

基于模糊数据挖掘的灌区水资源配置研究

包志炎^{1,2} 郑高安¹ 王萱³ 邱雁² 魏杰²

(1. 浙江水利水电学院计算机应用技术研究所, 杭州 310018; 2. 浙江省水利信息中心, 杭州 310009;
3. 杭州钱塘新区建设投资集团有限公司, 杭州 311228)

摘要: 在分析灌区大数据资源构成的基础上,构建了面向水资源配置的数据挖掘算法;汇聚某行政区域水资源、经济、人口、行业用水等数据,运用模糊层次聚类分析方法,对行政区域内行业间的配水案例进行分类及特征分析;采用固定步长穷举法对灌区配水相关参数进行动态赋权;运用模糊距离匹配相似灌区,通过构造加权影响因子和采用指数平滑法,对基于实例推理的灌区进行需水量预测。将建立的方法应用于2018年浙江省11个市(区、县)行业间配水特征分析和中型灌区需水量预测,结果表明:浙江省不同区域行业间配水被划分为4类,呈现出不同行业间的配水特征,灌区需水量预测相对误差均不大于9.39%,说明该方法有效,可为制定合理的区域行业间配水方案、估算灌区需水量提供决策支持。

关键词: 水资源配置; 灌区; 模糊数据挖掘; 相关性分析; 实例推理

中图分类号: TV213.9; TP391.1 **文献标识码:** A **文章编号:** 1000-1298(2020)08-0270-08

OSID:



Water Resources Allocation of Irrigation District Based on Fuzzy Data Mining

BAO Zhiyan^{1,2} ZHENG Gaoan¹ WANG Xuan³ QIU Yan² WEI Jie²

(1. Institute of Computer Application Technology, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, China
2. Water Resources Information Management Center of Zhejiang Province, Hangzhou 310009, China
3. Hangzhou Qiantang New Area Construction Investment Group Co., Ltd., Hangzhou 311228, China)

Abstract: Aiming to realize the reasonable allocation of water resources in irrigation district, the composition of big data resources of irrigation district was analyzed, and the corresponding data mining algorithms were proposed. The cosine correlation coefficient was introduced, and the fuzzy hierarchical cluster analysis was used to realize the case classification and characteristic analysis of department water distribution based on the data of regional water resources, economy, population and industrial water consumption. A variable step length exhaustive method was used to calculate the dynamic weight of relevant parameters of water distribution in irrigation district. Fuzzy distance was used to match the most similar irrigation district. Weighted influence factors and exponential smoothing method were used to estimate the water demand of irrigation area based on case-based reasoning. The proposed methodology demonstrated the effectiveness in the analysis of department water distribution characteristics and water demand prediction of medium-sized irrigation districts in 11 cities and administrative regions of Zhejiang Province in 2018. The results showed that water distribution among different regions in Zhejiang Province was divided into four categories, showing different characteristics of water distribution among different departments; the relative error of water demand prediction in irrigation area was not more than 9.39%. The established theoretical method could provide decision support for making reasonable regional inter industry water distribution scheme and estimating irrigation water requirement.

Key words: water resources allocation; irrigation district; fuzzy data mining; correlation analysis; case-based reasoning

收稿日期: 2020-04-12 修回日期: 2020-05-23

基金项目: 国家自然科学基金青年基金项目(41601461)、浙江省公益技术研究计划项目(LGF19E090011)和教育部人文社会科学研究青年基金项目(18YJCZH002)

作者简介: 包志炎(1982—),男,博士后,浙江省水利信息中心高级工程师,主要从事智慧水利研究,E-mail: bzy@zjwater.gov.cn

通信作者: 郑高安(1983—),男,讲师,主要从事数据挖掘和机电一体化研究,E-mail: 593300789@qq.com

0 引言

随着社会经济的发展、人口的增加以及人类对水资源需求的提高,水资源日益短缺。预计到 2050 年,全球人口将增长到 95 亿,届时粮食需增产 60%、农业用水量需增加 19%,才能满足基本需求^[1]。中国是农业大国,农业是最主要的用水部门。在区域水资源配置时,农业与工业、生活、服务业等其他行业一起参与用水分配。保障国家粮食安全,核心在灌区。灌区以占全国耕地 49% 的面积,生产了约占全国总量 75% 的粮食和 90% 以上的经济作物^[2]。因此,加强灌区水资源配置研究,科学、合理、有计划、有重点地用水显得尤为重要和迫切。

目前,国内外灌区水资源配置方法主要包括规划优化、系统模拟和智能计算。规划优化方法基于数学规划理论,在既定约束条件下,面向单目标^[3-5]或多目标^[6-9]实现,在多种方案中寻求最优方案,较为常用的有线性规划^[10]、动态规划^[7-8,11]和随机规划^[12-13]等,该方法需要构造较为复杂的目标函数和严格的约束条件。系统模拟方法既有面向农田的作物全生育期需水量模型^[14-15],也有面向区域、灌区的分布式水循环模型^[16-17]和水文模型^[18-19],需要模拟灌区水资源系统内部的复杂关系。智能计算方法将粗糙集、神经网络^[20]、粒子群算法^[21]、果蝇算法^[22]、人工鱼群^[23]等算法应用于灌区水资源配置中,需要进行收敛性分析,有时会出现局部最优解。数据挖掘侧重关联分析,不必进行严格的逻辑推理和收敛分析,从而能避开复杂的目标函数建模和边界条件分析过程,实现从海量数据中抽取感兴趣的、隐含的、有价值的知识,是当前水资源预测理论发展的重要趋势之一。EKASINGH 等^[24]基于数据挖掘建立了灌区农作物选择规则;THOMPSON 等^[25]和 MOHAN 等^[26]利用数据挖掘改进水资源日常管理,将决策过程简单化;SHAHANAS 等^[27]提出了智能水管理的技术架构。随着我国水资源监控能力建设项目实施^[28],全国各地水资源感知监测体系不断完善,灌区数据资源日益丰富,为数据挖掘在水资源配置中的应用创造了条件。蒋云钟等^[29]提出了水资源数据挖掘的体系框架,刘予伟等^[30]在分析水资源大数据采集的基础上提出构建水资源大数据云平台的思想,杨小柳等^[31]基于数据挖掘提出节水的精准化、差异化管理。目前的相关研究更多侧重于理论框架和设想建议。本文基于数据挖掘,融合相关性分析、聚类分析、相似性度量等算法,对浙江省行政区域行业间配水进行特征分析,

对中型灌区需水量进行预测,为科学合理配水提供决策支持。

1 灌区数据资源

构建灌区大数据,汇聚足够丰富的高质量灌区水资源数据资源是开展数据挖掘的前提。灌区数据资源既包括灌区占地、工程设施、作物种植结构等基本数据,也包括所在行政区域水资源、经济、人口、水文、气象等相关数据;既包括实时监测数据,也包括历年用水数据,如表 1 所示。表中数据均可以从区域水文、水资源、经济发展的统计年报或灌区感知监测数据库中获取。

表 1 灌区数据资源构成
Tab.1 Data resources of irrigation area

数据类别	数据项
灌区基础数据	设计灌溉面积、有效灌溉面积、实际灌溉面积、灌区类型、灌区规模、地形地貌、取水户、取水许可证、作物种植结构等
灌区工程数据	水源、渠系、灌溉设施、取水监测点、雨量站、水质监测站等
灌区监测数据	取水点水位、水量、流量、工情、墒情、水库蓄水量、作物遥感影像等
灌区用水数据	用水定额、作物用水量、毛用水量、净用水量、农田灌溉水有效利用系数等
行政区域水资源数据	水资源总量、降水量、地表水量、地下水量、耗水量等
行政区域行业用水数据	农田灌溉用水、林牧渔畜用水、工业用水、城镇公共用水、居民生活用水、生态环境用水、万元 GDP 用水量、单位面积用水量、人均综合用水量等
行政区域社会经济数据	区域面积、人口、产业结构、GDP、人均 GDP、水价等

2 面向灌区水资源配置的数据挖掘方法

2.1 区域行业间配水聚类分析

以行政区域为研究单元,当区域可供水量不能满足所需水量时,既要强调农业用水的基础性,又要保障区域社会经济目标实现。因此,在不同行业、不同部门之间就会产生竞争用水问题。由于区域间经济发展水平、产业结构和资源禀赋不一,各行业用水需求差异较大。在有限的区域水资源条件下,需要解决行业间如何合理配水问题。

针对某省所有行政区域,汇聚工业、生活以及生态环境等领域的历年用水数据,聚焦农业灌溉用水,引入模糊聚类算法,对区域行业配水案例进行聚类分析,从中发现同类区域行业配水的显著特征,为管理者开展区域行业间配水提供决策支持。区域行业配水聚类分析流程如下:

(1)建立行业配水案例集。假设某省共有 n 个

市级行政区,选取 m 个能反映某一时期在农业、工业、生活和生态等领域用水的相关指标,建立由 n 个区域行业配水案例组成的案例集,以各案例的指标值为内容构造数据矩阵

$$\mathbf{A} = [a_{ij}]_{n \times m} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \quad (1)$$

(2) 标准化数据矩阵。为消除各参数间不同量纲的偏差,通过“平移-标准差-极差”变换^[32]将数据转换到 $[0,1]$ 区间,进行数据标准化,计算得到矩阵

$$\mathbf{A}' = [a'_{ij}]_{n \times m} \quad (2)$$

其中

$$a'_{ij} = \frac{r_{ik} - \min_{1 \leq i \leq n} r_{ik}}{\max_{1 \leq i \leq n} r_{ik} - \min_{1 \leq i \leq n} r_{ik}}$$
$$r_{ik} = \frac{a_{ik} - \frac{1}{n} \sum_{i=1}^n a_{ik}}{\sqrt{\frac{1}{n} \sum_{i=1}^n \left(a_{ik} - \frac{1}{n} \sum_{i=1}^n a_{ik} \right)^2}} \quad (k = 1, 2, \cdots, m)$$

(3) 计算模糊相似矩阵。为了定量地进行分类,引入夹角余弦相关系数作为聚类统计量,将矩阵 $\mathbf{A}'$ 转换为一个模糊相似矩阵

$$\mathbf{A}'' = \left[\frac{\sum_{k=1}^m a'_{ik} a'_{jk}}{\sqrt{\sum_{k=1}^m a'^2_{ik}} \sqrt{\sum_{k=1}^m a'^2_{jk}}} \right]_{n \times n} \quad (3)$$

(4) 配水案例聚类分析。用平方法求 $\mathbf{A}''$ 的传递闭包 $t(\mathbf{A}'')$ 。使用 λ -截矩阵,使阈值 λ 由 1 减小到 0,则 $t(\mathbf{A}'')$ 的分类由细变粗,形成一个动态的聚类图。根据 λ 的不同取值,可以得到不同分类。 λ 取值应遵循“类间案例相似性最小化,类内案例相似性最大化”原则,分类过细或过粗均不利于共性特征提取,一般可分为 2~4 类,且应避免出现类间距的突变^[32]。在得到区域行业间配水的类别分组后,即可分析归入同一类的显著特征。

2.2 灌区配水相似实例匹配算法

灌区可供水量与该灌区所在区域降水、地表水、地下水、耗水以及其他行业竞争用水相关。灌区需水量与灌区类型、面积、种植结构等相关。在有限的区域水资源条件下,面向不同特征的灌区,需要解决灌区间如何合理配水问题。

以灌区为研究对象,建立一批典型灌区,汇聚其历年可供水量、需水量的相关数据,形成灌区配水案例集。对于预先给定的一个新灌区(或改造灌区),在灌区配水案例集中匹配出最相似灌区,为该灌区配水提供决策支持。

(1) 建立灌区配水案例集。假设共有 n 个典型灌区,选取 m 个与可供水量、需水量相关的指标,建立由 n 个灌区配水案例组成的案例集,以各案例的指标值为内容构造数据矩阵 $\mathbf{B}$,待估灌区记为 $\mathbf{b}_x$,即

$$\mathbf{B} = [b_{ij}]_{n \times m} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{bmatrix} \quad (4)$$

$$\mathbf{b}_x = (b_{x1}, b_{x2}, \cdots, b_{xm}) \quad (5)$$

(2) 计算差异矩阵。对 $\mathbf{B}$ 的行(每一案例)与 $\mathbf{b}_x$ 进行差运算,并进行归一化处理,计算得到其差异矩阵

$$\mathbf{B}' = \left[\frac{|b_{xj} - b_{ij}|}{\sum_{j=1}^m (b_{xj} - b_{ij})} \right]_{n \times m} \quad (6)$$

(3) 确定参数权重。权重反映某参数对于目标指标的重要程度。以灌区需水量为预测指标,确定其他各参数的权重。通过对各参数权重在 $[0,1]$ 区间内按一定的步长穷举,使用穷举权重,把每一个灌区视作新灌区,用剩余的 $n-1$ 个灌区预测需水量,并与实际值对比。当相对误差均在预先设定的范围内时,此时权重即为所求的解。若有两个及以上的解,相对误差总和最小的为最优解^[33]。当灌区配水案例集发生变化,权重会随着更新,这样的赋权方法实现了动态赋权,且具有客观性。权重向量记为 $\mathbf{w} = (w_1, w_2, \cdots, w_m)$ 。

(4) 获取相似实例。考虑到各参数之间的权重差异,引入加权的欧氏距离,计算新灌区与其他灌区间的相似度 $S(\mathbf{b}_x, \mathbf{b}_i)$,最大相似度所对应的灌区即为所求相似灌区。如果将相似度进行降序排序,可以获得排名靠前的若干灌区组成相似实例组。相似度计算式为

$$S(\mathbf{b}_x, \mathbf{b}_i) = 1 - \sqrt{\sum_{j=1}^m w_j^2 \left(\frac{|b_{xj} - b_{ij}|}{\sum_{j=1}^m (b_{xj} - b_{ij})} \right)^2} \quad (7)$$

式中 w_j ——第 j 个参数权重

2.3 基于实例推理的灌区需水量预测

修正 2.2 节获取的相似实例(或相似实例组)的需水量,可预测新灌区需水量。两种修正算法如下:

构造加权影响因子法,计算式为

$$R_x = R_i \left(1 + \sum_{j=1}^m w_j \frac{b_{xj} - b_{ij}}{b_{ij}} \right) \quad (8)$$

式中 R_x ——新灌区需水量

R_i ——相似灌区需水量

其中 $\sum_{j=1}^m w_j \frac{b_{x_j} - b_{ij}}{b_{ij}}$ 为影响因子。

指数平滑法, 计算式为

$$R_x = \alpha R_i + \alpha(1 - \alpha) R_j + \alpha(1 - \alpha)^2 R_k + (1 - \alpha)^3 \frac{R_i + R_j + R_k}{3} \quad (9)$$

式中 R_j, R_k ——相似实例组成员的需水量
 α ——平滑系数, 其取值方法详见文献[34]

3 应用案例

浙江省目前拥有 150 个大中型灌区, 分布在 11 个市行政区域, 其中中 I 型灌区 91 个, 中 II 型灌区 48 个, 大型灌区 11 个。表 2 为 2018 年浙江省各市工业、农业、城镇公共、居民生活以及生态环境等领域配水数据; 表 3 为 2018 年金华、丽水、衢州、台州、温州等地区部分中型灌区可供水量、毛需水量的相关数据。按灌区规模, 分别用数字 1~4 表示小型、中 I 型、中 II 型和大型灌区。为方便说明问题, 以乌溪江引水灌区、横锦水库灌区、罗岙水库灌区作为待估灌区, 如表 4 所示, 运用本文建立的方法, 分析浙江区域行业配水特征, 匹配每一个待求灌区的相似灌区, 并在此基础上预测其毛需水量。

基于表 2, 建立 2018 年浙江省各市行政区域行业配水案例集, 构造相应的数据矩阵

表 2 2018 年浙江省各市行政区域行业配水量

Tab. 2 Department water distribution cases in administrative regions of Zhejiang in 2018										
区域	单位面积 水资源量/ 万 m ³	单位面积 用水量/ 万 m ³	万元 GDP 用水量/ m ³	人均生活 用水量/ m ³	农田灌溉 单位面积 用水量/m ³	农田灌溉 水有效利 用系数	行业间用水占比/%			
							农田灌溉	工业	城镇公共 及居民生活	生态及 其他
杭州	87.00	19.60	24.00	116.90	405.00	0.600	27.72	28.18	35.29	8.81
宁波	88.90	22.20	19.30	88.30	258.00	0.604	31.55	28.56	34.92	4.97
温州	116.90	15.10	29.70	71.40	349.00	0.590	31.39	24.42	37.11	7.08
嘉兴	61.00	47.60	28.50	75.00	387.00	0.660	50.74	24.13	18.98	6.15
湖州	69.40	24.10	51.50	85.40	354.00	0.628	47.97	16.56	18.49	16.98
绍兴	76.30	21.90	33.40	90.60	300.00	0.590	39.86	25.15	25.26	9.73
金华	84.00	14.90	39.70	59.70	267.00	0.577	39.07	28.69	20.58	11.66
衢州	115.20	13.20	79.20	69.30	421.00	0.529	53.13	28.84	13.13	4.90
舟山	54.00	11.30	12.30	67.70	126.00	0.693	11.11	35.19	48.77	4.93
台州	95.70	16.30	31.70	67.00	358.00	0.579	40.41	23.32	26.62	9.65
丽水	110.30	4.00	49.50	62.50	335.00	0.578	62.61	13.48	20.00	3.91

表 3 2018 年浙江典型中型灌区配水量

Tab. 3 Water distribution cases of typical medium-sized irrigation areas of Zhejiang in 2018							
灌区名称	区域单位面 积水资源量/ 万 m ³	区域农田灌溉 单位面积用水 量/万 m ³	有效灌溉面积/ m ²	实际灌溉面积/ m ²	灌区规模	多年平均农田 灌溉水有效 利用系数	毛需水量/ 万 m ³
南江水库灌区 b_1	84.00	267.00	6 734	6 734	3	0.548	3 521.45
安地灌区 b_2	84.00	267.00	4 297	4 297	3	0.556	1 913.42
清溪口水库灌区 b_3	84.00	267.00	1 760	1 333	2	0.557	1 195.84
杨卜山中型灌区 b_4	84.00	267.00	2 933	2 852	2	0.586	884.01
碧湖镇碧湖灌区 b_5	110.30	335.00	3 534	3 534	3	0.512	1 362.05
松阳县江北灌区 b_6	110.30	335.00	2 616	2 616	2	0.518	1 033.63
碗窑灌区 b_7	115.20	421.00	22 598	17 668	3	0.514	11 585.19
铜山源水库灌区 b_8	115.20	421.00	16 341	12 667	3	0.520	8 546.85
牛头山水库灌区 b_9	95.70	358.00	16 808	16 117	3	0.570	9 509.45
里石门水库灌区 b_{10}	95.70	358.00	8 534	7 730	3	0.564	4 407.36
里林灌区 b_{11}	95.70	358.00	1 133	1 119	2	0.580	571.86
桥墩灌区 b_{12}	116.90	349.00	16 375	15 714	3	0.546	6 349.06
温瑞灌区 b_{13}	116.90	349.00	10 048	8 894	3	0.539	3 229.49
江北灌区 b_{14}	116.90	349.00	3 305	2 932	2	0.617	1 153.62
北引灌区 b_{15}	116.90	349.00	1 343	1 343	2	0.569	739.22

表 4 待估灌区参数

Tab. 4 Parameters of irrigated area to be estimated

灌区名称	区域单位面积 水资源量/ 万 m ³	区域农田灌溉 单位面积用 水量/万 m ³	2018 年有效 灌溉面积/m ²	2018 年实际灌溉 面积/m ²	灌区规模	多年平均农田 灌溉水有效 利用系数
乌溪江引水灌区 b_{x1}	115.20	421.00	17 048	14 061	3	0.494
横锦水库灌区 b_{x2}	84.00	267.00	7 160	6 374	3	0.547
罗岙水库灌区 b_{x3}	95.70	358.00	947	845	2	0.558

$$A = \begin{bmatrix} 87.00 & 19.60 & 24.00 & 116.90 & 405.00 & 0.600 & 27.72 & 28.18 & 35.29 & 8.81 \\ 88.90 & 22.20 & 19.30 & 88.30 & 258.00 & 0.604 & 31.55 & 28.56 & 34.92 & 4.97 \\ 116.90 & 15.10 & 29.70 & 71.40 & 349.00 & 0.590 & 31.39 & 24.42 & 37.11 & 7.08 \\ 61.00 & 47.60 & 28.50 & 75.00 & 387.00 & 0.660 & 50.74 & 24.13 & 18.98 & 6.15 \\ 69.40 & 24.10 & 51.50 & 85.40 & 354.00 & 0.628 & 47.97 & 16.56 & 18.49 & 16.98 \\ 76.30 & 21.90 & 33.40 & 90.60 & 300.00 & 0.590 & 39.86 & 25.15 & 25.26 & 9.73 \\ 84.00 & 14.90 & 39.70 & 59.70 & 267.00 & 0.577 & 39.07 & 28.69 & 20.58 & 11.66 \\ 115.20 & 13.20 & 79.20 & 69.30 & 421.00 & 0.529 & 53.13 & 28.84 & 13.13 & 4.90 \\ 54.00 & 11.30 & 12.30 & 67.70 & 126.00 & 0.693 & 11.11 & 35.19 & 48.77 & 4.93 \\ 95.70 & 16.30 & 31.70 & 67.00 & 358.00 & 0.579 & 40.41 & 23.32 & 26.62 & 9.65 \\ 110.30 & 4.00 & 49.50 & 62.50 & 335.00 & 0.578 & 62.61 & 13.48 & 20.00 & 3.91 \end{bmatrix}$$

归一化后,计算得到模糊相似矩阵

$$A'' = \begin{bmatrix} 1 & 0.990\,1 & 0.990\,3 & 0.989\,8 & 0.992\,5 & 0.996\,2 & 0.987\,5 & 0.981\,7 & 0.933\,6 & 0.993\,0 & 0.979\,7 \\ 0.990\,1 & 1 & 0.991\,0 & 0.974\,6 & 0.981\,5 & 0.994\,7 & 0.990\,2 & 0.974\,9 & 0.963\,9 & 0.986\,5 & 0.980\,5 \\ 0.990\,3 & 0.991\,0 & 1 & 0.981\,8 & 0.986\,8 & 0.991\,6 & 0.995\,5 & 0.990\,2 & 0.932\,1 & 0.997\,4 & 0.992\,6 \\ 0.989\,8 & 0.974\,6 & 0.981\,8 & 1 & 0.994\,3 & 0.989\,2 & 0.984\,0 & 0.984\,3 & 0.895\,8 & 0.991\,5 & 0.978\,8 \\ 0.992\,5 & 0.981\,5 & 0.986\,8 & 0.994\,3 & 1 & 0.995\,7 & 0.992\,4 & 0.992\,5 & 0.909\,8 & 0.994\,1 & 0.988\,4 \\ 0.996\,2 & 0.994\,7 & 0.991\,6 & 0.989\,2 & 0.995\,7 & 1 & 0.994\,9 & 0.987\,3 & 0.940\,6 & 0.993\,8 & 0.987\,6 \\ 0.987\,5 & 0.990\,2 & 0.995\,5 & 0.984\,0 & 0.992\,4 & 0.994\,9 & 1 & 0.995\,1 & 0.931\,2 & 0.996\,1 & 0.995\,5 \\ 0.981\,7 & 0.974\,9 & 0.990\,2 & 0.984\,3 & 0.992\,5 & 0.987\,3 & 0.995\,1 & 1 & 0.897\,0 & 0.994\,4 & 0.995\,6 \\ 0.933\,6 & 0.963\,9 & 0.932\,1 & 0.895\,8 & 0.909\,8 & 0.940\,6 & 0.931\,2 & 0.897\,0 & 1 & 0.916\,5 & 0.906\,6 \\ 0.993\,0 & 0.986\,5 & 0.997\,4 & 0.991\,5 & 0.994\,1 & 0.993\,8 & 0.996\,1 & 0.994\,4 & 0.916\,5 & 1 & 0.994\,0 \\ 0.979\,7 & 0.980\,5 & 0.992\,6 & 0.978\,8 & 0.988\,4 & 0.987\,6 & 0.995\,5 & 0.995\,6 & 0.906\,6 & 0.994\,0 & 1 \end{bmatrix}$$

按 2.1 节的聚类算法,生成动态聚类图,如图 1 所示。由图可知,当阈值 $\lambda = 0.994$ 、 0.992 、 0.987 时,案例分别被划分为 5、4、2 类,考虑到在 0.992 之前类间距约为 0.002 ,下一个值将出现 0.005 的较大间距,因此 λ 取 0.992 。

当 $\lambda = 0.992$ 时,各市行政区配水被分为 4 个组别:Ⅰ类{丽水,衢州,金华,台州,温州},Ⅱ类{湖州,嘉兴},Ⅲ类{绍兴,宁波,杭州},Ⅳ类{舟山}。可见,Ⅰ类地区单位面积水资源量高,农田灌溉水有效利用系数较低,万元 GDP 用水量偏高;Ⅱ类地区单位面积水资源量较少,农业用水占比最高,单位面积用水量偏高;Ⅲ类地区工农业用水基本相当,万元 GDP 用水量处于较低水平,人均生活用水、城镇公共及居民生活用水偏高;Ⅳ类地区单位面积水资源量和用水量均处于最低水平,农业用水占比也最低。

基于表 3,建立 2018 年浙江部分灌区配水案例

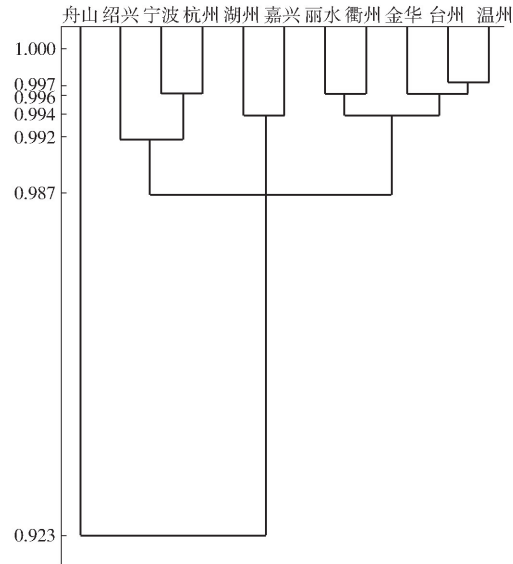


图 1 区域配水动态聚类图

Fig. 1 Dynamic cluster view of regional water allocation cases

集,构造相应的数据矩阵

$B =$

84.00	267.00	6 734	6 734	3	0.548
84.00	267.00	4 297	4 297	3	0.556
84.00	267.00	1 760	1 333	2	0.557
84.00	267.00	2 933	2 852	2	0.586
110.30	335.00	3 534	3 534	3	0.512
110.30	335.00	2 616	2 616	2	0.518
115.20	421.00	22 598	17 668	3	0.514
115.20	421.00	16 341	12 667	3	0.520
95.70	358.00	16 808	16 117	3	0.570
95.70	358.00	8 534	7 730	3	0.564
95.70	358.00	1 133	1 119	2	0.580
116.90	349.00	16 375	15 714	3	0.546
116.90	349.00	10 048	8 894	3	0.539
116.90	349.00	3 305	2 932	2	0.617
116.90	349.00	1 343	1 343	2	0.569

待估的 3 个灌区分别记为

$b_{x1} = (115.20, 421.00, 17\,048, 14\,061, 3, 0.494)$

$b_{x2} = (84.00, 267.00, 7\,160, 6\,374, 3, 0.547)$

$b_{x3} = (95.70, 358.00, 947, 845, 2, 0.558)$

以罗岙水库灌区为例,计算 B 的差异矩阵

$B' =$

0.058 6	0.159 1	0.055 6	0.063 4	0.083 3	0.028 6
0.058 6	0.159 1	0.032 2	0.037 2	0.083 3	0.006 0
0.058 6	0.159 1	0.007 8	0.005 3	0.041 7	0.002 5
0.058 6	0.159 1	0.019 1	0.021 6	0.041 7	0.081 1
0.073 1	0.040 2	0.024 8	0.028 9	0.083 3	0.130 3
0.073 1	0.040 2	0.016 0	0.019 1	0.041 7	0.113 2
0.097 6	0.110 1	0.207 9	0.181 1	0.083 3	0.125 3
0.097 6	0.110 1	0.147 8	0.127 3	0.083 3	0.108 6
0	0	0.152 3	0.164 4	0.083 3	0.035 8
0	0	0.072 8	0.074 1	0.083 3	0.017 7
0	0	0.001 8	0.003 0	0.041 7	0.064 0
0.106 1	0.015 7	0.148 1	0.160 1	0.083 3	0.032 6
0.106 1	0.015 7	0.087 4	0.086 7	0.083 3	0.052 1
0.106 1	0.015 7	0.022 6	0.022 5	0.041 7	0.170 0
0.106 1	0.015 7	0.003 8	0.005 4	0.041 7	0.032 2

按 2.2 节权重动态赋权法,以灌区需水量为预测指标,取固定步长为 0.01 穷举权重组合,设定相对误差阈值为 15%,可计算得到权重向量 $w = (0.05, 0.20, 0.25, 0.35, 0.10, 0.05)$ 。

采用加权欧氏距离,按式(7)计算待估灌区与其他灌区间的相似度,结果如表 5 所示。由此,可得

表 5 相似度 $S(b_x, b_i)$ 计算结果
Tab.5 Similarity calculation results

待估 灌区	灌区														
	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}
b_{x1}	0.961 4	0.955 2	0.947 8	0.951 7	0.958 6	0.956 3	0.983 6	0.990 5	0.984 3	0.972 5	0.952 0	0.984 8	0.976 1	0.957 8	0.952 7
b_{x2}	0.991 4	0.984 4	0.970 4	0.977 7	0.975 7	0.972 8	0.921 9	0.948 4	0.941 2	0.979 0	0.963 7	0.943 8	0.976 0	0.971 6	0.965 3
b_{x3}	0.957 8	0.963 6	0.967 7	0.966 3	0.981 8	0.986 3	0.914 3	0.937 1	0.930 5	0.967 2	0.994 6	0.932 0	0.961 1	0.985 1	0.992 1

到 b_{x1} (乌溪江引水灌区)、 b_{x2} (横锦水库灌区)、 b_{x3} (罗岙水库灌区)的相似灌区分别为 b_8 (铜山源水库灌区,相似度为 0.990 5)、 b_1 (南江水库灌区,相似度为 0.991 4)、 b_{11} (里林灌区,相似度为 0.994 6)。

因灌区配水案例集中成员较少,不宜采用指数平滑法,选择构造加权影响因子法,按式(8)计算待估灌区的毛需水量,结果如表 6 所示,相对误差均不大于 9.39%,说明本文方法有效。

表 6 需水量估算结果
Tab.6 Water demand estimation results

待估灌区	实际毛需水量/ 万 m ³	预测毛需水 量/万 m ³	相对 误差/%
乌溪江引水灌区	9 092.98	8 377.22	7.87
横锦水库灌区	3 514.25	3 276.37	6.77
罗岙水库灌区	518.28	469.63	9.39

需要指出的是,合理选择典型灌区,建立有效的灌区配水案例集是实现科学预测的前提。当案例集

中灌区案例不够丰富,案例指标值分布不合理,未能覆盖待估灌区的行政区域或类似灌区规模时,待估灌区往往无法获取到足够相似的历史灌区案例,预测结果也会不理想。

4 结论

(1)针对行政区域行业间配水问题,汇聚区域水资源、经济、人口、行业用水等数据,聚焦农业灌溉用水,运用模糊聚类分析,构建区域水资源行业配水案例分类及特征分析模型,并应用于 2018 年浙江省各市行业配水案例分析。结果显示,浙江省不同区域行业间配水被划分为 4 类,均呈现出较明显的特征。

(2)针对灌区需水量预测问题,采用固定步长穷举法对灌区配水相关参数进行动态赋权,运用模糊距离匹配相似灌区,通过构造加权影响因子和采用指数平滑法,对基于实例推理的灌区进行需水量

预测,并将该方法应用于2018年浙江省中型灌区的需水量预测,预测相对误差均不大于9.39%,验证了方法的有效性。

(3)与规划优化、系统模拟和智能计算等其他水资源配置方法相比,本文方法基于模糊数据挖掘,可以对各类数据直接进行相关分析,避开了复杂的

函数建模、边界条件设置和收敛性分析等过程,运用相对简单、易行。但也存在一定局限性:要求涉及相关指标可量化,且各指标已经汇聚相对比较丰富的历史数据;因数据挖掘结果反映的是相关关系,未经严格逻辑验证,存在不确定性,需要在实际应用中不断验证。

参 考 文 献

- [1] AREGAY F, YAO L, ZHAO M. Spatial preference heterogeneity for integrated river basin management: the case of the Shiyang River Basin, China[J]. Sustainability, 2016, 8(10): 970–986.
- [2] 靳晓莉, 王君勤, 高鹏. 中国灌区水资源优化配置研究进展[J]. 人民珠江, 2018, 39(3): 62–65.
JIN Xiaoli, WANG Junqin, GAO Peng. Research advance on water resources optimal allocation in irrigation district in China[J]. Pearl River, 2018, 39(3): 62–65. (in Chinese)
- [3] TAN Q, ZHANG T. Robust fractional programming approach for improving agricultural water-use efficiency under uncertainty[J]. Journal of Hydrology, 2018, 564: 1110–1119.
- [4] FLEMING R A, ADAMS R M, KIM C S. Regulating groundwater pollution: effects of geophysical response assumptions on economic efficiency[J]. Water Resources Research, 1995, 31(4): 1069–1076.
- [5] CARLOS P, GIDEON O, ABRAHAM M. Optimal operation of regional system with diverse water quality sources[J]. Journal of Water Resources Planning and Management, 1997, 123(2): 105–115.
- [6] REN C, GUO P, TAN Q, et al. A multi-objective fuzzy programming model for optimal use of irrigation water and land resources under uncertainty in Gansu Province, China[J]. Journal of Cleaner Production, 2017, 164: 85–94.
- [7] 付强, 肖圆圆, 崔嵩, 等. 基于多目标模糊规划的灌区多水源优化配置[J/OL]. 农业机械学报, 2017, 48(7): 222–227, 221.
FU Qiang, XIAO Yuanyuan, CUI Song, et al. Multi-water resources optimal allocation of irrigation district based on fuzzy multi-objective programming[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(7): 222–227, 221.
http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20170728&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2017.07.028. (in Chinese)
- [8] 谭倩, 候天宇, 张田媛, 等. 基于鲁棒规划方法的农业水资源多目标优化配置模型[J]. 水利学报, 2020, 51(1): 1–12.
TAN Qian, GOU Tianyu, ZHANG Tianyuan, et al. A multi-objective robust programming model for agricultural water management with uncertain weights[J]. Journal of Hydraulic Engineering, 2020, 51(1): 1–12. (in Chinese)
- [9] LI M, GUO P. A multi-objective optimal allocation model for irrigation water resources under multiple uncertainties[J]. Applied Mathematical Modelling, 2014, 38(19): 4897–4911.
- [10] WILLIS R. Multiple-criteria decision-making: a retrospective analysis[J]. IEEE Transactions on Systems Man Cybernetics-Systems, 1987, 15(3): 213–220.
- [11] WESTROM I, MESSING I, LINNAR H, et al. Controlled drainage effects on drain outflow and water quality[J]. Agricultural Water Management, 2001, 47(2): 85–100.
- [12] 陈红光, 王琼雅, 李晓宁, 等. 基于区间两阶段鲁棒优化模型的灌区水资源优化配置[J/OL]. 农业机械学报, 2019, 50(3): 271–280.
CHEN Hongguang, WANG Qiongya, LI Xiaoning, et al. Optimal water resources planning based on interval parameter two stage robust stochastic programming model[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(3): 271–280. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190329&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2019.03.029. (in Chinese)
- [13] 岳琼, 郭萍, 王友芝, 等. 基于区间两阶段模糊可信性约束模型的灌区水资源配置[J/OL]. 农业机械学报, 2019, 50(4): 228–235.
YUE Qiong, GUO Ping, WANG Youzhi, et al. Optimal water allocation of irrigation district based on interval-parameter two-stage stochastic fuzzy credibility constrained programming[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(4): 228–235. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190426&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2019.04.026. (in Chinese)
- [14] 李亮, 马良瑞, 熊康宁. 基于作物需水量的城市农业水资源评估——以贵州省贵阳市为例[J]. 水土保持通报, 2019, 39(4): 300–305.
LI Liang, MA Liangrui, XIONG Kangning. Evaluation of agriculture water resource in urban area based on crop water demand—a case study in Guiyang City in Guizhou Province[J]. Bulletin of Soil and Water Conservation, 2019, 39(4): 300–305. (in Chinese)
- [15] PHILIP J R. Plant water relations: some physical aspects[J]. Annual Review of Plant Physiology, 1966, 17: 245–268.
- [16] 王浩, 王建华, 秦大庸, 等. 基于二元水循环模式的水资源评价理论方法[J]. 水利学报, 2006, 37(12): 1496–1502.

- WANG Hao, WANG Jianhua, QIN Dayong, et al. Theory and methodology of water resources assessment based on dualistic water cycle model[J]. *Journal of Hydraulic Engineering*, 2006, 37(12): 1496–1502. (in Chinese)
- [17] 邹进. 基于二元水循环及系统熵理论的城市用水配置[J]. *水利水电科技进步*, 2019, 39(2): 16–20.
ZOU Jin. Allocation of urban water resources based on dualistic water cycle theory and system entropy[J]. *Advances in Science and Technology of Water Resources*, 2019, 39(2): 16–20. (in Chinese)
- [18] 张俊娥, 陆垂裕, 秦大庸, 等. 基于分布式水文模型的区域“四水”转化[J]. *水科学进展*, 2011, 22(5): 595–604.
ZHANG Jun'e, LU Chuiyu, QIN Dayong, et al. Regional “four-water” transformation based on distributed hydrological model [J]. *Advances in Water Science*, 2011, 22(5): 595–604. (in Chinese)
- [19] FORTES P S, PLATONOV A E, PEREIR A L S. GISAREG—a GIS based irrigation scheduling simulation model to support improved water use and environmental control[J]. *Agricultural Water Management*, 2005, 77(1): 159–179.
- [20] 刘倩倩, 陈岩. 基于粗糙集和 BP 神经网络的流域水资源脆弱性预测研究——以淮河流域为例[J]. *长江流域资源与环境*, 2016, 25(9): 1317–1327.
LIU Qianqian, CHEN Yan. Vulnerability prediction of basin water resources based on rough set and BP neural network—a case of Huaihe Basin[J]. *Resources and Environment in the Yangtze Basin*, 2016, 25(9): 1317–1327. (in Chinese)
- [21] QU Guodong, LOU Zhanghua. Application of particle swarm algorithm in the optimal allocation of regional water resources based on immune evolutionary algorithm[J]. *Journal of Shanghai Jiaotong University (Science)*, 2013, 18(5): 634–640.
- [22] 李春梅, 罗健, 吴云, 等. 改进果蝇算法在多目标水资源配置中的应用研究[J]. *人民长江*, 2019, 50(11): 107–112.
LI Chunmei, LUO Jian, WU Yun, et al. Application and research on improved FOA in multi-objective water resource allocation [J]. *Yangtze River*, 2019, 50(11): 107–112. (in Chinese)
- [23] 刘彬, 沙金霞. 改进人工鱼群算法在水资源优化配置中的应用[J]. *人民黄河*, 2017, 39(8): 58–62.
LIU Bin, SHA Jinxia. Application of improved artificial fish swarm algorithm optimal allocation of water resources[J]. *Yellow River*, 2017, 39(8): 58–62. (in Chinese)
- [24] EKASINGH B, NGAMSOMSUK K, LETCHER R, et al. A data mining approach to simulating farmers' crop choices for integrated water resources management[J]. *Journal of Environmental Management*, 2005, 77(4): 315–325.
- [25] THOMPSON K, KADIYALA R. Leveraging big data to improve water system operations[J]. *Procedia Engineering*, 2014, 89: 467–472.
- [26] MOHAN S, RAMSUNDRAM N. Data-mining models for water resource applications [J]. *ISH Journal of Hydraulic Engineering*, 2013, 19(3): 211–218.
- [27] SHAHANAS K M, SIVAKUMAR B. Framework for a smart water management system in the context of smart city initiatives in India [J]. *Procedia Computer Science*, 2016, 92: 142–147.
- [28] 金喜来, 徐瑶. 国家水资源监控能力建设项目数据库标准规范应用实践[J]. *中国水利*, 2015(19): 59–61.
JIN Xilai, XU Yao. Application of database specifications in national project of capacity building for water resources monitoring [J]. *China Water Resource*, 2015(19): 59–61. (in Chinese)
- [29] 蒋云钟, 张小娟. 数据挖掘技术在水资源领域的应用方法研究[J]. *人民黄河*, 2007, 29(9): 43–44, 47.
JIANG Yunzhong, ZHANG Xiaojuan. Research on the application method of data mining technology in the field of water resources [J]. *Yellow River*, 2007, 29(9): 43–44, 47. (in Chinese)
- [30] 刘予伟, 刘东润, 陈献耘. 大数据在水资源管理中的应用展望[J]. *水资源研究*, 2015, 4(5): 470–476.
LIU Yuwei, LIU Dongrun, CHEN Xianyun. Prospect for big data applications in water resources management[J]. *Journal of Water Resources Research*, 2015, 4(5): 470–476. (in Chinese)
- [31] 杨小柳, 范佳慧. 数据挖掘技术在节水管理中的应用[J]. *长江科学院院报*, 2019, 36(7): 1–6.
YANG Xiaoliu, FAN Jiahui. An application of data mining to water saving management [J]. *Journal of Yangtze River Scientific Research Institute*, 2019, 36(7): 1–6. (in Chinese)
- [32] 谢季坚, 邓小炎. 现代数学方法选讲[M]. 北京: 高等教育出版社, 2003: 54–55.
- [33] 肖刚, 张立彬, 高飞. 基于实例库自主聚类的自适应产品快速报价策略研究[J]. *中国机械工程*, 2010, 21(1): 80–84.
XIAO Gang, ZHANG Libin, GAO Fei. Research on strategy of adaptive rapid product quoting based on self-clustering of case library[J]. *China Mechanical Engineering*, 2010, 21(1): 80–84. (in Chinese)
- [34] 黎锁平, 刘坤会. 平滑系数自适应的二次指数平滑模型及其应用[J]. *系统工程理论与实践*, 2004(2): 95–99.
LI Suoping, LIU Kunhui. Quadric exponential smoothing model with adapted parameter and its applications [J]. *System Engineering Theory and Practice*, 2004(2): 95–99. (in Chinese)