



基于大数据的绞吸挖泥船参数 自主寻优方案设计

赵春峰^{1,2}, 陈定^{1,2}, 李鹏超^{1,2}, 刘洪公^{1,2}

(1. 中交天津航道局有限公司, 天津 300461; 2. 天津市疏浚工程技术企业重点实验室, 天津 300457)

摘要: 绞吸挖泥船的施工过程非常复杂, 具有较多的不确定性和随机性。针对绞吸挖泥船生产过程中疏浚产量受施工环境影响较多、稳定性不足、生产率不高的问题, 以绞吸挖泥船传输管道中的泥浆为研究对象, 分析挖泥施工的特性与影响因素, 将疏浚工程历史大数据进行预处理, 比较 T-S 模型和历史数据最近邻的两种不同的流量预测方案, 采用偏差反馈控制器进行基于大数据的施工参数自主寻优。结果表明: 由于数据进行了预处理, 第 2 种流量预测方案准确率更高; 按寻优参数进行反馈控制调节的流量比原始流量更高。

关键词: 绞吸挖泥船; 大数据; 施工参数寻优

中图分类号: U 616

文献标志码: A

文章编号: 1002-4972(2022)S2-0119-06

Design of autonomous optimization scheme of cutter suction dredger parameters based on big data

ZHAO Chun-feng^{1,2}, CHEN Ding^{1,2}, LI Peng-chao^{1,2}, LIU Hong-gong^{1,2}

(1.Tianjin Key Laboratory of Dredging Engineering Enterprises, Tianjin 300461, China;
2.CCCC Tianjin Dredging Co., Ltd., Tianjin 300457, China)

Abstract: The construction process of cutter suction dredger is very complex, with more uncertainty and randomness. Regarding the problems of many impacts in dredging output, insufficient stability and low productivity, we take the slurry in the transmission pipeline of cutter suction dredger as the research object, analyze the characteristics and influencing factors of dredging construction, preprocess the historical big data of dredging engineering, compare two different flow prediction schemes of T-S model and historical data nearest neighbor, and use deviation feedback controller to independently optimize construction parameters based on big data. The results show that the accuracy of the second scheme is higher because the data are preprocessed. The flow adjusted by feedback control according to the optimization parameters is higher than the original flow.

Keywords: cutter suction dredger; big data; construction parameters optimization

绞吸挖泥船的施工过程非常复杂, 具有不确定性、复杂性和随机性, 施工参数需要技术娴熟和有经验的专业人员进行人工调节^[1]。随着工业化和信息化的融合, 大数据分析、智能优化及计算机技术越来越多地运用到绞吸挖泥船的施工优化与施工技术的升级改造中。

朱师伦等^[2]研究疏浚作业过程中泥浆浓度的控

制方法, 构建泥浆浓度过程特征模型; 李明超等^[3]提出一种基于施工循环的绞吸挖泥船疏浚作业平稳度分析评价方法, 可以为绞吸挖泥船的疏浚作业提供有效指导; 朱逸峰等^[4]通过数据挖掘工具对疏浚工艺点进行聚类研究, 并根据聚类结果给出了初步的疏浚性能评估; 刘骏等^[5]提出一种改进的建模及优化方法, 提高 Takagi-Sugeno (简称

T-S)模糊模型的辨识精度和速度。

本文以绞吸挖泥船传输管道中固液两相流为研究对象,从疏浚工程的历史大数据出发,对影响施工的特征参数进行分析筛选,并进行数据滤波处理。设计出两种绞吸挖泥船泥浆流量预测方案,分别利用实船施工数据进行泥浆流量变化趋势的预测,比较两种方案的优缺点。最后,通过流量的预测偏差设计反馈环节,实现参数的调节,达到基于大数据的参数自主寻优的目的。

1 绞吸挖泥船特性参数

1.1 基本性能

1)公称生产能力是额定条件下的生产率,一般在实际施工过程中小于这个值。

2)挖深越大,船舶主要尺度就要加大,其挖掘浓度随之降低。

3)排距是挖泥船将从挖泥区挖掘的疏浚土用管线输送到卸泥区的管线长度。在其他工况完全相同的情况下,排距越远所需泥泵功率越大。

4)绞刀功率越大,船舶挖掘能力和适应能力越强,生产效率也越高。配备大功率绞刀的船舶整体强度较大,相应的泥泵功率也较大。

1.2 施工过程主要影响因素

1)疏浚土质决定了施工船舶类型及施工工艺的选择。在实船疏浚过程中,疏浚土由某一或某几种基本土类型构成,其密度、强度及颗粒度会影响绞吸挖泥船的挖掘能力,从而影响泥浆的形成及输送。

表 1 特征与流量的相关系数

水下泵吸入真空	1#泥泵功率	右横移绞车速度	左绞车横移速度	横移速度	2#泥泵功率	舱内泵转速 1
0.663	0.649	0.646	0.643	0.646	0.630	0.613
舱内泵转速 2	绞刀转速	水下泵转速	舱内泵吸入压力 1	舱内泵吸入压力 2	桥架角度	绞刀深度
0.616	0.602	0.598	0.440	0.579	0.570	0.566

2.2 泥浆流量数据平滑处理

泥浆流量需要由流速与浓度计算获得,浓度由密度计测量得到,带有随机性和正负偏差。在船舶施工过程中,实际控制的参数为泥浆浓度,需要根据每段时间内采样点上的累计效应采取适当的控制行为。另外,数据平滑处理可以消除偶

2)泥泵性能对疏浚作业有很大的影响,可以通过控制泥泵功率调整泥浆体积占比。在排泥管长度和缩口大小不变的情况下增大泥泵转速,泥浆流量与施工产量随之增大。

3)绞吸挖泥船绞刀的挖掘切削能力直接影响施工生产效率,疏浚土被绞刀头切削得越多,泥泵吸入的泥浆体积比越高,施工产量就越大。

2 绞吸挖泥船施工数据预处理

本文以“天骥”船某次施工记录的 4.3 万条历史数据作为分析处理对象。首先对原始施工数据记录的 253 个特征参数进行施工影响因素分析,选取其中的 75 个特征。

由于施工数据量过于庞大且找出影响施工过程的主要因素就可以进行流量预测,因此需要对数据进行筛选。另外,原始施工数据并不直接记录泥浆流量,因此本文试验预测的泥浆流量需要由流速与浓度间接获得,即泥浆流量为流速与浓度的乘积。泥浆流量在整个施工过程中是缓慢变化的,相比于每个测量时刻的具体数值,更关心施工时泥浆流量的变化趋势,因此对这项数据进行平滑处理。

2.1 特征筛选

各个特征与绞吸挖泥船泥浆流量的相关系数通过计算协方差矩阵获得,在 MATLAB 软件中调用 cov 函数计算协方差矩阵。去掉强相关特征,并综合考虑影响绞吸挖泥船施工效率的因素,共筛选出 15 个特征用于预测流量,它们与流量的相关系数见表 1。

然因素的影响,提高预测的准确度。

2.2.1 中值滤波

MATLAB 信号工具箱中的 medfilt1 函数可以对信号数据进行一维中值滤波。泥浆流量数据一般为等时间间隔采集的离散信号数据,故可以用该函数进行平滑处理。

medfilt1 函数应用形式为: $y = \text{medfilt1}(x, n)$, 其中 x 、 y 分别为滤波前、后的数据。当窗宽参数 n 取 8 时, 泥浆流量数据经过中值滤波平滑前后的效果如图 1 所示。可以看出, 经过滤波处理的数据仍有细小的波动。

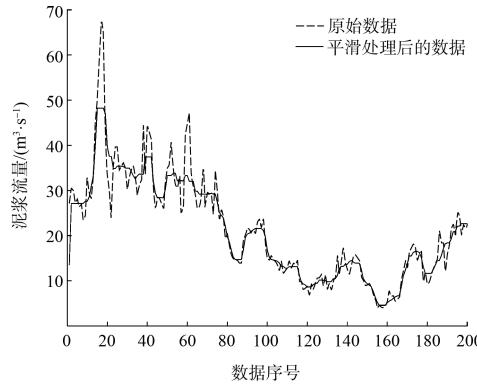


图 1 中值滤波平滑处理效果

2.2.2 移动平均值法

移动平均滤波是最常用的数据平滑方法之一, 它是基于统计规律, 在实时测量时, 将连续采样数据记录为长度固定为 N 的数组, 在下一次测量之后去掉首数据, 数据依次前移, 新的数据置于最后, 然后再次计算, 将结果作为测量结果。对移动平均滤波的首尾数据计算方法进行改进, 经过多次试验, 选取 5 点平均值, 4 次平滑, 处理效果较好。泥浆流量数据使用移动平均值法进行平滑前后的效果如图 2 所示。

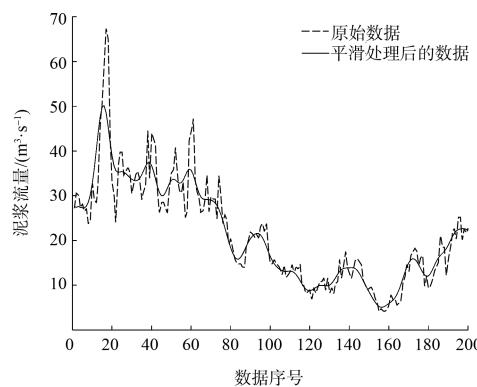


图 2 移动平均值法平滑处理效果

由以上 2 种数据平滑方法处理结果对比可得, 移动平均值法平滑效果优于中值滤波, 因此选用移动平均值法平滑数据。

2.3 试验数据的产生

将原始数据进行特征筛选, 得到包含泥浆流量在内的 16 维数据, 剔除泥浆流量中的无意义数据, 共有 2.6797 万条历史记录。为了除去不同量纲对数据的影响, 将前面 15 维数据进行标准化, 然后在 MATLAB 中调用 std 函数, 调整数据期望为 0, 方差为 1。将泥浆流量的数据进行平滑处理, 替代原始流量数据, 得到新的数据集, 即试验所用的数据集。

3 绞吸挖泥船泥浆流量预测方案设计

3.1 基于 T-S 模型的流量预测

绞吸挖泥船的施工数据记录量十分庞大, K -means 算法具有良好的时间复杂度, 能够高效地处理大数据集, 此次试验使用传统的 K -means 算法对疏浚历史数据进行聚类分析, 并用戴维森堡丁(DBI) 指数确定 K -means 聚类的参数 K 。随后, 根据聚类结果在类内运用线性回归方法, 每一类得到一个回归方程, 作为 T-S 模型每条规则的后件, 用于预测流量。

3.1.1 K -means 聚类算法

K -means 算法是一种先给定分类簇数 K 的迭代求解聚类算法, 该算法将原始数据划分为 K 类, 使类间距离尽可能大, 类内距离尽可能小。算法流程为: 首先给定样本集 $X = \{x_1, x_2, \dots, x_m\}$ 、聚类的簇数 K , 最大迭代次数 N 。步骤 1: 从数据集 X 中随机选择 k 个样本作为初始的 k 个质心向量 $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$; 步骤 2: 将簇划分 C 初始化为 $C_t = \emptyset, t = 1, 2, \dots, k$; 步骤 3: 对于 $i = 1, 2, \dots, m$, 计算样本点 x_i 和各个质心向量 $\mathbf{u}_j (j = 1, 2, \dots, k)$ 的距离, $d_{ij} = \|x_i - \mathbf{u}_j\|_2^2$, 将 x_i 标记最小的为 d_{ij} 所对应的类别 λ_i 。此时更新 $\{C'_{\lambda_i}\} = \{C_{\lambda_i}\} \cup \{X_i\}$; 步骤 4: 对于 $j = 1, 2, \dots, k$, 对 $\{C_j\}$ 中所有的样本点重新计算新的质心 $\mathbf{u}_j = \frac{1}{C_j} \sum_{x \in |C_j|}$; 步骤 5: 计算簇内方差

$E = \sum_{k=1}^m \sum_{x \in |C_k|} |x - \mathbf{u}_k|^2$; 步骤 6: 如果方差 E 收敛, 则输出簇划分 $C = \{C_1, C_2, \dots, C_k\}$; 否则转至步骤 2。

3.1.2 DBI 指数

DBI 指数又称为分类适确定性指标, 由 Davies 和 Bouldin 提出, 可以作为聚类算法优劣的评估指标。设有 m 个时间序列, 这些时间序列一共聚类成 n 个簇。 m 个时间序列组成输入矩阵 \mathbf{X} , n 个簇类设为 N 作为参数传入算法。使用下面的公式表示每个簇类最大相似度的平均值:

$$DBI = \frac{1}{N} \sum_{i=1}^N \max \left(\frac{\bar{S}_i + \bar{S}_j}{\|\omega_i - \omega_j\|_2} \right) \quad (1)$$

式中: N 为聚类的簇数; ω_i 、 ω_j 为第 i 、 j 类的质心位置; \bar{S}_i 、 \bar{S}_j 为类内数据到簇质心的平均距离, 能够表示簇类 i 、 j 中各时间序列的分散程度。

3.1.3 T-S 模型

针对多维模糊推理中往往含有许多推理规则的问题, Takagi 和 Sugeno 于 1985 年正式提出一种新模糊推理模型, 被命名为 T-S 模型, 其是将系统输出映射成输入的函数。对于一组输入 (x_1, x_2, \dots, x_p) (其中 $x_1 \in X_1, x_2 \in X_2, \dots, x_p \in X_p$), 经过模糊推理, 并采用重心法精确计算后得到的输出 u' 为:

$$u' = \frac{\sum_{j=1}^n \omega_j f_j(x_1, x_2, \dots, x_p)}{\sum_{j=1}^n \omega_j} \quad (2)$$

式中: ω_j 为输入变量对第 j 条规则的匹配程度, 可使用 Max-Min 推理方法或者 Sum-Product 等推理方法计算。

3.1.4 多元线性回归

本研究对象的数据包括多个变量, 采用 MATLAB 多元线性回归计算中的逐步回归法。逐步回归法是将自变量逐一引入, 并计算对输出的影响, 在引入新的自变量的同时, 对原有自变量进行逐一比对, 将影响小的自变量剔除。当回归方程不能再引入新的自变量, 也不能再剔除自变量时, 此时的回归方程则是最优的。

设对应随机变量 y 的自变量为 x_1, x_2, \dots, x_p , 则其多元线性回归模型为:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_p x_p \quad (3)$$

式中: $a_0, a_1, a_2, \dots, a_p$ 为回归模型估计系数。在 MATLAB 中调用 regress 函数可实现多元线性回归。

3.1.5 实现过程及结果

1) 将处理好的数据集中 2.679 7 万条数据按 4:1 的比例随机分为训练集与测试集, 训练集的前 15 维数据作为聚类对象。对于训练集 2.143 7 万条输入数据, 用 DBI 指标初步确定的最佳聚类簇数为 120 类。

2) 在 MATLAB 环境下编程, 将数据用 K -means 法划分为 120 类, 设置最大迭代次数为 100, 此时算法已经能很好地收敛; 设置聚类中心初始化次数为 5, 由于 K -means 算法对初始聚类中心的选取特别敏感, 应多选取几次聚类中心, 并将其中较好的作为最终结果。

3) 将聚类结果作为 T-S 模型规则建立的依据。实际试验发现预测采用全部规则 ($N=120$) 时, 平均相对误差超过 1, 考虑到数据分布密集的影响, 减小 N 值可以减少误差, 一般当 N 取 1 时, 平均相对误差最小。

由于 K -means 初始质心具有随机性, 重复试验 5 次, 平均相对误差均在 0.35 左右, 5 次程序运行结果的平均相对误差见表 2。

表 2 基于 T-S 模型的流量预测平均相对误差

试验次数	1	2	3	4	5
平均相对误差	0.327 3	0.336 3	0.355 4	0.348 9	0.337 1

测试集 5 360 条数据, 对每条数据进行流量预测, 其中前 100 条数据的预测值与实际值对比如图 3 所示。

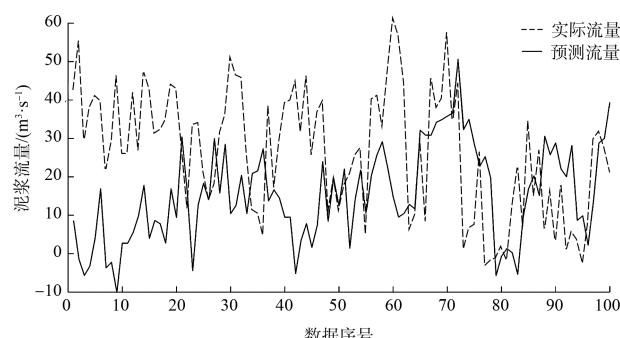


图 3 基于 T-S 模型的前 100 条数据流量预测值与实际值对比

3.2 基于历史数据最近邻的流量预测

采用数据库模式匹配的方法, 对于输入能够找到类似的数据用于预测。预测过程不需要建立

输入与输出的函数关系, 新的数据可以成为历史数据加入到数据库中。

绞吸挖泥船大量的历史施工数据记录可以保证样本的完备性, 充分体现不同工作状态下输入与输出的对应关系。为了更直观地反映泥浆流量变化趋势, 该方案仍使用经过平滑处理的泥浆流量数据。

3.2.1 实现过程

将疏浚施工历史数据建立数据库并用于预测流量。训练过程首先通过寻找每条数据的近邻来计算信任度, 根据信任度大小分配权重。其中数据信任度是指单位输入、输出的变化。由训练过程得到数据库, 测试集进行流量预测, 由测试误差判断方法可行性。

3.2.2 实现结果

测试集 5 360 条数据, 对每条数据进行流量预测, 前 100 条数据的预测值与实际值对比如图 4 所示。

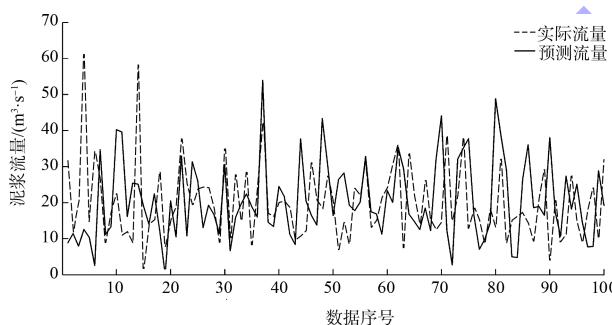


图 4 基于历史数据最近邻的前 100 条数据流量预测值与实际值对比

3.3 方案对比

方案 1 为采用传统 *K*-means 算法对疏浚数据进行聚类, 以聚类结果生成 T-S 模型的规则, 最终用于预测; 方案 2 为对每条数据找到最近邻计算数据信任度作为权重, 预测结果取输入数据的 5 个最近邻的对应输出的加权平均。

方案 1 的 *K*-means 算法优于初始质心选取是随机的, 具有很大的不确定性, 相比于方案 2 平均相对误差也较大, 但是多元线性回归用于预测效果较好; 方案 2 方法简单, 但是较依赖于数据质量, 如果存在太多异常数据, 会严重影响预测效果。由于本文对试验数据做了预处理, 选择方

案 2 用于泥浆流量预测获得的准确率较高。

4 基于预测偏差反馈的作业参数调节

绞吸挖泥船在特定工况下, 较为满意的输出范围可以被认为是最优输出, 这些输出数据与其对应输入数据组成最优集。对于给定输入, 需要在最优集中找到它的最近邻, 将它的最近邻输出与输入的预测输出值进行比较, 将差值作为预测偏差。参数调节的幅度需要预测偏差确定。

反馈环节作为闭环系统的关键部分, 在实际工程应用中可以起到提高系统自动化程度、增强系统鲁棒性的作用。可以选用偏差作为反馈量设计控制器, 并进行参数调节。设计思路为: 在训练集建立好的数据库中, 按输出降序排列, 取前 15% 作为最优集 Z , 泥浆流量值 (m^3/s) 的范围 ($\min Z, \max Z$) = (30.0, 73.0)。

为了获得最佳作业参数, 如果预测值在最优输出范围内, 则不调整输入, 并认为它是当前的最佳作业参数; 如果预测值不在最优输出范围内, 则根据预测误差获得反馈系数进行输入参数的调整, 直到预测输出达到最优。

4.1 基于数据的参数调节

从测试集中每次随机选择 50 个输入数据, 进行 5 次试验, 平均调节时间见表 3。

表 3 5 次试验的平均调节时间

试验次数	1	2	3	4	5
平均调节时间/s	11.08	10.85	12.17	12.09	12.31

第 5 次试验的参数调节前后泥浆流量值对比如图 5 所示。

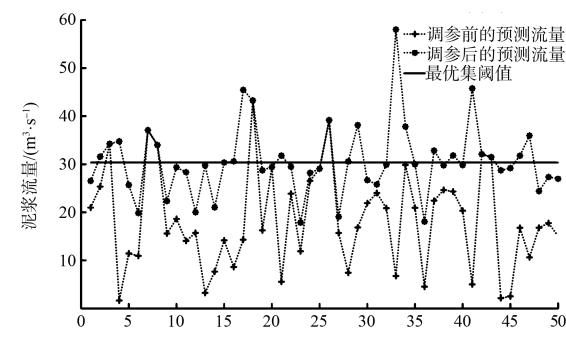


图 5 参数调节前后泥浆流量预测值与实际值对比

4.2 反馈控制流量走势

选取 1 500 个原始流量数据，并将流量记成浓度与流速的乘积，经过最近邻负反馈调整之后的曲线如图 6 所示。

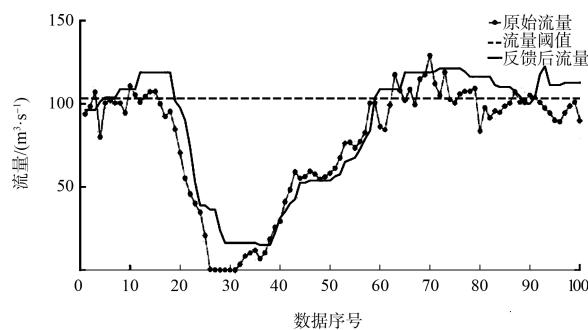


图 6 反馈后前 100 条数据的流量与原始流量对比

输入前 1 000 个流量数据，最大值为 $216.86 \text{ m}^3/\text{s}$ ，平均值为 $78.78 \text{ m}^3/\text{s}$ ，最小值为 $0 \text{ m}^3/\text{s}$ ；前 35% 设为高流量值，流量阈值为 $105.97 \text{ m}^3/\text{s}$ ；控制延时 10 s ；流量提升总量为 2.74 万 m^3/s ；调整次数为 644 次；平均每次提升 $42.56 \text{ m}^3/\text{s}$ 。

依照上述数据，疏浚产量总体提升 15% 左右。

5 结语

1) 经过对绞吸挖泥船输送过程中影响泥浆流量的主要特征因素进行分析，设计出两种不同预

测方案进行泥浆产量的预测，筛选出准确率更高的预测方案。

- 2) 建立对应于产量最优的施工工况的最优参数集合，随机从测试集中抽取数据模拟参数寻优，结果证明，应用该方法能够找到更优的参数。
- 3) 通过流量的预测偏差设计反馈环节实现参数的调节，达到基于大数据的参数自主寻优的目的。

参考文献：

- [1] 窦建强.“电天牛”系列绞吸船挖岩施工工艺[J].水运工程, 2019(6):212-214.
- [2] 朱师伦, 高岗, 徐合力, 等. 基于特征模型的疏浚过程中泥浆浓度控制系统设计[J]. 中国航海, 2020, 43(2): 74-79.
- [3] 李明超, 孔锐, 秦亮, 等. 基于施工循环的绞吸挖泥船疏浚作业平稳度分析[J]. 水利水运工程学报, 2018(6): 98-104.
- [4] 朱逸峰, 苏贞, 叶树霞. 基于 Hadoop 技术的挖泥船大数据平台架构研究[J]. 电子设计工程, 2021, 29(1): 50-54, 61.
- [5] 刘骏, 殷晓明, 顾幸生. 一种改进的 T-S 模糊模型建模及优化方法[J]. 华东理工大学学报(自然科学版). 2016, 42(2): 233-239.

(本文编辑 王璁)

· 消息 ·

四航局参与的联合体中标四川亭子口灌区工程

近日，四航局参与的联合体中标四川省亭子口灌区一期工程设计-采购-施工(EPC)第Ⅳ标段。中标金额约 27.15 亿元，工期 59 个月。该工程是国务院确定的“172”重大水利项目，是四川省水资源配置总体布局“五横六纵”的重要骨干工程。项目涉及广元、南充、广安、达州 4 市 13 个县(区)的农业灌溉和城乡生活及工业供水，建设内容包括渠系工程、囤蓄水库工程、提灌工程，以及泵站、控制建筑物、排水建筑物等配套工程。项目建成后，将彻底扭转灌区水资源分布不均的状况，解决沿线乡镇缺水问题。

https://www.ccccltd.cn/news/jcxw/jx/202207/t20220707_173516.html (2022-07-07)