德	7
3	刘业森	11	13	苑希民	7
4	李鸿雁	10	14	金菊良	7
5	刘媛媛	9	15	解建仓	6
6	张俊	8	16	石朋	6
7	胡彩虹	8	17	魏一鸣	6
8	金保明	7	18	陈华	6
9	许月萍	7	19	江衍铭	6
10	张小峰	7	20	王俊	6

20的高频研究机构共发文500篇,约占总论文数量的38%。其中,河海大学以123篇论文的数量,压倒性的成为发文量最多的研究机构;其次为武汉大学和中国水利水电科学研究院,分别发文51篇和42篇。根据研究机构的性质可以发现,多数为高等院校,但也有不少研究机构为科研院所;此外,还有一家为规划设计单位。这可侧面反映出该研究领域不仅存在显著的理论研究意义,并且具有较好的工程实践应用价值。

表2 1986—2024年TOP20的高频研究机构 单位:篇

序号	机构名称	数量
1	河海大学	123
2	武汉大学	51
3	中国水利水电科学研究院	42
4	西安理工大学	23
5	北京师范大学	22
6	中山大学	20
7	大连理工大学	20
8	四川大学	19
9	南京信息工程大学	19
10	浙江大学	18
11	天津大学	18
12	清华大学	16
13	中国水利部防洪抗旱减灾工程技术研究中心	15
14	中国科学院地理科学与资源研究所	15
15	南京水利科学研究院	15
16	郑州大学	15
17	长江水利委员会水文水资源勘测局	14
18	长江勘测规划设计研究院	12
19	中国科学院大学	12
20	华北水利水电学院	11

### 2.3 主要发文期刊特征

对研究成果的发文期刊类型进行统计整理,1 320 篇研究成果发布于共 200 种期刊。TOP20 的高频发文期刊的成果数量均大于 10 篇(见表 3),合计发表论文 453 篇,约占总论文数量的 34%。其中,《中国防汛抗旱》期刊发文量最高,达 51 篇,其次为《人民长江》和《水电能源科学》,分别发文 33 篇和 31 篇。对高频发文期刊进行梳理可以发现,排名前 20 的期刊中,水利相关领域的期刊数量高达 16 种,发文量为 386 篇,约占 TOP20 期刊发文量的 85%。另有 2 种期刊涉及灾害领域,剩下 2 种期刊则分别属于气象与自然科学领域。对研究成果所发的期刊类型进行统计可以发现(见图 2),其中有 614 篇文章发表于北大核心期刊,占总发文量的 47%。分别有 320 篇和 35 篇论文发表于 CSCD 与 CSSCI 期刊,合计占总发文量的 27%。

表 3 1986—2024 年 TOP20 的高频发文期刊 单位:篇

序号	期刊名称	数量	序号	期刊名称	数量
1	中国防汛抗旱	51	11	水文	20
2	人民长江	33	12	水利学报	19
3	水电能源科学	31	13	人民珠江	18
4	中国农村水利水电	28	14	水科学进展	18
5	人民黄河	28	15	自然灾害学报	17
6	中国水利	27	16	水力发电	15
7	水利信息化	26	17	水资源保护	15
8	水利水电技术	25	18	河海大学学报 (自然科学版)	14
9	灾害学	25	19	气象	11
10	水力发电学报	21	20	江苏水利	11

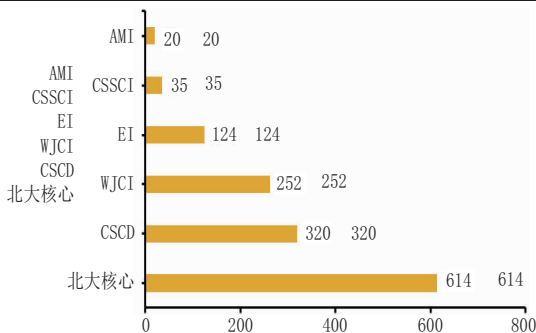


图 2 1986—2024 年研究成果所发期刊的类型数量

### 2.4 研究热点及趋势

#### 2.4.1 关键词聚类分析

为对关键词进行全面深入的探究,本研究运用 LLR 算法对关键词进行聚类分析,生成关键词聚类知识图谱(见图 3 和图 4)。参考相关研究可知<sup>[13,17-18]</sup>,

利用 CiteSpace 进行关键词聚类分析,结合图谱的各项数值可知,本研究得到的关键词聚类的模块值 Modularity(Q 值)为 0.674 (>0.3),说明聚类结构显著;关键词聚类平均轮廓值 Silhouette(S 值)为 0.895 1 (>0.7),说明结果是可信的。因此,本研究分析得到的关键词聚类效果显著,结果是合理可靠的。

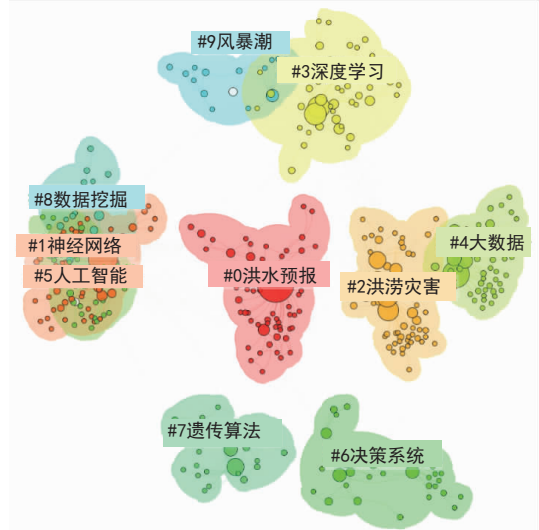


图 3 1986—2024 年关键词聚类图谱

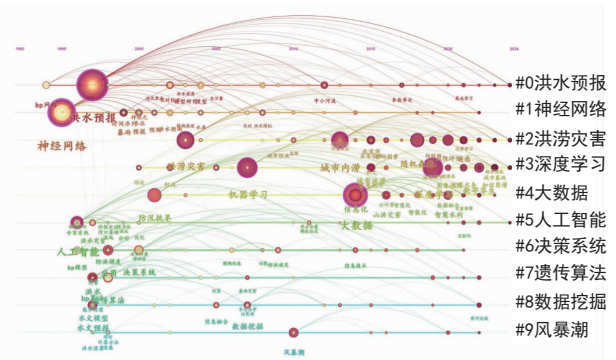


图 4 关键词聚类时间线

机器学习在城市洪涝灾害研究领域的应用可被大致分为 10 个聚类标签,由热点最高到最低分别为 #0 洪水预报、#1 神经网络、#2 洪涝灾害、#3 深度学习、#4 大数据、#5 人工智能、#6 决策系统、#7 遗传算法、#8 数据挖掘、#9 风暴潮。通过 CiteSpace 对每个关键词聚类的统计分析,可以识别出每个关键词聚类对应的大小(Size)、聚类平均轮廓值(Silhouette)、平均年份(Year)以及 LLR 算法下的高度关联的关键词(见表 4)。综合关键词聚类图谱与关键词聚类信息统计结果,可深入分析机器学习在城市洪涝灾害研究中的热点主题。

#0 洪水预报、#1 神经网络、#2 洪涝灾害、#7 遗传算法及 #9 风暴潮聚类侧重于通过 BP 神经网络、随机森林及 LSTM 神经网络模型等机器学习算法实现洪涝灾害的风险评估、演算模拟和预警预报。#0 洪

表4 1986—2024年关键词聚类信息统计表

ID	大小(Size)	聚类平均轮廓值(Silhouette)	年份(Mean)	LLR
0	64	0.799	2009	洪水预报;模型;实时校正;BP网络;模型研究
1	63	0.955	2005	神经网络;暴雨;预测;水位;洪水预测
2	61	0.833	2017	洪涝灾害;随机森林;城市内涝;洪水预报;风险评估
3	52	0.832	2018	深度学习;机器学习;智慧水务;数字孪生;城市暴雨
4	52	0.828	2018	大数据;信息化;防汛抗旱;智能化;云计算
5	33	0.938	2006	人工智能;专家系统;洪水灾害;颗粒组成;养分含量
6	31	0.948	2005	决策系统;bp模型;防洪减灾;应用;水位流量关系
7	27	0.942	2007	遗传算法;洪水;BP算法;灾害;暴雨灾害
8	21	0.929	2008	数据挖掘;水文预报;水文模型;数据仓库;信息融合
9	16	0.982	2007	风暴潮;洪水演算;漫堤;热带气旋;LSTM神经网络模型

水预报聚类主要关注于依托BP神经网络等机器学习方法对洪涝模拟模型进行研究,包括模型参数的实时校正等。如黄一昕等人将实时校正方法总结为终端误差校正和过程误差校正,并系统阐述了5种代表性的洪水预报实时校正技术的特点及适应性<sup>[19]</sup>。#1神经网络聚类重点关注通过神经网络方法对洪涝过程的水文等要素进行预测模拟。如侯精明等人基于AI技术,结合物理机制模型,构建洪涝过程的快速模拟预报机器学习模型,并且结果表明AI模型在精度相似的水平上,较物理机制模型可实现速度提升300多倍<sup>[20]</sup>。#2洪涝灾害聚类则表现为通过随机森林算法对洪涝灾害的风险进行模拟评估。如徐溱等人使用随机森林等机器学习方法对不可移动文物遭受暴雨灾害的风险进行研究<sup>[21]</sup>。#7遗传算法聚类主要是通过遗传算法或BP算法开展暴雨与洪涝灾害的分析与评估。如符洪恩等人使用遗传算法参数优化的支持向量机对城市暴雨洪涝灾害危险性进行预测分析<sup>[22]</sup>。#9风暴潮聚类是多以LSTM神经网络模型为技术方法开展风暴潮漫堤研究。如谢文鸿等人使用卷积长短时记忆网络(ConvLSTM)机器学习算法进行风暴潮漫滩的预报分析<sup>[23]</sup>。

#3深度学习、#4大数据及#8数据挖掘聚类多强调机器学习等新技术如何助力城市洪涝灾害的研究,具体来说包括大数据、云计算、数字孪生、数据挖掘、信息融合等新技术手段如何助力推进城市洪涝治理的信息化、数字化及智慧化,加快智慧水务的发展。#3深度学习聚类关注于通过机器学习模型针对城市暴雨洪涝灾害的研究,加快推进智慧水务建设。如张然等人对传统机器学习模型和深度学习模型进

行了划分与阐述<sup>[24]</sup>。#4大数据聚类体现的是通过大数据挖掘与分析、云计算等多种技术手段增强防汛抗旱能力,推进信息化与智能化建设。如韩岳麒指出空间遥感大数据可提供更准确、及时且丰富的空间信息,可为防灾减灾提供重要的数据支撑<sup>[25]</sup>。#8数据挖掘聚类强调通过数据挖掘技术增强水文模型的预警预报性能。如刘和昌等人基于数据挖掘技术构建了洪水预报模型并实现了较高的预报精度<sup>[26]</sup>。

#5人工智能和#6决策系统则强调机器学习与人工智能技术在防洪减灾领域的应用。#5人工智能聚类主要侧重于结合人工智能与专家系统应用到防洪减灾决策支持系统中。如阎俊爱结合城市防洪减灾特点,设计了基于GIS的耦合人工智能技术和专家系统的智能型防洪减灾决策支持系统<sup>[27]</sup>。#6决策系统聚类主要是构建综合模拟分析、诊断评估及防洪减灾优化的辅助决策系统。如戚晓明等人开发了淮河涉河工程的防洪影响评价辅助决策系统,实现了淮河蚌埠段涉河桥梁工程防洪管理的信息化水平<sup>[28]</sup>。

#### 2.4.2 研究趋势分析

关键词的突现分析是指短时期内出现频次突然增长的关键词,可反映相关领域具有突出推动作用的技术方法和理论概念等研究热点及其演变,进而揭示研究领域前沿的动态变化趋势<sup>[13,29]</sup>。本文主要选取排名前15的突现关键词进行分析,了解机器学习在城市洪涝灾害研究的前沿动态(见图5)。

从突现强度、持续时间和出现时间三个维度对突现关键词进行总结可以发现,(1)突现强度最强的突现词为“大数据”,于2016年开始突现,于2021年热度下降。这主要是因为大数据概念在2010年左右

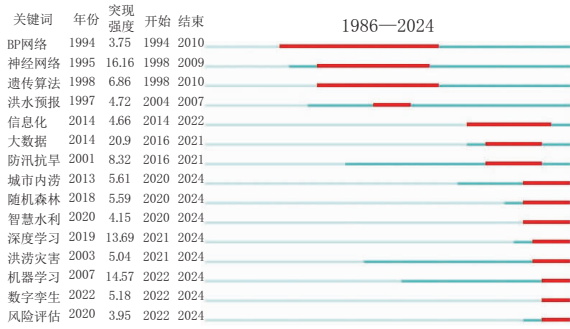


图5 前15个突现关键词图谱

的快速兴起,并且国务院于2015年8月印发了《促进大数据发展行动纲要》,促进了大数据技术在城市洪涝治理研究中的应用。(2)持续时间最长的突现词为“BP网络”,即“BP神经网络”,一直从1994年兴起持续到2010年。BP神经网络算法是1986年被首次提出,后被广泛应用于数据挖掘与机器学习等领域。(3)出现时间最晚的突现词为“机器学习”、“数字孪生”以及“风险评估”,三者均为2022年开始兴起并持续至今。这也反映了当前领域的最前沿研究热点关注于机器学习和数字孪生等技术的应用以及对洪涝灾害风险的评估。此外,也需关注到“城市内涝”、“随机森林”、“智慧水利”、“深度学习”以及“洪涝灾害”也是近5a突现的关键词并且持续至今。

此外,也可将识别出来的突现词划分为三个类别:技术方法、理论概念、研究内容。从技术方法来看,主要包括“BP神经网络”、“遗传算法”、“随机森林”及“数字孪生”;可看出BP神经网络、遗传算法以及随机森林是机器学习在城市洪涝灾害研究中的热门算法。从理论概念来看,主要涉及“信息化”、“大数据”、“智慧水利”、“深度学习”及“机器学习”;这凸显出当前国内较为热门的概念应用。从研究内容来看,主要为“洪水预报”、“防汛抗旱”、“城市内涝”、“洪涝灾害”及“风险评估”。这体现出机器学习方法在城市洪涝中的具体应用主要围绕洪涝预报和灾害评估等方面。

根据关键词突变图谱,发现机器学习在洪涝灾害风险评估中的应用或将是未来研究的热点,将依托于机器学习的数据驱动模型和传统水文水动力机理模型相耦合,根据实际需求不断研发城市洪涝灾害风险模拟与评估的新模型、新方法,是未来研究的重要内容。

### 3 结论与展望

本文基于可视化图谱分析技术对机器学习在城市洪涝灾害研究中的应用进行回顾与分析,围绕研

究成果数量、研究作者、研究机构、发文期刊以及关键词等,揭示机器学习在国内城市洪涝灾害中的研究热点、演进脉络及未来趋势。

研究发现:(1)机器学习在国内城市洪涝灾害中的研究成果数量呈现出整体上升的趋势,经历平稳-升温-波动-快速上升4个阶段。(2)从研究作者和机构来看,有4位作者的成果数量达到10篇以上,其中李致家的发文数量最多,达到12篇;机器学习在城市洪涝灾害研究中的应用主要集中在高校,其次为科研院所,河海大学的研究成果数量以压倒性优势占据第一,达到123篇。(3)从发文期刊来看,11个期刊发文数量超过20篇,其中《中国防汛抗旱》期刊发文量最多,达51篇,约有一半的论文发表在北大核心期刊上,而在CSCD和CSSCI期刊上的发文量约占总发文量的27%。(4)本文将机器学习应用于城市洪涝灾害研究中的关键词分为10个聚类。其中持续时间最长的关键词为“BP神经网络”,出现时间最晚的关键词为“机器学习”、“数字孪生”以及“风险评估”。研究热点趋势可划分为三类,以BP神经网络为主的技术方法、以信息化为主的理论概念及以洪水预报为主的研究内容。机器学习在城市洪涝灾害研究中呈现内容多样化特征,从以往洪水预报研究为主逐步向洪涝灾害风险评估为趋势转变。

总的来说,实现依托于机器学习的数据驱动模型和传统水文水动力机理模型的耦合,根据实际需求不断研发城市洪涝灾害风险模拟与评估的新模型、新方法,是未来研究的重要内容。

#### 参考文献:

- [1] 徐宗学,陈浩,任梅芳,等.中国城市洪涝致灾机理与风险评估研究进展[J].水科学进展,2020,31(5):713-724.
- [2] Aerts J. C., Botzen W. J., Clarke K. C., et al Integrating human behaviour dynamics into flood disaster risk assessment[J].Nature Climate Change,2018,8(3):193-199.
- [3] Piadeh, F., Behzadian, K., Alani, A. M. A critical review of real-time modelling of flood forecasting in urban drainage systems[J]. Journal of Hydrology,2022(607):127476.
- [4] Liao, Y., Wang, Z., Chen, X., et al. Fast simulation and prediction of urban pluvial floods using a deep convolutional neural network model[J]. Journal of Hydrology, 2023,624,129945.
- [5] Piadeh F., Behzadian K., Chen A. S., et al. Enhancing urban flood forecasting in drainage systems using dynamic ensemble-based data mining[J]. Water Research,2023(247):120791.
- [6] 夏军强,董柏良,李启杰,等.近年城市洪涝致灾的水动力学机理分析与减灾对策研究[J].中国防汛抗旱,2022,32(4):66-71.

