

DOI:10.3880/j.issn.1004-6933.2023.06.018

基于 XGBoost 算法的跨境河流水资源争端影响因素分析

施国良^{1,2}, 薛 航^{1,2}

(1. 河海大学商学院, 江苏 南京 211100; 2. 江苏省“世界水谷”与水生态文明协同创新中心, 江苏 南京 211100)

摘要:为分析跨境河流水资源争端的关键影响因素对争端的影响规律,基于跨境淡水争端数据库(TFDD)构建了跨境河流水资源争端评价指标体系,并采用 XGBoost 算法计算得到评价指标的权重;分析了国家人口密度、人均国内可再生淡水资源量、地区因素、事件主要问题类型、流域面积和事件发生的月份对跨境河流水资源争端的影响,并选取 6 条代表河流进行了争端事件聚焦分析,统计了每条代表河流的主要问题类型,分析了引发跨境河流水资源争端的代表性问题。结果表明:涉事国家人口密度、人均国内可再生淡水资源量和地区因素对跨境河流水资源争端的影响最大;引发跨境河流水资源争端的代表性问题包括地区气候特征、水污染以及基础设施建设等。

关键词:跨境河流;水资源争端;水资源管理;XGBoost 算法

中图分类号:TV213.4 **文献标志码:**A **文章编号:**1004-6933(2023)06-0137-08

XGBoost-based analysis of factors affecting water resources disputes in transboundary rivers//SHI Guoliang^{1,2}, XUE Hang^{1,2} (1. Business School of Hohai University, Nanjing 211100, China; 2. Jiangsu Provincial Collaborative Innovation Center of World Water Valley and Water Ecological Civilization, Nanjing 211100, China.)

Abstract: In order to analyze the impact of key factors on transboundary river water resource disputes, an evaluation index system of transboundary river water resource disputes is established based on the transboundary river dispute database (TFDD), and the weight of evaluation index was calculated by XGBoost algorithm; the impacts of national population density, per capita domestic renewable fresh water resources, regional factors, event types, basin area and event month on transboundary river water resource disputes are analyzed; six representative rivers were selected for the focused analysis of disputes. The main problem types of each representative river are counted, and the representative problems leading to transboundary river water resource disputes are analyzed. The results show that the population density of the countries involved, the amount of domestic renewable freshwater resources per capita, and regional factors have the greatest impact on water resource disputes in transboundary rivers; representative issues that trigger water resource disputes in transboundary rivers include regional climate characteristics, water pollution, and infrastructure construction, etc.

Key words: transboundary river; water resources disputes; water resources management; XGBoost algorithm

跨境河流一直是各国水资源争夺的焦点,是地缘政治中的敏感问题。近年来,跨境河流的水资源争端时有发生^[1]。跨境河流水资源争端事件是指因为跨境河流水资源的不合理分配、水质污染、水利工程的不合理建设等问题影响沿岸国家淡水资源的正常取用而产生的国家之间的争端,这些争端可能是政治、经济、军事方面的冲突;非争端事件则是指国家之间关于跨境河流展开的合作和协商,这些合作和协商可能包括共同管理河流、共同开发水资源和共同应对水污染等。随着国家之间关于跨境河流

问题的接触日益频繁,水资源冲突爆发的频率也显著提高^[2]。分析水资源竞争与冲突要点,有利于开展国际河流研究,提高跨境河流水资源安全风险管控能力;而这种风险管控能力对于跨境河流流域的政治稳定与流域国家间的涉水合作具有重要的理论和现实意义。

目前对跨境河流水资源争端影响因素的研究包括单因素研究^[5-10]和评价指标体系研究^[2,11-13]。Toset 等^[5]研究发现,两国共享的河流越多,发生冲突的可能性就越大;此外,当至少有一个国家的淡水

基金项目:中央高校基本科研业务费专项资金资助项目(B200207036);国家自然科学基金项目(71974053)

作者简介:施国良(1966—),男,副教授,博士,主要从事自然语言处理与数据挖掘研究。E-mail:shigl@hhu.edu.cn

供应不足时,上下游关系的国家之间最可能发生冲突。Toset 等^[5]同时强调了两个国家边界线长度可能是冲突产生的影响因素,但没有对两国边界长度与冲突产生的关系进行分析。Furlong 等^[6]利用与 Toset 等^[5]研究中相同的数据集检验了 Toset 的理论研究,并增加了国家间陆地边界线长度的数据,证实了 Toset 等^[5]研究的可靠性。Gleditsch 等^[7]的研究表明,一个流域的水资源量与水资源冲突相关,即流域越大,冲突越少,因为流域越大,水资源越充足;研究结果还表明,经济发展水平较高的国家之间发生冲突的风险较低。Gizelis 等^[8]指出,政治制度会对水资源短缺情况下水资源争端的发生与否产生影响。Brochmann 等^[9]将研究对象限制在共享河流的国家,结果表明,上下游关系的国家发生冲突的可能性更高;流域大小与冲突可能呈负相关关系。Beck 等^[10]使用包含详细水文政治依赖关系的数据集来重现 Brochmann 等^[9]的研究,但得到了不同的结论:上下游国家之间的不对称关系对合作或冲突的影响比较小。李浩等^[11]利用经济理论分析了跨境水资源冲突的动因。张晖等^[2]基于社会冲突理论,将跨境水资源冲突的诱发因素分为物质利益和权利两个层面,并对这两个层面的因素进行了分析。Wolf 等^[12]根据俄勒冈州立大学创建的跨境淡水争端数据库(Transboundary Freshwater Dispute Database, TFDD)分析并确定了基于历史数据的跨境河流水资源争端风险指标,并应用这些指标来确定未来 10 年可能发生冲突的流域。相比之下,Rai 等^[13]认为 TFDD 缺乏对发展和霸权参数的考虑,通过将模糊综合评价技术与风险评价方法相结合,对发展和霸权参数进行表征和量化,构建了更加严密的水资源冲突和合作预测框架。

综上所述,以往的研究对跨境河流水资源争端的影响因素及其背后的规律分析还不够充分,而评价指标体系研究要么未能反映影响因素之间的重要性关系^[2,11-12],要么仅根据专家意见确定影响因素之间的重要性关系^[13]。由于权重反映各评价指标的相对重要程度,权重越大,该指标对评价结果的影响越大^[14],因此需要更客观的方法来确定评价指标体系中的指标权重,分析指标对应的因素对争端的影响规律。为此,本文构建了跨境河流水资源争端评价指标体系,采用 XGBoost 算法^[15]计算指标的权重,并分析各指标对跨境河流水资源争端的影响规

律,以期为解决跨境河流水资源争端问题提供参考。

1 研究方法

1.1 研究数据

1.1.1 数据来源

本文采用的数据主要为 TFDD 事件数据库的事件信息。此外,流域面积数据来自跨境水评估计划(Transboundary Waters Assessment Programme)^[16],综合国力信息来自 The Correlates of War 项目(1918—2016 年)^[17],人均国内可再生淡水资源量数据来自联合国粮农组织(FAO)^[18],人口密度信息来自世界银行的数据库^[19]。

TFDD 事件数据库记录了从 1948—2008 年的跨境河流水事件。按照水资源争端是否发生,以及冲突的严重程度将跨境河流水资源争端事件分为 -7~7 共 15 个等级,级别小于 0 代表争端事件,级别大于或等于 0 代表非争端事件。争端事件中,级别为 -1 的事件冲突程度最低,代表在国际互动中表现出言语交流上的冲突;级别为 -7 的事件冲突程度最高,代表正式宣战。非争端事件中,级别为 0 代表中立态度;级别 1~7 分别代表国家之间不同程度的合作或互相支持,并且级别越高,合作或者互相支持的程度越高。

TFDD 事件数据库记录了每个事件涉及的国家、发生争端的河流位置、事件发生时间和主要问题类型、冲突/合作级别以及事件的文本描述,其中每一条记录涉及两个国家。图 1 为 TFDD 事件数据库主要信息示例。

1.1.2 数据预处理

由于各个国家的人口密度和人均国内可再生淡水资源量的数据并不是每年都进行统计,因此采用线性回归法填补各国缺失的人均国内可再生淡水资源量数据,采用带常数项的二次函数拟合并填补各国人口密度缺失数据。将置信度设置为 95%,得到各国人均国内可再生淡水资源量回归系数 t 检验的 p 值均小于 0.05、判定系数 R^2 均大于 0.8,各国人口密度回归系数 t 检验的 p 值均小于 0.05、 R^2 均大于或等于 0.699,表明对各国人均国内可再生淡水资源量数据和人口密度数据拟合较好,对缺失数据的

时间	河流位置	国家 1	国家 2	冲突等级	事件描述	主要问题类型
1948/4/1	Indus	India	Pakistan	-3	The "Standstill Agreement" between	Water Quantity
1948/12/7	JORDAN	Egypt, Arab Rep.	Israel	-6	Egypt destroys to Israel some wate	Infrastructure/Development

图 1 TFDD 事件数据库主要信息示例

Fig. 1 Examples of main information of TFDD event database

填补结果较合理。

利用 TFDD 事件数据并融合其他数据后得到 2 793 条有效数据,其中有 2 308 条为非争端事件数据,485 条为争端事件数据。

以 1996—1999 年发生的 701 条事件数据来测试本文构建的跨境河流水资源争端评价指标体系预测未来跨境河流水资源争端的能力,以 1996 年之前发生的 2092 条事件数据来训练 XGBoost 算法,随机选择训练 XGBoost 算法数据的 80% 作为训练集,20% 作为测试集。

1.2 跨境河流水资源争端评价指标体系

如图 2 所示,选取以往研究已经证明或根据经验能够推断出对跨境河流水资源争端有影响的 10 个指标初步构建跨境河流水资源争端评价指标体系,在此基础上,采用 XGBoost 算法对指标进行进一步筛选并且计算指标的权重。

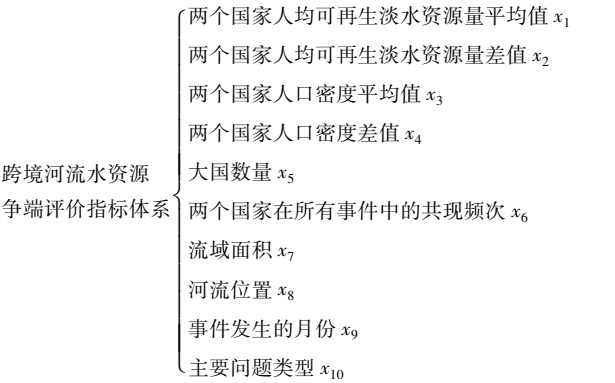


图 2 初始跨境河流水资源争端评价指标体系

Fig.2 Evaluation index system for initial transboundary river water resource disputes

图 2 评价指标体系包含涉事国家、跨境河流以及事件 3 个方面的指标。

a. 涉事国家方面包括反映人均国内可再生淡水资源量、国家人口密度、国家实力以及国家所属地区的指标。Gizelis 等^[8]的研究表明涉事国家的人均国内可再生淡水资源量对跨境河流水资源争端的发生具有影响。一个国家的可用淡水资源越少,就越可能因为争夺水资源而与其他国家发生冲突。因此,选取两个国家人均国内可再生淡水资源量平均值(x_1)和两个国家人均国内可再生淡水资源量差值(x_2)作为反映人均国内可再生淡水资源量的指标。Wolf 等^[12]以流域人口密度为指标研究跨境河流水资源争端,并且证明该指标与国家之间关于跨境河流的合作与争端存在关系。除了流域人口密度外,国家人口密度也与淡水需求密切相关,因此有理由认为其会影响到跨境河流水资源争端的发生。因此,选取两个国家人口密度平均值(x_3)和两个国家

人口密度差值(x_4)作为反映国家人口密度的指标。Toset 等^[5]的研究表明,两个国家中大国数量的多少会影响跨境河流事件的性质(合作或冲突)。因此,使用 The Correlates of War 项目^[17]中的 major powers 变量值作为反映国家实力的指标——大国数量(x_5)。大国数量描述了两国的国力,按照一些特定的标准(主要是按照国家实力和国际影响力,标准来自 Small 和 Singer 的研究^[20])来衡量一个国家是否是 大国,将两个国家中的大国数量相应编码为 0、1、2。根据涉事国家名称可以推断国家所属的地区,不同地区干旱程度不同,发生水资源争端的风险也不同。因此,以两个国家在所有事件中的共现频次(x_6)作为反映涉事国家所属地区的指标, x_6 计算方法是将一条事件记录中的两个国家名称合并后的值在整个数据集中出现的次数除以数据集记录总数。

b. 跨境河流方面包括反映流域面积和河流所属地区的指标。Gleditsch 等^[7]的研究表明流域面积也与争端的发生有关,因此选择流域面积(x_7)作为评价指标。不同地区的河流发生水资源争端的概率不同,干旱地区的概率更高,因此选取事件发生的河流位置(x_8)作为反映河流所属地区的指标。

c. 事件方面包括反映事件发生的月份和事件的主要问题类型的指标。事件发生的时间信息中包含了月份信息,不同月份的降水量有所不同,因此选取事件发生的月份(x_9)作为评价指标。此外,事件的主要问题类型包括河流水污染、大坝建设等可能会对各国淡水资源取用造成影响的问题。例如,河流污染会导致可用淡水资源减少,而大坝建设则会导致河流流向或流量改变,影响沿岸国家取水用水。因此,选取主要问题类型(x_{10})作为评价指标。

1.3 基于 XGBoost 算法的指标权重计算及指标筛选

1.3.1 XGBoost 算法

XGBoost 算法是经典的基于决策树的集成学习方法。使用指标数据集训练 XGBoost 算法,训练完成后可以计算每个指标对跨境河流水资源争端的权重,并且可以根据训练后的 XGBoost 算法的效果判断评价指标体系的合理性,实现对指标的筛选。

1.3.2 算法的评价指标

为评估评价指标体系的合理性,使用训练后的 XGBoost 算法预测测试集中的跨境河流事件是否为争端事件。采用受试者工作特性(ROC)曲线下的面积(AUC)和精确率指标评价 XGBoost 算法的预测

效果。AUC 可以用于评价 XGBoost 算法对争端事件和非争端事件的分类能力,并且正负样本不均衡对 AUC 影响较小,可以对算法进行较为客观的效果评价。精确率可以评估 XGBoost 算法对争端事件的正确分类能力,因此,在 AUC 指标之外通过识别争端事件样本的精确率来评价 XGBoost 算法的预测效果。

AUC 值的变化范围为 0.5~1,精准率的变化范围为 0~1。AUC 值和精确率越大,能够说明 XGBoost 算法对于实验数据集中的数据特征学习越充分,建立的评价指标体系越合理。

1.3.3 XGBoost 算法训练

在训练过程中,XGBoost 算法不断地对训练集数据进行建模学习,使得 XGBoost 算法的损失函数值不断降低。当损失函数值不再降低时,表明 XGBoost 算法已经对训练集中的数据特征学习充分,训练已经完成。

训练 XGBoost 算法的过程就是确定最优特征进行数据集划分构建子树的过程^[15]。训练的过程如下:训练 XGBoost 算法时,使用贪心算法从根节点开始递归选择节点分裂的最优指标特征,将数据集不断分割,最终形成多个叶子节点,并保证单个叶子节点中的事件属于同一个类别(争端事件/非争端事件)。

以 XGBoost 运行过程中一棵子树中一条路径(路径1)为例(图3)说明节点分裂过程:首先对训练集中所有指标进行排序;然后遍历每个指标的所有可能切分点,计算每个切分点带来的信息增益,选择信息增益最大的切分点作为该指标的最优切分点;最后从所有指标的最优切分点中选择信息增益最大的作为该分裂节点的最优切分点。依照以上原理,对于跨境河流水资源争端事件集合A,选出的最优划分指标是流域面积(x_7),切分点为 75 029 km²,

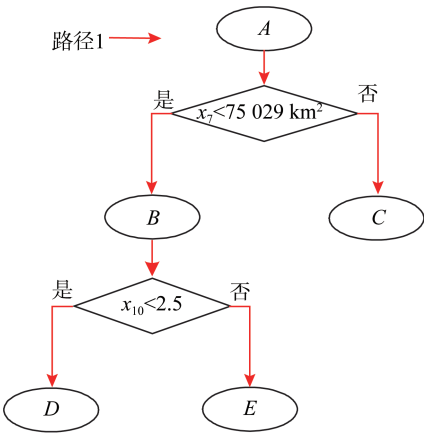


图3 XGBoost 训练时子树分裂过程

Fig.3 Sub-tree splitting process during XGBoost training

找出事件集合A中流域面积小于 75 029 km²的事件得到子集B,找出流域面积大于 75 029 km²的事件得到子集C。对于子集B,选出的最优划分指标是主要问题类型(x_{10}),切分点为 2.5,找出子集B中主要问题类型值小于 2.5 的事件到子集D,找出主要问题类型值大于 2.5 的事件得到子集E。

1.3.4 指标筛选和指标权重确定

XGBoost 算法训练完成后,计算每个指标在所有分裂中带来的平均信息增益,最后将所有指标的平均信息增益值进行归一化处理,作为最终的指标权重。使用训练后的 XGBoost 算法预测测试集中的跨境河流事件是否为争端事件,并计算 AUC 值和精确率。从评价指标体系中去掉权重远低于其他指标的指标,重新训练 XGBoost 算法并比较去掉该指标前后计算得到的 AUC 值和精确率,如果 AUC 值和精确率变高,表明该指标作为评价指标体系的组成不合理,不当放入评价指标体系。

2 结果与分析

2.1 指标权重计算结果

根据初始的评价指标体系训练 XGBoost 算法,并计算各个指标的权重,其中指标 x_5 的权重仅为 0.009,相对其他指标而言,其对预测跨境河流水资源争端的贡献较小。尝试将指标 x_5 移除,重新训练 XGBoost 算法,在测试集上评估的 AUC 值达到 0.71,精确率达到 88%,预测效果相较于移除指标 x_5 前更优,表明指标 x_5 对跨境河流水资源争端的影响较小,作为评价指标体系的组成指标是不合理的。

使用训练后的 XGBoost 算法对 1996—1999 年的跨境河流水资源争端事件进行预测,精确率达到 78%,表明构建的评价指标体系比较合理。最终确定的评价指标体系及各指标的权重见图4(括号内数据为指标权重)。

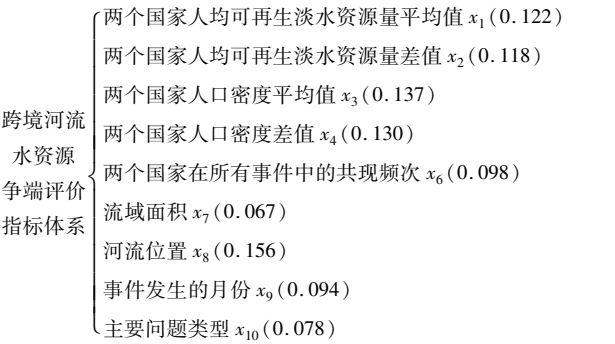


图4 跨境河流水资源争端评价指标体系及指标权重

Fig.4 Evaluation index system and index weight of transboundary river water resource disputes

2.2 争端影响因素分析

对计算得到的指标权重进行排序,可以看出国

家人口密度(指标 x_3 和 x_4)、人均国内可再生淡水资源量(指标 x_1 和 x_2)和地区因素(指标 x_6 和 x_8)是影响跨境河流水资源争端的主要因素,流域面积、事件的主要问题类型和事件发生的月份对跨境河流水资源争端的发生也有不同程度的影响。

a. 国家人口密度和人均国内可再生淡水资源量。相较于非争端事件,争端事件记录中两个涉事国家的人均国内可再生淡水资源量平均值和人均国内可再生淡水资源量差值的绝对值集中在更低的水平,即当两个国家的国内可用淡水资源量都处于较低水平时,更可能因为争夺水资源而发生冲突。分析结果表明人均国内可再生淡水资源量会影响跨境河流水资源争端,与 Gizelis 等^[8]的研究结论相吻合。两个涉事国家的人口密度平均值集中在更高的水平,而两个涉事国家人口密度差值的绝对值集中在更低的水平,说明当两个国家的人口密度都比较高时,对于水资源的需求更高,更可能因为争夺水资源而发生冲突。

b. 地区因素。地区因素体现为河流所属地区(指标 x_8)和涉事国家所属地区(指标 x_6)。图 5 为争端事件涉及的国家分布情况,发现 1948—2006 年发生争端频率最高的是中东国家,例如以色列、土耳其和叙利亚。大多数跨境河流水资源争端发生在约旦河和底格里斯河,这两条河流都位于中东地区。中东地区整体较为干旱,水资源比较匮乏,而约旦河和底格里斯河是重要的淡水来源,也是沿岸国家淡水资源争端的焦点。因此,地区因素实际上反映了

- 1—巴基斯坦
- 2—黎巴嫩
- 3—尼泊尔
- 4—以色列
- 5—土耳其
- 6—阿塞拜疆
- 7—亚美尼亚
- 8—印度
- 9—阿根廷
- 10—英国
- 11—泰国
- 12—巴西
- 13—匈牙利
- 14—埃及
- 15—叙利亚
- 16—墨西哥
- 17—约旦
- 18—巴拉圭
- 19—伊拉克
- 20—美国
- 21—孟加拉国

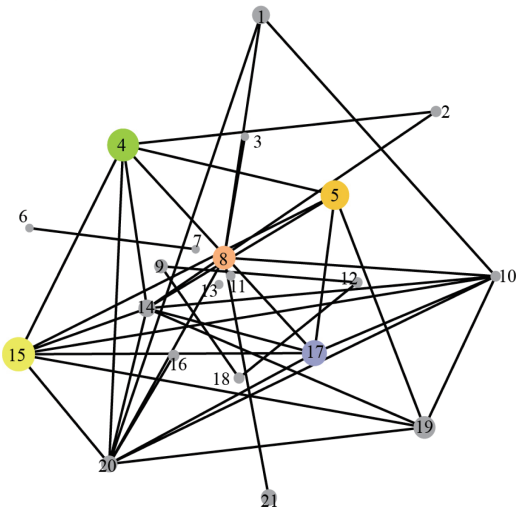


图 5 争端事件涉及的国家分布

Fig. 5 Distribution of countries involved in dispute

跨境河流或者沿岸国家所处地区的缺水 and 干旱程度。

c. 主要问题类型和流域面积。主要问题类型(指标 x_{10})也是争端是否发生的影响因素之一,本文数据覆盖的时间范围中,争端事件主要问题类型的分布如表 1 所示。由表 1 可见,271 件跨境河流水资源争端事件与水资源量有关,67 件与大坝等基础设施建设有关,13 件与水质问题有关。争端涉及的其他具体问题包括流域水资源联合管理、渔业、航运和农业灌溉等。根据重要性排序结果,流域面积也会影响跨境河流水资源争端的发生,数据分析结果显示越小的河流越容易发生争端事件。该结果支持 Brochmann 等^[9]的观点,即流域规模与冲突可能性负相关。

表 1 争端事件主要问题类型出现次数统计

Table 1 Statistics on occurrence times of issue type in dispute events

争端事件的主要问题类型	争端事件中出现的次数
淡水量 (water quantity)	271
基础设施建设 (infrastructure/development)	67
水质 (water quality)	13
河流边界 (border issues)	11
联合管理 (joint management)	10
水电 (hydro-power/hydro-electricity)	6
防洪救灾 (flood control/relief)	5
技术合作/援助 (technical cooperation/assistance)	4
航海、航运、港口 (navigation)	3
农业灌溉 (irrigation)	1
河流的领土归属 (territorial issues)	1

d. 事件发生的月份。不同地区在相同月份具有不同的气候特征,会极大地影响淡水储量。各国在雨季淡水储量会增加,在淡水资源充足的情况下,与他国发生淡水资源争夺的可能性比较小,反之,旱季发生争端的可能性会增大。

2.3 争端事件聚焦分析

2.3.1 跨境河流主要问题分析

对每条跨境河流发生争端事件的数量进行统计,选取争端事件发生数量超过 10 次的跨境河流作为代表河流,得到 6 条代表河流的主要问题类型如表 2 所示。

由表 2 可以看出,导致不同跨境河流水资源争端的主要问题不同。对于中东地区的跨境河流约旦河、底格里斯河、幼发拉底河和尼罗河,普遍存在的主要问题是淡水量和基础设施建设问题。此外,约旦河、底格里斯河和幼发拉底河都存在水质问题。根据世界银行组织 2018 年的报告^[21],当时中东水资源缺乏问题严重,并且淡水供给量不足、水质问题仍然存在。淡水供给量不足主要是因为中东地区常

表2 代表河流的主要问题类型

Table 2 Main issue types of representative rivers

跨境河流	争端事件数量	争端事件的主要问题类型
约旦河(Jordan River)	97	淡水量,水质,基础设施建设
底格里斯河(Tigris)	40	淡水量,水质,基础设施建设
幼发拉底河(Euphrates)	34	淡水量,水质,基础设施建设
印度河(Indus)	15	淡水量,基础设施建设
尼罗河(Nile)	11	淡水量,基础设施建设
恒河(Ganges)	10	淡水量,基础设施建设

年干旱,导致该地区可用淡水资源严重缺乏,需求量大而供给量严重不足。中东地区的水质问题来源于两个方面:一是中东地区有大量废水未经处理直接排放,污染淡水资源;二是地表水过度抽取导致水体中污染物浓度上升,地下水过度抽取导致水源中盐分过高,造成水污染。同时水污染也导致可用淡水资源减少,增加了淡水资源争端发生的风险。基础设施建设问题指修建大坝和供水通道等基础设施满足自己国家的供水需求,进而会影响其他国家的取水,还可能对生态环境造成影响。当上游国家修建大坝对水资源进行拦截,或者过度取水时,下游国家可从跨境河流中获取的水资源量就会大大减少,可能出现淡水供给不足的问题。

位于南亚地区的印度河、恒河都存在淡水量和基础设施建设问题。根据争端事件内容分析,印度河流域主要争端在印度与巴基斯坦之间发生,巴基斯坦有大量农田需要印度河河水浇灌,但印度河上游的大部分位于印度境内,巴基斯坦的取用水受到印度供水的限制。而在恒河争端事件中,法拉卡(Farakka)大坝曾是印度和孟加拉国之间的一个争议点。印度在恒河修建了法拉卡大坝,大坝在旱季截留了大部分恒河水,导致孟加拉国水资源缺乏,影响航运和灌溉,在多雨季节放水时可能会加剧洪水泛滥,造成破坏。

结合以上对代表性跨境河流的主要问题分析,可以发现:①中东地区的特殊气候导致其淡水资源严重缺乏,该地区大部分淡水资源供给需要依赖于跨境河流,这种情况长期存在。该地区的国家可以采取做好雨水收集和存储,做到对水资源的充分利用。同时对于农业灌溉用水,可以推广滴灌等更先进的灌溉方式,提高水资源的利用率。②水污染会导致可用淡水资源减少,这不仅是中东跨境河流需要关注和解决的问题,同时也是其他跨境河流需要关注的问题。对于水污染问题要完善相关的法规和加强监管。法规和监管不应该限制在国家内部,同一跨境河流的沿岸国家应当进行协商,就水污染治理和监管问题达成一致意见,并互相监督。③6

条代表性河流都存在基础设施建设问题,可见大坝等设施的建设会影响沿岸国家淡水资源的取用。因此,对大坝等会影响沿岸国家取水用水的设施建设,应当经由跨境河流沿岸国家协商一致,充分考虑可能引发的问题。

2.3.2 争端事件关键词分析

除了以上争端影响因素外,关键词也可以预示争端事件的发生。TFDD事件数据库包含对每个事件的描述,采用TF-IDF算法^[22-23]分别提取争端事件和非争端事件相关的关键词,从争端事件中提取的关键词涉及国家(例如Arab、Iraqi)和河流名称(例如Tigris、Yarmuk),证明了地区因素的重要性。提取的关键词还包括territory、dams、right、share等与争端事件主要问题类型相关的特征词,同样可以预示争端发生的风险。

为了进行关键词共现分析并发现关键词与所反映的事实之间的联系,使用Spacy工具进一步对TFDD事件数据库中的事件描述文本进行命名实体识别,然后对命名实体进行共现分析。提取的命名实体的标签属性为PERSON、NORP、FAC、ORG、LOC、GPE、EVENT、LAW、DATE。命名实体属性的标签解释如表3所示。

表3 命名实体标签解释

Table 3 Label descriptions of named entity

命名实体标签	解释
PERSON	人名(包括虚构的)
NORP	国籍、宗教或政治团体
FAC	建筑物、机场、高速公路、桥梁等
ORG	公司、机构
LOC	地点(不包括国家名、城市名和州名)、山脉、水体
GPE	国家、城市、州
EVENT	已命名的飓风、战争和体育赛事等
LAW	法律法规名
DATE	时间

分析结果表明,共现次数最多的国家是以色列和叙利亚,图6为争端事件描述中为“Israel”和“Syria”同时出现的争端事件记录构建的命名实体共现网络^[24]。图6涉及的命名实体标签为LOC、GPE,表明在“Israel”和“Syria”同时出现的争端事件中,涉及的国家包括以色列、叙利亚、土耳其、黎巴嫩、埃及、巴勒斯坦、伊拉克等国家,涉及约旦河和胡莱河。

分析争端事件的文本描述,以色列和叙利亚在水资源分配方面存在严重冲突,而约旦河是冲突的焦点。卷入这场冲突的国家包括土耳其、黎巴嫩等国家。以色列和叙利亚同为中东国家,都面临缺水的困境。对事件主要问题类型分析发现,以色列和叙利亚之间87%的争端都是由于缺水引起的。

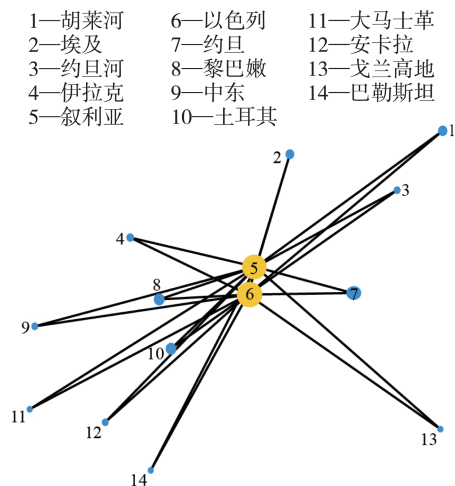


图6 国家和地区实体共现关系

Fig. 6 Co-occurrence relationship between state and place entities

3 结 论

a. 人均国内可再生淡水资源量、国家人口密度、地区因素(河流或国家所处地区的干旱程度)、流域面积和月份都会影响跨境河流水资源争端的发生风险,且影响程度不同,人均国内可再生淡水资源量、国家人口密度和地区因素影响程度最大。

b. 中东地区降水量少,因气候原因可用淡水资源稀缺,导致国家之间水资源争端频发。

c. 水污染问题会减少可用淡水资源量,增加跨境河流水资源争端风险。水污染问题可能由工业废水排放和过度抽取地下水引起,是各个跨境河流流域国家都需要重点关注和解决的问题。

d. 基础设施建设引发跨境河流水资源争端的现象普遍存在。大坝和水电站等基础设施建设可能会改变河流流向或者改变流向下游的水量,导致下游国家出现水资源短缺的问题,引发争端。

参考文献:

[1] 钟勇. 基于互惠合作的跨界河流开发利用博弈模型研究[D]. 北京:清华大学,2016.

[2] 张晖,雷丽萍. 跨界水冲突诱发因素研究述评[J]. 湖北理工学院学报(人文社会科学版),2020,37(6):71-77. (ZHANG Hui, LEI Liping. Review of inducing factors of transboundary water conflict [J]. Journal of Hubei Polytechnic University (Humanities and Social Sciences), 2020,37(6):71-77. (in Chinese))

[3] 何大明,苟俊华,KUNG H T. 国际河流(湖泊)水资源的竞争利用、冲突和求解[J]. 地理学报,1999,54(增刊1):38-46. (HE Daming, GOU Junhua, KUNG H T. Competitive utilization, conflicts and solutions of water resources in international rivers (lakes) [J]. Acta Geographica Sinica, 1999, 54 (Supl): 38-46. (in

Chinese))

[4] 何大明,刘恒,冯彦,等. 全球变化下跨境水资源理论与方法研究展望[J]. 水科学进展,2016,27(6):928-934. (HE Daming, LIU Heng, FENG Yan, et al. Perspective on theories and methods study of transboundary water resources under the global change[J]. Advances in Water Science, 2016,27(6):928-934. (in Chinese))

[5] TOSET H P W, GLEDITSCH N P, HEGRE H. Shared rivers and interstate conflict [J]. Political Geography, 2000,19(8):971-996.

[6] FURLONG K, GLEDITSCH N P, HEGRE H. Geographic opportunity and neomalthusian willingness: boundaries, shared rivers, and conflict[J]. International Interactions, 2006,32(1):79-108.

[7] GLEDITSCH N P, FURLONG K, HEGRE H, et al. Conflicts over shared rivers: resource scarcity or fuzzy boundaries? [J]. Political Geography, 2006, 25 (4): 361-382.

[8] GIZELIS T I, WOODEN A E. Water resources, institutions, & intrastate conflict[J]. Political Geography, 2010,29(8):444-453.

[9] BROCHMANN M, GLEDITSCH N P. Shared rivers and conflict-a reconsideration[J]. Political Geography, 2012, 31(8):519-527.

[10] BECK L, BERNAUER T, SIEGFRIED T, et al. Implications of hydro-political dependency for international water cooperation and conflict: insights from new data[J]. Political Geography, 2014,42:23-33.

[11] 李浩,刘陶,黄薇. 跨界水资源冲突的动因分析[J]. 中国水利,2010(3):12-14. (LI Hao, LIU Tao, HUANG Wei. Analysis of motivation for trans-boundary water conflict[J]. China Water Resources, 2010(3):12-14. (in Chinese))

[12] WOLF A T, YOFFE S B, GIORDANO M. International waters: identifying basins at risk[J]. Water Policy, 2003,5 (1):29-60.

[13] RAI S P, SHARMA N, LOHANI A K. Risk assessment for transboundary rivers using fuzzy synthetic evaluation technique [J]. Journal of Hydrology, 2014, 519: 1551-1559.

[14] 陈达,钟菁,吴腾,等. 江苏省长江岸线利用率影响因素分析[J]. 水资源保护,2022,38(3):17-24. (CHEN Da, ZHONG Jing, WU Teng, et al. Factors influencing utilization rate of Yangtze River shoreline in Jiangsu Province[J]. Water Resources Protection, 2022,38(3): 17-24. (in Chinese))

[15] 刘敬文,白金朋,王建,等. 基于 XGBoost 的初始地应力场修正反演法[J]. 水利水电科技进展,2022,42(2):101-108. (LIU Jingwen, BAI Jinpeng, WANG Jian, et al. A modified inversion method of initial geo-stress field based on XGBoost algorithm[J]. Advances in Science and Technology of Water Resources, 2022,42(2):101-108. (in Chinese))

- [16] UNEP. TWAP rivers portal-indicators application [EB/OL]. [2023-02-08]. <http://twap-rivers.org/indicators/>.
- [17] The Correlates of War. State system membership (v2016) [EB/OL]. [2023-02-08]. <https://correlatesofwar.org/data-sets/state-system-membership/>.
- [18] Food and Agriculture Organization of the United Nations. Aquastat data [EB/OL]. [2023-02-08]. https://tableau.apps.fao.org/views/ReviewDashboard-v1/country_dashboard?%3Aembed=y&%3AisGuestRedirectFromVizportal=y.
- [19] The World Bank. Population density (people per sq. km of land area) [EB/OL]. [2023-02-08]. <https://data.worldbank.org/indicator/EN.POP.DNST?view=chart>.
- [20] SMALL M, SINGER J D. Resort to arms: international and civil wars, 1816-1980 [M]. California: Beverly Hills SAGE Publications, 1982.
- [21] World Bank Group. Beyond scarcity: water security in the Middle East and North Africa [M]. Washington, D. C.: The World Bank, 2018: 25-51.
- [22] 武永亮, 赵书良, 李长镜, 等. 基于 TF-IDF 和余弦相似度的文本分类方法 [J]. 中文信息学报, 2017, 31(5): 138-145. (WU Yongliang, ZHAO Shuliang, LI Changjing, et al. Text classification method based on TF-IDF and cosine similarity [J]. Journal of Chinese Information Processing, 2017, 31(5): 138-145. (in Chinese))
- [23] 张振豪, 过弋, 韩美琪, 等. 基于关键词相似度的短文本分类方法研究 [J]. 计算机应用研究, 2020, 37(1): 26-29. (ZHANG Zhenhao, GUO Yi, HAN Meiqi, et al. Research on short text classification based on keyword similarity [J]. Application Research of Computers, 2020, 37(1): 26-29. (in Chinese))
- [24] 孙海生. 作者关键词共现网络及实证研究 [J]. 情报杂志, 2012, 31(9): 63-67. (SUN Haisheng. Author keyword co-occurrence network analysis: an empirical research [J]. Journal of Intelligence, 2012, 31(9): 63-67. (in Chinese))

(收稿日期: 2023-02-10 编辑: 熊水斌)

(上接第 136 页)

- [12] 缪昭旺, 吴成国, 崔毅, 等. 水资源空间均衡评价的联系数-耦合协调度模型及应用 [J]. 华北水利水电大学学报 (自然科学版), 2021, 42(3): 86-95. (MIAO Zhaowang, WU Chengguo, CUI Yi, et al. Connection number and coupling coordination degree model and application of water resources spatial equilibrium evaluation analysis [J]. Journal of North China University of Water Resources and Electric Power, 2021, 42(3): 86-95. (in Chinese))
- [13] 杨亚锋, 闫佳伟, 王红瑞, 等. 长江经济带水环境承载力时空变化特征 [J]. 南水北调与水利科技 (中英文), 2022, 20(4): 714-723. (YANG Yafeng, YAN Jiawei, WANG Hongrui, et al. Spatial-temporal variation of water environment carrying capacity in the Yangtze River Economic Belt [J]. South-to-North Water Transfers and Water Science & Technology, 2022, 20(4): 714-723. (in Chinese))
- [14] 程乾生. 属性识别理论模型及其应用 [J]. 北京大学学报 (自然科学版), 1997(1): 14-22. (CHENG Qiansheng. Attribute recognition theoretical model with application [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 1997(1): 14-22. (in Chinese))
- [15] 郭旭宁, 郦建强, 李云玲, 等. 京津冀地区水资源空间均衡评价及调控措施 [J]. 水资源保护, 2022, 38(1): 62-66. (GUO Xuning, LI Jianqiang, LI Yunling, et al. Water resources spatial equilibrium evaluation and regulation measures in Beijing-Tianjin-Hebei Region [J]. Water Resources Protection, 2022, 38(1): 62-66. (in Chinese))
- [16] 夏帆, 陈莹, 窦明, 等. 水资源空间均衡系数计算方法及其应用 [J]. 水资源保护, 2020, 36(1): 52-57. (XIA Fan, CHEN Ying, DOU Ming, et al. Calculation method and application of spatial equilibrium coefficient of water resources [J]. Water Resources Protection, 2020, 36(1): 52-57. (in Chinese))
- [17] 江东, 林刚, 付晶莹. “三生空间”统筹的科学基础与优化途径探析 [J]. 自然资源学报, 2021, 36(5): 1085-1101. (JIANG Dong, LIN Gang, FU Jingying. Discussion on scientific foundation and approach for the overall optimization of “Production-Living-Ecological” space [J]. Journal of Natural Resources, 2021, 36(5): 1085-1101. (in Chinese))
- [18] 安莉娜, 范国福, 吴迪, 等. 基于水资源承载力与城市经济社会发展水平耦合协调度模型的遵义市发展评价 [J]. 河海大学学报 (自然科学版), 2023, 51(2): 35-41. (AN Lina, FAN Guofu, WU Di, et al. Development evaluation of Zunyi City based on coupling coordination degree model of water resources carrying capacity and urban economic and social development level [J]. Journal of Hohai University (Natural Sciences), 2023, 51(2): 35-41. (in Chinese))
- [19] 高玉琴, 吴迪, 刘海瑞, 等. 城市化影响下区域水资源承载力评价 [J]. 水利水电科技进步, 2022, 42(3): 1-8. (GAO Yuqin, et al. Evaluation of regional water resources carrying capacity under urbanization [J]. Advances in Science and Technology of Water Resources, 2022, 42(3): 1-8. (in Chinese))
- [20] 潘辛伟, 金菊良, 刘晓薇, 等. 水资源利用系统脆弱性机理分析与评价方法研究 [J]. 自然资源学报, 2016, 31(9): 1599-1609. (PAN Zhengwei, JIN Juliang, LIU Xiaowei, et al. Mechanism and evaluation method of water resources utilization system vulnerability [J]. Journal of Natural Resources, 2016, 31(9): 1599-1609. (in Chinese))

(收稿日期: 2022-11-06 编辑: 彭桃英)