

基于拥挤距离和时空数据挖掘的农田水资源 自适应配置算法

周爱民[†]

(巨野县农业农村局, 山东 菏泽 274900)

摘要:在农田水资源自适应配置过程中,受到环境和水源复杂不确定性的影响,配置算法的计算误差比较明显,需要进一步加强算法的适用性。为此,提出了基于拥挤距离和时空数据挖掘的农田水资源自适应配置算法。利用时空数据挖掘技术从农田的历史供水数据中筛选出关联数据,经过支持度和关联度的计算,获得特征数据集,通过拥挤距离计算公式计算出目标数据的拥挤度,并按照一定顺序排序,在此基础上,计算农田的需水量,确定需水量的上下限,结合拥挤度等参数,建立水资源自适应配置模型,实现农田水资源的灵活配置。实验结果表明:提出的基于拥挤距离和时空数据挖掘的配置算法计算误差小,在水资源利用上,节水效果好,该算法的适用性得到加强。

关键词:拥挤距离;时空数据挖掘;农田;水资源;自适应配置

中图分类号:TP393

文献标识码:A

An Adaptive Allocation Algorithm of Farmland Water Resources Based on Congestion Distance and Spatiotemporal Data Mining

ZHOU Aimin[†]

(Juye County Agriculture and Rural Bureau, Heze, Shandong 274900, China)

Abstract: In the process of farmland water resources adaptive allocation, due to the complex uncertainty of environment and water source, the calculation error of the allocation algorithm is obvious, and the applicability of the algorithm needs to be further strengthened. To solve this problem, an adaptive allocation algorithm of farmland water resources based on congestion distance and spatiotemporal data mining was proposed. Spatiotemporal data mining technology is used to screen out the associated data from the historical water supply data of farmland, and through the calculation of support degree and correlation degree, the characteristic data set is obtained. The crowding degree of the target data is calculated through the crowding distance calculation formula, and sorted in a certain order. On this basis, the water demand of farmland is calculated, and the upper and lower limits of water demand are determined. The adaptive allocation model of water resources was established to realize the flexible allocation of farmland water resources. The experimental results show that the proposed allocation algorithm based on congestion distance and spatiotemporal data mining has small calculation error and good water-saving effect on water resources utilization, and the applicability of the algorithm has been strengthened.

Key words: crowding distance; spatio-temporal data mining; farmland; water resources; adaptive configuration

农田水资源配置是农业可持续发展的重要保障^[1-3],具有重要的现实意义。根据农田水资源条件和作物生长需求^[4-6],对区域农田进行合理配置,能够充分发挥区域水资源的生产潜力,提高农作物产量和质量^[7-10]。

文献[11]采用基于自适应精确罚函数的分布式资源分配算法,利用距离函数实现局部约束求解,并通过信息交互优化资源分配。使用跟踪技术设计算法,结合凸函数和非光滑分析法进行严谨的收敛性分析,以确保算法在资源分配过程中收敛到最优解。文献[12]采用水循环过程和不确定性条件下的多目标模型,以高效配水方案促进灌区精准灌溉和节水增效。通过分析水量分配、经济效益和产量之间的关系,优化灌溉策略。然而,以上农业水资源配置方法存在较大的局限性。

因此,在拥挤距离和时空数据挖掘技术的支持下,构建了一种基于拥挤距离和时空数据挖掘的农田水资源自适应配置算法。根据农田灌溉需求和水资源条件变化,利用该算法动态调整灌溉用水量,将传统的固定灌溉定额转变为动态调配灌溉定额;在保证作物生长所需用水的基础上,实现农田水资源的高效利用。

1 基于拥挤距离和时空数据挖掘的农田水资源自适应配置算法设计

1.1 获取农田水文地质特征数据

在农田水资源自适应配置中,关键在于获得农田的需求量,这就需要获取农田的水文地质参数,了解农田区域的地表水、地下水等水资源的分布情况,以及需要供水区域的分布特征^[13,14]。

利用时空数据挖掘技术从农田水资源配置历史数据中挖掘各区域的特征数据,利用拥挤距离计算公式计算出各个区域的拥挤度,并按照大小顺序排序,将其作为水资源自适应配置的依据之一^[15]。

农田水资源配置包括农田灌溉用水和林牧渔业用水,从历史配置数据中筛选出相关时空数据,组成数据集 Q ,数据集中每个数据 q 均有 n 个属性,用来描述数据。同时,每个属性也是一个类,每个属性 a_i 都有唯一的取值,一个属性与其对应的取值组成项,在农田水资源配置时空数据挖掘过程中,利用数据的类和项挖掘出需要的特征数据。

上述中的类为农田的位置信息,扫描数据库 Q ,遍历农田所有区域内的数据,以类为中心,计算类与类之间的支持度和置信度。计算公式如下:

$$\sup(a_i \Rightarrow a_j) = P(a_i \cup a_j) \quad (1)$$

$$\text{conf}(a_i \Rightarrow a_j) = \frac{\sup \text{port}(a_i \cup a_j)}{\sup \text{port}(a_i)} \quad (2)$$

公式中, $\sup(a_i \Rightarrow a_j)$ 表示类 a_i 与类 a_j 之间的支持度; $P(a_i \cup a_j)$ 表示类 a_i 与类 a_j 在数据中同时存在的概率; $\text{conf}(a_i \Rightarrow a_j)$ 表示类 a_i 与类 a_j 之间的置信度; i 和 j 均为常数。在获得所有类的置信度和支持度后,选择满足最小支持度和最小置信度的数据,将其加入新的数据集中,作为特征数据集。

利用拥挤距离计算公式对特征数据集中的数据进行计算。公式如下:

$$D_m = \frac{x_m - x_{m-1}}{x_{\max} - x_{\min}} \quad (3)$$

公式中, D_m 表示第 m 个数据的拥挤度; x_m 表示第 m 个数据的属性值; $x_{\max}$ 和 $x_{\min}$ 分别表示数据集中所有数据中属性的最大值和最小值。依据公式(3),计算出所有特征数据的拥挤度,按照从大到小的顺序排列组成新的数据集,拥挤度越大的数据拥有的优先级越高,将这一规则作为水资源自适应配置的条件之一。从特征数据集中各个数据的拥挤度可以分析出农田各个区域的需水量优先级,在此基础上,计算出具体需水量。

1.2 计算农田需水量

在计算农田需水量前,确定农田供水线,由于地理位置不同,农田的实际供水情况也不同。在计算时,以极限供水情况为例,极限供水操作线如图1所示。

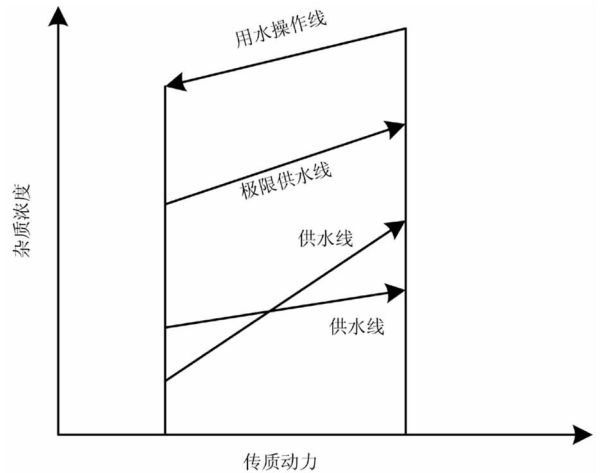


图1 极限供水操作线

依据上述极限情况,采用额定法估算农田需水量,计算公式为:

$$f_c = \sum_k \sum_l S_c^k \cdot u_c^k \quad (4)$$

公式中, f_c 表示第 c 年农业需水估算值; S_c^k 表示区域 k 第 c 年农业产品种植面积; u_c^k 表示区域 k 第 c 年农业产品的用水定额。依据上述计算结果, 确定水资源自适应配置模型参数并建立模型。

1.3 建立水资源自适应配置模型

在建立水资源自适应配置模型前, 要确定模型参数。由于农田水资源配置需要多种水源联合供水, 因此, 需要确定各个水源的优先程度。一般情况下, 供水水源包括地下水、蓄水工程、河道引水和污水回收。设供水水源的优先程度为 r , 其计算公式为:

$$r_i^k = \frac{1 + y_{\max}^k - y_i^k}{\sum_{j=1}^J [1 + y_{\max}^k - y_j^k]} \tag{5}$$

公式中, y_i^k 表示 k 区域 i 水源的供水次序序号; $y_{\max}^k$ 表示供水次序序号的最大值。为保证水资源的充分利用, 还需要设置农田需水量的上限和下限, 约束水资源自适应配置模型。表达式如下:

$$g_i^k = s_a^k \times T^k \tag{6}$$

$$e_i^k = s_s^k \times T^k \tag{7}$$

公式中, g_i^k 表示农田 k 区域的需水量上限; e_i^k 表示农田 k 区域的需水量下限; s_a^k 表示农田 k 区域的有效灌溉面积; s_s^k 表示农田 k 区域的保证灌溉面积; T^k 表示农田 k 区域的综合灌溉定额。除了以上参数外, 模型的建立还需要区域权重系数和区域公平系数, 其中区域权重系数与农田当地的经济情况有关, 根据经济的重要性确定权重; 区域公平系数则是指农田内某个区域相对于其他区域优先得到供水的重要性程度, 由供水的次序确定。

农田水资源自适应配置模型中, 存在多个区域和多个水源, 因此在构建模型时, 以 h_i^k 表示第 i 个水源供水给 k 区域, 在保证区域供水净效益最大、污染物排放最小的情况下, 构建配置模型。表达式为:

$$Z(h) = \text{opt} \{ z_1(h), z_2(h) \} = \begin{cases} \max \sum_{k=1}^K \left[\sum_i^M (g_i^k - e_i^k) h_i^k r_i^k \beta_s^k \right] \omega_k \\ \min \sum_{k=1}^K 0.01 D^k p^k \left(\sum_i^M h_i^k \right) \end{cases} \tag{8}$$

公式中, r_i^k 表示区域水资源利用系数; β_s^k 表示区域公平系数; ω_k 表示区域权重系数; p^k 表示污染物排放系数。污染物排放系数根据当地具体情况确定, 通过建立的农田水资源自适应模型即可实现对农田的合理供水。至此, 基于拥挤距离和时空

数据挖掘的农田水资源自适应配置算法设计完成。

2 基于拥挤距离和时空数据挖掘的农田水资源自适应配置算法实验研究

2.1 实验准备

在上述农田水资源自适应配置算法设计完成后, 针对以往水资源配置中存在的适应性差的问题, 设计实验方案, 实验中将文献[11]和文献[12]算法作为对照, 设计两组对比研究方案。为保证实验的公平公正, 在实验前期准备中, 采用相同的参数的实验样本作为实验对象。具体的参数设置如表 1 所示。

表 1 实验参数设置

项目	样本 1	样本 2	样本 3	样本 4	样本 5
大气温度/℃	-6.184	-9.435	-13.462	-15.542	-12.093
表层土壤温度/℃	-1.205	-4.873	-9.712	-11.439	-14.305
太阳辐射/(W/m ²)	-65.421	-83.046	-93.291	-86.725	-54.307
光合有效辐射/(mmol/s)	-0.117	-0.153	-0.174	-0.168	-0.099
大气相对湿度/%	1.278	9.563	19.351	8.405	10.452
风速/(km/h)	-1.373	3.512	2.855	3.792	0.931
气压/hPa	6.553	4.684	6.201	3.984	7.823

以上数据为人工设置的固定数据, 在后续实验中, 利用计算机软件生成大量随机样本数据, 增加实验的多样性。依据设置的实验参数, 开展地下水位计算误差实验和节水率实验, 在实验结束后, 通过这两组实验的实验结果分析各个水资源配置算法的适用性。

2.2 地下水位计算误差实验结果及分析

在农田水资源自适应配置中, 涉及的水资源包括地表水、地下水、雨水等多种资源, 其中地下水的利用往往更加复杂, 需要准确掌握地下水位, 才能保证水资源配置的灵活性。因此, 涉及地下水位计算误差实验方案, 通过对各个算法地下水位计算误差的分析, 判断算法的实际适用能力。实验中, 以混合了固定样本和随机样本的样本数据作为输入, 利用计算机软件执行各个配置算法, 在算法执行完毕后, 统计各个算法的执行结果, 将各个算法计算的地下水位数据与实测数据对比, 分析各个算法的

性能。实验结果如图 2 所示。

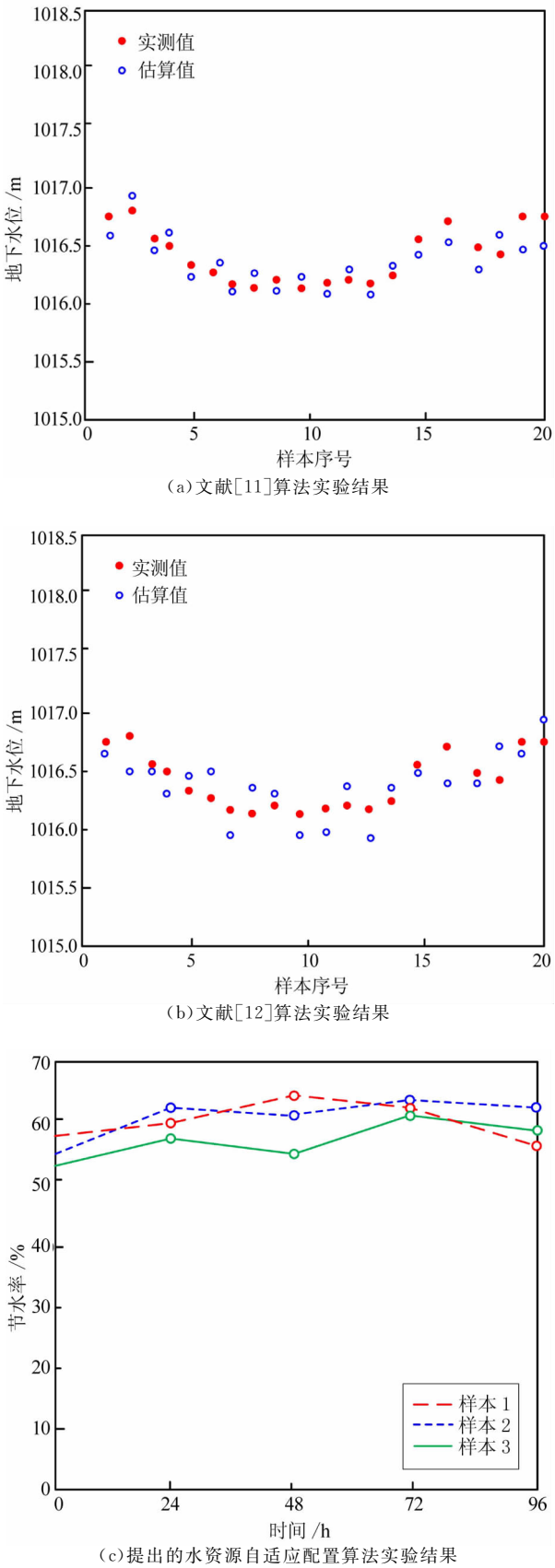


图 2 不同算法的地下水位计算误差实验结果

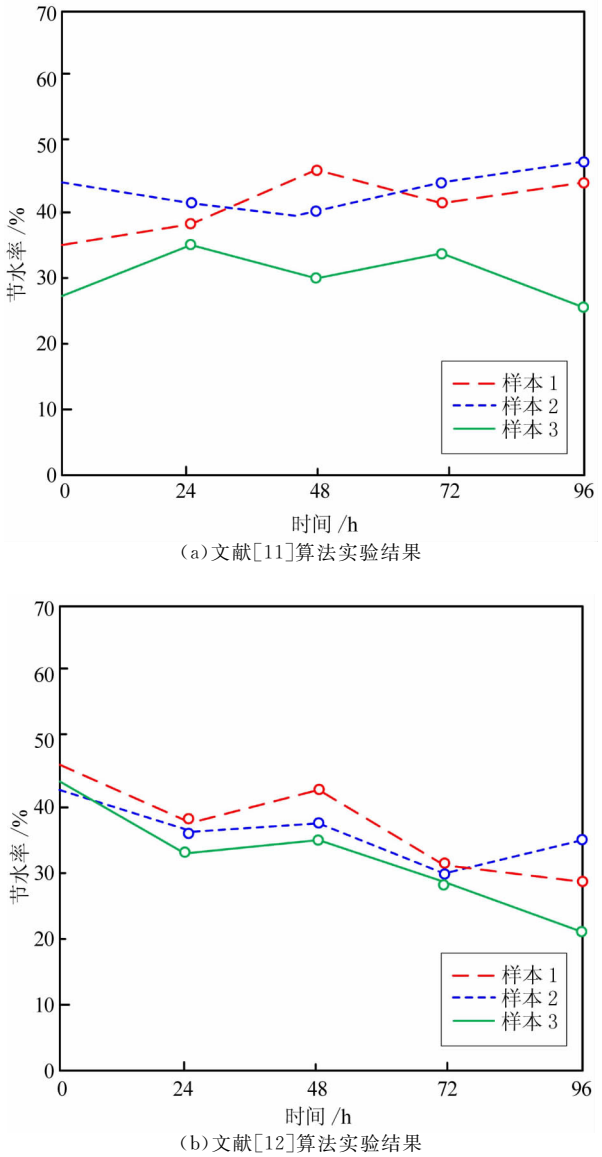
通过图 2 的实验结果可以看出,文献[11]算法

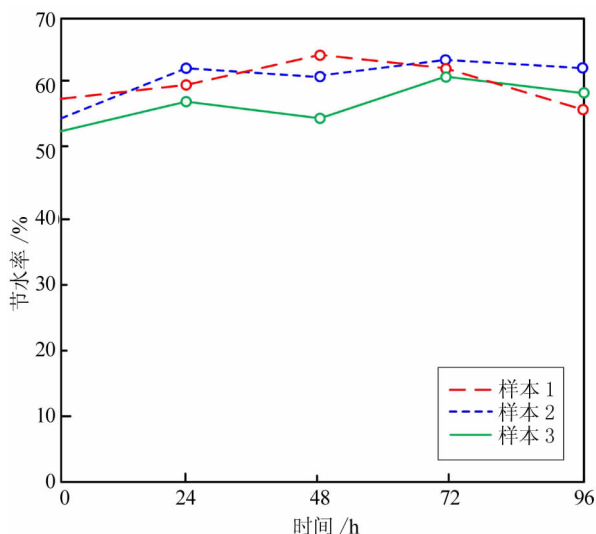
和文献[12]算法的地下水位估算值与实测值存在一定差距,计算误差较明显。相比之下,提出的水资源自适应配置算法的地下水位计算值与实测值基本一致,计算误差小,在标准要求范围内,说明提出的基于拥挤距离和时空数据挖掘的农田水资源自适应配置算法计算能力更好。

2.3 节水率实验结果及分析

在节水率实验中,主要验证水资源配置算法在执行过程中是否能够高效利用水资源以及水量调配是否合理。使用计算机软件执行各个水资源配置算法后,利用统计软件统计出各个算法在配置前后的用水效率,获得相应的节水率。实验结果如图 3 所示。

从图 3 的实验结果可以看出,在三组实验结果中,提出的水资源自适应配置算法节水率最高,各个样本的节水率均在 50% 以上,节水效果好,水资源配置合理。而其他两组实验的节水率均在 50% 以下,并且在有效的实验时间内,有





(c)提出的水资源自适应配置算法实验结果

图 3 不同算法的节水率实验结果

明显的下降趋势,节水效果不佳,水资源配置并不理想。结合地下水位计算误差实验结果综合分析可知,提出的基于拥挤距离和时空数据挖掘的农田水资源自适应配置算法计算误差小、节水效果好,整体适用性优于其他配置算法。

3 结 论

以农田水资源自适应配置作为研究重点,根据区域水资源条件和作物生产需求,建立了农田水资源配置模型,采用基于拥挤距离和时空数据挖掘的农田水资源自适应配置算法对该模型进行求解,并将结果与常见的水资源配置算法进行对比,证明了提出的水资源配置算法的适用性。由于农田水资源配置模型具有时空异质性和不确定性等特点,本文提出的基于拥挤距离和时空数据挖掘的农田水资源自适应配置算法虽然取得了不错成果,但是,考虑到时间和空间数据之间存在着较强的非线性关系,在数据处理方面还需要进一步研究和完善。

参考文献

- [1] 齐学斌.加强农业水资源配置与利用研究,为粮食安全及水安全提供坚强支撑[J].灌溉排水学报,2022,41(8):2.
- [2] 王平波,刘杨.基于改进自适应 IMM-UKF 算法的水下目标跟踪[J].电子与信息学报,2022,44(6):1999—2005.
- [3] 赵鹏程,宋保维,毛昭勇,等.基于改进的复合自适应遗传算法的 UUV 水下回收路径规划[J].兵工学报,2022,43(10):2598—2608.
- [4] 党兴东,程立平,刘文兆.黄土高原沟壑区小流域雨季末期土壤水资源状况及其空间分异[J].水土保持研究,2021,28(6):52—57.
- [5] 梅鹏,孙振新,马沙沙.基于机动检测的水下目标自适应跟踪算法[J].舰船科学技术,2022,44(6):114—120.
- [6] 王玉莲,黄林显,桑国庆,等.基于自适应罚函数遗传算法求解地下水管理模型[J].人民黄河,2022,44(3):69—73.
- [7] 田雪莹,王高旭,吴永祥,等.云水资源利用对北三河流域农田灌溉需水影响分析[J].水利水运工程学报,2022(3):23—33.
- [8] 王斌,李昕.融合动态残差的多源域自适应算法研究[J].计算机工程与应用,2022,58(7):162—166.
- [9] 吴佳楠,贺曼利,邸焕双,等.考虑电导率和能量的水下电场自适应调整多权值成簇算法[J].中南大学学报(自然科学版),2021,52(9):3223—3233.
- [10] 孙波.基于 GNSS 单天线技术的农田土壤湿度反演方法研究[J].测绘学报,2021,50(6):848.
- [11] 时侯圣,徐磊,杨涛.基于自适应精确罚函数的分布式资源分配算法[J].控制理论与应用,2022,39(10):1937—1945.
- [12] 李莱,曹凯华,付强,等.不确定条件下考虑水循环过程的灌区多水源高效配置[J].农业工程学报,2021,37(18):62—73.
- [13] 吴建华,戴鹏,胡烈云.一种面向多尺度面状居民地匹配的 Voronoi 图自适应构建算法[J].武汉大学学报(信息科学版),2022,47(2):304—312.
- [14] 屈省源.基于自适应调控技术的农田监测无人机设计[J].农机化研究,2021,43(9):210—214.
- [15] 赵杨,熊伟丽.基于多策略自适应差分进化算法的污水处理过程多目标优化控制[J].化工学报,2021,72(4):2167—2177.