

DOI:10.19322/j.cnki.issn.1006-4710.2018.03.010

基于 Spark 的水库群多目标调度粒子群并行化算法

马川惠¹, 李 璘¹, 黄 强¹, 李 凤²

(1. 西安理工大学 省部共建西北旱区生态水利国家重点实验室, 陕西 西安 710048;

2. 西安理工大学 计算机科学与工程学院, 陕西 西安 710048)

摘要: 随着社会经济的发展,在江河修建的水库数量和规模越来越大,水库群优化调度所需资料量庞大、数据关系复杂,应用常规技术进行优化调度存在一定的缺陷,如计算速度慢、存在“维数灾”等。基于此,本文的研究旨在弥补智能算法的缺陷,提高求解模型的速度。本文以黑河流域为例,应用基于 Spark 框架的水库群多目标调度粒子群并行化算法,并使用 Scala 语言开发了水库群多目标优化调度软件系统,大大提升了计算效率。研究对水库群多目标优化调度的并行编程发展与应用也有很好的现实意义与应用价值。

关键词: 水库群多目标优化调度; Spark; 计算速度; 粒子群算法; 并行化

中图分类号: TV 697.11 **文献标志码:** A **文章编号:** 1006-4710(2018)03-0309-05

Parallel particle swarm optimization algorithm based on Spark multi-objective optimal scheduling of reservoir group

MA Chuanhui¹, LI Ying¹, HUANG Qiang¹, LI Feng²

(1. State Key Laboratory of Eco-hydraulics in Northwest Arid Region of China,

Xi'an University of Technology, Xi'an 710048, China; 2. School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China)

Abstract: With the development of social economy, the number and scale of reservoirs built in rivers are getting ever greater. The optimal scheduling for reservoir group has a large amount of data with the data relationship being complex. The application of conventional technology to optimize scheduling calculation is inefficient, and there is “dimension disaster”, etc. The purpose of this paper is to make up for the shortcomings of the intelligent algorithm and improve the speed of solving the model. Taking the Heihe River basin as an example, multi-objective optimal scheduling of reservoir group for parallel particle swarm optimization algorithm based on Spark is used. The Scala language is used to develop the multi-objective optimal scheduling system of the reservoir group, with the high-speed operation realized. It is of great practical significance and application value for the parallel programming development and application of multi-objective optimization scheduling of reservoir group.

Key words: multi-objective optimal scheduling for reservoir group; Spark; calculation speed; particle swarm optimization; parallelization

传统的水库多目标调度模型求解是以单目标优化模型为基础,多采用目标约束法、主次目标法、目标加权法、理想点法等为主要求解方法。1985年, Yeh^[1]在水库多目标分析过程中,采用 ϵ -约束法得到了兼顾发电和供水两个目标的折中解。1987年,董子敖和阎建生^[2]提出了一种把多水库多目标的复杂优化问题经过多次分层后,转化为相对简单的子问题来求解的多目标多层次优化方法,为解决大规

模水库群优化调度的“维数灾”难题提供了新途径。1997年,黄强和沈晋^[3]提出了水库群多目标调度模型的求解方法,对解决复杂水库群联合优化调度问题具有实际参考价值。上述方法其本质上还是单目标优化方法,不足之处是计算工作量大、计算速度慢、存在“维数灾”等,不利于工程实际应用。

目前,智能优化算法在水库群多目标调度模型求解中正处于快速发展阶段,如遗传算法、蚁群算

收稿日期: 2018-05-21

基金项目: 国家重点研发计划资助项目(2017YFC0405901-3);国家自然科学基金资助项目(51879213)

作者简介: 马川惠,女,硕士生,研究方向为水资源调度。E-mail:2669174313@qq.com

通讯作者: 黄强,男,教授,博导,博士,研究方向为水库优化调度。E-mail:syhuangqiang@163.com

法、粒子群算法等已经取得了显著成果,具有并行计算特性,对解决复杂水库群多目标调度模型具有很大优势。1985年,Schaffer^[4]提出第一个设计多目标进化算法——向量评价遗传算法。2000年,由Deb等^[5]对非支配排序遗传算法(NSGA)进行了针对性改进,提出了基于快速分类的非支配排序遗传算法(NSGA-II)。2006年,Kim等^[6]针对Han River流域的4座水库2个目标调度问题,利用NSGA-II算法进行求解,获得了令人满意的效果。2007年,Reddy和Kumar^[7]提出一种用于求解水库群多目标调度问题的多目标差分进化算法,取得了比NSGA-II更好的优化效果。2007年,Li Chen等^[8]提出了一种用于确定以用水和发电为目标的综合利用水库调度图的基于宏进化的多目标遗传算法。2016年,杨延伟^[9]针对基于实数编码的遗传算法(RAGA)在实际水库调度应用中易于早熟的缺点,提出一种改进的RAGA。2017年,张翔宇等^[10]针对NSGA-II算法在进化过程中存在的不足,利用拟随机Halton序列及自适应调整的方法加以改进。

2006年,马细霞等^[11]以某综合利用水库优化调度为实例进行研究,表明基于粒子群优化算法的调度模型适用于年内水库优化调度规则的确定。2007年,杨俊杰等^[12]提出一种新的基于自适应网格技术的多目标粒子群优化算法,为求解大规模多目标优化问题提供了一种有效手段。2008年,Alexandre和Darrell^[13]建立了基于多目标粒子群算法的水库群优化调度模型,并指出该模型的缺点,如对算法参数的选择较为敏感等。2017年,张忠波等^[14]把下山搜索策略引入到粒子群智能算法中,函数测试证明该方法改进了算法的鲁棒性,提高了算法求解效率。

但是,随着水库群优化调度的规模和数据急剧增大,各智能算法存在搜索时间长、易陷入局部最优解等问题。为了提高智能算法的运行效率,本文建立了水库群多目标优化模型,提出了基于Spark框架的粒子群并行化算法,旨在弥补智能算法的缺陷,提高求解模型的速度,对更好地解决水库群多目标调度问题具有重要的理论意义和应用价值。

1 水库群多目标优化模型建立及求解

1.1 模型建立

本文以发源于祁连山的黑河为研究对象,黑河流域水资源短缺,综合利用矛盾突出,为了充分开发利用水资源,已修建了众多水库,如何科学、合理调度水库群,发挥水资源综合利用效益,已成为亟待解决的问题。黑河流域水库群调度具有多目标的特

点,既要兼顾上游发电目标、又要满足中游农业灌溉目标,同时还要考虑下游生态用水目标。因此,本文以灌溉缺水最小、生态缺水最小、发电量最大为目标,建立目标函数:

$$\min f(f_1, f_2, -f_3) \quad (1)$$

式中, f_1 、 f_2 、 f_3 分别为灌溉缺水、生态缺水、发电量。

目标1 灌溉缺水最小

$$\min f_1 = \sum_{t=1}^T \sum_{i=1}^I \alpha(t, i) [QW(t, i) - QS(t, i)] \quad (2)$$

$$\alpha(t, i) = \begin{cases} 0 & QW(t, i) \leq QS(t, i) \\ 1 & QW(t, i) > QS(t, i) \end{cases} \quad (3)$$

式中, T 为总时段数; I 为总取水口数; $\alpha(t, i)$ 为第 t 个时段第 i 个取水口的灌溉系数; $QW(t, i)$ 和 $QS(t, i)$ 分别为第 t 个时段第 i 个取水口的灌溉需水量和供水量。

目标2 生态缺水最小

$$\min f_2 = \sum_{t=1}^T \beta(t) [QD(t) - QP(t)] \quad (4)$$

$$\beta(t) = \begin{cases} 0 & QD(t) \leq QP(t) \\ 1 & QD(t) > QP(t) \end{cases} \quad (5)$$

式中, $\beta(t)$ 为第 t 时段的生态系数; $QD(t)$ 和 $QP(t)$ 分别为第 t 时段的生态需水量和供水量。

目标3 发电量最大

$$\max f_3 = \sum_{t=1}^T \sum_{m=1}^M N(t, m) \Delta t \quad (6)$$

$$N(t, m) = k(m) Q(t, m) H(t, m) \quad (7)$$

式中, M 为电站个数; $N(t, m)$ 为第 t 个时段第 m 电站的出力; Δt 为最小计算时段; $k(m)$ 为第 m 电站出力系数; $Q(t, m)$ 为第 t 个时段第 m 电站的发电流量; $H(t, m)$ 为第 t 个时段第 m 站的发电水头。

上述目标函数应满足的约束条件为:

约束1 水量平衡约束

$$V(t+1, m) = V(t, m) + (QO(t, m) - QI(t, m)) \Delta t \quad (8)$$

约束2 流量连续约束

$$QI(t, m+1) = QO(t, m) + q(t, m+1) \quad (9)$$

约束3 蓄水位约束

$$Z_{\min}(t, m) \leq Z(t, m) \leq Z_{\max}(t, m) \quad (10)$$

约束4 库容约束

$$V_{\min}(t, m) \leq V(t, m) \leq V_{\max}(t, m) \quad (11)$$

约束5 下泄流量约束

$$QO_{\min}(t, m) \leq QO(t, m) \leq QO_{\max}(t, m) \quad (12)$$

约束6 出力约束

$$N_{\min}(m) \leq N(t,m) \leq N_{\max}(m) \quad (13)$$

约束 7 边界条件约束

$$Z(1,m) = Z_c, Z(T+1,m) = Z_e \quad (14)$$

约束 8 控制断面流量约束

$$QS(t,i) \geq QW(t,i), QP(t) \geq QD(t) \quad (15)$$

式中, $V(t+1,m)$ 、 $V(t,m)$ 分别为第 $t+1$ 时段初(第 t 时段末)、第 t 时段初第 m 电站的库容; $QO(t,m)$ 、 $QI(t,m)$ 分别为第 t 时段第 m 电站的出库流量、入库流量; $q(t,m+1)$ 为第 t 时段第 m 电站与第 $m+1$ 电站之间的区间入流; $Z_{\min}(t,m)$ 、 $Z_{\max}(t,m)$ 分别为第 t 时段第 m 电站的最低、最高水位; $Z(t,m)$ 为第 t 时段第 m 电站的水位; $V_{\min}(t,m)$ 、 $V_{\max}(t,m)$ 分别为第 t 时段第 m 电站的最小、最大库容; $V(t,m)$ 为第 t 时段第 m 电站的库容; $QO_{\min}(t,m)$ 、 $QO_{\max}(t,m)$ 分别为第 t 时段第 m 电站的最小、最大出库流量; $QO(t,m)$ 为第 t 时段第 m 电站的出库流量; $N_{\min}(m)$ 、 $N_{\max}(m)$ 分别为第 m 电站的最小出力、装机容量; Z_c 、 Z_e 分别为起调水位和调度期末蓄水位。

式(1)~(15)是本文建立的水库群多目标优化调度模型,该模型具有多目标、多维、非线性、动态等特点,模型输入资料涉及社会、经济、水文、气象、生态、水利工程、水资源综合利用等方面,数据量庞大、关系复杂。因此,寻求求解模型的算法、处理庞大的数据量与保证计算速度成为难点。本文采用基于 Spark 框架的粒子群并行化算法求解该模型,旨在克服“维数灾”,提高计算速度。

1.2 模型求解

水库群多目标优化调度模型的求解可用主次目标、目标加权、理想点等方法,把多目标模型转化为单目标模型,对单目标优化模型采用动态规划、遗传算法、蚁群算法等进行求解。但这些方法计算工作量大、速度慢,存在“维数灾”。本文利用粒子群算法(PSO)对多目标优化模型进行整体一次求解。该算法是从随机解出发,通过迭代寻找最优解,它通过适应度来评价解的品质,通过追随当前搜索到的最优值来寻找全局最优。粒子群算法是一种成熟的算法,其具体步骤见文献[15]。

2 基于 Spark 平台的粒子群算法并行化

2.1 Spark 平台

Spark 支持的大规模并行分布处理算法实现了基于内存的分布式计算,基于弹性分布式数据集(RDD)成功构建起了一体化、多元化的大数据处理体系。使用 Spark SQL、Spark Streaming、MLlib、

GraphX 等工具,可以解决大数据批处理、流处理、图处理、机器学习、即席查询与关系查询等关键技术问题,只需一个框架就可满足各种使用场景的需求。2003 年以来,Spark 逐渐获得了学术界与工业界的认可,成为大数据处理技术的研究热点、新一代大数据处理平台的首选。RDD 是 Spark 并行处理的基础,所有 Spark 的操作都围绕 RDD 进行。应用 RDD 可以进行并行数据集之间的互相转换、缓存等操作。基于 Spark 进行并行化编程时,要将输入数据分解成一个个批处理片段,然后再将这些数据段都转换成 Spark 中的 RDD,即将数据封装在 RDD 中,通过进行 RDD 的并行操作,实现数据处理的并行执行。因此,基于 Spark 平台并行计算模型编程实现粒子群算法的基本步骤为:把粒子群封装为 RDD,并将其初始化为大小规模均等的多个小种群;再对这些小种群粒子群进行并行处理,构建可行解^[16-20]。

2.2 基于 Scala 语言的粒子群并行编程方法

Scala 语言是具有面向对象和函数式编程特性的多范式编程语言,是 Spark 主要支持的高级计算机语言,面向分布式并程序序设计;Spark 中大多数代码都是通过 Scala 语言编写的。

本文运用 Scala 并行编程方法实现了并行 PSO 算法在水库群优化调度中的应用,具体流程图见图 1。

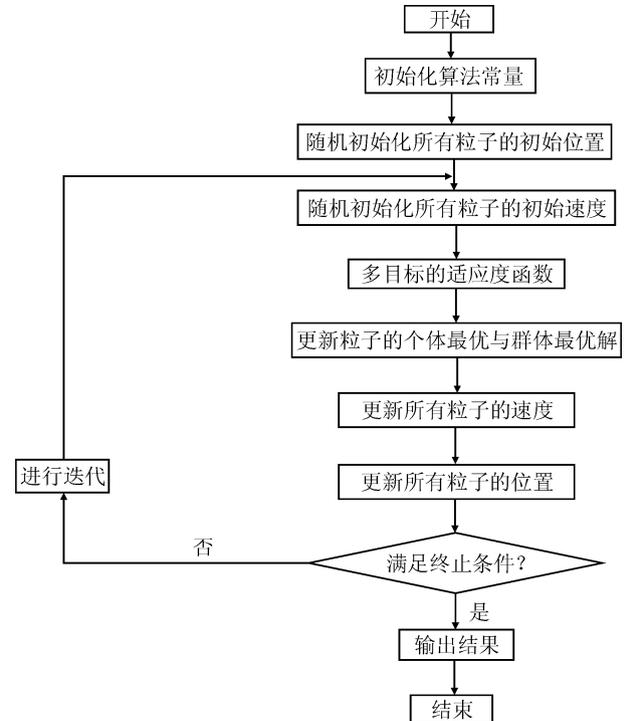


图 1 并行 PSO 算法流程图

Fig. 1 Parallel PSO algorithm process

3 应用实例

3.1 黑河水库群多目标调度实例

黑河是我国西部地区较大的一条内陆河,发源于祁连山与大通山之间,流经青海、甘肃、内蒙三省区,流域面积约为 14.29 万 km^2 ,干流河长 821 $\text{km}^{[21]}$ 。莺落峡以上为干流上游,地处山区,分东西两岔,河道比降大,水量较为充沛,不仅是黑河流域的主要产流区,也是该流域的梯级水电开发基地。莺落峡至正义峡之间为中游,地势相对平坦,光热资源丰富,沿岸分布多个灌区,是该流域的主要用水区。正义峡以下为下游,降水稀少,蒸发强烈,生态系统脆弱,是该流域的主要生态耗水区。黑河干流上游主要有 8 座梯级水电站水库,分别是黄藏寺(在建)、宝瓶河、三道湾、二龙山、大孤山、小孤山、龙首二级和龙首一级。本研究采用资料包括气象、水文、水利枢纽、河渠机井、中游灌区、生态需水、分水方案及相关社会经济等方面的海量数据。

采用基于 Spark 并行计算框架的粒子群算法,求解水库群多目标优化调度模型,获得了较满意的结果:中游灌区灌溉保证率为 50.5%,居民生活和工业生产供水保证率为 95.1%;狼心山断面多年平均下泄生态水量满足多年平均应下泄水量的要求;黑河上游 8 座梯级水电站多年平均发电总量为 21.85 亿 $\text{kW}\cdot\text{h}$,比设计值减少 3.8%。说明黑河流域水库群多目标调度牺牲了发电效益,满足了供水和生态效益,符合“以水定电”的调度策略和黑河流域的实际状况。

3.2 计算性能比较

系统性能测试在 64 位单机 Windows 7 操作系统、8 GB 内存的环境下运行。为了对比分析,本文分别采用基于串行和并行化的粒子群算法求解水库群多目标优化模型,且逐步增加了输入的数据条数来测试这个程序的性能,测试结果如图 2 所示。其中,串行粒子群算法应用 Visual Basic 语言编程,并行化粒子群算法用 Scala 语言编程。

数据条数从 2 052 增加至 20 520,增至原来的 10 倍,在单机条件下,串行运行时间从 26.90 min 增加至 364.60 min,并行运行时间从 3.08 min 增加至 53.20 min,并行比串行运行速度有了很大提升,提升速度约 6~8 倍。

为了测试单机与多机的运行速度,本文分别在 Ubuntu Linux 操作系统环境下,采用单机和 10 台联想服务器并行计算,且逐步增加程序运行的数据条数(从 2 052 增加至 20 520),得到了更加理想的

结果,如图 3 所示。

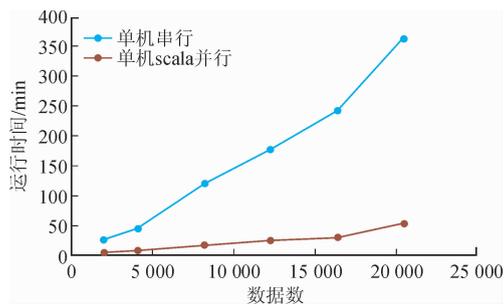


图2 单机串行与单机并行程序运行时间比较图
Fig. 2 Run time comparison between single machine serial and single machine parallel program

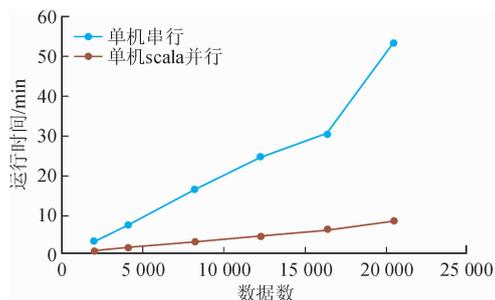


图3 单机与多机并行程序运行时间比较图
Fig. 3 Run time comparison between single machine and multi-machine parallel program

与上一步所得的单机并行运行时间进行比较,多机并行运行时间在数据条数增加的过程中,从 0.80 min 增加至 8.27 min,多机并行比单机并行计算速度有了大幅提升,约提升了 4~7 倍,且随着输入数据条数的增加,速度提升优势明显,体现出 Spark 平台处理大量数据的优势。

4 结语

随着社会经济的发展,在江河修建的水库数量和规模越来越大,水库群调度所需资料、数据量庞大,关系复杂。由于水库群之间的水力联系更加复杂化和调度规模更加庞大,进一步加剧了水库调度问题求解的复杂性,应用常规技术进行优化调度计算难度较大、速度慢,且存在“维数灾”等。此外,水库运行积累的数据也呈现出指数增长的趋势,形成了海量数据源。因此,水库群多目标调度对海量数据处理有着强烈的需求;同时,需要并行计算技术为水库群调度及其优化计算提供强大的计算能力。

本文以黑河流域的水资源调度问题为背景,建立了水库群多目标优化模型,提出了基于 Spark 框架求解模型的粒子群并行化算法。实例表明,将粒子群算法与 Spark 并行计算框架结合使用,满足了当前水库群调度多目标的需求,提高了程序的运行

速度,说明基于 Spark 框架求解模型的粒子群并行化算法具有很好的性能。研究旨在弥补智能算法的缺陷,提高求解模型的速度,对水库群多目标优化调度的并行编程发展与应用具有重要的参考价值。

参考文献:

- [1] YEH W W. Reservoir management and operations models: a state-of-the-art review [J]. *Water Resources Research*, 1985, 21(12): 1797-1818.
- [2] 董子敖, 阎建生. 计入径流时间空间相关关系的梯级水库群优化调度的多层次法 [J]. *水电能源科学*, 1987, 5(1): 29-40.
DONG Ziao, YAN Jiansheng. Multilevel method of optimal operation of reservoirs in series taking into account correlations of river flows in time and space [J]. *International Journal Hydroelectric energy*, 1987, 5(1): 29-40.
- [3] 黄强, 沈晋. 水库联合调度的多目标多模型及分解协调算法 [J]. *系统工程理论与实践*, 1991, (1): 75-82.
HUANG Qiang, SHEN Jin. Multi-purpose and multi-model of joint reservoir operation as well as decomposition and coordination algorithm [J]. *Systems Engineering-Theory & Practice*, 1991, (1): 75-82.
- [4] SCHAFFER J D. Multiple objective optimization with vector evaluated genetic algorithms [C]// *International Conference on Genetic Algorithms*. L. Erlbaum Associates Inc., 1985: 93-100.
- [5] DEB K, PRATAP A, AGARWAL S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II [J]. *IEEE Transactions on Evolutionary Computation*, 2002, 6(2): 182-197.
- [6] KIM T, HEO J H. Application of multi-objective genetic algorithms to multireservoir system optimization in the Han River basin [J]. *Hydrological Processes*, 2010, 20(9): 2057-2075.
- [7] REDDY J M, KUMAR N D. Multiobjective differential evolution with application to reservoir system optimization [J]. *Journal of Computing in Civil Engineering*, 2007, 21(2): 136-146.
- [8] CHEN L, MCPHEE J, YEH W G. A diversified multiobjective GA for optimizing reservoir rule curves [J]. *Advances in Water Resources*, 2007, 30(5): 1082-1093.
- [9] 杨延伟. 一种改进的基于实数编码的遗传算法以及在水库调度中的应用 [J]. *中国水运*, 2017, 17(1): 165-166.
YANG Yanwei. An improved genetic algorithm based on real number coding and its application in reservoir scheduling [J]. *China Water Transport*, 2017, 17(1): 165-166.
- [10] 张翔宇, 董增川, 马红亮. 基于改进多目标遗传算法的小浪底水库优化调度研究 [J]. *水电能源科学*, 2017, 35(1): 65-68.
ZHANG Xiangyu, DONG Zengchuan, MA Hongliang. Study on optimization operation of Xiaolangdi Reservoir based on improved multi-objective genetic algorithm [J]. *Water Resources and Power*, 2017, 35(1): 65-68.
- [11] 马细霞, 储冬冬. 粒子群优化算法在水库调度中的应用分析 [J]. *郑州大学学报(工学版)*, 2006, 27(4): 121-124.
- MA Xixia, CHU Dongdong. Application analysis on reservoir operation by particle swarm optimization [J]. *Journal of Zhengzhou University (Engineering Science)*, 2006, 27(4): 121-124.
- [12] 杨俊杰, 周建中, 方仍存, 等. MOPSO 算法及其在水库优化调度中的应用 [J]. *计算机工程*, 2007, 33(18): 249-250, 264.
YANG Junjie, ZHOU Jianzhong, FANG Rengcun, et al. Multi-objective particle swarm optimization and its application in optimal regulation of reservoir [J]. *Computer Engineering*, 2007, 33(18): 249-250, 264.
- [13] ALEXANDRA M B, DARRELL G F. Use of multi-objective particle swarm optimization in water resources management [J]. *Journal of Water Resources Planning and Management*, 2008, 134(3): 257-265.
- [14] 张忠波, 何晓燕, 耿思敏, 等. 改进的粒子群算法在水库优化调度中应用 [J]. *中国水利水电科学研究院学报*, 2017, 15(5): 338-345.
ZHANG Zhongbo, HE Xiaoyan, GENG Simin, et al. The application of improved particle swarm algorithm to reservoir operation optimization [J]. *Journal of China Institute of Water Resources and Hydropower Research*, 2017, 15(5): 338-345.
- [15] 解阳阳. 基于径流预报的黑河流域水资源调配研究 [D]. 西安: 西安理工大学, 2017.
XIE Yangyang. Research on water resources allocation of Heihe River basin based on runoff prediction [D]. Xi'an: Xi'an University of Technology, 2017.
- [16] 王家林. 大数据 Spark 企业级实战 [M]. 北京: 电子工业出版社, 2015.
- [17] 夏俊鸾, 程浩, 邵赛赛, 等. Spark 大数据处理技术 [M]. 北京: 电子工业出版社, 2015.
- [18] 邱荣财. 基于 Spark 平台的 CURE 算法并行化设计与应用 [D]. 广州: 华南理工大学, 2014.
QIU Rongcai. The parallel design and application of the CURE algorithm based on Spark platform [D]. Guangzhou: South China University of Technology, 2014.
- [19] 李凤. 基于 Spark 的粒子群算法并行编程及其在水库调度中的应用 [D]. 西安: 西安理工大学, 2017.
LI Feng. Parallel programming of particle swarm optimization algorithm based on Spark and its application in reservoir scheduling [D]. Xi'an: Xi'an University of Technology, 2017.
- [20] 彭安帮, 彭勇, 周惠成. 跨流域调水条件下水库群联合调度图的多核并行计算研究 [J]. *水利学报*, 2014, 45(11): 1284-1292.
PENG Anbang, PENG Yong, ZHOU Huicheng. Multi-core parallel computation for deriving joint operating rule curves in multi-reservoir system under the condition of inter-basin water transfer [J]. *Journal of Hydraulic Engineering*, 2014, 45(11): 1284-1292.
- [21] 王渊. 黑河流域梯级水电站联合发电优化调度的研究 [J]. *科技传播*, 2013, (19): 96-97.
WANG Yuan. Study on optimal dispatching of cascade hydropower stations in Heihe River basin [J]. *Public Communication of Science & Technology*, 2013, (19): 96-97.