

引用格式:宋闯,韩伟,杜兴伟,等.基于Spark的电网继电保护整定计算方法研究[J].电力科学与技术学报,2025,40(2):50-58.

Citation: SONG Chuang, HAN Wei, DU Xingwei, et al. Relay protection setting calculation method for power grid based on Spark[J]. Journal of Electric Power Science and Technology, 2025, 40(2): 50-58.

基于Spark的电网继电保护整定计算方法研究

宋闯¹, 韩伟¹, 杜兴伟², 王敬军²

(1. 国网河南省电力公司电力科学研究院, 河南 郑州, 450052; 2. 国网河南省电力公司, 河南 郑州, 450052)

摘要:为使电网适应智能化以及调控云技术路线的要求,提出一种基于分布式并行计算的电网继电保护整定计算研究方法。首先,介绍Spark分布式计算平台的集群架构,并对分布式并行计算的关键问题,如负载的平衡问题、系统的容错问题等进行分析,基于此设计基于Spark的电网继电保护整定计算系统;其次,对计算系统中的超高压电网整定计算进行分析,并对保护的原则以及整定计算的原则进行总结;再次,为实现对输入系统的电网初始数据实现预处理操作,通过改进帝王蝶优化算法实现对数据特征选择;最后,对某地区具体实例的整定计算进行仿真分析,以此验证系统的有效性。仿真结果证明,该计算系统能够使电网的整定计算适应智能化以及调控云的发展,能够有效增加计算速度,提高电网运行的可靠性。

关键词:智能化电网;分布式并行计算;整定计算;帝王蝶优化算法

DOI: 10.19781/j.issn.1673-9140.2025.02.006 中图分类号: TM 文章编号: 1673-9140(2025)02-0050-09

Relay protection setting calculation method for power grid based on Spark

SONG Chuang¹, HAN Wei¹, DU Xingwei², WANG Jingjun²

(1. State Grid Henan Electric Power Research Institute, Zhengzhou 450052, China;

2. State Grid Henan Electric Power Company, Zhengzhou 450052, China)

Abstract: To adapt the grid to the requirements of intelligentization and the dispatching and control cloud technology route, this paper proposes a relay protection setting calculation method for power grid based on distributed parallel computing. First, the cluster architecture of the Spark distributed computing platform is introduced, and the key issues of distributed parallel computing, such as load balancing, system fault tolerance, etc. are analyzed. On this basis, a computing system for relay protection setting calculation based on Spark is designed. Secondly, the extra-high-voltage power grid setting calculation in the computing system is analyzed, and the principles of protection and setting calculation are summarized. Next, to realize the preprocessing operation of the initial power grid data input to the system, data feature selection is realized by improving the monarch butterfly optimization algorithm. Finally, simulation analysis was conducted on specific instances in a region to verify the effectiveness of the system. The simulation results prove that the computing system enables the setting calculation of the power grid to accommodate intelligentization and the development of dispatching and control cloud, which can effectively increase the calculation speed and improve the reliability of power grid operation.

Key words: intelligent grid; distributed parallel computing; setting calculation; monarch butterfly optimization algorithm

电网整定计算是保障电网安全稳定运行的核心技术,也是确保继电保护装置快速切除故障、应

对突发故障工况的基础^[1-4]。近年来,随着中国电力事业的快速发展,电力系统网络覆盖范围持续扩

收稿日期:2021-11-08;修回日期:2022-01-22

基金项目:国家电网有限公司总部管理科技项目(521702200017)

通信作者:宋闯(1991—),男,硕士,中级工程师,主要从事继电保护方面的研究;E-mail:kmdwth@163.com

大,省、地、县三级电网节点数量快速增长,导致整定计算工作量激增,传统集中式计算模式面临严峻挑战。因此,优化整定计算流程、提升计算效率已成为电力系统安全领域亟待解决的课题。

分布式并行计算技术通过整合网络中的计算资源,实现算力聚合与任务协同分配,以提升任务处理能力。研究表明,该技术在电力系统领域应用广泛。文献[5]基于分布式计算解决了智能电网负载激增引发的计算性能瓶颈,显著提升了实时数据处理能力;文献[6]提出故障电压分布式筛选算法,通过节点优先级排序实现故障精准定位;文献[7]构建的计量故障识别模型利用分布式架构将数据处理效率提升超过 40%;文献[8]通过分布式算法实现微电网功率信息同步,其计算结果与集中式算法误差小于 0.5%;文献[9]则解决了电网分层管理与数据隔离难题,构建了跨区域协同计算机制。

目前针对分布式计算在电网继电保护整定计算中的应用研究仍较缺乏,超高压电网领域的相关研究尤为不足。在少数应用案例中,分布式计算存在整定计算效率低、负载均衡不足及系统容错性不足等关键技术问题,现有解决方案难以满足电网运行要求。因此,本文提出了一种基于 Spark 分布式并行计算的电网整定计算方法,首先,利用 Spark 在处理大量数据拥有更优性能、更低的成本、更高的计算效率等特点,建立了一个基于 Spark 电网继电保护整定计算的云计算;其次,为提高分布式计算系统性能,通过对负载均衡以及分布式系统容错性等问题进行研究并采用更优解决方案进行处理;再次,为确保该分布式计算系统能够面向各等级电网,对超高压电网整定计算原则进行了分析研究;最后,通过仿真计算对该计算系统的有效性进行验证。计算结果表明,基于 Spark 电网继电保护整定计算系统的计算效率更优、准确率更高,且符合电网云计算要求。

1 基于 Spark 的分布式并行计算系统

1.1 Spark 主要架构

Spark 是 AMP 公司所研发的一款基于内存计算的分布式并行计算引擎,其工作原理为将海量的数据放入内存中进行计算,以此来减少磁盘的读写,增强在处理大规模数据时的性能。Spark 的计算速度比 Hadoop 中的 MapReduce 的计算速度要快数十倍甚至上百倍,并且具有更强的兼容性,操作

更加简单,也更加容易使用。

Spark 集群架构沿用了传统分布式算法的主从架构模式,在该架构模式中,主节点的功能:对集群中的每一个计算节点进行中央协调以及调度;对集群的工作状态进行实时维护;保证集群运行的正常稳定。从节点的功能:接收由主节点发送的工作命令;向主节点实时发送工作状态。Spark 的集群结构示意如图 1 所示。

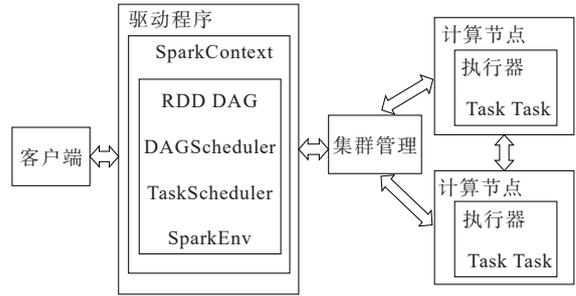


图 1 Spark 集群架构

Figure 1 Spark cluster architecture

图 1 中,集群管理主要功能为调度与管理集群中的计算资源;计算节点主要功能为执行管理下放命令中的相关任务;客户端作为与用户对接的端口,主要功能为将用户的各种需求提交到系统;驱动程序主要功能为调度 Spark 中的具体任务,其工作流程如图 2 所示。Spark 的具体工作流程如下:

- 1) 创建新的 SparkContext, 集群管理模块将计算资源分配给新创建的任务;
- 2) 驱动程序根据 DAG 逻辑关系将任务划分为不同阶段,并向执行器发送命令;
- 3) 执行线程收到命令后,开始执行任务,并将关键数据放入内存;
- 4) 驱动程序汇总结果,结束后开始下一阶段。

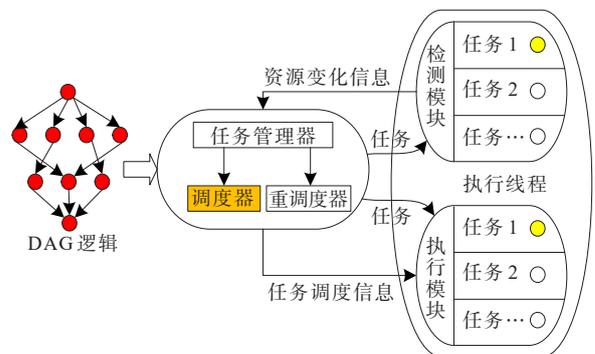


图 2 Spark 工作流程

Figure 2 Flow chart of Spark

在 Spark 工作流程中,最核心的部分为弹性分布式数据集(resilient distributed dataset, RDD)。通过使用不同的 RDD 对不同的数据集进行保存,再调用 RDD 的各类操作算子完成数据计算,实现分布式并行计算。

通过 RDD 数据结构,Spark 就能够在编程中通过操作本地集合来实现操作分布式集中数据,且 RDD 能够直接将计算结果缓存到内存中,在使下一个计算节点时能够直接从内存中读取计算结果,从而减少了数据读写中会占用大量磁盘的 IO 操作,降低了数据落地时间,提升了整体任务执行的计算效率;而且 RDD 自带的自动容错性能够充分保证计算结果的准确性,在保证计算效率的基础上,保证整个任务最终结果的准确。

1.2 关键技术问题

为了更好地发挥计算系统的性能,需要对系统中的负载平衡以及系统容错等问题进行研究。

1.2.1 分布式系统的容错性问题

分布式并行计算系统与传统单机系统有一个关键点不同,传统单机系统则会因为组件的故障而直接无法继续工作,进而系统直接崩溃^[10-11];而分布式并行计算系统在计算过程中会发生部分故障,即当系统内部一个组件发生故障时,分布式系统可能会受影响而不能正常进行计算或者可能完全不受影响。这个关键特性是分布式并行计算系统特别设计的,其目的是使系统部分组件发生故障时,系统能够从部分故障中自动恢复过来,并且不影响系统整体性能。因此,在故障发生的时候,计算系统应当能够在从故障中恢复的同时继续进行不影响结果的计算,这意味着巫妖系统能够容忍错误的发生,并且能够在错误产生的时候继续进行计算,且最终结果能够满足要求。

对于系统容错性,通常采用容错算法进行操作。容错算法主要分为检查点和回滚设置与模块复制 2 种方法。检查点和回滚设置容错算法又分为独立检查点以及一致性检查点 2 种模式,前者无法检查系统整体的一致性状态,后者则无法检查独立组件的状态;模块复制则可以分为主动复制及被动复制 2 种模式,无论是哪一种模式,都十分占用系统资源,不适用分布式并行计算系统。

因此,本文采用了将检查点和回滚设置与模块复制算法结合的分布式对象和检查点容错算法,该算法可以在客户端的通信连接发生失效时,允许客

户端重新能够调用请求到一个可连接的对象上,并且能够在客户程序中设置检查点,保证客户端发生崩溃或主动退出后再进入时仍能从最近保存的结果开始。

1.2.2 分布式系统的负载平衡问题

在 Spark 系统中,在对数据进行清洗时通常会出现数据存储容器产生数据倾斜而导致负载不平衡。为解决这类负载不平衡问题,通常需要运用到 Load Balance 技术。

Load Balance 的工作原理为^[12]:将负载较重的执行器转移一部分工作负载到负载较轻的执行器上,使得负载能够均匀分配给各个执行器,充分发挥每个执行器的计算性能,避免出现负载不均衡导致系统性能下降的问题。

处理负载平衡问题时,通常需要运用到分布式负载平衡算法。目前,分布式负载平衡算法最常使用的有:概率调度算法、梯度调度算法以及邻接契约算法^[13-16]。将这些算法按照调度的驱动对象划分,则可以分为发送方驱动、接收方驱动以及双方驱动。

发送方驱动的原理为各个计算节点之间的任务调度由创建任务的节点来进行操作,当任务被创建时,创建任务的节点就将任务分配给一个轻负载节点来执行;接收方驱动则与之相反,其工作原理为轻载节点执行任务时,会请求程序向相邻节点请求任务,以实现负载平衡;而双方驱动则是,当发送方与接收方都能够分配任务,系统负载轻时,发送方更容易进行驱动;系统负载较重时,接收方则更容易进行驱动。

本文为解决负载平衡问题采用了双向驱动以及一种新型的中间结果数据块处理算法:

在 Spark 集群框架中,在执行 Task 任务时,输入的初始数据会被加载进入文件系统中,每个文件中都包含多个大小相等的数据块,将其称之为输入分区。每一个数据块(cluster)都对应一个键(key)以及其对应的值(value)的集合,且只是一个子集:

$$C_{key} = \{(k, v) \in I\}, k \in K, v \in V \quad (1)$$

式中, $I \subseteq K \times V$ 为 Task 执行过程的中间结果; K 为 key 的集合; V 为 value 的集合。

将输入分区用分区函数进行表示:

$$\prod: K \rightarrow \{1, 2, \dots, n\} \quad (2)$$

式(2)表示在数据清洗过程中将 Task 的中间结果 I 分为 n 个互不相同的分区。由于输入分区包含数个

cluster,因此将分区定义为

$$P(j) = \bigcup_{k \in K; \prod(k)=j} C(k) \quad (3)$$

中间结果数据块处理算法根据分区的定义,对数据清洗中的分区划分进行重新规划,通过在数据清洗中将中间结果的数据进行重新划分,将数据块重组为 n 个大小相同的分区,达到分区均衡的效果,从而解决数据倾斜导致的负载不平衡问题。

1.2.3 基于 Spark 的整定计算系统

结合 Spark 集群构架及 Spark 关键技术问题,建立继电保护整定计算系统。对于 Spark 关键技术问题,负载均衡算法选择双向驱动,更加快速、有效地保证系统的负载能够分配均衡;而容错性则由 Spark 中的 RDD 数据结构保证,RDD 能够通过设置数据检查点或是将数据的更新记录下来实现对计算结果正确性的保证。

继电保护整定计算系统架构如图 3 所示。客户端为软件级服务层 (software-as-a-service, SaaS) 部分的客户端,可以通过网页浏览器或是通过安装智能客户端来使用系统内的应用服务;服务器层主要提供整定计算、数据管理、元数据以及实体逻辑等;云平台为系统核心,主要有 SaaS、平台级服务层 (platform-as-a-service, PaaS) 以及基础设施级服务层 (infrastructure-as-a-service, IaaS);物理基础设施主要由服务器,网络,存储器,电源以及冷却等物理设施构成,是整个计算系统的基础。

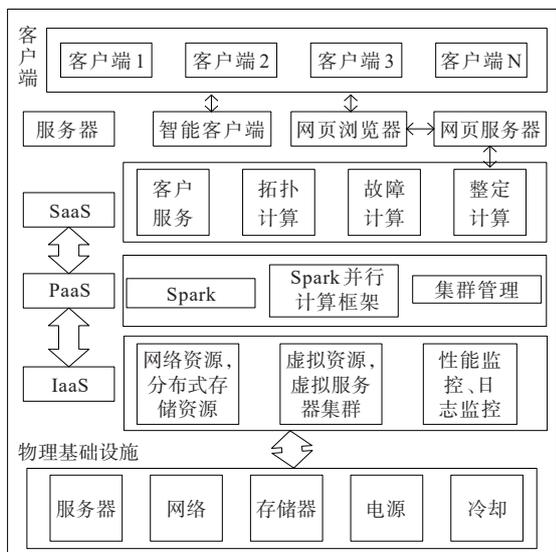


图 3 基于 Spark 的继电保护整点计算系统框架

Figure 3 Structure of computing system for relay protection setting calculation based on Spark

其中,IaaS 一方面主要是将网络资源、分布式存储资源、虚拟资源以及虚拟服务器集群聚集到一起进行统一的管理以及分配,另一方面,就是进行性能监控以及日志监控,保证计算系统运行状态的正常;PaaS 主要利用 Spark 平台下的并行计算框架,分配不同的 Task 来分别对数据进行处理,达成一种多线程计算模式,提升资源的利用率,加快计算效率;SaaS 主要提供客户服务以及各种计算服务,与用户对接,方便用户能够随时在网页或客户端提交需求、使用拓扑计算、故障计算以及整定计算等服务。

2 超高压继电保护整定计算

在电网的继电保护整定计算中,超高压电网与中低压电网的整定计算有着较大的区别。因此,需要对其继电保护整定计算分别进行分析,研究其保护以及整定计算原则。

2.1 超高压电网下的继电保护

对超压电网下的继电保护进行研究,对其区别于中低压电网继电保护的特点进行分析,包括:系统稳定、长距离线路的重负荷、线路中的分布电容以及串联补偿电容等。

1) 系统稳定。超高压电网由于传输功率比中低压线路的传输功率更大,稳定储备不高。因此,要求远/近故障点侧在切除故障时所耗费的时间不得超过 0.1 s/0.15 s,以此保证系统的稳定。为达到这个要求,继电保护的動作时间应在 20~40 ms。

2) 长距离线路的重负荷。在有重负荷的长距离线路上,有可能会出出现小于负荷电流的短路电流,导致继电保护的拒动,为保证系统的可靠性,要求继电保护的灵敏性要更高。

3) 线路中的分布电容。330 kV 及 330 kV 以上的超高压线路要求一定要考虑分布电容增大所带来的影响。

4) 串联补偿电容。一般要求串联补偿电容到超高压长线电路中,来保证系统运行的稳定性。因补偿电容对继电保护的影响较大,故补偿电容的补偿值通常要不超过线路总电抗的 50%,且安装位置在线路始/末端的变电站内。

对于超高压电网中的特殊问题进行分析解决后,需要对超高压电网的继电保护原则进行分析,使得在超高压电网继电保护的安全性和可靠性得到保证的前提下,电网继电保护整定计算系统能够对超高压电网下的继电保护进行整定计算。

首先,由于超高压电网的独特性,需要实现双重主保护,要求设置两套完善且相互独立的速动主保护以及后备保护,并且在有需要的情况下必须安装辅助保护;其次,以单相重合闸为线路重合闸主要部分,但是对于能够使用三相重合闸的部分,优先采用三相重合闸或综合重合闸;最后,需要对相邻的高压电压进行检查,保证其继电保护水平能够满足要求;否则,对其进行对应的改造,使其达到要求。

2.2 超高压继电保护整定计算原则

为保证超高压电网下整定计算结果的准确率,需要对其计算原则进行研究。在考虑计算原则之前,需要先对继电保护定值进行合理的选择,并且要保证定值与继电保护之间的配合协调,从而使得发生故障时能够快速、有效切除故障,以此维护超高压电网运行的安全稳定。在确定继电保护定值后,就需要对其整定计算原则进行深入考究,需要考虑的原则主要有2个:灵敏系数校验与短路计算。

1) 灵敏系数校验。首先,需要对异常运行方式中出现故障时的继电保护灵敏系数进行校验,其中,异常运行方式指一条输电线路或线路上的设备在检修时的运行方式;其次,当相邻元件发生故障导致保护拒动时,须按其他灵敏度足够的分支相跳闸时的条件对继电保护进行校验;最后,在工程设计中,对灵敏系数进行校验时,必须采用金属性短路作为条件进行校验。

2) 短路计算。忽略如发电机、变压器、电缆线路等阻抗的电阻部分,且当旋转电机为负序阻抗时,假定其为正序阻抗;将电源电势认定为恒定并其标幺值取1;短路电流的衰减不作考虑;变压器电压的分接头实际位置产生变动的情况不作考虑;线路分布电容以及负荷电流不作考虑;故障位置的相间电阻、接地电阻都不作考虑;短路暂态电流的非周期分量不作考虑;过渡电阻不作考虑。

2.3 超高压继电保护整定计算

1) 瞬时电流速断保护。动作电流需要按照能够躲过该线路末端的最大短路电流值进行整定:

$$I_{\text{set}} = K_{\text{rel}} I_{\text{k.max}} \quad (4)$$

式中, K_{rel} 为可靠系数,取值为1.2~1.3; $I_{\text{k.max}}$ 为在系统的最大运行方式下该线路末端发生三相短路时的电流。

其最小保护范围必须大于受保护线路整体长

度的15%~20%,最小保护范围表达式为

$$I_{\text{k.L.min}} = \frac{\sqrt{3}}{2} \frac{U}{Z_{\text{s,max}} + Z_1 L_{\text{min}}} \quad (5)$$

式中, U 为系统相电动势; L_{min} 为最小保护范围的长度; Z_1 为单位长度线路的正序阻抗; $Z_{\text{s,max}}$ 为最小运行方式时系统的等效阻抗。

2) 限时电流速断保护。该保护通常与下一线路的I段保护配合整定,灵敏度不够时,与II段保护配合整定,其表达式为

$$I_{\text{set}} = K_{\text{rel}} I_{\text{set.2}} \quad (6)$$

式中,当与I段保护配合整定时, $I_{\text{set.2}}$ 为I段保护整定值;当与II段保护配合整定时, $I_{\text{set.2}}$ 为II段保护整定值。

为保证保护的选择性,该保护动作时限应比下一级保护动作时限长一个 Δt (通常取0.5s),其表达式为

$$t_2 = t_1 + \Delta t \quad (7)$$

式中,在与I段保护配合整定或与II段保护配合整定时, t_1 、 t_2 分别为I段保护整定时间、II段保护整定时间。

该保护的灵敏度必须满足要求,其表达式为

$$K_{\text{sen}} = \frac{I_{\text{k.min}}}{I_{\text{set}}} \quad (8)$$

式中, $I_{\text{k.min}}$ 为最小运行方式下系统线路末端的二相短路电流整定值。

3) 后备保护。电流整定值的计算表达式为

$$I_{\text{set}} = \frac{K_{\text{rel}} K_{\text{M}}}{K_{\text{re}}} I_{\text{f.max}} \quad (9)$$

式中, K_{rel} 取值为1.15~1.25; K_{M} 为电机自启动系数; K_{re} 为返回系数; $I_{\text{f.max}}$ 为系统的最大负荷电流。

其动作时限为

$$t_2 = t_1 + \Delta t \quad (10)$$

式中,在本保护作为近后备保护与该线路II段保护配合时, t_1 为II段保护整定时间;在本保护作为远后备保护与该线路III段保护配合时, t_1 为III段保护整定时间。

后备保护的灵敏度 K_{sen} 要求不小于1.2,其表达式为

$$K_{\text{sen}} = \frac{I_{\text{k.min}}}{I_{\text{set}}} \quad (11)$$

3 基于帝王蝶算法的特征选择

电网设备采集到的初始数据并不完整且杂乱无章,无法直接使用。为保证计算结果的准确性,需要对数据先进行预处理,本文采用了改进帝王蝶

优化算法对电网初始数据进行特征选择,从而实现预处理操作。

3.1 帝王蝶优化算法

帝王蝶优化算法(monarch butterfly optimization, MBO)是一种受帝王蝶的迁徙行为启发而诞生的一种群智能优化算法,该算法采用“迁移算子”与“调整算子”对帝王蝶2个子种群进行更新。

帝王蝶种群会在区域1(Land¹)与区域2(Land²)分为2个子种群,分别记为 N_1 和 N_2 ,2个子种群的总体数量为 N , p 为 N_1 子种群的数量占比,其种群数量可设置为

$$N_1 = \text{ceil}(p \cdot N) \quad (12)$$

式中,ceil为向上取整函数。

N_2 子种群数量可设置为

$$N_2 = N - N_1 \quad (13)$$

迁移算子的数学模型为

$$\begin{cases} x_{i,k}^{t+1} = x_{r_1,k}^t, r \leq p \\ x_{i,k}^{t+1} = x_{r_2,k}^t, r > p \end{cases} \quad (14)$$

式中, $x_{i,k}^{t+1}$ 为在第 $t+1$ 次迭代时 x_i 的第 k 维,即帝王蝶 i 的位置; $x_{r_1,k}^t$ 与 $x_{r_2,k}^t$ 分别为在第 t 次迭代时 x_{r_1} 与 x_{r_2} 的第 k 维,即帝王蝶 r_1 和 r_2 的位置;标量 $r = \text{rand} \cdot \text{peri}$,rand为从 $[0,1]$ 之间随机选择的数, p 为帝王蝶种群迁移周期。

调整算子会使 N_2 子种群的位置更新方式发生改变。若 $\text{rand} \leq p$,则其位置更新的数学模型为

$$x_{j,k}^{t+1} = x_{\text{best},k}^t \quad (15)$$

式中, $x_{j,k}^{t+1}$ 为在第 $t+1$ 次迭代时 x_j 的第 k 维,即帝王蝶 j 的位置; $x_{\text{best},k}^t$ 为在第 t 次迭代时 x_{best} 的第 k 维,即全局最优帝王蝶个体位置。

若 $\text{rand} > p$,则其位置更新的数学模型为

$$x_{j,k}^{t+1} = x_{r_3,k}^t \quad (16)$$

式中, $x_{r_3,k}^t$ 为第 t 次迭代时从 N_2 子种群中随机挑选的 x_{r_3} 的第 k 维,基于此情况,若 $\text{rand} > B_{\text{AR}}$,则 $x_{j,k}^{t+1}$ 的位置可再次更新为

$$x_{j,k}^{t+1} = x_{j,k}^{t+1} + \alpha(d_{x_k} - 0.5) \quad (17)$$

式中, B_{AR} 为调整率,通常设置 $B_{\text{AR}} = p$; α 为权重因子,通常由 $\alpha = S_{\text{max}}/l^2$ 计算得来; S_{max} 为最大步长; d_x 为帝王蝶的随机步长,其数学模型为

$$d_x = \text{Levy}(x_j^t) \quad (18)$$

3.1 改进帝王蝶优化算法

在实际搜索中,通常帝王蝶的步长过大会导致无法搜索到最优解而陷入局部最优解的情况,为避

免这种境况的发生,需对步长的权重因子进行调整和扰动,从而帮助算法跳出局部最优的情况,达到全局最优的效果。对 α 进行改进:

$$\alpha = \frac{(t_{\text{max}} - t)^2 \cdot S_{\text{max}}}{t_{\text{max}}^2 \cdot t^2} \quad (19)$$

对权重因子改进之后,需要对帝王蝶的种群划分进行改进。采用适应度函数对帝王蝶个体的适应度 F 进行计算,获取种群个体中的最优适应度 F_{best} 与 F_{worst} ,种群划分的数学模型为

$$\begin{cases} D_{i,\text{best}} = |F_i - F_{\text{best}}| \\ D_{i,\text{worst}} = |F_i - F_{\text{worst}}| \end{cases} \quad (20)$$

式中, $D_{i,\text{best}}$ 、 $D_{i,\text{worst}}$ 分别为第 i 个帝王蝶个体与最优个体、最差个体之间的适应度差值,若满足:

$$\log_{10} \left(\frac{D_{i,\text{worst}}}{D_{i,\text{best}}} \right) > \sqrt{\frac{t}{t_{\text{max}}}} \quad (21)$$

则第 i 个帝王蝶个体直接进入下一次迭代;否则,继续在本次迭代中进行种群划分。

根据改进过的帝王蝶优化算法,为进行特征选择,需要将调整算子模型进行修改,使用tanh函数将结果转变为二进制值。故式(17)应改进为

$$x_{j,k}^{t+1} = \begin{cases} 1 - x_{j,k}^{t+1}, & \text{rand} < |\tanh \alpha(d_{x_k} - 0.5)| \\ x_{j,k}^{t+1}, & \text{rand} \geq |\tanh \alpha(d_{x_k} - 0.5)| \end{cases} \quad (22)$$

适应度函数表达式为

$$F = \chi C + \rho \frac{|A|}{|B|} \quad (23)$$

式中, $|A|$ 为确定的特征子集的模; $|B|$ 为特征个数,其中 $A \subseteq B$; $\chi \in [0,1]$ 为分类质量; $\rho = 1 - \chi$ 为子集长度权重。

其算法步骤如下:

1) 初始化帝王蝶种群参数: $N, p, B_{\text{AR}}, p, S_{\text{max}}, t_{\text{max}}$,设置 $t = 1$;

2) 迭代开始,通过适应度函数计算种群个体的适应度值,并计算个体的特征数量,根据式(20)、(21)对个体进行选择;

3) 未进入下一迭代的个体按照 p 分为种群1与种群2,种群1个体按照式(14)更新,种群2个体按照式(15)、(16)、(22)进行更新;

4) 迭代次数加一,判断迭代次数是否达到最大或最优个体是否满足要求,若否返回步骤2),若是则到下一步

5) 输出最优个体位置 x_{best} 及其适应度 $F(x_{\text{best}})$ 。

4 试验分析

结合超高压电网继电保护特点以及其整定计算原则,对电网初始数据采用改进帝王蝶优化算法进行数据预处理,接着结合预处理后的数据,并通过 Spark 分布式并行计算的继电保护整定计算系统,输入某地区超高压电网继电保护参数进行计算,评估该系统的有效性。

4.1 预处理操作

通过改进帝王蝶优化算法进行特征选择,对电网初始数据进行预处理操作,算法参数: $N=150, p=0.5, B_{AR}=0.5, \rho=1.2, S_{max}=1.0, t_{max}=20\ 000$ 。

电网数据预处理流程如图 4 所示。首先,输入电网数据,采用改进帝王蝶算法进行特征选择对其进行清洗,将其中的异常/空白数据进行剔除,并实现对清洗后的所有不同来源不同类型的电网数据集成;其次,将集成后的数据统一管理,并对数据形式进行转换,将其转换为统一且更易进行数据挖掘的形式;最后,提取出统一、准确的最终数据。

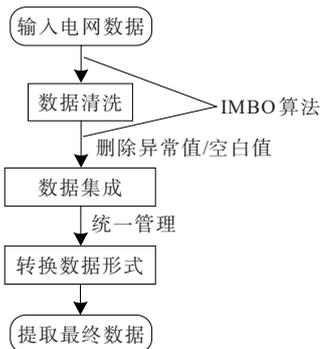


图 4 电网数据预处理流程

Figure 4 Flow chart of power grid data preprocessing

4.2 仿真分析

将某地区超高压电网数据先进行数据的预处理,再将预处理得到的最终数据输入整定计算系统中进行计算,继电保护整定计算系统的计算流程如图 5 所示。

本文为搭建 Spark 集群采用了 9 台硬件设施的主机进行物理平台搭建,每台主机服务器所搭建的硬件设施为:内存 512 G,主频 2.6 GHz,核心数 72,并在每台主机服务器上安装 VMware vSphere 软件。每台主机服务器都对 2 个计算节点进行管理,共 18 个计算节点,每个计算节点都以桥接的连接方式接入局域网,并且所接入局域网为同一局域网,在该局域网内,计算节点能够互通信息。每个计算

节点占用内存为 256 G,核心数为 36,其分布式计算框架为 Spark-3.0.0。

为验证基于 Spark 分布式并行计算的继电保护整定计算系统的有效性,通过将数据输入到不同计算平台下的整定计算系统,对比各个计算过程所耗费时间。为使对比结果之间的差别更加显著,分别对数据进行 50、200、500、1 000 次计算,计算节点都选择为 5 个,对比结果如图 6、7 所示。

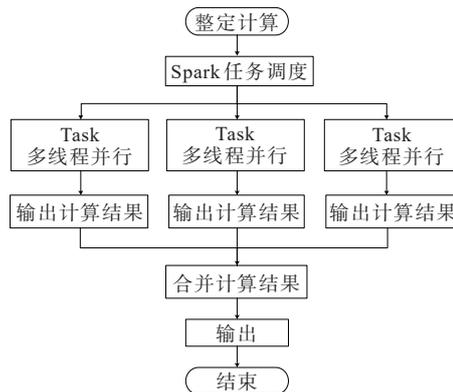


图 5 整定计算流程

Figure 5 Flow chart of setting calculation

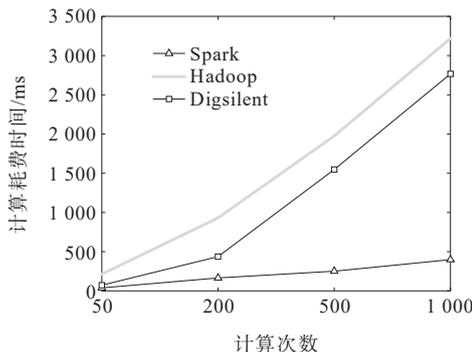


图 6 不同平台计算耗时对比

Figure 6 Comparison of time consumed by different platforms

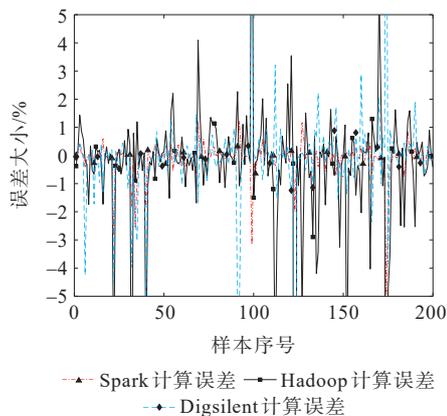


图 7 不同平台计算误差大小对比

Figure 7 Comparison of calculation errors of different platforms

为使结果更加明晰易懂,通过将结果对比值进行处理计算,并对比计算耗费时间以及计算结果准确度,两者的对比结果如表1、2所示。

表1 不同平台整定计算性能对比

Table 1 Performance comparison among setting calculation platforms

计算次数	计算耗费时间/ms		
	Spark	Hadoop	Digsilent
50	38	213	72
200	165	931	437
500	251	1 976	1 547
1 000	398	3 219	2 767

表2 不同平台整定计算结果对比

Table 2 Output comparison among different setting calculation platforms

平台	平均准确度/%
Spark	99.76
Hadoop	98.35
Digsilent	99.13

由表1、2可以看出,对比Hadoop以及Digsilent平台,基于Spark分布式并行计算的继电保护整定计算系统的计算效率以及计算结果平均准确度都更高。并且就平均准确度来说,Spark平台的可靠率在99.50%以上,满足电网整定计算要求;从计算耗费时间来看,计算性能相比于其他平台有着较大的提升。这些都验证了基于Spark分布式并行计算的继电保护整定计算系统的有效性,并且该系统能够保证对大规模电网数据的处理,能够有效降低计算时间,提升整定计算效率。

5 结语

本文提出了一种基于分布式并行计算的电网整定计算方法研究:

1) 设计了基于Spark分布式并行计算的继电保护整定计算系统,并就分布式并行计算的关键技术问题进行了分析研究;

2) 对特高压电网继电保护整定计算的特点进行了分析,并对其整定计算原则进行了分析,保证整定计算系统能够对其进行合理计算;

3) 通过改进调整因子,以及种群划分策略,对帝王蝶优化算法实现改进效果,并通过该算法进行特征选择对输入继电保护整定系统的电网初始数据进行预处理操作;

4) 通过对某地区特高压电网继电保护进行实例分析,对比不同平台下继电保护整定计算的性能,最终验证了基于Spark分布式并行计算的继电保护整定计算系统的可行性。

参考文献:

- [1] 陈田田,赵振廷,李银红.不依赖边界元件的柔性直流电网反时限过流保护配置方案[J].电力系统保护与控制,2023,51(16):48-58.
CHEN Tiantian, ZHAO Zhenting, LI Yinhong. Inverse time overcurrent protection configuration scheme for a flexible DC power grid independent of boundary components[J]. Power System Protection and Control, 2023,51(16):48-58.
- [2] 郭鹏,王文焕,詹荣荣,等.基于Petri网出现序列的继电保护业务模型[J].电力科学与技术学报,2020,35(4):169-175.
GUO Peng, WANG Wenhuan, ZHAN Rongrong, et al. Research on model of protection relay procedure based on the Petri net occurrence sequence[J]. Journal of Electric Power Science and Technology, 2020,35(4):169-175.
- [3] 李铁成,范辉,张卫明,等.基于5G通信的有源配电网新能源送出线路纵联保护[J].中国电力,2024,57(11):139-150.
LI Tiecheng, FAN Hui, ZHANG Weiming, et al. Pilot protection of new energy transmission line in active distribution network based on 5G communication[J]. Electric Power, 2024,57(11):139-150.
- [4] 薛士敏,李翔宇,陈文杰,等.一种基于修正电压的新型暂态量保护原理[J].电力建设,2024,45(4):89-99.
XUE Shimin, LI Xiangyu, CHEN Wenjie, et al. Novel transient protection principle based on corrected voltage [J]. Electric Power Construction, 2024,45(4):89-99.
- [5] 李媛.面向智能电网监测的分布式数据流处理[J].电测与仪表,2023,60(8):121-124+163.
LI Yuan. Distributed data stream processing for smart grid monitoring[J]. Electrical Measurement & Instrumentation, 2023,60(8):121-124+163.
- [6] 刘洋,伍双喜,朱誉,等.计及隐私保护的多智能体深度强化学习有源配电网电压控制策略[J].智慧电力,2024,52(12):104-111.
LIU Yang, WU Shuangxi, ZHU Yu, et al. Multi-agent deep reinforcement learning voltage control strategy for active distribution networks considering privacy protection[J]. Smart Power, 2024,52(12):104-111.
- [7] 黄公跃,付婷婷,林思远,等.用电信息采集系统电能计量数据异常识别研究[J].电网与清洁能源,2023,39(4):25-30+46.

- HUANG Gongyue, FU Tingting, LIN Siyuan, et al. A study on abnormal identification of electric energy measurement data in electric energy information acquisition system[J]. *Power System and Clean Energy*, 2023,39(4):25-30+46.
- [8] 刘硕,吴旭,马速良,等.基于模型预测控制的光-氢-储耦合系统的功率优化分配方法研究[J]. *高压电器*,2024,60(7):23-33.
- LIU Shuo, WU Xu, MA Suliang, et al. Research on power optimization allocation method for solar-hydrogen-storage coupled system based on model predictive control[J]. *High Voltage Apparatus*, 2024,60(7):23-33.
- [9] 于学良,方立华,曹永吉,等.基于量测和热平衡方程的输电线路温度分段实时估计[J]. *电力工程技术*,2024,43(6):223-234.
- YU Xueliang, FANG Lihua, CAO Yongji, et al. Real-time segmented temperature estimation of transmission lines based on measurement and heat balance equation[J]. *Electric Power Engineering Technology*, 2024,43(6):223-234.
- [10] 王文焕,郭鹏,祝洁,等.基于故障树及贝叶斯网络的继电保护系统风险评估及故障定位方法[J]. *电力科学与技术学报*,2021,36(4):81-90.
- WANG Wenhuan, GUO Peng, ZHU Jie, et al. Fault positioning and risk assessment method of relay protection based on fault tree and Bayesian network[J]. *Journal of Electric Power Science and Technology*, 2021,36(4):81-90.
- [11] TANG X Y, SHI C, DENG T, et al. Parallel random matrix particle swarm optimization scheduling algorithms with budget constraints on cloud computing systems[J]. *Applied Soft Computing*, 2021,113:107914.
- [12] 颜琬滢,段建东,陈童,等.基于综合电流幅值比的永磁直驱风力发电机送出线路纵联保护[J]. *供用电*,2023,40(4):47-54.
- YAN Wanying, DUAN Jiandong, CHEN Tong, et al. Pilot protection based on comprehensive current amplitude ratio for outgoing transmission line with permanent magnet synchronous wind power generator[J]. *Distribution & Utilization*, 2023,40(4):47-54.
- [13] DE JESÚS JARAMILLO SERNA J, LÓPEZ-LEZAMA J M. Calculation of distance protection settings in mutually coupled transmission lines: a comparative analysis[J]. *Energies*, 2019,12(7):1290.
- [14] KHOSHBAKHT E, NAMDARI F, DOOSTIZADEH M. Effects of protection settings on optimal performance of reconfigurable smart distribution systems[J]. *IET Renewable Power Generation*, 2021,15(8):1678-1692.
- [15] XIA T, ZHANG W, CHIU W S, et al. Using cloud computing integrated architecture to improve delivery committed rate in smart manufacturing[J]. *Enterprise Information Systems*, 2021,15(9):1260-1279.
- [16] 杨捷,孙哲,苏辛一,等.考虑振荡型功率的直流微电网储能系统无互联通信网络的多目标功率分配方法[J]. *发电技术*,2024,45(2):341-352.
- YANG Jie, SUN Zhe, SU Xinyi, et al. A wireless multi-objective power sharing method for energy storage system in DC micro-grid considering oscillatory-type power[J]. *Power Generation Technology*, 2024,45(2):341-352.