

基于智能优化算法的饮用水污染源定位方法研究综述

龚瑾玉, 颜雪松[†], 胡成玉, 龚文引

(中国地质大学(武汉) 计算机学院, 湖北 武汉 430078)

摘要: 近年来, 突发性饮用水污染事件频繁发生, 严重危害居民生活健康, 通过在饮用水供水管网中布置传感器对水质进行监测, 能快速识别污染源的位置、质量、发生时间等特征, 有利于有关部门采取措施控制污染扩散, 饮用水污染源定位研究具有重要的实际意义. 随着智能优化算法在工程问题中的广泛应用, 运用模拟-优化方法求解污染源定位问题成为当前学者们研究的热门领域. 本文首先给出污染源定位问题的模型和据此抽象出的数学模型, 并对模拟-优化法求解该问题的一般方法进行了描述. 通过对问题的深入分析, 归纳出污染源定位问题具有多模性、昂贵性和不确定性, 围绕这3个特性, 重点综述智能优化算法在饮用水污染源定位问题中的代表性研究成果, 最后指出有待于进一步研究的若干方向和内容.

关键词: 供水管网; 水污染; 污染源定位; 智能计算; 多模优化; 不确定性

引用格式: 龚瑾玉, 颜雪松, 胡成玉, 等. 基于智能优化算法的饮用水污染源定位方法研究综述. 控制理论与应用, 2021, 38(9): 1313 – 1323

DOI: 10.7641/CTA.2021.00825

Survey on methods for drinking water contamination source identification based on intelligent optimization algorithm

GONG Jin-yu, YAN Xue-song[†], HU Cheng-yu, GONG Wen-yin

(College of Computer Science, China University of Geosciences (Wuhan), Wuhan Hubei 430078, China)

Abstract: In recent years, sudden water pollution incidents have occurred frequently, seriously endangering residents' health. By placing sensors in the water supply networks to monitor the water quality, some characteristics of contamination source including the location, quality and occurrence time can be quickly identified, which is conducive to the relevant departments to deal with the diffusion of pollution. Therefore, the research on the contamination source identification has important practical significance. With the wide application of intelligent optimization algorithms in engineering problems, using simulation-optimization method to solve contamination source identification problem has become a hot research field. In this paper, the model of this problem and the abstract mathematical model are given, and then the general method of solving the problem with simulation-optimization method is described. Through in-depth analysis of this problem, the paper concludes that it is multimodal, computing expensive and uncertain. Based on these three characteristics, this paper focuses on the representative research results of intelligent optimization algorithms in contamination source identification problem, and finally points out some directions and contents to be further studied.

Key words: water supply network; water contamination; contamination source identification; intelligent computing; multimodal optimization; uncertainty

Citation: GONG Jinyu, YAN Xuesong, HU Chengyu, et al. Survey on methods for drinking water contamination source identification based on intelligent optimization algorithm. *Control Theory & Applications*, 2021, 38(9): 1313 – 1323

1 引言

水资源是人类赖以生存的条件之一, 饮用水的安全对居民的生活健康至关重要, 供水系统安全在维护公共健康和工业增长方面具有重要的战略地位^[1]. 而供水系统因其开放性、易访问、基础设施易老化等原因很容易受到外界的意外侵害和故意侵害, 导致饮用

水被污染, 使公众健康面临风险. 2016年5月, 北京饮用水供给系统因为管道连接失误的问题, 意外导致灰水入侵到饮用水管道中; 同年5月31日, 香港饮用水被检测出铅含量超标, 且远超世界卫生组织的最高级别. 这两种情况都是饮用水被污染了而没有及时被检测到而导致的. 由于污染物在供水管网中传播的速度相

收稿日期: 2020-11-20; 录用日期: 2021-04-23.

[†]通信作者. E-mail: yanxs@cug.edu.cn; Tel.: +86 18908632572.

本文责任编辑: 王凌.

国家自然科学基金项目(U1911205, 62073300, 61673354)资助.

Supported by the National Natural Science Foundation of China (U1911205, 62073300, 61673354).

对较快,且不可预测,因此快速识别污染源的位置和相关特性对有关部门采取措施控制和遏制污染具有十分非常重要的实际意义^[2].

传感器网络技术能检测到供水管网中的污染并可以获得有用的观测数据来识别和管理污染事件.目前常用的传感器分为固定传感器和移动传感器:固定传感器一般放置在固定的节点位置,而移动传感器在投放到管网节点后会随着水流进行移动,位置是动态变化的.目前,大部分有关供水管网水质检测的研究都是基于固定传感器进行的.受到成本和预算的限制,不可能在供水管网的每个节点都布置传感器,传感器所在的节点位置将直接影响污染物检测与污染源识别,有关传感器部署的文献已有很多^[3-5].在现有的研究中,学者们解决水污染源定位问题时直接指定传感器位置,并且这些传感器能够立即检测到由于污染事件引起的任何水质变化. Propato等人为了将传感器的观测数据与污染物可能发生的时间地点联系起来,构建了水质IO模型,该模型便于后来的学者们将污染源识别问题转化为反演问题或参数估计问题进行求解^[6].

饮用水污染源定位问题通常以固定传感器传回的污染物监测数据为依据,结合相应的算法,找出污染源的的特征,这些特征包括:污染源位置、污染开始时间、污染持续时间、污染物浓度.目前用于解决污染源识别问题的方法大致可分为3类:粒子回溯法、机器学习法、模拟-优化法.粒子回溯法是早期最常使用的方法,使用粒子反向追踪,根据传感器的检测信息反向推断前一时间各个节点的状态.此后,得益于机器学习算法的发展和研究,研究者们利用机器学习算法,如贝叶斯算法等,计算管网节点被污染的概率,其中被污染概率最大的管网节点就是最容易发生污染事件的节点.近年来,启发式算法也逐渐被用于解决污染源识别问题,通过模拟-优化的思想将饮用水污染源定位问题看作是一个根据水质传感器收集到的信息逆向寻找污染源的反演问题,指定时间步长与管网中传感器的数量,计算管网中所有传感器在每个时间步长上的估计污染浓度与观测污染浓度之间的变化,优化的目的是尽量减少观测数据与所建立模型的计算数据之间的差异^[7].

以上提到的3种方法中,模拟-优化法寻找污染源位置的精确度最高,但因其需要调用仿真软件模拟管网水质和污染源事件,计算量比另外两种方法更大.

本文主要对目前运用模拟-优化法解决饮用水污染源定位问题的相关文献进行综述,归纳总结智能优化算法在饮用水污染源定位问题中的应用.文章结构如下:第2节介绍饮用水污染源定位问题的数学模型和运用模拟-优化法解决污染源定位问题的优化模型,第3节重点综述智能优化算法在饮用水污染源定位问

题中的代表性研究成果,第4节给出智能优化算法在解决污染源定位问题上未来的可发展方向.

2 污染源定位问题模型

如图1所示为一个简化的供水管网,包含一个水源、一个水泵、一个水塔和多个节点及管道,其中黑色箭头代表水的流向.在A点模拟一个污染事件的发生,注入污染物,污染物会随水的流向扩散到管网中其他位置,并被位于E点的水质传感器检测到.从图中可以直观地看出受污染的水主要有3个流向,第1条从A点经B点流向水塔,第2条从A点经D点流向G点,第3条是从A点经D点流向E点.当水质传感器检测到流过E点的水中污染物浓度超过阈值,传感器就会示警,代表本次污染事件被检测到了.整个污染事件从污染注入到污染被检测的过程都可以由EPANET软件进行模拟^[8].

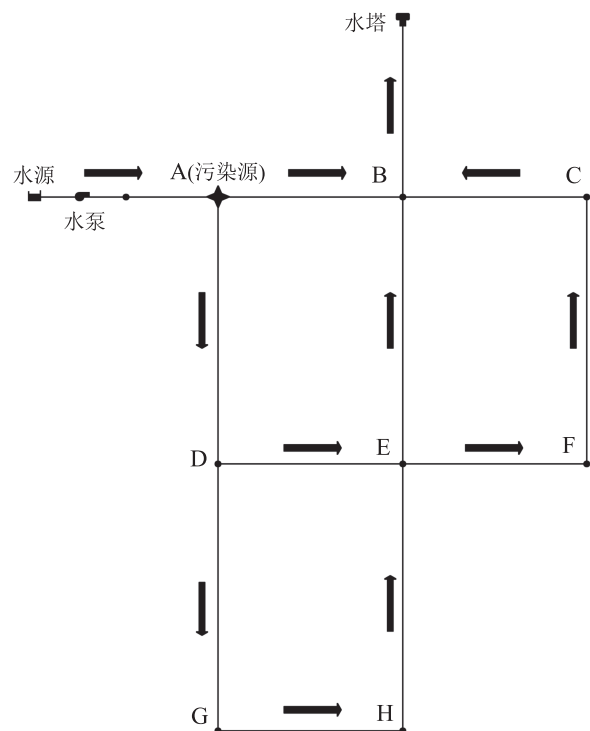


图1 供水管网模型

Fig. 1 Model of water supply network

EPANET是美国环境保护署开发的一款水力、水质分析软件,其源代码是开源的,由C语言编写,其开发的目的是为了改善对水管网系统中物质迁移转化规律的理解,它可以实现许多不同类型的给水管网系统水力水质分析,对管网水源的改变与配置,对水泵的提升,对水池注水/放水时间的调度安排,对水处理的补充措施(例如蓄水池中重新加氯)等,其中水质模拟是获取传感器数据的关键.水质模拟提供以下功能:模拟供水管网中不发生反应的示踪剂随时间的变化的过程;模拟供水管网中反应物质的变化情况,如消毒副产物可以随时间而浓度增长,而余氯随时间浓度

降低; 模拟整个管网的水龄; 模拟主流水体和管壁处的反应; 在管网中任意节点, 允许任意的浓度的污染物输入; 模拟蓄水池中不同的混合模型. 利用上述的水质模拟特性, EPANET可以实现以下的水质现象: 使来源不同的水源进行混合、模拟整个系统的水龄、余氯的损失、消毒副产物的增长、跟踪污染事件.

EPANET为每个节点都能设置源头水质, 通过设置源头水质可以模拟供水管网污染源入侵的场景. 污染物可能从管网的任何节点进入, 向管网节点注射进某一质量或者某一浓度的污染物, 然后调用EPANET进行水力、水质模拟获取污染事件对应的水质传感器检测的污染物浓度, 便实现污染事件入侵模拟.

污染事件中, 污染物可能在任意时间在管网的任何节点进入, 通过向管网中某个节点中注射一定浓度的污染物, 然后通过EPANET模拟污染物入侵时管网数据的动态变化, 通过这些数据获取到污染事件矩阵. 通过设置源头水质来模拟给水管网的污染物入侵场景, 入侵节点表示增强源头, 在不考虑节点需水量的条件下, 污染物质在该节点中直接注射到管网. 常见的污染模拟入侵方式分为: 质量注入、流量步进注入和设置点注入.

当使用模拟-优化法解决污染源识别问题时, 通常定义以下数学模型^[9], 以确定传感器检测到污染后任何时刻的污染物位置、开始时间和释放历史, 在单节点污染源入侵事件中, 污染源定位问题模型描述如下:

$$\begin{aligned} & \text{Find} \{L, M_{t_c}, T_0\}, \\ & \min F = \\ & \sqrt{\frac{\sum_{t=T_0}^{t_c} \sum_{i=1}^{N_s} [C_{it}^{\text{obs}} - C_{it}(L, M_{t_c}, T_0)]^2}{N_s * t_c}}, \quad (1) \end{aligned}$$

其中: F 表示模拟污染场景与真实污染场景间的误差, L 表示污染源位置, T_0 表示污染事件开始时间, t_0 表示传感器检测到污染的时间, t_c 表示当前时间步, M_{t_c} 表示从 T_0 到 t_c 的污染浓度矩阵, 即传感器从 T_0 到 t_c 的示数矩阵, 它代表从开始时间到当前时间各传感器检测到的污染物浓度变化, C_{it}^{obs} 表示传感器 i 在时间步 t 观测到的浓度, $C_{it}(L, M_{t_c}, T_0)$ 表示传感器 i 应当在时间步 t 观测到的浓度, 即真实浓度, i 表示传感器位置, t 表示观测时间步, N_s 表示传感器数目.

在给定一组传感器网络的浓度观测值后, 将水分布模拟模型结合优化算法, 可以构造一个反问题来识别污染源位置、浓度和释放历史特性. 该反问题的解决方案是最小化网络中传感器节点的模型浓度和真实观测浓度之间的误差, 即满足式(1).

模拟-优化法在求解饮用水污染源定位问题时的基本思路如下: 首先由优化算法迭代产生污染事件, 并用EPANET模拟器进行水力水质模拟, 得出该污染

事件的污染浓度矩阵. 通过比较真实观测到的污染浓度矩阵和模拟得到的污染浓度矩阵, 如果误差小于设定阈值则视优化算法最优解为污染源, 否则继续迭代求解, 直到程序结束.

3 基于智能优化算法的污染源定位

智能优化算法在解决工程优化问题上已表现出明显的优势, 智能优化算法作为一种启发式算法, 提供了一种全局搜索机制来有效识别大型非线性优化问题的最优解. 目前, 智能优化算法已在饮用水污染源定位问题上取得一定进展.

在模拟-优化法解决污染源定位问题中, 是以EPANET作为模拟器, 优化算法作为优化器进行求解, 优化的目标是在迭代的过程中逐渐最小化式(1)中的目标函数. 在早期运用模拟-优化法的文献中, Guan等人^[10]提出一种仿真优化法解决非线性污染源识别和释放历史定位问题, 该方法以EPANET作为模拟器, 通过在水管网中指定的任意位置的潜在污染源释放历史, 通过仿真软件在任意选择的检测位置产生对应的浓度, 将这些信息用于连续优化预测-校准算法中识别污染源及释放历史. 在该研究中, 假设供水管网在已知的水力运行模式下运行. Pries和Ostfeld^[11]提出了一种结合启发式搜索技术的污染源定位方法, 通过水质传感器检测到的污染物浓度状态信息来匹配随机污染矩阵中的污染事件, 由此逆向搜索污染源的位置和注入质量等等, 这项研究能解决计算量过大的问题, 但是算法假设管道内的流量是已知的, 这和实际情况不符.

之后又有学者运用遗传算法等智能优化算法进一步提高了定位的准确性, Yan等人^[12]运用改进的遗传算法, 根据问题变量的特性提出混合编码方案, 提高了算法的收敛速度和定位精度. Leichombam等人^[13]将基于梯度的局部搜索策略和遗传算法相结合提出一种混合的优化方法用于污染源识别, 该算法主要是对遗传算法进行了改进, 以不同的方式处理位置和通量的源变量, 并通过仿真实验证明了该方法在污染源定位问题的有效性和适用性.

以上方法都是将污染源定位问题转化为一个优化问题, 再运用优化算法进行问题求解. 而供水管网的真实情况的复杂性使饮用水污染源定位问题具有3个复杂的特性: 1) 解的非唯一性. 考虑到经济等因素, 一般在城市给水管网中放置的水质传感器的数量远少于管网节点数目, 因此利用有限的水质传感器的检测信息预测整个管网污染物注入状态, 可能会得到多个污染物注入场景, 但是, 真实的污染物注入事件往往只在一个或两个节点发生. 该问题是一个典型的多模函数优化问题, 即虽然是单目标优化问题, 但适应值空间却存在一个或多个全局极值点. 2) 计算代价昂贵

性. 饮用水供水管网节点规模巨大, 并且污染源头可能不止一个, 造成了决策空间巨大, 一般算法难以求解. 该问题具有数据密集和计算密集的特点, 利用优化算法求解该问题时, 需要较多的计算和存储资源, 是一个典型的昂贵优化问题. 3) 解的不确定性. 在实际的饮用水供水管网中, 水流的方向和速度会随着居民的需水量而动态变化, 如果考虑到污染物为非保守物质, 有可能与水体、管壁发生化学反应, 则实际的污染源位置则更难预测, 该问题是一个典型的动态优化问题. 针对问题的这些特性, 学者们将污染源定位问题分别转化成相应的优化问题进行研究, 得到了一系列新的研究成果.

3.1 多模优化

前文列举的研究成果都是将污染源定位问题看作是一个单目标优化问题, 通过对该问题进行深入的研究和分析, 该问题还可被看作是一个多模优化问题.

现实生活中很多问题并不是只有一个全局最优, 它需要找到所有的全局最优和局部最优, 这一类问题被称为多模问题, 例如机器学习中的聚类问题和地震波的反演问题^[14]. 优化算法需要找到搜索空间中的多个最优解, 这么做有两大好处: 通过定位多个最优解, 提高寻找全局最优的可能性; 面对工程问题时, 有时受限于时间或预算, 算法并不需要找到全局最优, 满足条件的次优解也可作为最后采纳的解^[15].

多模优化是智能计算中最具挑战性的任务之一. 近年来, 发展了许多进化多模态优化算法. 要想成功地解决多模态问题, 就必须解决以下两个问题: 如何识别多个全局/局部最优解以及如何将已识别的最优解保持到搜索结束.

在进化算法的基础, 学者们将小生境方法融入其中来定位多个最优解. 小生境是将每一代种群中的个体根据相似性分成若干子种群, 在每个子种群中找出适应值比较好的个体作为这一子种群的优秀代表组成一个群, 再在群内和各个子种群间进行选择操作、交叉操作和突变操作产生下一代种群. 小生境优化技术^[16]的本质是通过在一次运行中保持种群的多样性来寻找并保持尽可能多的全局最优和局部最优. 与传统的优化算法相比, 小生境优化算法在提高种群多样性的同时, 降低了陷入局部最优的概率. 其优势在于能搜索和定位多个全局/局部峰值, 缺点在于要找到全局最优通常需要更多的函数评估. 小生境技术具有相对简单、有效和通用的特性, 因此可以作为进化算法强有力的组成部分. 它可促使群体内个体间协同合作, 使算法易于找出优化问题的所有局部最优解和全局最优解. 小生境已经被证明能够非常有效的解决多解问题^[17].

近年来, 多模多目标优化成为多模优化研究领域的热点问题. Yue等人^[18]基于指数的环形拓扑结构产

生稳定的小生境, 从而能够识别更多的帕累托最优解, 并在决策空间和目标空间中采用特殊拥挤距离的概念作为密度度量, 提出了一种新的粒子群优化算法, 用于求解具有多个帕累托最优解的多模多目标优化问题. 该算法不仅能定位和保持大量的帕累托最优解, 而且在决策空间和目标空间都能得到良好的分布. Liu等人^[19]提出了一种新的多模态多目标进化算法, 采用双存档和重组策略. 该算法首先分析了决策变量的性质和它们之间的关系, 以指导进化搜索. 然后采用融合档案和多样性档案两个档案的总体框架来协同解决这些问题. 同时采用聚类策略保证目标空间的多样性, 采用基于小生境的清除策略提高决策空间的多样性, 在进化过程结束时, 对收敛性和多样性档案中的解进行重组, 得到大量的多重帕累托最优解. Liang等人^[20]提出了一种多模态多目标差分进化优化算法用于求解帕累托多模问题. 该算法的主要贡献是制定了一个决策变量预选方案, 在决策空间和目标空间促进了解决方案的多样性. 同时还引入了一种新的突变结合过程, 作为差分进化方法中经典突变方案的补充, 该方案给予搜索边界之外的后代第2次突变机会, 从而减少了搜索空间边界上的个体密度. Qu等人^[21]提出了一种基于自组织物种形成的多目标粒子群优化算法, 用于求解多模态多目标问题的多帕累托最优解. 该方法利用物种形成策略形成稳定的生态位, 并对这些生态位/亚种群进行优化, 以并行地搜索和保持帕累托最优解. 在此基础上, 提出了一种自组织机制, 以提高物种形成的效率和算法的性能, 为了保持决策空间和目标空间解的多样性, 将非支配排序方案和特殊拥挤距离技术相结合. Lin等人^[22]提出了一种在决策空间和目标空间中具有对偶聚类的多模态多目标进化算法. 在决策空间中运行一个聚类来聚集邻近的解, 将解划分为多个局部聚类. 首先选择每个局部聚类内的非支配解以保持局部帕累托集, 然后选择在目标空间内具有较好的收敛性的非支配解, 形成一个有 N 个以上解的临时种群(N 为种群规模). 然后, 对该临时种群在目标空间进行第2次聚类, 最终得到目标空间中具有良好多样性的 N 个聚类. 最后, 在上述聚类上反复运行一个剪枝过程, 直到每个聚类只有一个解决方案, 每次都从目标空间最拥挤的聚类中删除决策空间中最拥挤的解决方案. 这样, 决策空间的聚类可以区分所有的帕累托集, 避免局部帕累托集的损失, 而目标空间的聚类则可以保持目标空间的多样性. Liang等人^[23]设计了基于聚类的特殊拥挤距离方法来计算决策空间和目标空间的综合拥挤程度. 同时引入基于距离的精英选择机制来确定不同个体的学习范例. 在样本周围生成新的个体, 从而在决策空间和客观空间中获得均匀分布的种群. 在此基础上提出了一种基于聚类技术和精英选择机制的差分进化算法来

解决多模态多目标问题.

利用反演思想求解污染源定位问题时存在着潜在的解的非唯一性问题, 主要的原因在于由于经济因素, 一般在管网中布置的传感器数量远少于管网节点数目, 不同来源的污染物排放特性显著不同, 但可能预测误差相似, 因此该问题具有多个全局极值, 这导致所得解的精度大大下降. 对于该类优化问题, 算法的目标是在一次执行中搜索到所有的极值点. 污染源定位问题已经被证明是一个多模函数优化问题^[24].

针对污染源定位问题解的非唯一性的特征, Yan等人^[25]将该问题转化为一个多模优化问题, 在式(1)的基础上对问题模型进行了改进, 描述如下:

$$\left\{ \begin{array}{l} \min_{M, n, t_I} f = \sum_{j=1}^{N_s} \sum_{t=1}^{T_s} (c_j(t) - c_j^*(t))^2, \\ \text{s.t. } M = \{m_1, m_2, \dots, m_k\}; m_i \geq 0, \\ n \in \{1, \dots, N\}, \\ t_I \leq T_s, \\ \varepsilon = \sqrt{\sum_{j=1}^{N_s} \sum_{t=1}^{T_s} (c_j^*(t))^2}, \end{array} \right. \quad (2)$$

其中: N 是管网的节点总数, N_s 表示传感器的数目, T_s 表示仿真周期, M 表示污染物注入向量, n 表示污染源注入的管网节点序号, t_I 表示注入污染物的初始时间, $c_j(t)$ 表示在时间 t 时传感器 j 的污染物浓度, 它是 (M, n, t_I) 的函数, $c_j^*(t)$ 表示在时间 t 时传感器 j 的实际检测的污染物浓度. EPANET模拟个体对应的污染事件, 当水质传感器检测不到任何污染物时, 此时 $f = \sum_{j=1}^{N_s} \sum_{t=1}^{T_s} (c_j^*(t))^2$; 当污染事件偏离真实污染事件时, 水质传感器检测的浓度与真实传感器浓度相差更大, 致使个体适应度值 $f > \sum_{j=1}^{N_s} \sum_{t=1}^{T_s} (c_j^*(t))^2$; 当污染事件接近真实污染事件时, 水质传感器检测的浓度与真实浓度更相近, 个体适应度值 $f < \sum_{j=1}^{N_s} \sum_{t=1}^{T_s} (c_j^*(t))^2$. 随着优化算法的逐渐收敛, 个体的适应度值不断减小, 通过实验观察, 当适应度 $f < \sqrt{\sum_{j=1}^{N_s} \sum_{t=1}^{T_s} (c_j^*(t))^2}$ 时, 表示算法足够收敛, 即该污染事件可以当作满足条件的污染事件. 在改进的问题模型下, 迭代条件由取得更小的适应度值变为取得满足阈值条件的适应度值.

文章采用动态小生境遗传算法, 通过算法的一次执行计算出多个潜在的污染源, 为筛选真实污染源提供了可能性, 该研究认为, 当污染事件在传感器处的累积模拟浓度和实际累积检测浓度最小方差小于阈值 ε 时, 则认为该污染事件的注入节点为实际的污染源, 最终的实验结果表明小生境遗传算法在解决污染

源定位这个多模问题时是有效的. 然而, 文章并未在大规模管网中进行实验, 也没有考虑管网中各种动态因素, 如变化的用户水需求等, 因此该算法在求解大规模污染源定位问题时的有效性还有待验证.

3.2 昂贵优化

在实际工程问题中, 许多目标函数不能用解析式表达出来, 优化模型也十分复杂, 这时就需要运用仿真模型来模拟或评价目标函数, 而这种仿真模型调用一次计算量往往大于简单函数计算, 甚至一些特殊问题的仿真模型本身造价就高, 多次使用仿真模型将耗费高昂的时间代价和经济代价, 这类问题被称为昂贵优化问题.

在模拟-优化法求解污染源定位问题的模型中, 需要使用EPANET模拟器模拟污染事件, 进而计算预测误差. 而调用EPANET需要消耗一定时间成本, 当使用优化算法求解时, 作为一种靠迭代寻优的方法, 无疑会消耗大量的计算时间和资源. 以管网BWSN2^[26]为例(该管网包含12527个节点、2个水库、2个池塘和20个传感器), 模拟一个污染事件, 计算适应度值大约需要3 s. 当使用种群规模100的遗传算法运行100代求解时, 需要花费329 min, 接近5.5 h. 在目前的研究中, 小型管网如Rossmann2000 (92个节点)所需计算时间较少, EPANET模拟时间也较少, 然而现实情况中的供水管网节点数量通常比较大, 必然会消耗更多的时间, 这在实际应用中是不被允许的. 因此, 污染源定位问题是一个昂贵优化问题.

普通的优化问题可以通过迭代计算得出最优解, 然而昂贵优化问题受限于高昂成本, 不能简单地依靠大量迭代计算解决, 优化的主要目的变成了如何尽可能少地使用仿真模型同时不损失解得精度. 为了克服这些障碍, 基于代理的智能优化算法已经被广泛使用. 该方法使用一个计算花费小的近似模型去替代部分昂贵的适应度函数评价, 近似模型也被称作为元模型或者代理模型.

代理模型的种类有很多, 选择合适的代理模型对算法非常重要. 常用的代理模型^[27]大致有以下几类:

1) 多项式响应曲面法 (polynomial response surface, PRS). 响应面方法 (response surface methodology, RSM)^[28]采用统计技术进行回归和方差分析, 以获得响应的最小方差, 多项式的简单性使它们成为近似大多数多项式响应曲面的好方法.

2) 克里金法(Kriging), 也称高斯代理模型. 高斯随机过程^[29]是随机变量的集合, 其中任意数量的样本点具有联合高斯分布. 高斯随机过程能模拟出函数的分布情况, 它不仅能进行回归预测, 而且还能计算未知区域的不确定性. 一般的回归模型, 仅仅能够进行预测, 无法判断该预测是否正确, 而高斯随机过程能够通过不确定性反映预测值的可信度. 由于这一特性

使得高斯随机过程在很多领域得到了发展,也成为代理模型中最为广泛的模型之一。

3) 径向基函数(radial basis function, RBF). 径向基函数方法由Hardy^[30]在1971年提出, RBF是一个实值函数,其值仅取决于从输入到神经元中心的距离,只要符合一定条件的函数,包括线性、立方、多元二次或高斯函数都可以作为核函数。

4) 支持向量机(support vector machine, SVM). 支持向量机^[31]从统计学习理论中汲取灵感,是一种相关的监督学习方法,用于分析数据和识别模式. SVM在高维中构造超平面或一组超平面空间,可用于分类和回归。

5) 人工神经网络 (artificial neural network, ANN). 人工神经网络是大量神经元互连而成的复杂网络. 网络中的每个神经元都能够接收输入信号,处理它们并发送输出信号. 它由一组加权的突触、一个加法器和一个激活函数组成. 加法器用于计算输入信号经神经元相应突触加权的和,激活函数用于限制神经元输出的幅度. 网络架构有两种不同的神经网络,包括多层前馈网络和循环网络. Komogorov定理表明,具有3层的多层感知器可以准确的模拟任何线性和非线性函数. 人工神经网络也经常用于近似复杂、耗时的目标函数,或者耗时的仿真实验。

在优化算法框架中嵌入代理模型,根据不同的情况选择使用代理模型还是真实昂贵函数的策略称为模型管理. 模型管理同样会影响算法的性能,管理方式分为离线与在线. 离线模型在训练好模型后就不再更新模型,如果数据量足够多,那么问题变成了如何自适应地选择数据子集来减少计算时间^[32],如果数据量很少且有噪声,就需要相应的策略来建立精准的模型;在线模型可以通过获得的新数据不断更新代理模型,这种模型管理方式会让模型更精确,但同时更新模型也会消耗计算资源. 代理模型的应用可以根据在优化过程中是否更新大致分为两种:一种是代理模型建立之后在优化过程中保持不变,一直使用已建立的模型;另一种是在优化过程中代理模型不断更新,每次迭代就再次建立代理模型. 面对不同的实际问题,选择不同的应用方式:当样本量较少时,就需要优化过程中的数据来更新模型,以提高模型精确度;而样本量比较大时,如果采用以上方式就会耗费大量的计算时间,因此根据大量样本建立高精度模型,并在优化过程中使用已建立的高精度模型. 近年来,代理模型相关的算法和应用越来越受到关注,许多研究人员获得了丰硕的成果. Chugh等人^[33]使用Kriging来近似目标函数来降低计算成本. 利用Kriging模型得到预测点的不确定性,平衡算法的多样性与收敛性,并且他们还提出了模型样本点选取策略. Yang等人^[34]在高维优化问题中,采用分而治之的解决方案,提出子问

题的搜索应被视为昂贵优化问题,并通过代理模型解决. Chen等人^[35]将代理模型应用在运输网的最优收费方案中,建立近似的反应曲面,减少求解时间,从结果表明:代理模型的使用,使得预测的最优收费可以让社会受益. Pan等人^[36]将代理模型引入到高维多目标优化问题中,算法使用人工神经网络预测候选解与参考解之间的支配关系,既考虑个体的优势关系,也考虑模型的精确度. Yu等人^[37]提出一个基于代理模型的分层粒子群优化器来求解高维问题,该算法是标准粒子群优化算法和社会学习粒子群优化算法的结合,使用这两种算法在搜索空间中一起来进行搜索,同时增强代理模型全局和局部的搜索性能. Wang等人^[38]提出了一种综合考虑全局勘探和局部开采两种能力的演化抽样辅助优化方法. 利用变异算子和交叉算子,采用差分进化方法产生后代. 建立了全局径向基函数代理模型,对后代的目标函数值进行预筛选,并筛选出最优的目标函数值,用真函数对目标函数值进行评价. 当其函数值小于亲本时,其最佳后代将取代亲本在种群中的位置. 然后使用选定的当前最佳解决方案构建局部代理模型. Li等人^[39]提出了一种求解高维计算昂贵问题的代理辅助多种群优化算法,该算法包括两个种群:第1个是使用基于教学优化的学习者阶段来增强探索,第2个是使用粒子群优化来加快收敛速度,这两个种群可以互相学习. 同时提出了一种动态种群大小调整方案来控制进化过程. 为了进一步提高粒子群算法在不同功能场景下的搜索效率,采用了两个坐标系来生成粒子群算法的理想位置. 此外,提出了一种新的预筛选准则来选择有潜力的个体进行精确的功能评估. Cai等人^[40]提出了一种代理引导的差分进化算法. 与其他代理辅助的元启发式算法不同的是,该算法实现了差分进化算法与代理算法之间的融合,而不仅仅是作为加速元启发式算法收敛的额外工具. 具体而言,该算法将全局和局部代理预测的最优值与变异算子相结合,使变异算子引导差分进化算法的变异方向,从而使差分进化算法快速收敛. Cai等人^[41]还提出了一种高效的代理辅助的粒子群优化算法,该算法有时需要进行昂贵的仿真分析. 与其他几种代理辅助的元启发式算法不同之处在于该算法在优化过程中能够有效平衡代理模型的预测能力和粒子群优化的全局搜索能力. 该算法利用建立在整个设计空间的全局代理模型和建立在个人历史最佳粒子周围的邻近区域的局部代理模型来更新粒子的速度. 需要强调的是,用于获得最优局部代理模型的邻域划分策略是该算法的一个重要方面,该策略有助于获得邻域局部代理的预测最优解,以指导优化过程中粒子群优化的搜索。

将污染源定位问题看作是一个昂贵优化问题后,在原有的模拟-优化算法基础上就是以代理模型来代

替部分EPANET软件计算误差的过程,即用简单的数学模型替代复杂的仿真软件,达到节约计算资源的目的.基本的思路如下:首先建立一个与目标函数相近的模型,使用样本集建立一个误差较低的代理模型.代理模型的目的是代替真实的评价函数对个体进行适应度评价,在保证算法精确度的情况下,尽可能的减少时间成本较大的真实目标函数的使用.与普通的优化算法不同,昂贵优化算法的模型中,在种群产生新的后代后,可以选择代理模型代替EPANET模拟器评价新的个体,或者仍然使用原始目标函数.如何平衡高斯代理模型和EPANET模拟器间使用对算法的优化过程有着直接的影响.如果使用代理模型的次数过多,模拟评价个体适应度的误差较大,直接影响算法的收敛方向;如果使用EPANET模拟器次数过多,又不能达到减少计算代价的目的.因此,在优化算法求解污染源定位问题中,不仅需要根据问题特性建立代理模型,而且需要对代理模型使用进行设计.

此时问题转化为如何平衡EPANET和代理模型之间的使用,在不影响求解精度的同时尽可能减少计算时间. Gong等人^[42]通过实验得出使用高斯代理模型和不使用高斯代理模型分别调用EPANET的次数,前者约为后者的 $\frac{1}{5}$.

Preis和Ostfeld^[43]在2006年首次提出了一种利用耦合模型树-线性规划算法进行供水系统污染源识别的方法,模型树取代了EPANET模拟器,线性规划算法用模型树来识别污染源,对于一个给定的网络,模型树结构模拟网络对不同随机污染事件的响应,该方法提供了对时间、位置和污染注入浓度的估计. Gong等人^[42]提出了一种基于协同的昂贵优化算法,该算法根据供水管网的特点,针对不同的变量分别提出不同策略来指导算法的搜索方向,算法尽可能采用高斯代理模型,以保证识别精度,最后通过仿真实验,验证了所提算法的有效性和稳定性. Yan等人^[44]根据供水管网的特性,建立了管网中每个节点的高斯代理模型,提出了一种基于高斯代理模型的昂贵优化算法来解决供水管网中的污染源定位问题,并通过仿真实验证明了算法的有效性和高效性.

代理模型虽然代替部分EPANET,减少了算法运行时间,但代理模型与真实评估函数间仍存在差距,可能会引导算法向错误的方向收敛,导致算法陷入局部最优,代理模型使用比例越高,误导率也会越高.因此在基于代理的优化算法需要平衡代理模型与EPANET仿真软件之间的使用,既不损失解精度,保证算法朝正确的方向收敛,又能尽可能减少计算时间.

3.3 不确定性

现实生活中很多实例都存在不确定性,例如不断有新任务到达的调度问题.不确定性使优化问题的复

杂性大幅度提升,传统的静态优化算法已不能满足问题需求.常见的不确定性因素包括适应度值发生变化、问题实例变化和约束条件变化,这些因素都使问题转化为一个动态优化问题(dynamic optimization problems, DOPs)^[45],其定义如下:

$$\text{DOP} = \begin{cases} \min f(x, t), \\ \text{s.t. } x \in X(t) \subseteq S, t \in T, \end{cases} \quad (3)$$

其中: $S \in \mathbb{R}^n$, S 是搜索空间, t 是时间, $S \times T \rightarrow \mathbb{R}$ 是目标函数, x 是可行解, $X(t)$ 是在 t 时间内可行解集合. 对于一个最小化问题, DOPs的目标是寻找一系列随时间变化的最优个体集合 $X^* = \{x_t^*\}$, $t \in T$ 满足 $f(x_t^*, t) \leq f(x, t), \forall x \in X(t)$.

有关动态优化问题的研究已经成为优化算法研究领域的一个热点,传统的进化算法(evolutionary algorithms, EA)在解决动态优化问题时,一旦收敛,失去了探索新区域的多样性,当新的环境到来时,就不能跟踪到新环境里继续搜索了.为了解决DOPs,静态优化算法需要在演化过程中对变化的环境做出及时的反应并跟踪最优值.2005年后有大量动态环境进化算法被提出,学者们关注的问题主要在于两个方面:环境监测;保持群体多样性.检测通常是通过重新评估解的适应度值来判断环境是否变化的,除了在环境发生变化后增加群体多样性,一些学者还通过在算法运行过程中维持多样性、设置记忆机制、预测机制、多种群机制等方法设计进化算法^[46].

前面提到的污染源定位问题研究都假设供水管网系统的水力特性是已知的,例如压力、流量、消耗、水箱水位等,而实际情况是只有部分信息能被获取.当污染事件发生时,供水运营商需要在不确定的环境下做出决策,他们只能利用有限的信息,并且水管网的水力因素也会因消费者的反应动态变化,这些都使得管网系统的动态性更加复杂,导致优化目标函数可能随着时间改变,问题的最优解也在随时改变. Ostfeld在文献[47]中总结了供水系统中包括的不确定性和风险,并提出了这些问题带来的相应挑战. Preis等人^[48]提出一种量化不确定性的方法,在先前开发的污染源模型的基础上,将水需求和传感器传输的数据中的不确定性纳入其中.

考虑不确定性的污染源定位问题具有两个复杂的特性:

1) 不确定的用水需求数据. 供水管网污染源定位系统的不确定性包括3个方面:由于抽样的偏差和检测的错误导致的观测误差和传感器测量误差;由于水利水质模型中的简化和不正确的假设导致的模型误差;由于未知的即时水需求的波动导致的不确定的用水需求,用水消费水平的固有变化,消费者节点水需求是不确定的主要来源.由以上存在的不确定性,给

水管网污染源定位问题不能使用确定性方法(假设系统在完全已知的条件下运行)解决。

2) 突发性污染源的实时特性. 在采用模拟-优化的方法求解污染源定位问题时, 为了尽量减小污染物对公共健康的危害, 需要在发生突发性事件时能够实时定位污染源位置, 即当水质传感器开始检测到污染获取了水质信息后开始模拟, 需要尽可能快的定位污染源位置。

Liu等人^[49]提出一种基于多种群的动态优化算法, 该方法利用流数据可来自自适应估计源的特性, 在检测到污染事件后, 源特性的估计值会随着新的观测数据不断更新, 从而动态更新预测误差函数, 改变优化模型中的目标函数. 随着更多观测数据的添加, 多种群方案会逐渐寻找到更好地解决方案, 同时减少解的非唯一性程度. Laird等人^[50]提出了一种基于子域的动态优化方法, 以减少过多的计算时间, 研究表明子域方法减少了用于优化的可用信息数量, 因此减少了计算时间, 同时提高了解的质量. Vankayala等人^[51]采用高斯模型和自回归模型综合生成随机水需求, 将随机遗传算法和噪声遗传算法应用于同一问题进行比较. 结果表明噪声遗传算法比随机遗传算法具有更好的鲁棒性和更小的计算开销, 自回归模型比高斯模型更能代表污染源识别过程中的不确定性。

Yan等人分别采用高斯模型^[52]、自回归模型^[53]和齐次泊松模型^[54]来模拟用户的需水量, 然后提出一种改进的遗传算法来解决不确定需水量条件下的饮用水污染源定位问题, 并在不同规模的供水管网上进行仿真实验对比。

1) 高斯模型是一个加入简单白噪声干扰的完全不相关模型, 被应用在缺少先验概率的情形下, 也就是在没有大量的观测到的用水需求数据下使用, 在高斯模型中, 在每个节点每个时间步长下的用水需求被加入一个简单的高斯扰动。

2) 自回归模型是一种随机的、线性的需求变化模型, 是用自身做回归变量的过程, 即利用前期若干时刻的随机变量的线性组合来描述以后某时刻随机变量的线性回归模型^[55], 当已知大量的用水需求数据时, 适合使用自回归模型建立水需求变化模型。

3) 泊松分布模型是一种统计与概率学里常见的离散机率分布, 与前两种模型不同, 泊松分布可以给出管网中任意节点的瞬时水需求数据。

文章分别用上述3种模型产生的若干组用水需求数据, 与基础用水需求数据的曲线作对比, 结果显示: 高斯模型具有很大的随机性, 自回归模型更贴近真实的用水需求, 泊松的随机性居中; 但是3种不同的水需求变化模型产生的水需求符合居民每日用水变化规律. 作者将不确定的用水数据作为模拟-优化模型的输入, 在式(1)的基础上加入不确定的居民用水需求以

及模拟时间, 采用最小最大化准则求解不确定水需求的污染源定位问题. 不确定水需求污染源定位模型如下:

$$\begin{cases} \text{Find}\{L, C_0, t\}, \\ \min F = \max_{k=1, \dots, N_s} \left\{ \sum_{t=1}^{T_s} [C_k^{\text{obs}}(t) - C_k^*(L, C_0, t)]^2 \right\}, \\ \text{s.t. } M = \{m_1, m_2, \dots, m_k\}; m_i \geq 0, \\ n \in \{1, \dots, N\}, \\ t_I \leq T_s, \end{cases} \quad (4)$$

其中: L 表示污染源位置节点, C_0 表示在已知的污染注射时间内源污染物浓度, t 表示当前的时间步长, 需要求解得到注入污染的污染源位置节点和注入污染物的质量; F 表示预测误差, 目标函数是最小化预测误差, N_s 表示传感器的数量, T_s 表示仿真周期, $C_k^{\text{obs}}(t)$ 表示在时间 t 时传感器 k 处的观测污染物浓度, 它是 (M, n, t_I) 的函数, $C_k^*(L, C_0, t)$ 表示在时间 t 时传感器 k 处的实际污染物浓度; M 表示污染物注入向量, n 表示污染源注入的管网节点序号, t_I 表示注入污染物的初始时间。

文章改进基础遗传算法求解该模型, 实验结果证明了提出改进的遗传算法具有更高的准确性和更好的鲁棒性, 最后给出了对供水管网突发性污染源实时定位问题的有效求解方法. 但文章只展现了两个管网中的实验数据, 且并未在大规模管网中进行研究, 在建立自回归模型时并未使用真实的居民用水需求数据, 而使用由高斯模型产生的与基准居民用水曲线具有相同规律的不确定居民用水数据, 因此文中的工作还可基于真实用水数据进行验证, 并尝试运用在大规模管网中以贴近现实城镇管网情况。

除了需水量不确定, 突发性污染也是管网中的不确定性因素. Yan等人^[56]完善了这方面的研究, 他们在研究需水量不确定的情况下, 突发污染源的实时定位问题, 当污染源被检测到后, 用提出的实时定位算法对污染源进行跟踪, 实验结果表明, 与传统的污染源定位方法相比, 文章提出的方法能够以较少的传感器数据在较短时间内找到真实的污染事件。

4 展望

在环境问题日益突出的今天, 研究饮用水污染源定位问题对环境保护有重要意义. 饮用水污染源定位问题与居民生活安全息息相关, 一直以来都是亟需解决的热点问题, 随着智能优化算法的发展, 目前将模拟-优化方法用于污染源定位问题的研究成果已有许多, 但仍有许多工作需要研究与拓展, 下面就问题、算法和应用等层面给予展望。

1) 问题层面.

一方面,目前的基于模拟-优化方法的污染源定位研究通常对问题假设过多,与实际环境差距较大。许多研究都假设污染物为非衰减类,且不与水中其他物质发生化学生物等反应。但在实际供水管网中的真实情况并不是这样的,因此,研究污染物具有衰减性,且与水中其他物质发生化学生物反应的饮用水污染源定位问题更具有现实意义。同时,对于供水管网入侵污染物流量问题,现有的研究是假定了污染物入侵流量的多少并不影响该管网的水力工况,然而供水管网实际突发污染事故的情况是不可预测的,污染源的个数、污染物浓度、侵入方式、持续时间等均不可预知。因此,研究不确定环境下的污染源定位问题更具挑战性。研究应考虑供水管网的真实情况和规模,充分研究不确定性对问题带来的影响。对问题的建模可考虑加入约束条件以模拟现实因素,使研究成果更具现实意义。另一方面,目前针对污染源定位的理论研究很少,而且大多针对单污染源定位问题。围绕污染源定位问题,分析问题的特性,剖析解空间的特征,研究高效的编码和解码策略、快速性能评价机制、定位过程中的有效调整策略,无疑将有助于推动污染源定位的理论与方法的发展。

2) 算法层面。

污染源定位问题的求解复杂性随着管网规模的增大将急剧增加,目前的算法大多局限于机器学习方法和智能优化方法,而基于数学模型和仿真软件的方法只适于小规模问题。对于智能优化方法而言,目前的方法主要是遗传算法和演化策略。一方面,针对污染源定位问题的特征提出特定的智能优化算法,包括基于人工智能的方法;另一方面,将目前流行的群体智能算法、差分进化、多模优化、昂贵优化算法等推广应用于污染源定位,均是很有意义的研究工作。智能优化算法的关键在于遗传操作的设计、参数的设定以及全局搜索与局部搜索的均衡。因此,研究的重点在于合理的搜索策略、对问题特征的知识挖掘、基于问题特征的高效的编码和解码策略、参数的自适应机制、有效的遗传操作的混合智能优化算法以及协同智能优化算法。另外,求解精度仍需进一步提高,其中污染源位置是最主要的评判指标,后续研究在保证位置准确的情况下可进一步提高时间与污染物浓度求解的精确程度。每种智能优化算法都有其缺陷,使用单一算法无法避免,可以考虑使用集成算法。当然,智能优化算法的计算复杂性、收敛性、下界等理论问题也值得关注。

3) 应用层面。

开展理论研究主要是为了能指导实际应用,同时应用研究也对理论研究有促进作用。虽然污染源定位问题的相关理论和方法在某些实际问题中得到了初步应用,但离实际应用还有不小的差距。一方面,要加

强对实际问题的理解和深入研究,问题特征分析、模型构建、算法设计等工作均要跟实际问题紧密结合;另一方面,要对现有应用研究的范围进行拓展,比如:充分考虑到公众健康的需求,污染源定位的时间仍需要进一步缩短,尽可能快地为政府部门提供污染源信息以便及时制定应对措施。可在保证求解精度的同时尽量缩短算法求解的时间;针对污染源定位问题的3个特性:多模性、昂贵性和不确定性,现有的研究都是侧重解决某一类问题,后续研究可以将其两个特性同时考虑,或3个特性同时考虑,完善问题的数学模型,使之更符合实际情况;现有的研究中求解污染源定位问题时都是假设传感器已经部署在设定点上,没有考虑传感器位置的最优布置,后续研究可以结合传感器部署问题进行,以更优化的传感器位置改善污染源定位问题的求解精度和求解时间。

参考文献:

- [1] CRISTO C D, LEOPARDI A. Pollution source identification of accidental contamination in water distribution networks. *Journal of Water Resources Planning and Management*, 2008, 134: 197 – 202.
- [2] RASEKH A, BRUMBELOW K. Adaptive emergency response to water distribution system contamination events. *World Environmental and Water Resources Congress 2012: Crossing Boundaries*. United States: American Society of Civil Engineers, 2012: 2981 – 2986.
- [3] SCHWARTZ R, LAHAV O, OSTFELD A. Optimal sensor placement in water distribution systems for injection of chlorpyrifos. *World Environmental and Water Resources Congress*. United States: American Society of Civil Engineers, 2014: 485 – 494.
- [4] HU C, REN G, LIU C, et al. A spark-based genetic algorithm for sensor placement in large scale drinking water distribution systems. *Cluster Computing*, 2017, 20(2): 1 – 12.
- [5] HU C, SHU X, YAN X, et al. Inline wireless mobile sensors and fog nodes placement for leakage detection in water distribution systems. *Software Practice and Experience*, 2020, 50(7): 1152 – 1167.
- [6] PROPATO M, CHEUNG P B, PILLER O. Sensor location design for contaminant source identification in water distribution systems. *Water Distribution System Analysis Symposium*, 2006, DOI: 10.1061/9780784409411.
- [7] LAIRD C, BIEGLER L, WAANDERS B. Contamination source determination for water networks. *Journal of Water Resources Planning and Management*, 2005, 131(2): 125 – 134.
- [8] ROSSMAN L A. Epanet 2 users manual. *Laboratory Office of Research and Development U.S. Environmental Protection Agency*, 2000, 19(1): 115 – 118.
- [9] YAN X, GONG W, WU Q. Contaminant source identification of water distribution networks using cultural algorithm. *Concurrency and Computation: Practice and Experience*, 2017, 29, DOI: 10.1002/cpe.4230.
- [10] GUAN J, ARAL M, MASLIA M. Identification of contaminant sources in water distribution systems using simulation-optimization method: case study. *Journal of Water Resources Planning and Management*, 2006, 132(4): 252 – 262.
- [11] PREIS A, OSTFELD A. Multiobjective sensor design for water distribution systems security. *Water Distribution System Analysis Symposium*. United States: American Society of Civil Engineers, 2006.
- [12] YAN X, ZHAO J, HU C, et al. Contaminant source identification in water distribution network based on hybrid encoding. *Journal of*

- Computational Methods in Sciences and Engineering*, 2016, 16(2): 379 – 390.
- [13] LEICHOMBAM S, BHATTACHARJYA R K. New hybrid optimization methodology to identify pollution sources considering the source locations and source flux as unknown. *Journal of Hazardous, Toxic, and Radioactive Waste*, 2019, 23(1).
- [14] QU B, LIANG J, SUGANTHAN P. Niching particle swarm optimization with local search for multi-modal optimization. *Information Sciences*, 2012, 197(none): 131 – 134.
- [15] PARROTT D, LI X. Locating and tracking multiple dynamic optima by a particle swarm model using speciation. *IEEE Transactions on Evolutionary Computation*, 2006, 10(4), 440 – 458.
- [16] LIU Q, DU S, WYK B J V. Niching particle swarm optimization based on Euclidean distance and hierarchical clustering for multimodal optimization. *Nonlinear Dynamics*, 2020, 99(3): 2459 – 2477.
- [17] LI X, EPITROPAKIS M, DEB K, et al. Seeking multiple solutions: an updated survey on niching methods and their applications. *IEEE Transactions on Evolutionary Computation*, 2016, 99: 1 – 21.
- [18] YUE C, QU B, LIANG J. A multiobjective particle swarm optimizer using ring topology for solving multimodal multiobjective problems. *IEEE Transactions on Evolutionary Computation*, 2018, 22(5): 805 – 817.
- [19] LIU Y, YEN G G, GONG D. A multi-modal multi-objective evolutionary algorithm using two-archive and recombination strategies. *IEEE Transactions on Evolutionary Computation*, 2019, 23(4): 660 – 674.
- [20] LIANG J, XU W. Multimodal multiobjective optimization with differential evolution. *Swarm and Evolutionary Computation*, 2019, 44: 1028 – 1059.
- [21] QU B, LI C, LIANG J, et al. A self-organized speciation based multi-objective particle swarm optimizer for multimodal multi-objective problems. *Applied Soft Computing*, 2020, 86: 105886.
- [22] LIN Q, LIN W, ZHU Z, et al. Multimodal multi-objective evolutionary optimization with dual clustering in decision and objective spaces. *IEEE Transactions on Evolutionary Computation*, 2021, 25(1): 130 – 144.
- [23] LIANG J, QIAO K, YUE C, et al. A clustering-based differential evolution algorithm for solving multimodal multi-objective optimization problems. *Swarm and Evolutionary Computation*, 2021, 60: 100788.
- [24] HU C, ZHAO J, YAN X. A mapReduce based parallel niche genetic algorithm for contaminant source identification in water distribution network. *Ad Hoc Networks*, 2015, 35: 116 – 126.
- [25] YAN X, ZHAO J, HU C, et al. Multimodal optimization problem in contamination source determination of water supply networks. *Swarm and Evolutionary Computation*, 2019, 47: 66 – 71.
- [26] OSTFELD A, UBER J, SALOMONS E, et al. The battle of the water sensor networks (BWSN): a design challenge for engineers and algorithms. *Journal of Water Resources Planning and Management*, 2008, 134(6): 556 – 568.
- [27] DÍAZ-MANRÍQUEZ A, GREGORIO T, HUGO B. A review of surrogate assisted multiobjective evolutionary algorithms. *Computational Intelligence and Neuroscience*, 2016, DOI: 10.1155/2016/9420460.
- [28] GUNST R. Response surface methodology: process and product optimization using designed experiments. *Technometrics*, 2008, 38(3): 284 – 286.
- [29] DONALD R, MATTHIAS SCHONLAU, WILLIAM J. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 1998, 13(4): 455 – 492.
- [30] HARDY, ROLLAND L. Multiquadric equations of topography and other irregular surfaces. *Journal of Geophysical Research*, 1971, 76(8): 1905 – 1915.
- [31] VAPNIK V. Statistical learning theory. *Encyclopedia of the Ences of Learning*, 1998, 41(4): 3185 – 3185.
- [32] WANG H, JIN Y, JANSEN J. Data-driven surrogate-assisted multi-objective evolutionary optimization of a trauma system. *IEEE Transactions on Evolutionary Computation*, 2016, 20(6): 939 – 952.
- [33] CHUGH T, JIN Y, MIETTINEN K, et al. A surrogate-assisted reference vector guided evolutionary algorithm for computationally expensive many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 2018, 22(1): 129 – 142.
- [34] YANG P, TANG K, YAO X. Turning high-dimensional optimization into computationally expensive optimization. *IEEE Transactions on Evolutionary Computation*, 2018, 22(1): 143 – 156.
- [35] WANG H, JIN Y, SUN C, et al. Offline data-driven evolutionary optimization using selective surrogate ensembles. *IEEE Transactions on Evolutionary Computation*, 2019, 23(2): 203 – 216.
- [36] PAN L, HE C, TIAN Y, et al. A classification-based surrogate-assisted evolutionary algorithm for expensive many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 2019, 23(1): 74 – 88.
- [37] YU H, TAN Y, ZENG J, et al. Surrogate-assisted hierarchical particle swarm optimization. *Information Sciences*, 2018, 454/455: 59 – 72.
- [38] WANG X, WANG G, SONG B, et al. A novel evolutionary sampling assisted optimization method for high dimensional expensive problems. *IEEE Transactions on Evolutionary Computation*, 2019, 23(99): 815 – 827.
- [39] LI F, CAI X, GAO L, et al. A surrogate-assisted multiswarm optimization algorithm for high-dimensional computationally expensive problems. *IEEE Transactions on Cybernetics*, 2020, 99: 1 – 13.
- [40] CAI X, GAO L, LI X, et al. Surrogate-guided differential evolution algorithm for high dimensional expensive problems. *Swarm and Evolutionary Computation*, 2019, 48: 288 – 311.
- [41] CAI X, QIU H, GAO L, et al. An efficient surrogate-assisted particle swarm optimization algorithm for high-dimensional expensive problems. *Knowledge-Based Systems*, 2019, 184: 1 – 17.
- [42] GONG J, YAN X, HU C, et al. Collaborative based pollution sources identification algorithm in water supply sensor networks. *Desalination and Water Treatment*, 2019, 168: 123 – 135.
- [43] PREIS A, OSTFELD A. Contamination source identification in water systems: a hybrid model trees-linear programming scheme. *Journal of Water Resources Planning and Management*, 2006, 132(4): 263 – 273.
- [44] YAN X, YANG K, HU C. Pollution source positioning in a water supply network based on expensive optimization. *Desalination and Water Treatment*, 2018, 110: 308 – 318.
- [45] CHEN Li, DING Lixin. Survey on dynamic optimization algorithms. *Journal of Wuhan University (Natural Science Edition)*, 2011, (3): 255 – 264.
(陈莉, 丁立新. 动态优化算法综述. 武汉大学学报: 理学版, 2011, (3): 255 – 264.)
- [46] NGUYEN T T, YANG S, BRANKE J. Evolutionary dynamic optimization: a survey of the state of the art. *Swarm and Evolutionary Computation*, 2012, 6: 1 – 24.
- [47] OSTFELD A. Uncertainty and risk inclusions in water distribution systems management: review and challenges. *International Conference on Vulnerability and Risk Analysis and Management*. 1985: 1980 – 1985.
- [48] PREIS A, PERELMAN L, OSTFELD A. Uncertainty quantification of contamination source identification. *World Environmental and Water Resources Congress*. 2008: 1 – 6.
- [49] LIU L, RANJITHAN S R, MAHINTHAKUMAR G. Contamination source identification in water distribution systems using an adaptive dynamic optimization procedure. *Journal of Water Resources Planning and Management*, 2010, 137(2): 183 – 192.

- [50] WAANDERS B, BIEGLER L, LAIRD C. Real-time, large scale optimization of water network systems using a subdomain approach. *Journal of Indian Prosthodontic Society*, 2005, 8(2): 116 – 118.
- [51] VANKAYALA P, SANKARASUBRAMANIAN A, RANJITHAN S. Contaminant source identification in water distribution networks under conditions of demand uncertainty. *Environmental Forensics*, 2009, 10(3): 253 – 263.
- [52] YAN X, SUN J, HU C. Research on contaminant sources identification of uncertainty water demand using genetic algorithm. *Cluster Computing*, 2017, 20(2): 1007 – 1016.
- [53] YAN X, ZHU Z, LI T. Pollution source localization in an urban water supply network based on dynamic water demand. *Environmental Science and Pollution Research*, 2019, 26(18): 17901 – 17910.
- [54] YAN X, GONG J, WU Q. Pollution source intelligent location algorithm in water quality sensor networks. *Neural Computing and Applications*, 2021, 33: 209 – 222.
- [55] LAIRD C, BIEGLER L, WAANDERS B. Mixed integer approach for obtaining unique solutions in source inversion of water networks. *Journal of Water Resources Planning and Management*, 2008, 132(173): 242 – 251.
- [56] YAN X, LI T, HU C. Real-time localization of pollution source for urban water supply network in emergencies. *Cluster Computing*, 2019, 22: 5941 – 5954.

作者简介:

龚瑾玉 硕士研究生, 研究方向为智能优化及应用, E-mail: christinegjy@hotmail.com;

颜雪松 博士, 副教授, 博士生导师, 研究方向为智能优化及其应用、人工智能及其应用、大数据及其应用, E-mail: yanxs@cug.edu.cn;

胡成玉 副教授, 硕士生导师, 研究方向为智能优化及其应用、强化学习, E-mail: huchengyu@cug.edu.cn;

龚文引 博士, 教授, 博士生导师, 研究方向为智能优化及其应用, E-mail: wygong@cug.edu.cn.