

DOI: 10.19783/j.cnki.pspc.190231

基于流式计算的暂态电压扰动并行实时监测技术

康瑞¹, 齐林海¹, 王红¹, 潘爱强², 周健²

(1. 华北电力大学控制与计算机工程学院, 北京 102206; 2. 国网上海市电力公司电力科学研究院, 上海 200437)

摘要: 暂态电压扰动模式识别面临两个挑战, 一是局限于单一监测点的扰动识别不能准确解释复杂扰动的完整过程, 另一是离线分析很难满足辅助决策实时性的要求。提出基于 Storm 流式计算框架, 结合 logstash 和 Kafka 消息中间件, 构建面向多监测点的实时数据监测处理平台。采用滑动时间窗口算法, 实现 Storm 编程逻辑拓扑。通过设置基本时间窗口大小和数量, 实现面向区域电网的多时空尺度、多业务模型的暂态电压扰动模式识别。实验结果表明, 合理设置 Storm 组件的任务数目能够最大限度发挥并行处理能力。通过仿真数据测试得到的吞吐量和平均处理延迟结果, 能够满足电网对流数据实时处理的高吞吐量、可扩展性、实时性和准确性的要求。

关键词: 暂态电压扰动; 流式计算; 并行滑动窗口; 模式识别

Parallel real-time monitoring technology for transient voltage disturbance based on flow calculation

KANG Rui¹, QI Linhai¹, WANG Hong¹, PAN Aiqiang², ZHOU Jian²

(1. School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China;
2. Electric Power Research Institute, SMEPC, Shanghai 200437, China)

Abstract: Transient voltage disturbance pattern recognition faces two challenges. First, disturbance recognition limited to a single monitoring point cannot accurately explain the complete process of complex disturbance, and the other is that off-line analysis cannot meet the requirements of real-time decision-making. Based on the Storm streaming computing framework, combined with logstash and Kafka message middleware, this paper builds a real-time data monitoring and processing platform for multiple monitoring points. The sliding time window algorithm is used to implement the Storm programming logic topology. By setting the basic time window size and quantity, it realizes the identification of transient voltage disturbance patterns for multi-temporal scale and multi-service models for regional power grids. The experimental results show that the number of tasks for properly setting the Storm component can maximize the parallel processing capability. The throughput and average processing delay results obtained by the simulation data test can meet the requirements of high throughput, scalability, real-time and accuracy of real-time processing of grid convection data.

This work is supported by Fundamental Research Funds for the Central Universities (No. 2018MS030) and Science and Technology Project of State Grid Corporation of China “Big Data Deep Analysis and Application Technique Research for Power Quality of Urban Power Grid” (No. 52094018001C).

Key words: transient voltage disturbance; flow calculation; parallel sliding window; pattern recognition

0 引言

电能质量对电网安全稳定运行和用户优质可靠用电至关重要。电能质量问题给电力用户造成的经济损失越来越大,对电网安全运行的威胁日益严重。

随着计算机应用技术、自动化控制技术、大功率技术和大功率电力电子技术的迅速发展,用电负荷已进入电压敏感时代^[1-2];智能电网中新能源发电和直流输电技术的推广,使得电力电子技术在电力系统“源”、“网”、“荷”的应用日益广泛,电网具备了显著的电力电子化特征^[3],从而导致系统侧和用户侧均受到劣质电能的严重影响。其中,最为突出的是暂态电压扰动(包括暂降、暂升和短时中断),其在所有电能质量扰动中所占比重及造成损失均最大^[4-5]。

基金项目: 中央高校基本科研业务专项资金项目资助(2018MS030); 国家电网公司科技项目资助(52094018001C)
“城市电网电能质量大数据深化分析及应用技术研究”

在复杂电网环境下，造成暂态电压扰动的因素增多；扰动在源-荷之间、网-荷之间、荷-荷之间传播与叠加，扰动之间的关联性增大。针对复杂的暂态电压扰动事件，局限于单一监测点的扰动识别，不能全面解释复杂扰动的完整过程和综合信息，须面向区域电网，选择特定监测点集群，基于同一时间截面，进行多监测点并行的扰动识别和综合分析，才能形成准确的判断。暂态电压扰动识别的目标是基于幅值、持续时间和相位跳变等基本要素，实现扰动源、扰动分类和传播路径的分析及判定，从而为扰动治理提供辅助决策支持。电能作为一个特殊的产品，它的生产和消费是瞬间同时完成的，因此，扰动事件及时发现、尽早处理，才能将损失降到最低。现有的电能质量监测系统虽然可以实现实时监测^[6-9]，但存在局限性：1) 只能完成基于扰动监测数据的基本指标实时计算，很难实现实时模式识别；2) 局限于单一监测点的实时监测，没有实现基于区域电网的多个监测点并行实时高效计算；3) 现有监测系统大多通过前置机定时上传监测数据到主站数据库来支持离线方式的扰动模式识别。采用离线定时处理方式，只能实现事后分析处理，与实际生产需求脱节，扰动识别结果在辅助决策的实时性方面失去意义。

大数据流式计算为应对上述挑战提供了解决途径。大数据处理技术分批量计算和流式计算两类。批量计算按先存储、后计算方式，适用于大规模数据的分布式存储和离线处理。Hadoop 平台属于批量计算框架，先将数据分布存至计算机集群节点，再以数据切片方式并行批量处理节点数据^[10-14]。由于进行数据库的存取操作，无法满足数据即时到达即时处理的实时性需求。流式计算作为新兴的计算模式，采用全内存计算，可以满足即时处理的要求^[15-18]。流式计算同样采用计算机集群节点模式，通过分布式内存并行计算，避免了单个计算机内存容量小的限制，从而实现大规模监测点海量数据并行实时计算处理。

局限于单一监测点的扰动识别不能全面准确解释复杂扰动的完整过程及综合信息和离线分析无法满足辅助决策实时性要求。本文提出采用流式计算框架和滑动时间窗口算法，实现了多个监测点暂降扰动分类识别的并行实时高效计算处理^[19-20]。为面向区域电网电能质量综合分析评估提供实时高效计算支撑^[21-22]。

1 流式计算框架与编程模型

Storm 是由 Twitter 开源的一个分布式、高容错

的实时流式计算框架。Storm 集群由一个主节点和多个工作节点组成。主节点运行 Nimbus 守护进程，用于分配代码、布置任务及故障检测。工作节点运行 Supervisor 守护进程，用于监听工作，启动并终止工作进程。ZooKeeper 用于管理和协调集群中的不同组件^[23]。

Storm 的编程模型可以抽象为一种拓扑 (Topology)，不同业务需求定义不同的拓扑。拓扑中的组件有喷嘴 (Spout) 和螺栓 (Bolt)，元组 (tuple) 是拓扑的最小单位，是组成数据流的基本单元。喷嘴负责从某个队列中不断读取数据元素并装配为元组发送给下游螺栓，形成数据流。螺栓完成对具体业务的逻辑实现，是处理数据的核心模块。喷嘴和螺栓之间是订阅关系。它们之间源源不断传递的元组被抽象为流 (Stream)，在传递元组时可以制定不同的分组策略。在拓扑中分组策略可以指定数据流的流向，实现业务逻辑的灵活制定。

拓扑运行在 Storm 流式框架下，如图 1 所示。喷嘴和螺栓组件被 Zookeeper 协调映射到 Storm 不同节点机器，当一个节点宕机，并不会影响集群工作，Zookeeper 会将喷嘴和螺栓重新进行分发使拓扑能够流畅运行。

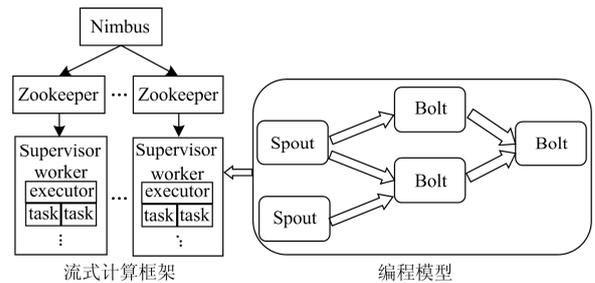


图 1 流式计算框架与编程模型关系图

Fig. 1 Flow computing framework and programming model diagram

2 大规模监测点并行实时处理框架

基于流数据处理的整体框架分为四个流程：实时数据获取、数据转换、协调控制和业务逻辑处理，如图 2 所示。

在实时数据获取中，每个监测点均部署 logstash 软件，实现数据实时收集、分析、格式处理，并将接收到监测点的数据流转发到数据转换模块；数据转换模块主要由 Kafka 集群组成，Kafka 是一种高吞吐量的分布式发布订阅消息系统，把上游 logstash 传输过来的海量监测点数据暂存起来，以保证在业务逻辑处理时能够稳定地处理监测点数据，实现实

时数据获取模块与业务处理模块之间数据传输流转和自动推送; Zookeeper 构成协调控制模块, 实现业务逻辑拓扑与 Storm 集群映射的协调控制, 它便于将分布式协调的职责从应用中分离出来, 从而可以大大减少系统之间的耦合性, 而且能够显著提高系统的可扩展性; 业务处理则是构造的拓扑, 进行暂降模式识别和业务逻辑处理。

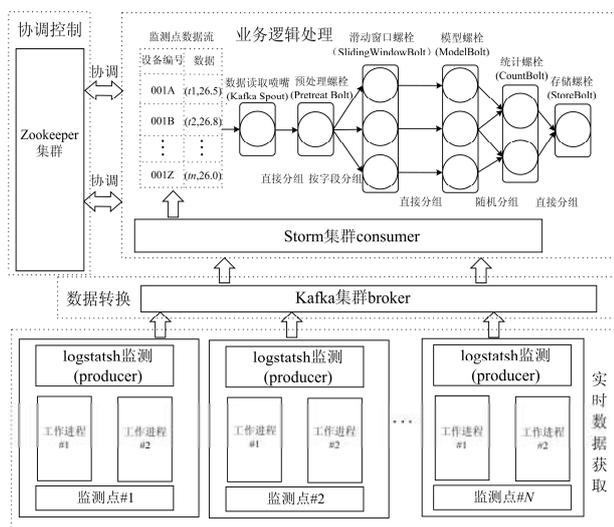


图2 大规模监测点并行实时处理框架

Fig. 2 Large-scale monitoring point parallel real-time processing framework

业务逻辑处理模块是整体框架的核心。该模块采用基于滑动时间窗口的算法, 通过设置窗口的大小满足不同敏感特性设备对时间颗粒度要求; 通过设置窗口个数实现不同业务需求的监测点规模控制; 将算法模型与逻辑拓扑融合构建灵活的业务体系。

3 实时数据获取(图3)

logstash 包括输入、过滤器和输出三部分。其中, 输入组件实时监测数据产生, 除监测数据外, 同时写入监测点位置信息; 过滤器为监测点数据记录标签; 输出组件写入 Kafka 集群的地址, 同时依据记录标签为不同类型数据创建 topic(主题)名称, 以便于后续业务分类处理。

Kafka 的数据输入来自于 logstash 的推送。当 logstash 监测到有数据传送来时, 将数据从输入组件通过过滤器打上标签传送到输出组件, 输出组件根据配置的 Kafka 集群的地址和 topic 名称将监测点的数据推送到对应的 Kafka 集群。Kafka 集群作为一个蓄水池, 将传送过来的数据暂存到对应的 topic 消息队列中。

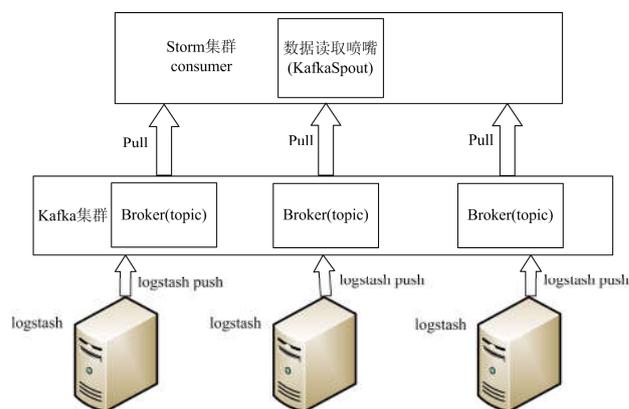


图3 实时数据获取

Fig. 3 Real-time data acquisition

作为 Kafka 的消费者, 当 Storm 发现 Kafka 中有待处理的数据后, 将数据拉取到数据读取喷嘴, 按照拓扑逻辑进行业务处理。

4 基于滑动窗口算法多时空尺度的灵活控制

基于大规模监测点并行实时模式识别是实现复杂暂态电压扰动准确识别和实时分析的有效途径。基于时间角度, 针对敏感度不同的负荷, 实现不同时间周期的实时计算; 基于空间角度, 面向区域电网的不同监测点集群, 实现覆盖不同空间对象; 基于不同的应用需求, 将相应的算法模型嵌入拓扑结构的螺栓组件中, 实现不同的业务逻辑功能。本文在 Storm 流式框架基础上, 基于时间滑动窗口算法, 结合业务逻辑拓扑, 提出适合于不同监测点集群, 满足不同时间颗粒度的可靠数据传输策略, 实现覆盖多时空尺度大规模监测点并行实时暂态电压扰动模式识别。

滑动窗口一般分为两种类型: 基于势的和基于时间的。基于势的滑动窗口, 窗口大小固定, 窗口内保存最近到来的 k 个元组; 基于时间的滑动窗口存储的是最近 T 时间内到达的元组^[24]。本文采用基于时间的滑动窗口算法。随着时间的流逝以及新流元素的到达, 滑动窗口分别向前滑动, 从而使窗口中的流元素不断发生变化。流数据处理目的就是实时处理最新到达的数据, 通过构造的拓扑进行模式识别。对监测数据的实时处理是一个连续计算的过程, 其实质可以分解为短小的批处理作业, 而时间滑动窗口模型能够很好地满足这样的应用需求。

基于时间的并行滑动窗口结构如图4所示。时间基本窗口是滑动窗口结构中的基本单元。依据监测数据标签每个基本窗口与不同监测点对应, 同时与逻辑拓扑中螺栓组件映射, 基本窗口时间区间到

达的数据将被统一批量处理；连续多个基本窗口构成基于时间的滑动窗口。滑动窗口的大小反映出处理能力。每一个基本窗口对应一个监测点，整个滑动窗口体现出覆盖监测点规模。通常情况下，窗口滑动距离与滑动窗口的时间区间大小一致。

图 4 中基本窗口设置为 1 s，滑动窗口设置有 60 个基本窗口，每个基本窗口对应 1 个监测点计算单元，整个滑动窗口覆盖 60 个监测点。电能质量监测系统的多个监测点的实时监测数据以数据流方式不断地通过窗口，60 个计算单元并行进行识别分析计算，每 1 s 可以输出 60 个监测点暂态电压扰动的识别结果。

根据电网运行管理的实际需求以及不同负荷敏感度，调整基本窗口时间区间大小和滑动窗口中基本窗口数量，就可以实现不同时间尺度、不同监测点规模的并行实时模式识别。

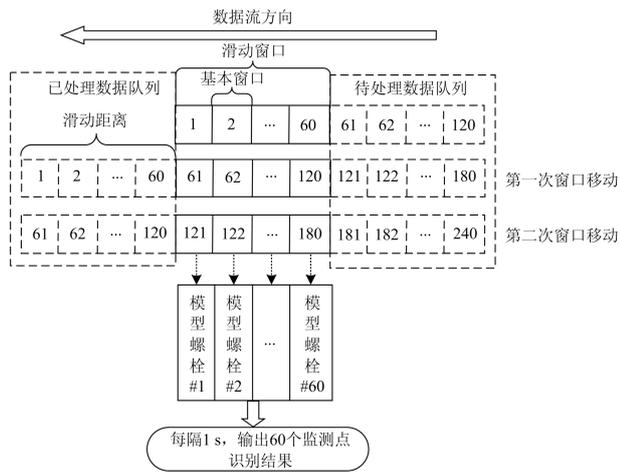


图 4 并行滑动窗口模型

Fig. 4 Parallel sliding window model

5 并行实时模式识别拓扑实现

5.1 拓扑设计

流式计算的实时处理本质上是一种连续“微批量计算”，滑动时间窗口算法的技术特点正好符合这一本质需求。Storm 的并行实时处理机制与滑动时间窗口流程设置也是相吻合的。因此，采用滑动时间窗口算法实现不同时间尺度的暂态电压扰动模式识别是可行的。

实现并行滑动窗口拓扑图如图 5 所示。拓扑由数据读取喷嘴和不同功能的螺栓组成，数据元组根据分组策略控制在螺栓中传递，形成数据流。首先数据读取喷嘴(KafkaSpout)接收从 Kafka 集群中来的监测数据，并将数据流以直接分组的方式发送给

预处理螺栓，实现数据类型转换和格式化处理；预处理螺栓(PretreatBolt)接着以字段分组的方式将监测数据发送给滑动窗口螺栓(SlidingWindowBolt)；滑动窗口螺栓再以直接分组的方式将同一监测点的数据发送给模型螺栓(ModelBolt)，并将模型螺栓分类、预测的结果以随机分组打乱的方式发送到统计螺栓(CountBolt)；最后多个统计螺栓将结果发送给存储螺栓(StoreBolt)由其将结果存储到数据库中。

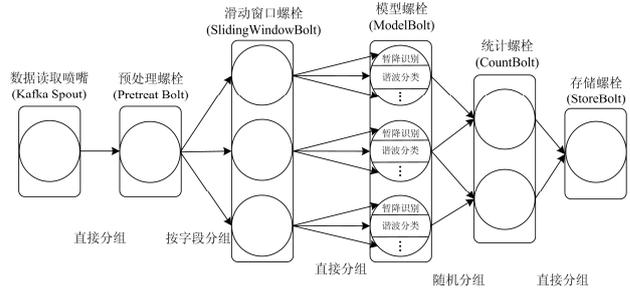


图 5 并行滑动窗口拓扑图

Fig. 5 Parallel sliding window topology

具体实现时，为了使每一个时间基本窗口对应一个监测点，数据流从预处理螺栓至滑动窗口螺栓必须是按字段 id 分组，即同一监测点的数据由唯一的一个滑动窗口螺栓处理，而滑动窗口螺栓的个数表示覆盖监测点规模。通过模型螺栓的配置，可以实现不同业务逻辑的灵活制定。模型螺栓可以是暂降模式识别模型，实现即时的暂降扰动模式识别。也可以是谐波分类模型。模型螺栓也可以是多个不同判别模型融合，在实现时，只需在监测点数据中加一字段标识流数据具体做哪种处理。

5.2 拓扑实现

拓扑实现分为 6 个步骤：1) 创建拓扑构造器 TopologyBuilder 对象；2) 设置 Zookeeper 集群的 ip 地址；3) 设置 SpoutConfig 里面的参数包括：Zookeeper 集群，Kafka 建立的 topic 名称，Zookeepr 的根目录和自定义 id 值；4) 构建数据读取喷嘴、预处理螺栓和滑动窗口螺栓，其执行器的并发度均为 1；5) 构建模型螺栓，其执行器的并发度为 n；6) 构建统计螺栓和存储螺栓，其执行器的并发度分别为 Y 和 1。并行滑动窗口拓扑实现编码如图 6 所示。

通过构建以上拓扑实现每 1 s 发送一个数据，滑动窗口模型的滑动距离为 ns，滑动窗口大小为 ns，则滑动窗口分为 n 个时间基本窗口。各个时间基本窗口是独立并行处理，统计螺栓可以根据电网中业务对时间敏感性的要求设置并发度 Y，最终实现每 1 s 统计最近 ns 监测数据。

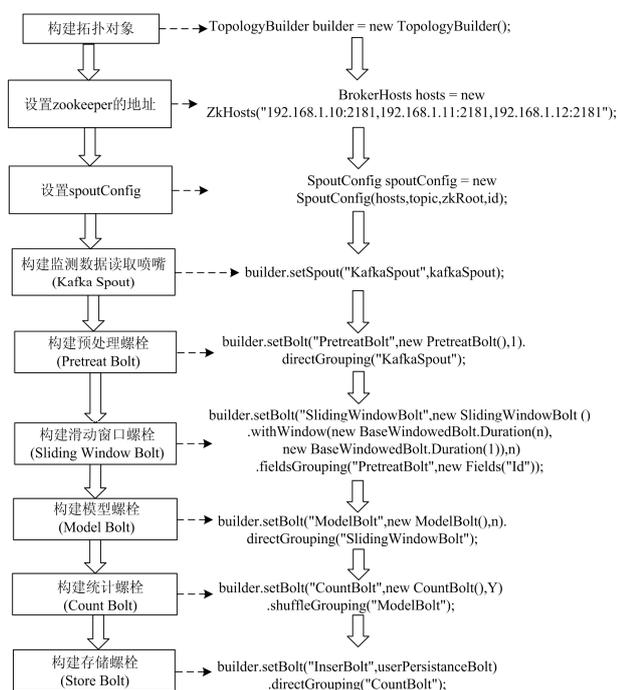


图 6 并行滑动窗口拓扑实现代码

Fig. 6 Parallel sliding window topology implementation code

5.3 模型螺栓构建

暂态电压扰动模式识别是计算机通过人工神经网络的方法在暂态情况下根据样本的特征对样本进行分类,是故障诊断方法之一。本文模型螺栓采用文献[25]深度学习融合模型,基于模型融合的电压暂降源识别研究整体结构如图 7 所示。首先训练 CNN 模型,获取 CNN 模型中的卷积层和全连接层的参数,接着固定 CNN 模型卷积层的参数,利用训练集经过卷积层后的输出数据对 DBN 模型进行无监督的预训练,得到 DBN 模型中各 RBM 层的参数,下一步改 CNN 模型的全连接层为 DBN 模型的 RBM 层,迁移预训练后 CNN 模型的卷积层参数和 DBN 模型中 RBM 层的参数。并利用带标注的电压暂降信号数据对整个模型进行有监督的微调,从而生成训练好的融合模型。最后将训练好的融合模型放入模型螺栓中。

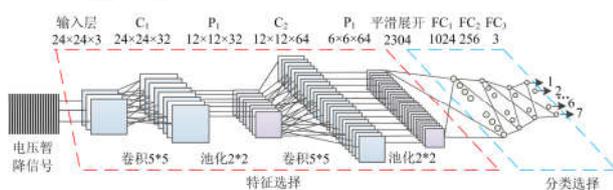


图 7 基于模型融合的电压暂降源识别方法框架

Fig.7 Framework of voltage sag sources' recognition method based on model fusion

6 实验设计与结果分析

6.1 实验环境说明

本文建立了由 1 个主控节点和 2 个工作节点构成的 Storm 集群,仿真模拟出 50 万条实验数据,为模拟 60 个监测点,给每条实验数据打对应的监测点标签,将处理好的数据统一通过程序模拟监测点发送数据 1 s 发送一条,60 个数据同时输入 Storm 集群构成的并行实时处理框架中。统计螺栓执行器并发度设置为 10。实验中,采用文献[26]中的基于多模型融合的电压暂降识别方法对复合电压暂降进行并行实时识别验证,3 个节点的机器配置如表 1 所示。

表 1 计算设备及实验平台

Table 1 Computing equipment and experimental platform

计算机设备配置	操作系统	JDK 环境	实验平台	数量/台
IntelCore i5-6200; CPU/8.00 G RAM	Centos 6.5	jdk1.8.0_111-b14	Storm1.1.1; Zookeeper3.4.6; Python 3.5.4	3

6.2 仿真实验与数据采样

电压暂降源数据参照文献[26]采用的仿真方法。本文在 Matlab/Simulink 平台中建立了如图 8 所示的仿真系统模型用于模拟四种复合电压暂降源对应的信号。仿真实验设置基波频率为 50 Hz,设置仿真总时长为 0.2 s,示波器的录波长度为 10 个周波,分别录制 A、B、C 三相电压的录波数据。图中变压器 T1、T2 变比为 110 kV/10.5 kV;变压器 T3、T4 变比为 10 kV/0.4 kV;K1、K2 表示三相开关;F 为故障点,M 为测量点;负载 L2 附近的 M1 为感应电机。通过改变仿真系统模型中线路负荷大小、故障发生位置、故障起止时刻、变压器及感应电动机容量等参数,获得短路故障引起的多级电压暂降 C1、单相接地与大型感应电动机启动的复合 C2、单相接地与变压器投切的复合 C3 和大型感应电动机启动与变压器投切的复合 C4 四种电压暂降源的信号。电压暂降源的信号各 300 组样本数据,每组样本数据的采样点 625 个,每类取 225 组样本为训练数据,剩余的 75 组样本作为测试数据,进行模型测试。最后将得到稳定的深度学习判别模型融入模型螺栓中,为验证整体框架的性能,还需通过以上方法仿真出 50 万数据量进行测试。

6.3 电压暂降源识别准确率对比

实验在不同噪声等级下分别对搭建好的网络进行迭代训练,对四种复合暂降原因类型进行分类,

并对生成的模型进行识别精度的测试。由表 2 可知,随着噪声等级的提高,融合模型的平均识别率从 99.1% 下降到 98.2%、96.8% 和 94.2%。

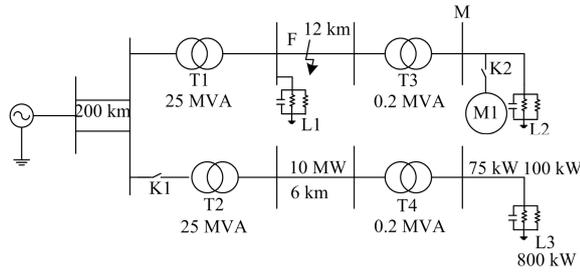


图 8 复合电压暂降源仿真系统模型

Fig. 8 Composite voltage sag source simulation system model

表 2 电压暂降源分类结果

Table 2 Classification results of voltage sag sources

暂降源类型	%			
	无噪声 准确率	40 dB 准确率	30 dB 准确率	20 dB 准确率
C1	100.0	99.3	98.0	95.7
C2	98.0	97.0	95.3	92.7
C3	99.3	98.3	96.7	94.3
C4	98.7	97.0	94.7	92.0

在 Storm 集群环境下和本地环境识别准确率都如表 2 所示。经过测试,流式计算并不影响模型识别的准确率,能够更有效地应用于实际工程中。

6.4 吞吐量测试

在保证电压暂降识别准确率情况下,将仿真的数据做打乱处理作为输入,观察工作进程和所有组件的任务数为 1 和 5,设置拓组件的并发度分别为 1, 1, 60, 60, 10, 1 时的吞吐量。分析在单机环境下和在集群环境下的吞吐量。吞吐量是指单位时间处理数据的规模。不断增加数据量,记录处理完成所需时间。比较结果如图 9 所示,所有的数据均是 3 次的实验数据取平均值。

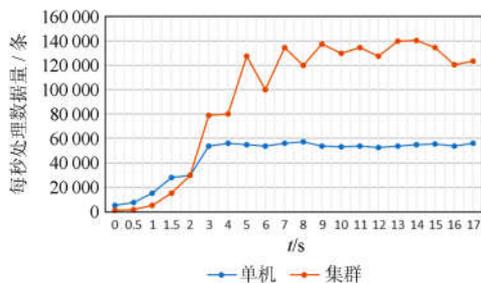


图 9 单机与集群吞吐量对比

Fig. 9 Stand-alone and cluster throughput comparison

结果表明:随着时间的推移,在单机环境下,每秒处理数据 5.5 万左右,并且运行相对稳定。在集群下,处理能力达到 13 万左右,稳定性差一些,由于集群中存在数据分发,扩展拓扑任务和数据的传输导致数据处理能力的不稳定。在输入的数据量小于 3 万条时,单机处理速度优于集群处理速度。数据 3 万条以上时,集群的优势体现出来了,相同时间处理数据量更多,Storm 集群还可以进一步扩展,满足更大数据量的处理需求。

6.5 处理延迟测试

在 Storm 集群节点、拓扑的进程数、各个组件任务数一定条件下,考虑增加模型螺栓的并发度,研究模型螺栓的并发度对集群处理的延迟影响。时间处理延迟定义是从监测点产生数据到其完全处理的时间,可以通过编写程序计算任务执行的延迟。本实验以 3 个节点为例,工作进程和所有组件的任务数设置为 1 和 5,模型螺栓的并发度分别设置为 15, 30, 60,即拓组件的并发度分别为 2, 1, 60, 15, 10, 1、2, 1, 60, 30, 10, 1 和 2, 1, 60, 60, 10, 1。分别统计三次拓扑任务执行时的数据处理延迟情况,如图 10 和图 11 所示。

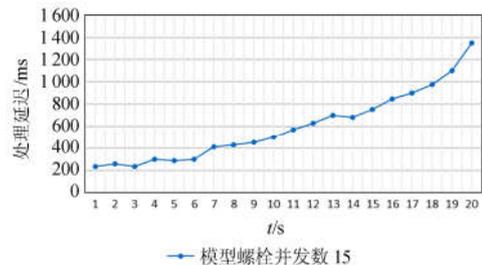


图 10 模型螺栓并发度 15 延迟性

Fig. 10 Model bolt concurrency 15 delay

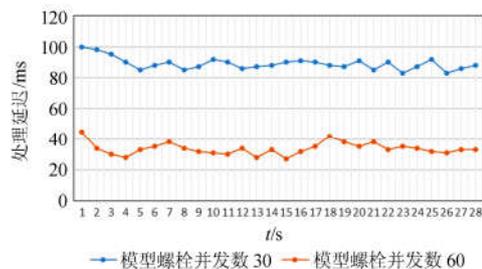


图 11 模型螺栓并发度 30 和 60 延迟对比

Fig. 11 Comparison of model bolt concurrency 30 and 60 delay

结果表明:模型螺栓并发度设置 15 延迟性折线图如图 10 所示,数据处理延迟越来越大,这是因为喷嘴不断产生新数据,模型螺栓无法及时处理掉数据,导致数据积累量越来越多,处理延迟也呈上涨趋势。

模型螺栓并发度设置 30 和 60 延迟性折线图如 11 所示, 当模型螺栓并发度设为 30 时, 绝大部分的处理延迟都稳定在 88.5 ms 左右, 表明喷嘴螺栓产生的数据可以及时得到处理。当模型螺栓并发度设为 60 时, 数据处理延迟都稳定在 32.5 ms 左右。平均处理延时降低了 63.3%, 分析认为实验的主要延迟性影响是在数据接入层 Kafka 接入 Storm 数据中转产生的。偶尔有一些处理延迟较大, 分析认为是因为拓扑任务调度引起的喷嘴线程在某一时刻获取了较多数据, 从而发送了较多的数据, 而后端的处理能力较强, 可以很快将这些数据处理完毕, 数据没有产生堆积, 但是数据处理延迟有所增长。

7 结论

通过搭建实时监测数据的 Storm 流计算集群, 利用 logstash 和 Kafka 消息中间件, 在流计算拓扑实例和拓扑并行编程模型上, 引入 Storm 中的并行快速滑动时间窗口算法, 满足不同监测点数据的时间颗粒度和监测点规模, 可以实现面向区域电网的多监测点并行实时暂态电压扰动模式识别。

通过实验得出, 流式计算并不影响模型识别的准确率。将训练好的深度学习判别模型融入不同时间窗格(模型螺栓)中, 可以实现业务逻辑的灵活制定, 满足更多的并行实时处理应用需求。合理设置各组件的任务数目是优化流式计算框架 Storm 性能的重要途径。当集群节点规模扩大, 处理数据量越多时, 吞吐量越大, 稳定性越好。根据实际的计算任务合理分配计算资源, 可以最大限度发挥 Storm 的并行处理能力。

参考文献

- [1] 王东旭, 乐健, 刘开培, 等. 复杂电网多重故障条件下的电压暂降分析[J]. 中国电机工程学报, 2012, 32(7): 101-106.
WANG Dongxu, LE Jian, LIU Kaipei, et al. Voltage sag analysis under complex fault conditions in complex power grids[J]. Proceedings of the CSEE, 2012, 32(7): 101-106.
- [2] VEGUNTA S C, MILANIOVIC J V. Estimation of cost of downtime of industrial processes due to voltage sags[J]. IEEE Transactions on Power Delivery, 2011, 26(2): 576-587.
- [3] 帅智康. 特约主编寄语[J]. 电力系统自动化, 2017, 41(8): 1-1.
SHUAI Zhikang. Special editor's message[J]. Automation of Electric Power Systems, 2017, 41(8): 1-1.
- [4] TAN R H G, RAMACHANDARAMURTHY V K. Voltage sag acceptability assessment using multiple magnitude duration function[J]. IEEE Transactions on Power Delivery, 2012, 27(4): 1984-1990.
- [5] 肖湘宁. 电能质量分析与控制[M]. 北京: 中国电力出版社, 2010.
- [6] 焦东升, 陆冬良, 应俊豪, 等. 动态电能质量实时监测系统的设计与实现[J]. 电网技术, 2011, 35(5): 110-114.
JIAO Dongsheng, LU Dongliang, YING Junhao, et al. Design and implementation of dynamic power quality real-time monitoring system[J]. Power System Technology, 2011, 35(5): 110-114.
- [7] 王玲, 康健, 邹宏亮, 等. 实时电能质量监测系统的构建及应用[J]. 电力系统保护与控制, 2011, 39(2): 108-111.
WANG Ling, KANG Jian, ZOU Hongliang, et al. Construction and application of real-time power quality monitoring system[J]. Power System Protection and Control, 2011, 39(2): 108-111.
- [8] 姜中华, 罗安, 赵伟, 等. 配电网电能质量实时监测分析与治理仿真[J]. 电网技术, 2010, 34(8): 18-24.
JIANG Zhonghua, LUO An, ZHAO Wei, et al. Real-time monitoring analysis and governance simulation of power quality in distribution network[J]. Power System Technology, 2010, 34(8): 18-24.
- [9] 董海艳, 贾清泉, 王宁, 等. 复杂电能质量关联扰动模式的可拓挖掘方法[J]. 中国电机工程学报, 2016, 36(5): 1269-1278.
DONG Haiyan, JIA Qingquan, WANG Ning, et al. Extension mining method for complex power quality correlation disturbance mode[J]. Proceedings of the CSEE, 2016, 36(5): 1269-1278.
- [10] 马天男, 王超, 彭丽霖, 等. 多源异构大数据下综合能源系统用户用能行为预测分析研究[J]. 智慧电力, 2018, 46(10): 86-95.
MA Tiannan, WANG Chao, PENG Lilin, et al. Prediction and analysis of user energy behavior of integrated energy systems under multi-source heterogeneous big data[J]. Smart Power, 2018, 46(10): 86-95.
- [11] 袁少光. 输变电设备监测云平台中任务调度算法的研究[D]. 北京: 华北电力大学, 2015.
YUAN Shaoguang. Research on task scheduling algorithm in cloud platform for power transmission and transformation equipment monitoring[D]. Beijing: North China Electric Power University, 2015.
- [12] CAO Z, LIN J, WAN C, et al. Hadoop-based framework for big data analysis of synchronised harmonics in active distribution network[J]. IET Generation, Transmission & Distribution, 2017, 11(16): 3930-3937.
- [13] 耿俊成, 马文栋, 郭志民, 等. 基于智能电表大数据分析的台区变压器三相不平衡治理[J]. 智慧电力, 2018, 46(9): 103-108.

- GENG Juncheng, MA Wendong, GUO Zhimin, et al. Three-phase unbalanced control of transformers in Taiwan based on big data analysis of smart meters[J]. Smart Power, 2018, 46(9): 103-108.
- [14] 林静怀. 基于大数据平台的电网运行指标统一管控方案[J]. 电力系统保护与控制, 2018, 46(4): 165-170.
LIN Jinghuai. Unified control and control scheme for power system operation index based on big data platform[J]. Power System Protection and Control, 2018, 46(4): 165-170.
- [15] 孙大为, 张广艳, 郑纬民. 大数据流式计算: 关键技术及系统实例[J]. 软件学报, 2014, 25(4): 839-862.
SUN Dawei, ZHANG Guangyan, ZHENG Weimin. Big data streaming computation: key technologies and system examples[J]. Journal of Software, 2014, 25(4): 839-862.
- [16] 韩德志, 陈旭光, 雷雨馨, 等. 基于 Spark Streaming 的实时数据分析系统及其应用[J]. 计算机应用, 2017, 37(5): 1263-1269.
HAN Dezhi, CHEN Xuguang, LEI Yuxin, et al. Real-time data analysis system based on Spark Streaming and its application[J]. Journal of Computer Applications, 2017, 37(5): 1263-1269.
- [17] 马帅. 基于 Spark Streaming 的反刷单系统的设计与实现[D]. 北京: 北京邮电大学, 2018.
MA Shuai. Design and implementation of anti-brushing system based on Spark Streaming[D]. Beijing: Beijing University of Posts and Telecommunications, 2018.
- [18] 刘少伟. 基于 Storm 框架的电网电力设备监测大数据并行诊断方法[D]. 北京: 华北电力大学, 2018.
LIU Shaowei. Parallel diagnosis method for grid power equipment monitoring big data based on Storm framework[D]. Beijing: North China Electric Power University, 2018.
- [19] 刘羽霄, 张宁, 康重庆. 数据驱动的电力网络分析与优化研究综述[J]. 电力系统自动化, 2018, 42(6): 157-167.
LIU Yuxiao, ZHANG Ning, KANG Chongqing. A review of data driven power network analysis and optimization research[J]. Automation of Electric Power Systems, 2018, 42(6): 157-167.
- [20] 周念成, 廖建权, 王强钢, 等. 深度学习在智能电网中的应用现状分析与展望[J]. 电力系统自动化, 2019, 43(4): 180-197.
ZHOU Niancheng, LIAO Jianquan, WANG Qianggang, et al. Analysis and prospect of application status of deep learning in smart grid[J]. Automation of Electric Power Systems, 2019, 43(4): 180-197.
- [21] 李志杰, 李元香, 王峰, 等. 面向大数据流的多任务加速在线学习算法[J]. 计算机研究与发展, 2015, 52(11): 2545-2554.
LI Zhijie, LI Yuanxiang, WANG Feng, et al. Multitask accelerated online learning algorithm for big data streams[J]. Computer Research and Development, 2015, 52(11): 2545-2554.
- [22] 王赛一, 余建平, 孙丰杰, 等. 电力大数据的价值密度评价及结合改进 k-means 的提升方法研究[J]. 智慧电力, 2019, 47(3): 8-15.
WANG Saiyi, YU Jianping, SUN Fengjie, et al. Evaluation of value density of power big data and improvement of improved K-means[J]. Smart Power, 2019, 47(3): 8-15.
- [23] 屈志坚, 王冬. 面向智能调度监测的流计算并行滑动窗口技术[J]. 电网技术, 2016, 40(7): 2219-2229.
QU Zhijian, WANG Dong. Flow-computing parallel sliding window technology for intelligent dispatch monitoring[J]. Power System Technology, 2016, 40(7): 2219-2229.
- [24] 王伟平. 数据流上基于时间滑动窗口的连接算法研究[C] // 第二十届全国数据库学术会议论文集(研究报告篇), 中国计算机学会数据库专业委员会, 2003.
WANG Weiping. Research on connection algorithm based on time sliding window on data stream[C] // Proceedings of the 20th National Database Conference (Research Report), China Computer Society Database Professional Committee, 2003.
- [25] 郑智聪, 王红, 齐林海. 基于深度学习模型融合的电压暂降源识别方法[J]. 中国电机工程学报, 2019, 39(1): 97-104, 324.
ZHENG Zhicong, WANG Hong, QI Linhai. A method for voltage sag source identification based on deep learning model fusion[J]. Proceedings of the CSEE, 2019, 39(1): 97-104, 324.
- [26] 李夏林, 刘雅娟, 朱武. 基于配电网的复合电压暂降源分类与识别新方法[J]. 电力系统保护与控制, 2017, 45(2): 131-139.
LI Xialin, LIU Yajuan, ZHU Wu. A new method for classification and identification of composite voltage sag source based on distribution network[J]. Power System Protection and Control, 2017, 45(2): 131-139.

收稿日期: 2019-03-05; 修回日期: 2019-07-10

作者简介:

康 瑞(1994—), 男, 硕士, 研究方向为大数据应用技术; E-mail: afanti@ncepu.edu.cn

齐林海(1964—), 男, 通信作者, 副教授, 研究方向为智能电网与电力信息化分析和电能质量分析控制工作; E-mail: qilin Hai@ncepu.edu.cn

王 红(1978—), 女, 博士, 讲师, 研究方向为大数据应用技术及电能质量分析控制。E-mail: wh@ncepu.edu.cn

(编辑 周金梅)