

DOI:10.3876/j.issn.1000-1980.2024.02.003

基于流域日降水量图的相似性搜索方法

余宇峰¹, 贺新固¹, 张潇², 万定生¹, 杨永杰¹

(1. 河海大学计算机与软件学院, 江苏 南京 211100; 2. 长江水利委员会水文局, 湖北 武汉 443010)

摘要: 为了提升降水量图相似性分析的精确度, 提出了一种基于流域日降水量图的相似性搜索方法, 该方法从降雨图像中提取日降水量、降雨空间分布和降雨中心特征, 并分别计算各特征的相似距离, 同时通过提出的归一化折旧累积增益改进粒子群优化的集合加权方法对3个特征的相似距离进行加权融合, 作为降雨图像的相似性度量。嘉陵江流域实例验证表明: 该方法能够更好地表征降水量图的时空特征, 可快速地从降水量图中检索出相似的降雨过程。

关键词: 降水量图; 特征提取; 相似性分析; 多元特征距离融合; 改进粒子群算法

中图分类号: TP39 文献标志码: A 文章编号: 1000-1980(2024)02-0019-09

A rainfall similarity search method based on daily precipitation images of watershed

YU Yufeng¹, HE Xingu¹, ZHANG Xiao², WAN Dingsheng¹, YANG Yongjie¹

(1. College of Computer Science and Software, Hohai University, Nanjing 211100, China;
2. Bureau of Hydrology, Changjiang Water Resources Commission, Wuhan 443010, China)

Abstract: In order to improve the accuracy of similarity analysis of precipitation images, a rainfall similarity search method based on daily precipitation images of watershed is proposed. The algorithm first extracts the daily precipitation, precipitation distribution, precipitation center characteristics of the precipitation images, and calculates the similarity distance of each characteristic respectively. Then, an ensemble weighting method of normalized discounted cumulative gain-improved particle swarm optimization is proposed to weight and fuse the three extracted features as the similarity measure of precipitation image. The similarity search experiments of daily precipitation images on the Jialing River Basin illustrate that the method proposed in this paper can better characterize the spatiotemporal characteristics of the precipitation image and quickly discover similar rainfall processes from precipitation images.

Key words: precipitation image; feature extraction; similarity analysis; multivariate feature fusion; improved particle swarm optimization

近年来, 极端降雨引起的山洪暴发已造成严重的社会和经济损失^[1]。受强度、分布等降雨因素的影响, 洪水发生的时间、地点、级别、过程等都存在很大的不确定性, 这给洪水的早期预警和预防造成了很大的困难。因此, 从历史降雨事件中提取降雨-径流过程的时空特征, 识别和分类降雨-径流过程, 发现相似的降雨-洪水模式, 从而为水文预报和水资源利用提供指导和技术支持, 已成为水文水资源应用领域的迫切任务^[2]。

降雨-洪水相似度分析采用模糊数学、数据挖掘和机器学习方法, 通过定义相似性度量方法, 从历史水文时间序列模式中识别与当前实时降雨-洪水序列相似(最接近)的序列对。降雨-洪水相似度分析最直接的应用是确定当前的水文过程是否与历史时期的过程相似^[3]。因此, 相似度搜索方法研究在降雨-径流过程预测、环境演变分析和水文规律发现等方面具有重要的潜力^[4-8]。

降雨相似度分析是洪水风险评估的重要组成部分, 它不仅可以发现历史上相似降雨-洪水模式的发生规律, 而且可以为降雨-洪水预报提供新的思路和技术支持。章龙飞等^[9]利用基于 Davies-Bouldin 指数的 K-means 聚类算法建立降雨站点相似模型, 再采用单降雨类型直方图相似模型对聚类结果进行分析, 得到相似

基金项目: 国家重点研发计划项目(2021YFB3900605); 江苏省水利科技项目(2021065, 2020014)

作者简介: 余宇峰(1979—), 男, 副教授, 博士, 主要从事智能数据处理、数据质量控制和智能洪水预报研究。E-mail: yfyu@hhu.edu.cn

引用本文: 余宇峰, 贺新固, 张潇, 等. 基于流域日降水量图的相似性搜索方法[J]. 河海大学学报(自然科学版), 2024, 52(2): 19-27.

YU Yufeng, HE Xingu, ZHANG Xiao, et al. A rainfall similarity search method based on daily precipitation images of watershed[J]. Journal of Hohai University(Natural Sciences), 2024, 52(2): 19-27.

站点,对水资源管理具有重要价值。Xiao 等^[3]提出了一种基于多尺度的降雨事件相似度分析模型,该模型从数量相似度、模式相似度、陆地移动距离和暴雨空间分布相似度等多个角度对两次暴雨事件进行相似度评价,结果表明,基于相似度分析的洪水预报方法是有效可行的。Ohno 等^[10]开发了基于天气预报降水量图的深度学习判断水位是否会超过“洪水”阈值的预报新技术,为利用天气预报降水量图延长洪水预报周期提供了新思路。

近些年来,随着信息技术的发展,降雨数据类型从传统的时间序列扩展到半结构化和非结构化。随着时间尺度和空间尺度的积累,从这些数据集中搜索类似的降雨过程变得越来越困难。因此,如何利用最新的机器学习和人工智能算法对积累的降雨图像大数据进行特征提取和快速相似度分析,为相似雨洪过程的识别和防洪提供技术支持,正成为一个有意义和热点的研究问题^[11-14]。

降雨是一个连续的过程,一场降雨过程可能包含着多个单日降雨,研究相似单日降雨是研究相似场次降雨的基础,能为场次降雨过程的相似性度量提供理论基础,并回答当前日降雨和历史上哪天的降雨相似的问题。因此,本文提出了一种基于流域日降水量图的相似性搜索方法(以下简称 NDCG-IPSO 方法),以期能为气象预报、洪水预警和水文模型校准等领域提供参考。

1 研究方法

1.1 图像特征提取

特征提取(feature extraction, FE)是模式识别、数字图像处理和机器视觉等研究的重要步骤^[15]。图像特征提取是用来提取图像数据集中呈现的最独特、最有用的信息,形成一个用低维特征空间表示和描述的图像,供搜索、浏览或检索使用。一般来说,图像特征大致分为一般特征(general features, GF)和领域特定特征(domain-specific features, DSF)。GF 是应用程序自主的特征,如颜色、形状和纹理^[16],而 DSF 是应用程序依赖的特征,包括概念特征和人脸特征。

图像的颜色特征通常采用基于 RGB、HMM、HSV 和 LUV 等颜色空间的方法来定义^[17],如颜色直方图(CH)^[18]、颜色矩(CM)^[19]和颜色相关图^[20]等。图像的纹理特征是一种反映图像中同质现象的视觉特征,只能从一组像素中测量,如基于 Gabor 滤波器的方法^[21]是用于提取图像纹理特征的一种技术。此外,图像的形状特征也是人类识别不同对象的关键要素。

对于图 1 所示的日降水量图像,同一区域不同日期的降水量具有相似的形状和纹理特征,仅需提取颜色特征即可表征不同的降雨特征。因此,本文研究基于颜色直方图的降雨图像的颜色特征提取方法,并根据每种颜色的实际意义计算流域的区域降雨。为更好地描述降水量图,本文将图像划分为 $m \times n$ 个网格,提取降雨空间分布、降雨中心等详细信息。

1.2 相似度搜索

相似度搜索(又称最近邻搜索)是指从搜索(参考)数据集中搜索与待查询的项最近的数据项,是聚类、分类等数据挖掘任务的基础,已被应用于时间序列预测、图像检索等领域。通常,相似度搜索依赖于量化两个元素在特征空间中的距离度量。元素越接近,它们之间的相似性越高。常用的距离度量包括欧氏距离、闵可夫斯基距离(Minkowski distance, 又称 L_h 距离)^[22]、动态时间弯曲距离(dynamic time warp distance, DTW)^[23]、编辑距离^[24]等,本文采用 L_h 距离。

L_h 距离定义时间序列 $Q' = \{q_1, q_2, \dots, q_n\}$ 和 $C = \{c'_1, c'_2, \dots, c'_n\}$ 间的距离 $D_{Q',C}$ 为

$$D_{Q',C} = \left(\sum_{i=1}^n |q_i - c'_i|^h \right)^{\frac{1}{h}} \quad (1)$$

式中 h 为一个变参数,为任意实数。

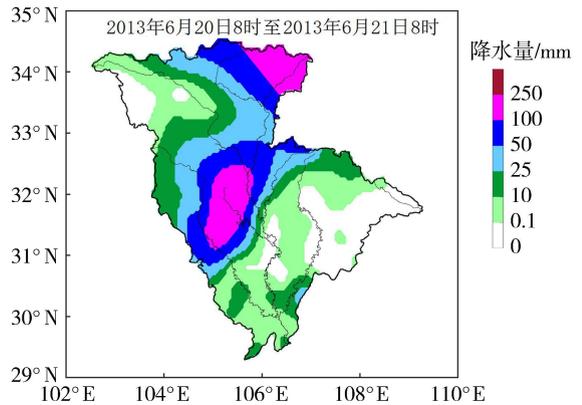


图 1 嘉陵江流域日降水量实况
Fig. 1 Actual precipitation image in the Jialing River Basin

h 常见的取值有 3 个:① $h=1$, L_h 距离简化为曼哈顿距离;② $h=2$, L_h 距离简化为欧氏距离,是时间序列相似性研究中应用最广泛的距离度量;③ $h=\infty$, L_h 距离简化为切比雪夫距离。

1.3 粒子群优化算法

粒子群优化 (particle swarm optimization, PSO) 算法是一种灵感来自鸟群朝着未知目的地 (适应度函数) 搜索食物资源的深度学习优化算法^[25]。PSO 算法用具有特定位置和速度的随机粒子 (鸟) 进行初始化,以计算优化问题的目标函数。PSO 算法的每一次迭代中,均计算出最佳的个人和全局适应度位置。根据计算出的适应度函数更新每只鸟的位置和速度,直到得到最优解^[26]。

2 NDCG-IPSO 方法

2.1 整体流程

图像搜索系统的性能很大程度上取决于特征表示和相似性度量。因此,降水量图相似度搜索的过程主要包括两个步骤:①从历史降水量图中提取日降水量特征、降雨空间分布特征和降雨中心特征并存储在降雨特征库中;②从待查询的降水量图中提取这 3 个特征,与降雨特征库中存储的降雨特征进行相似性匹配,并根据相似度进行排序。降水量图匹配的流程如图 2 所示。

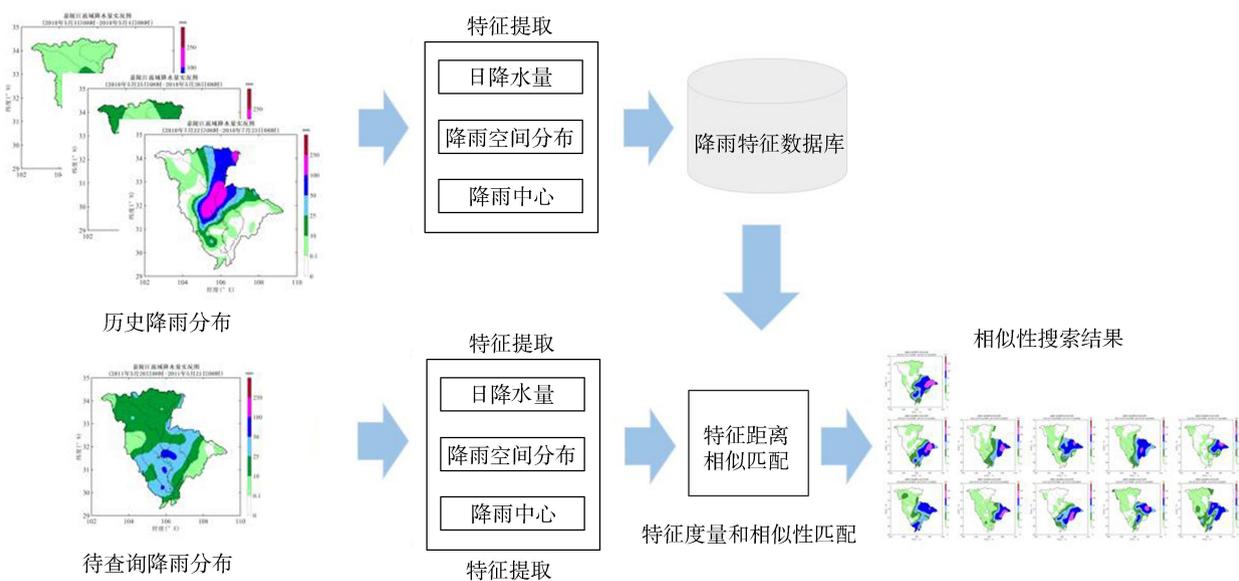


图 2 降水量图搜索的处理流程

Fig. 2 Process flow of precipitation image searching

2.2 特征提取

2.2.1 日降水量

日降水量代表某一特定时间内某一特定地区的降雨总量。在降水量图中,每个颜色对应一个降水量范围。因此,对颜色直方图进行加权,可以得到区域降雨特征。

设降雨图像对应的颜色直方图包含 K 种颜色,即 C_1, C_2, \dots, C_K , 图像中每种颜色的出现次数为 $T_{c_i} (1 \leq i \leq K)$, 每种颜色对应的降水量为 p_{c_i} (单位: mm), 则给定区域内的日降水量为

$$P = \sum_{i=1}^K T_{c_i} p_{c_i} \quad (2)$$

假设两幅降水量图中提取出来的流域日降水量分别为 P_1 和 P_2 , 则两幅降水量图的日降水量之间距离 D_p 可表示为

$$D_p = |P_1 - P_2| \quad (3)$$

2.2.2 降雨空间分布

日降水量可以大致代表某一区域内的降雨总量,但难以反映降雨的空间分布特征。因此,采用基于分块的颜色直方图,在截断冗余标注后,将流域最大外接矩形分割为如图 3 所示的 $m \times n$ 矩形块,即将降雨图像划

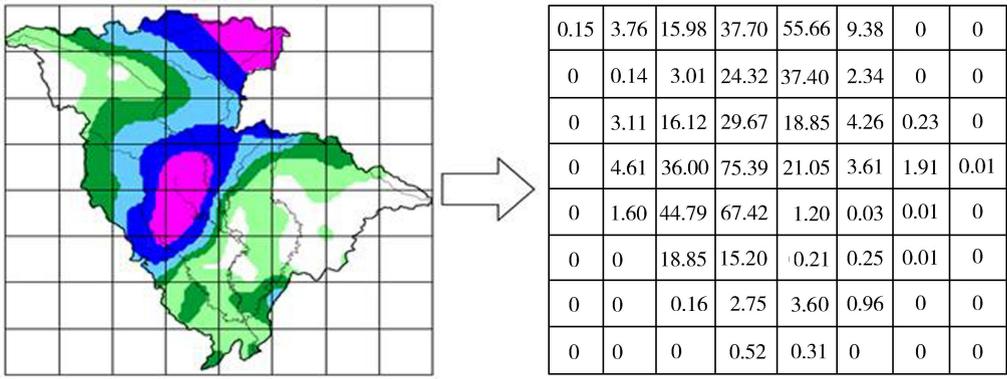


图3 降雨分布矩阵

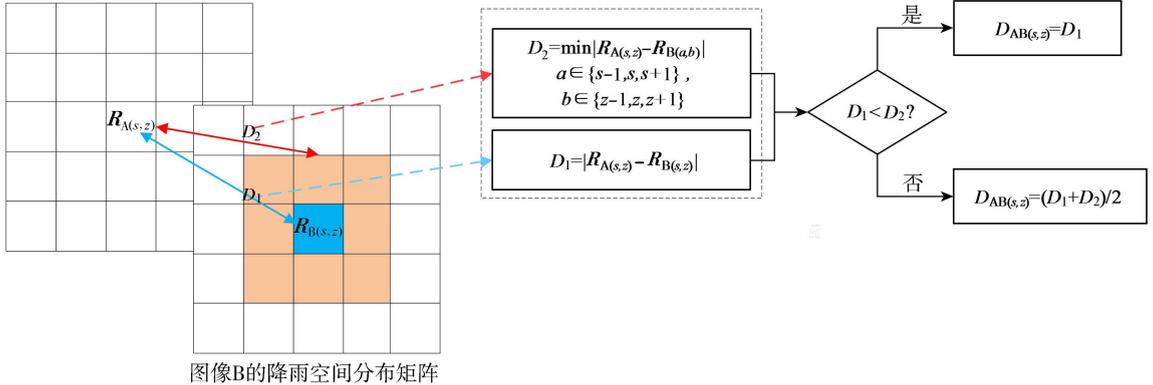
Fig. 3 Precipitation distribution matrix

分为 $m \times n$ 个小网格,各小格网的区域降雨可分别按式(2)计算。设 $P_{s,z}$ 为第 s 行第 z 列网格的区域降雨,采用降雨分布矩阵 R 表征降雨图像的空间分布特征:

$$R = \begin{bmatrix} P_{1,1} & \cdots & P_{1,n} \\ \vdots & & \vdots \\ P_{m,1} & \cdots & P_{m,n} \end{bmatrix} \quad (4)$$

如图4所示,设 $R_{B(a,b)}$ 是图像B中 $R_{(s,z)}$ 周围的一个块(图4中橙色对应的8个相邻网格),其中 $a \in \{s-1, s, s+1\}$, $b \in \{z-1, z, z+1\}$ 。设 $D_{AB(s,z)}$ 为降雨分布矩阵 R_A 和 R_B 第 s 行第 z 列的元素之间的距离,将 $R_{A(s,z)}$ 到 $R_{B(s,z)}$ 的距离记为 D_1 ,将 $R_{A(s,z)}$ 到 $R_{B(a,b)}$ 的距离的最小值记为 D_2 ,如果 $D_1 < D_2$, $D_{AB(s,z)}$ 以 D_1 表示,否则 $D_{AB(s,z)}$ 以 D_1 和 D_2 的均值表示。

图像A的降雨空间分布矩阵

图4 $D_{AB(s,z)}$ 距离的计算方法Fig. 4 Calculation method of $D_{AB(s,z)}$ distance

因此,两幅降水量图的降雨空间分布特征间的距离 D_R 为

$$D_R = \sum_{s=1}^m \sum_{z=1}^n D_{AB(s,z)} \quad (5)$$

2.2.3 降雨中心

降雨中心对洪水过程具有重要影响,因此将降雨中心作为降雨图像相似度研究的一个主要特征十分必要。结合实际降雨情况和图像相似度检索要求,将降雨图像分块后降水量最大的块作为降雨中心。

设 $P_{(s,z)}$ 是降雨图像中各块的最大降水量,则降雨中心 $C_{(s,z)}$ 可以表示为第 s 行第 z 列的网格位置。设 $C_{(s_1,z_1)}$ 和 $C_{(s_2,z_2)}$ 为两幅降雨图像的降雨中心。欧氏距离 D_C 定义为两个降雨中心的距离:

$$D_C = \sqrt{(s_1 - s_2)^2 + (z_1 - z_2)^2} \quad (6)$$

2.3 相似性搜索

降雨图像具有不同的特征,如果采用单一特征表示和度量降雨图像的相似度,可能会丢失其他特征信

息,降低相似度检索的精度。因此,需要融合多个特征的距离来表示两幅降雨图像之间的综合距离:

$$D = \gamma_1 D_p + \gamma_2 D_R + \gamma_3 D_C \quad (7)$$

式中 γ_1 、 γ_2 、 γ_3 分别为日降水量距离、降雨空间分布距离和降雨中心距离的待定权重系数。

确定待定系数有主观权重法、客观权重法和主客观综合权重法3种方法。主客观综合权重法可以将客观统计特征与主观经验相结合,成为解决多因素综合评价问题的首选方案^[27]。

NDCG(normalized discounted cumulative gain)-IPSO 是为提高降雨图像相似度搜索效率而提出的一种新的主客观综合权重方法,该方法利用 IPSO 调整多个指标的权重,使评价结果更接近专家根据主观经验得出的评价结果,再以 NDCG 为指标对多特征加权的图像搜索结果进行评价。该方法结合了客观加权法和主观加权法的优点,通过拟合和逼近的方式使加权结果满足要求。

2.3.1 评价指标

NDCG 作为评价图像搜索结果的指标,由多个特征加权代表归一化折扣累计收益。假设某批次相似性搜索任务是针对实体 E_1 、 E_2 、 \dots 、 E_n 的,对实体 E_l 的搜索结果为 $e_{l,1}$ 、 $e_{l,2}$ 、 \dots 、 $e_{l,j}$ 。结果中的每一项 $e_{l,j}$ 都是搜索系统认为与实体 E_l 相似的其他实体,且都有一个与 E_l 相似程度的真实得分,分值越小代表与 E_l 真实相似程度越低,搜索结果中的每一项与实体 E_l 的相似程度得分分别为 $r_{l,1}$ 、 $r_{l,2}$ 、 \dots 、 $r_{l,j}$ 。实体 E_l 的搜索结果前 Q 项的累计增益 C_Q 为

$$C_Q = \sum_{f=1}^Q r_{l,f} \quad (8)$$

折损累计增益 (discounted cumulative gain, DCG) 是指考虑排序顺序的因素,使得搜索结果列表中排名靠前的项对增益影响更大,即对排名靠后的项进行增益折损。实体 E_l 的搜索结果前 Q 项的折损累计增益 D_Q 为

$$D_Q = \sum_{f=1}^Q \frac{r_{l,f}}{\log_2(f+1)} \quad (9)$$

由于每一个实体的真实搜索结果增益通常并不相同,累计折损增益指标并不能用于对不同实体搜索结果的比较。这就需要对折损累计增益进行归一化处理。

NDCG 归一化折损累计增益就是对折损累计增益进行归一化后的指标。归一化相关系数为理想搜索结果的 D_Q 值,记作 I_Q 。实体 E_l 的搜索结果前 Q 项的归一化折损累计增益 N_Q 为

$$N_Q = \frac{D_Q}{I_Q} \quad (10)$$

2.3.2 参数优化

通过改变原始 PSO 算法中的惯性权重 w 、学习因子 c_1 和 c_2 ,调整式(7)中的权重系数 γ_1 、 γ_2 和 γ_3 ,从而找到用于降雨图像检索的最优权重。IPSO 随着迭代次数的增加自适应地改变 w 、 c_1 和 c_2 ,避免算法陷入部分最优解。IPSO 的改进主要包括对惯性权重 w 的改进、对学习因子 c_1 和 c_2 的改进。

2.3.2.1 对 w 的改进

w 表示粒子对上一次迭代时速度的保留程度,能够调节算法的全局和局部搜索能力, w 越大,算法的全局搜索能力越强而局部搜索能力越弱;反之则局部搜索能力越强而全局搜索能力越弱。因此, w 往往取值为一个接近 1 的适中值以兼顾算法全局和局部的搜索能力。基于 Shi 等^[28] w 迭代次数线性递减方法,优化 w :

$$w_k = w_{\max} - (w_{\max} - w_{\min}) \frac{k_{\max} - k}{k_{\max}} \quad (11)$$

式中: w_{\max} 为初始最大惯性权重; w_{\min} 为迭代至最大代数时的最小惯性权重; k_{\max} 为最大迭代次数; w_k 为迭代 k 次时的惯性权重值。

2.3.2.2 对 c_1 和 c_2 的改进

c_1 、 c_2 分别表示粒子向个体最优解和群体最优解学习的能力,其取值通常依据经验取 $[0,4]$ 上的相同值。在算法的前期, c_1 取较大值、 c_2 取较小值,可以强化粒子对个体最优解学习能力,弱化粒子对群体最优解学习能力,增强粒子的全局搜索能力;在进化后期, c_1 取较小值、 c_2 取较大值,可以强化粒子对群体最优解的学习能力,弱化粒子对个体最优解学习能力,增强粒子的局部搜索能力。采用对称线性变化思想^[29] 对 c_1 和 c_2

进行改进,使得 c_1 和 c_2 均以某一固定值为中间值,在初始值和终止值之间随着迭代次数进行线性变换,具体改进如下:

$$c_{1, \text{begin}} = c_{2, \text{end}} = c_{\text{mid}} + \Delta c \quad (12)$$

$$c_{2, \text{begin}} = c_{1, \text{end}} = c_{\text{mid}} - \Delta c \quad (13)$$

$$c_{1k} = c_{1, \text{begin}} - \frac{k}{k_{\text{max}}}(c_{1, \text{begin}} - c_{1, \text{end}}) = c_{\text{mid}} + \Delta c \left(1 - 2 \frac{k}{k_{\text{max}}}\right) \quad (14)$$

$$c_{2k} = c_{2, \text{begin}} - \frac{k}{k_{\text{max}}}(c_{2, \text{end}} - c_{2, \text{begin}}) = c_{\text{mid}} - \Delta c \left(1 - 2 \frac{k}{k_{\text{max}}}\right) \quad (15)$$

式中: $c_{1, \text{begin}}$ 、 $c_{2, \text{begin}}$ 分别为 c_1 、 c_2 初始值; $c_{1, \text{end}}$ 、 $c_{2, \text{end}}$ 分别为 c_1 和 c_2 的终止值; Δc 为调节量,表示 c_1 和 c_2 可变动的最大长度,取为 0.625; c_{mid} 为 c_1 、 c_2 的中间值,取 1.875; c_{1k} 、 c_{2k} 为第 k 次迭代的 c_1 、 c_2 。

随着迭代次数 k 递增, c_1 由 2.5 线性递减为 1.25, c_2 由 1.25 线性递增为 2.5。在算法过程的前期,由于 $c_1 > c_2$,粒子的运动趋势受个体最优位置影响较大,此时全局搜索能力较强;随着迭代次数的增加, c_1 逐渐减小,而 c_2 逐渐增大并大于 c_1 ,粒子的运动趋势受全局最优位置影响更大,此时局部搜索能力增强。

3 实例验证

3.1 数据准备和处理

嘉陵江是长江最大的支流,长约 1120 km。嘉陵盆地位于北纬 $29^{\circ}17'30'' \sim 34^{\circ}28'11''$ 和东经 $102^{\circ}35'36'' \sim 109^{\circ}01'08''$ 范围内,面积约 16 万 km^2 。属于湿润季风气候区,年平均降水量 931 mm。在正常年份,流域降雨引发的洪水多集中在 7 月和 8 月,具有波动快、持续时间短、流量快、洪峰高的特点,对沿江居民的生命财产构成了极大的威胁。

为了验证 NDCG-IPSO 方法用于图像相似度搜索的有效性,利用嘉陵江流域 2010 年 1 月 1 日至 2019 年 12 月 12 日的逐日降雨图像进行训练和验证。此外,为了评价基于 NDCG-IPSO 方法的图像相似度搜索的效率,还需要标记一些样本作为数据预处理。因此,选取了 30 幅不同降雨等级的降雨图像(6 幅小雨图像、9 幅中雨图像、15 幅大雨图像)作为查询样本,并从历史降雨图像中选择每个查询样本对应 10 个匹配样本。并根据专家经验给每个匹配样本从 0(完全不相似)到 2(非常相似)的相似度评分,以衡量查询样本与匹配样本之间的相似程度。查询样本按 2:1 的比例分为训练样本集和测试样本集。

3.2 结果分析

使用 NDCG-IPSO 方法对嘉陵江流域日降水量图进行单日降雨相似性搜索,其中改进的粒子群算法(IPSO 算法)中粒子数设为 30,迭代次数设为 80,根据训练样本获得降水量图多特征距离的最优权重。其中,流域日降水量距离的权重系数为 0.46,降雨空间分布距离为 0.12,降雨中心距离为 0.42。因此,设定 $\gamma_1 = 0.46$ 、 $\gamma_2 = 0.12$ 、 $\gamma_3 = 0.42$ 计算降水量图之间的综合距离。通过测试样本验证使用该权重进行降雨相似性搜索的效果(以 N_5 和 N_{10} 作为评判)。其中训练样本的平均 N_5 为 0.984,平均 N_{10} 为 0.978;测试样本的平均 N_5 为 0.978,平均 N_{10} 为 0.964。可以看出,NDCG-IPSO 方法在训练样本和测试样本都能够获得较高的指标值。该方法在测试样本上的指标都非常接近 1。结果表明:①选取的 3 个降雨特征一定程度上能够代表专家对降雨的关注点,这 3 个降雨特征的距离可以作为评判降雨相似程度的指标;②使用 NDCG-IPSO 方法对 3 个降雨特征距离加权的很好,可以快速地从降水量图中检索相似的图像。

图 5 展示了基于 NDCG-IPSO 方法的降雨图像相似度搜索结果。根据设定的待搜索图像日期,查询结果数等条件,显示搜索结果。结果展示区第一行是待查询的降水量图,第 2 行和第 3 行是最相似的 10 幅图像。对降水量图和降雨时间序列数据的分析结果表明,查询结果与待查询图像在日降水量、降雨空间分布、降雨中心等方面非常相似,证明本文方法能快速有效地检索出相似日降水量图像,为相似水文过程的分析提供技术支持。

3.3 对比分析

为了进一步说明 NDCG-IPSO 方法的有效性,将 NDCG-IPSO 方法的图像搜索结果与 Global CH^[30]、Block CH^[31]、主成分分析(PCA)^[32]、BORDA^[33] 和 NDCG-PSO 方法进行了比较,结果见表 1。

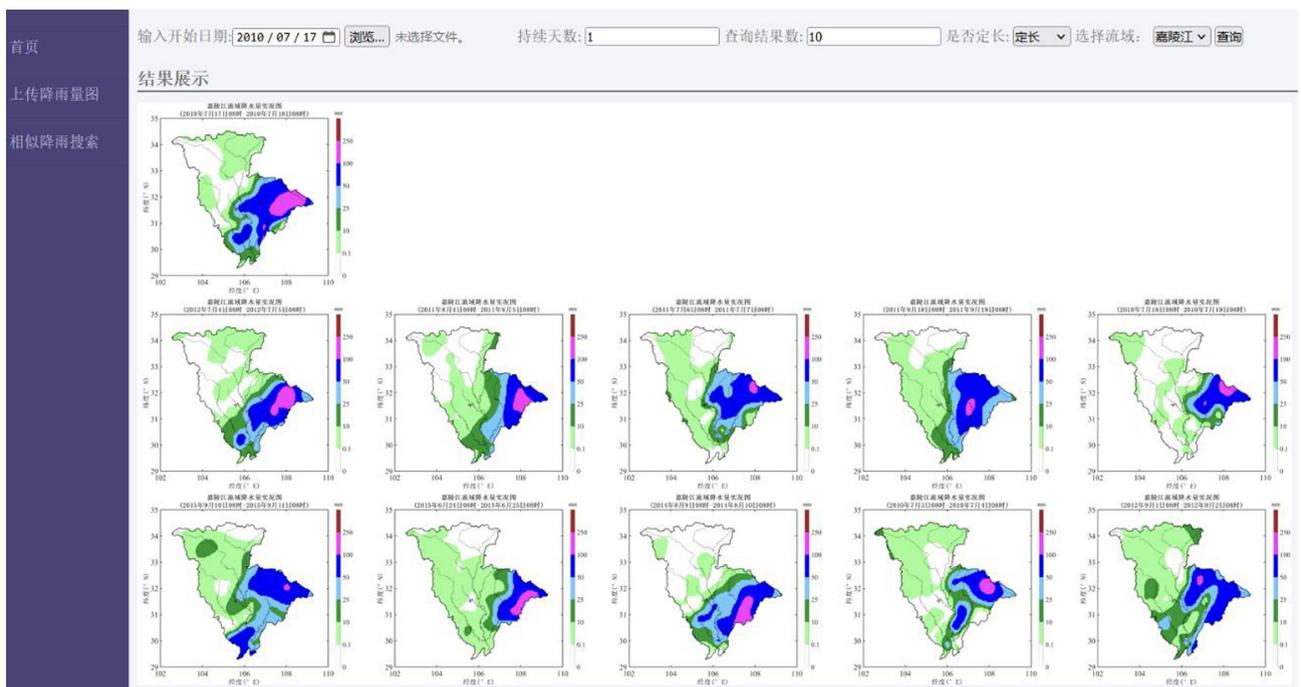


图 5 NDCG-IPSO 方法搜索结果

Fig. 5 Search results of NDCG-IPSO method

颜色直方图在图像检索系统中具有广泛应用,它通过提取图像中的颜色直方图计算直方图之间的距离来搜索相似的图像。降雨图像主要采用颜色特征来表示不同的降雨信息,因此颜色直方图是最适合降雨图像相似度搜索的方法。使用 Global CH 和 3×3 的 Block CH 搜索降雨图像的相似图像,结果如表 1 所示。由表 1 可知,Global CH 虽然可以很好地表征降雨图像的颜色特征,但它忽略了颜色特征的空间信息,导致相似度检索精度较低。Block CH 考虑了部分空间信息,搜索精度优于 Global CH,但 Block CH 没有考虑降雨图像上不同颜色的物理意义,导致搜索精度低于 NDCG-IPSO 方法。

表 1 基于不同方法的图像搜索结果

Table 1 Image search results based on different methods

方法	N_5	N_{10}
NDCG-IPSO	0.978	0.964
Global CH	0.721	0.662
Block CH	0.834	0.803
PCA	0.702	0.623
BORDA	0.836	0.796
NDCG-PSO	0.959	0.923

PCA 和 BORDA 是两种常用的多指标综合评价方法,也被广泛应用于多元水文相似度的挖掘。PCA 对 3 个特征距离进行主成分分析,将累计方差贡献率大于 85% 的特征作为主成分,根据方差贡献率对特征进行加权,排序后得到搜索结果。BORDA 对特征距离进行一次排序,通过 BORDA 综合这些特征距离的相似度搜索结果,得到最终的查询结果。虽然 PCA、BORDA 和 NDCG-IPSO 方法均提取降雨空间分布、降雨中心和日降水量,并综合考虑 3 个特征距离对降雨图像进行相似图像搜索,但 NDCG-IPSO 方法同时考虑了图像的物理意义,并使用机器学习算法进行参数优化,从而获得了比 PCA 和 BORDA 更好的结果。

NDCG-PSO 方法与 NDCG-IPSO 方法的唯一不同之处在于, NDCG-IPSO 方法改变了 PSO 算法中的惯性权值和两个学习因子,使其能够随着迭代次数的增加而自适应变化,从而提高了 NDCG-IPSO 方法的搜索性能和精度。图 6 显示了 NDCG-PSO 方法和 NDCG-IPSO 方法的适应度函数值 N_5 随迭代次数的变化。由图 6 可以看出, NDCG-IPSO 方法经过 18 次迭代,得到最优的粒子适应度 0.984;而 NDCG-PSO 方法在 27 次迭代后陷入局部最优 0.959。这表明 IPSO 算法可以提高图像相似度搜索

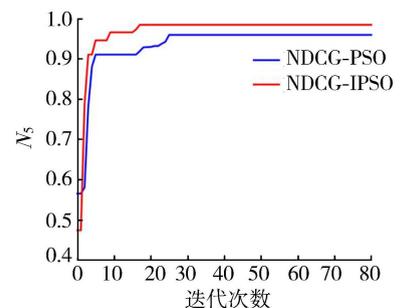


图 6 NDCG-PSO 方法和 NDCG-IPSO 方法结果对比

Fig. 6 Comparison of experimental results between NDCG-PSO method and NDCG-IPSO method

的搜索精度和优化速度,为提高相似度降雨图像检索提供支持。

4 结 语

本文提出了一种基于流域日降水量图的相似性搜索方法,该方法能有效检索出相似日降水量图像,其训练样本的平均 N_5 为 0.984,平均 N_{10} 为 0.978;测试样本的平均 N_5 为 0.978,平均 N_{10} 为 0.964。可以看出,使用 NDCG-IPSO 方法在训练样本和测试样本都能够获得较高的指标值。但场次降雨过程一般由多个单日降水量图组成。因此,如何基于 NDCG-IPSO 方法的相似度量方法进行降雨过程相似度搜索,从而构建降雨-洪水相似模式库,为水文预报和水资源利用提供指导和技术支持,是我们未来研究的方向。

参考文献:

- [1] 郑德凤,高敏,李钰,等. 基于 GIS 的大连市暴雨洪涝灾害综合风险评估[J]. 河海大学学报(自然科学版),2022,50(3):1-8. (ZHENG Defeng,GAO Min,LI Yu, et al. Comprehensive risk assessment of rainstorm-flood disaster in Dalian City based on GIS[J]. Journal of Hohai University (Natural Sciences),2022,50(3):1-8. (in Chinese))
- [2] 谢志高,贾文豪,王霞雨,等. 西江流域极端降水演变规律及其对洪水径流的影响[J]. 水利水电科技进展,2023,43(6):128-136. (XIE Zhigao,JIA Wenhao,WANG Xiayu, et al. Evolution characteristics of extreme rainfall and influence on flood runoff in Xijiang River Basin[J]. Advances in Science and Technology of Water Resources,2023,43(6):128-136. (in Chinese))
- [3] XIAO Zhangling,LIANG Zhongmin,LI Binqun, et al. New flood early warning and forecasting method based on similarity theory [J]. Journal of Hydrologic Engineering,2019,24(8):04019023.
- [4] 王京晶,徐宗学,赵刚,等. 基流分割对城市雨洪模拟不确定性分析的影响[J]. 水资源保护,2022,38(5):65-71. (WANG Jingjing,XU Zongxue,ZHAO Gang, et al. Impact of baseflow separation on uncertainty analysis of urban storm water simulation [J]. Water Resources Protection,2022,38(5):65-71. (in Chinese))
- [5] BARTHEL R,HAAF E,GIESE M, et al. Similarity-based approaches in hydrogeology: proposal of a new concept for data-scarce groundwater resource characterization and prediction[J]. Hydrogeology Journal,2021,29(5):1693-1709.
- [6] OUYANG R,REN L,CHENG W, et al. Similarity search and pattern discovery in hydrological time series data mining[J]. Hydrological Processes,2010,24(9):1198-1210.
- [7] LIANG Z,XIAO Z,WANG J, et al. An improved chaos similarity model for hydrological forecasting[J]. Journal of Hydrology,2019,577:123953.
- [8] FEOLI E,ZUCCARELLO V. Spatial pattern of ecological processes: the role of similarity in GIS applications for landscape analysis[C]//Proceedings of GISDATA Specialist Meeting on GIS & Spatial Analysis. Amsterdam:Taylor & Francis,1993:175-186.
- [9] 章龙飞,朱跃龙,李士进,等. 基于降雨类型直方图分析的降雨站点相似性研究[J]. 水文,2013,33(3):10-17. (ZHANG Longfei,ZHU Yuelong,LI Shijin, et al. Study on similarity model of precipitation series based on precipitation type histogram[J]. Journal of China Hydrology,2013,33(3):10-17. (in Chinese))
- [10] OHNO G,KAZUNORI I. Flood forecast based on deep learning using distribution MAP of precipitation[C]//Proceedings of the 22nd IAHR APD Congress. Sapporo: IAHR APD,2020:14-17.
- [11] WANG X,LIU Y,CHEN Y, et al. An adaptive density-based time series clustering algorithm: a case study on rainfall patterns [J]. ISPRS International Journal of Geo-Information. 2016,5(205):1-19.
- [12] 周聂,侯精明,陈光照,等. 基于机器学习的山洪灾害快速预报方法[J]. 水资源保护,2022,38(2):32-40. (ZHOU Nie,HOU Jingming,CHEN Guangzhao, et al. A rapid forecasting method for mountain flood disaster based on machine learning algorithms[J]. Water Resources Protection,2022,38(2):32-40. (in Chinese))
- [13] ZHU Y,FENG J,YAN L, et al. Flood prediction using rainfall-flow pattern in data-sparse watersheds[J]. IEEE Access,2020,8:39713-39724.
- [14] 方国华,王雪,方应学,等. 基于改进粒子群优化算法的区域水量水质联合配置模型[J]. 水资源保护,2022,38(3):58-64. (FANG Guohua,WANG Xue,FANG Yingxue, et al. Regional water quantity and water quality joint allocation model based on improved particle swarm optimization algorithm[J]. Water Resources Protection,2022,38(3):58-64. (in Chinese))
- [15] DING S,ZHU L,JIA W, et al. A survey on feature extraction for pattern recognition[J]. Artificial Intelligence Review,2012,37(3):169-180.
- [16] TIAN D. A review on image feature extraction and representation techniques [J]. International Journal of Multimedia and

- Ubiquitous Engineering,2013,8(4): 385-396.
- [17] STANCHEV P L, GREEN D, DIMITROV B. High level color similarity retrieval[J]. Information Theories & Applications,2003, 10(3): 283-287.
- [18] JAIN A K, VAILAYA A. Image retrieval using color and shape[J]. Pattern Recognition,1996,29(8): 1233-1244.
- [19] FLICKNER M, SAWHNEY H, NIBLACK W, et al. Query by image and video content: the QBIC system[J]. Computer,1995,28(9): 23-32.
- [20] HUANG J, KUMAR S R, MITRA M, et al. Image indexing using color correlograms[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York;IEEE,1997: 762-768.
- [21] MANJUNATH B S, MA W Y. Texture features for browsing and retrieving of large image data[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2011,33(1): 117-128.
- [22] DING H, TRAJCEVSKI G, SCHEUERMANN P, et al. Querying and mining of time series data: experimental comparison of representations and distance measures[J]. Proceedings of the VLDB Endowment,2008,1(2): 1542-1552.
- [23] HONG J Y, PARK S H, BAEK J G. SSDTW: shape segment dynamic time waring[J]. Expert Systems with Applications,2020, 150: 113291.
- [24] RISTAD E S, YIANILOS P N. Learning string-edit distance[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1997,20(5): 522-532.
- [25] KENNEDY J, EBERHART R. Particle swarm optimization[C]//Proceedings of ICNN'95-International Conference on Neural Networks. New York :IEEE,1995,4: 1942-1948.
- [26] KNIGHT J T, SINGER D J, COLLETTE M D. Testing of a spreading mechanism to promote diversity in multi-objective particle swarm optimization[J]. Optimization and Engineering,2015,16(2): 279-302.
- [27] PARAMANIK A R, SARKAR S, SARKAR B. OSWMI: an objective-subjective weighted method for minimizing inconsistency in multi-criteria decision making[J]. Computers & Industrial Engineering,2022,169: 108138.
- [28] SHI Y, EBERHART R. A modified particle swarm optimizer [C]//1998 IEEE International Conference on Evolutionary Computation Proceedings. New York :IEEE,1998: 69-73.
- [29] 冯克涛, 李晓毅, 钱璇, 等. 基于改进粒子群算法拟合大气光学湍流廓线模式的研究[J]. 激光与光电子学进展,2022,59(5):83-94. (FENG Ketao, LI Xiaoyi, QIAN Xuan, et al. Atmospheric optical turbulence profile model fitting based on improved particle swarm algorithm[J]. Laser & Optoelectronics Progress,2022,59(5): 73-84. (in Chinese))
- [30] LIU C, SUI X, KUANG X, et al. Optimized contrast enhancement for infrared images based on global and local histogram specification[J]. Remote Sensing,2019,11(7): 849.
- [31] SEO H, CHO G, KIM S, et al. Multievent histogramming TDC with pre-post weighted histogramming filter for CMOS LiDAR sensors[J]. IEEE Sensors Journal,2022,22(23): 22785-22798.
- [32] BOUDOU A, VIGUIER-PLA S. Principal components analysis and cyclostationarity[J]. Journal of Multivariate Analysis,2022, 189: 104875.
- [33] MAHAJNE M, VOLIJ O. Pairwise consensus and the Borda rule[J]. Mathematical Social Sciences,2022,116: 17-21.

(收稿日期:2022-12-14 编辑:刘晓艳)