

基于决策树挖掘算法的气象大数据云平台设计

王立俊^{1,2}, 杜建华^{1,2}, 刘骥超^{1,2}, 王双双^{1,2}, 谢寒生¹, 赵冰¹

(1. 海南省气象信息中心, 海口 570203;

2. 海南省南海气象防灾减灾重点实验室, 海口 570203)

摘要: 大数据、云计算技术的迅猛发展为挖掘气象数据丰富的科研和经济价值提供了技术支撑, 促进了 Hadoop 及其包含的文件存储系统 (HDFS, hadoop distributed file system) 和分布式计算模型在气象数据处理领域广泛应用; 由于气象数据具有大数据的 4V 特征, 还需要引入新的数据处理算法来提高气象数据处理效率; 通过对决策树算法原理的研究, 基于 Hadoop 云平台, 创建随机森林模型, 为数据挖掘算法在云平台上的应用提供一种新的可能性; 基于决策树 (CART, classification and regression trees) 挖掘算法的气象大数据云平台设计, 采用 Hadoop 系统架构和 MapReduce 工作流程, 对气象大数据云平台采用集群部署; 平台总体架构分为基础设施层、数据管理与处理层、应用层, 减少了决策树建立的时间, 实现了气象数据高效加工和挖掘分析等平台功能。

关键词: 气象数据; 气象大数据云平台; 决策树算法; Hadoop; MapReduce

Design of Meteorological Big Data Cloud Platform Based on Classification and Regression Trees Mining Algorithm

WANG Lijun^{1,2}, DU Jianhua^{1,2}, LIU Jichao^{1,2}, WANG Shuangshuang^{1,2},
XIE Hansheng¹, ZHAO Bing¹

(1. Meteorological Information Center of Hainan Province, Haikou 570203, China;

2. Key Laboratory of South China Sea Meteorological Disaster Prevention and Mitigation of Hainan Province, Haikou 570203)

Abstract: The rapid development of big data and cloud computing technology provides the technical support for mining the rich scientific research and economic value of meteorological data. It promotes the wide application of Hadoop and Hadoop distributed file system (HDFS) and distributed computing model in the field of meteorological data processing. Due to the 4V characteristics of big data, the new data processing algorithms need to be introduced to improve the efficiency of meteorological data processing. Through the research on the principle of classification and regression trees (CART) algorithm, based on Hadoop cloud platform, a random forest model is constructed, which provides a new possibility for the application of data mining algorithm on the cloud platform. The design of meteorological big data cloud platform based on CART mining algorithm adopts Hadoop system architecture and MapReduce workflow to deploy the meteorological big data cloud platform in clusters. The overall architecture of the platform is divided into the infrastructure layer, data management and processing layer, application layer, which reduces the time to establish the decision tree and realizes the functions of the big data cloud platform such as efficient processing and mining analysis of the meteorological data.

Keywords: meteorological data; meteorological big data cloud platform; classification and regression trees; Hadoop; MapReduce

0 引言

气象数据为一类典型的时序数据, 在防灾减灾以及服务农业、交通、旅游等领域都有重要意义, 气象数据的处理与分析准确高效直接关系到预报服务的质量^[1-5]。《“十四五”气象信息网络业务发展规划》明确我国未来气象业务的发展方向, 主要是优化结构流程, 提高气象数据的标准化和统一化水平, 促进信息网络的智能化发展, 建立起高效协同的数字气象业务体系^[6]。对于气象数据来说, 其生产过程中秒级数据流量达到 6 万次/秒, 大数据、云计算

技术在大量气象数据的分析领域的应用, 为解决上述问题提供支持^[7-9]。

Bilgin 等^[10]针对土耳其的气象数据进行特征提取, 确定出相关温度特性, 基于这些特性划分出土耳其的气候区。Reikard^[11]基于频域算法对复杂的气象数据进行处理, 且引入了神经网络工具, 在此基础上建立起基于时频域的 BP 网络气象预测模型。根据仿真分析结果表明这种模型的预测准确率达到较高水平。邱声春等^[12]提出一种数据挖掘和数据融合相结合的预测方法, 并将其运用到天气预报领域。

收稿日期: 2022-07-17; 修回日期: 2022-08-17。

基金项目: 国家自然科学基金(41775011); 海南省气象局科技创新项目(HNQXSJ202114)。

作者简介: 王立俊(1989-), 男, 海南海口人, 硕士, 工程师, 主要从事气象数据处理与应用方向的研究。

通讯作者: 刘骥超(1988-), 男, 河南沁阳人, 硕士, 工程师, 主要从事气象数据管理与应用方向的研究。

引用格式: 王立俊, 杜建华, 刘骥超, 等. 基于决策树挖掘算法的气象大数据云平台设计[J]. 计算机测量与控制, 2022, 30(11): 140-146.

Nui 等^[13]提出距离度量学习方法用于气象数据挖掘, 将其应用于 2016 年在肯尼迪等机场收集的天气数据集, 与其他传统方法相比, 在分类精度上显示出非常有潜力的优势。Farouk 等^[14]运用决策树、KNN 和 Naive Bayes 算法研究了数据挖掘技术在预测不同的大气现象中的应用, 并对每个模型进行评估。王丹妮等^[15]通过对广东省气象观测数据挖掘分析, 以广东省农业气象灾害中的高温为例, 预测可能存在的灾害及其等级。陈凯等^[16]在进行降雨预测时, 引入了加权最近邻算法, 基于这种算法进行预测时, 设置降雨量作为类, 气象因素设置为分类因子, 计算分析确定出二者的相关系数, 将所得结果赋予因子的权重, 据此建立起相关区域的降雨预报模型。柯秀文^[17]建立起海洋气象灾害预测系统, 并对这种系统的预测性能做了仿真研究。胡冬梅^[18]研究了城市空气污染问题, 并引入了气象、交通相关的因素建立起城市空气污染预测模型, 对变量之间相关性进行分析, 根据仿真分析结果表明, 这种系统表现出较高的应用价值。王春政等^[19]在对气象数据挖掘处理时, 引入了贝叶斯算法、模糊 K 邻近算法, 通过这些算法对一定时段内全美历史航班和气象数据进行分析, 其所得结果表明节点预测误差小于 5 分钟, 很好地满足航空气象预测要求。还有少部分学者为提高处理海量气象数据的效率, 将其与 Hadoop 平台结合。张晨阳等^[20]在一定简化处理基础上建立起基于 MapReduce 的数据约简算法与朴素贝叶斯分类算法, 对这种算法的性能进行研究, 发现其在处理海量气象数据方面的优势很显著, 且可以根据应用要求进行适当的扩展。王昊等^[21]为了提高传统朴素贝叶斯分类器对气象数据挖掘的精度, 拥有更高的处理海量数据的效率, 建立起基于离散贝叶斯网络的数据挖掘模型, 且进行了实证分析。

人工智能、大数据、云平台、物联网技术在气象数据处理中的应用, 极大地提升了气象业务能力和服务水平。传统的气象数据挖掘大部分为单机模式, 在数据量不断增加和算法更复杂形势下, 气象预测的实时性明显降低, 同时也会出现硬件故障等相关的损失^[22]。因此, 本文基于主流的开源云架构 Apache Hadoop, 建立气象大数据云平台, 在此基础上结合决策树 (CART, classification and regression trees) 数据挖掘算法, 并行创建随机森林模型, 对处理海量气象数据提供支持, 为相关研究起到参考作用。

1 基于决策树算法的随机森林模型

1.1 决策树算法原理

在数据挖掘领域 CART 算法被大量应用, 其可以高效的进行分类预测, 适应性较强, 对离散和连续数据都可以高效的分类。CART 算法主要是进行数据类型差异实现分类作用。设置判定指标时主要是基于如下方法:

1) 离散型数据——选择 GINI 值作为分类依据。

GINI 值和分类后的“纯净度”存在正相关关系, 其数值大则可判断选择的特征更优。一般条件下对于样本集 S , 可通过如下公式计算出 GINI:

$$GINI(S) = 1 - \sum P_i^2 \quad (1)$$

在上式当中, S 代表的为样本数据集; P_i 代表的为分类结果中第 i 个种类对应频率。

对含 N 个数据的集 S , 基于 A 的第 i 个属性值进行划分而分为两部分, 划分后 GINI 的表达式为:

$$GINI_{A,i}(S) = \frac{n_1}{N}GINI(S_1) + \frac{n_2}{N}GINI(S_2) \quad (2)$$

在上式当中, n_1, n_2 各自代表的是样本子集 S_1, S_2 对应样本数。

对于属性 A , 先带入数据确定出 GINI 系数, 然后进行排序确定出其中的最小值, 设置为此特征条件下的最优二分方案:

$$\min_{i \in A}(GINI_{A,i}(S)) \quad (3)$$

对集合 S , 带入数据计算出全部特征的最好二分方案, 进行排序后得到最小值, 将其设置为集合 S 的此种方案:

$$\min_{A \in Attribute} \min_{i \in A}(GINI_{A,i}(S)) \quad (4)$$

2) 连续型数据——选择 σ 值, 进行分类预测。

回归树方法在进行预测分析时, 设置 σ 来评价分类效果, 确定出 σ 最小的特征及值, 对应于最佳的分类特征和值。一般条件下 σ 值大则可判断出分类后子集的“差异性”大, 相反情况下则可判断出此特征下, 可取得更优分类结果。

对连续型集合 S , 总方差表达式为:

$$\sigma(S) = \sqrt{\sum (y_k - \mu)^2} \quad (5)$$

在上式当中: μ 代表的为样本集针对 S 中预测结果之下的均值; y_k 代表的为第 k 个样本对应预测结果。

集合 S 中含有 N 个样本条件下, 基于特征 A 的第 i 个特征值, 对其进行二分处理, 划分后可通过如下方法计算出 σ :

$$\sigma_{A,i}(S) = \sigma(S_1) + \sigma(S_2) \quad (6)$$

接着对特征 A , 带入数据进行分析确定出任意特征进行划分条件下的 σ , 然后排序得到最小值, 这样可确定出 A 下的最优分类方案:

$$\min_{i \in A}(\sigma_{A,i}(S)) \quad (7)$$

对集合 S , 带入数据进行处理确定出全部特征的最优分裂方案, 确定出其中最小者作为集合 S 的此类方案:

$$\min_{A \in Attribute} \min_{i \in A}(\sigma_{A,i}(S)) \quad (8)$$

这样处理后得到的特征 A 及值, 就表示为集合 S 的最优二分属性和相关的属性值。

1.2 构建随机森林模型

气象数据云平台存储的数据种类不断增加, 且数据类型也日益复杂, 涉及到雷达、卫星、地面、高空、海洋等相关的气象数据, 基于组织形式不同对这些数据进行划分, 可分为结构化和非结构化数据。分析气象数据特征, 设计对应的大数据云平台整体解决方案, 用于对天气状况进行预测、分类, 使其应用到农业、交通、旅游等领域。气象资料和服务数据大部分为视频、音频、文本, 在长时间运

行后对应的数据量急剧增长，符合大数据的 4 V 特点。

在大数据领域中，Hadoop 是最著名的大数据处理框架之一，它以可靠、高效、可伸缩的方式进行大数据的存储、处理和分析。在经济社会发展过程中，随着信息化水平不断提高，数据量与日俱增，在对这些数据进行处理时，传统的随机森林模型开始不满足应用要求，这样就需要引入新的数据分析技术。如为提高算法的处理性能，可引入并行化的决策树算法，这样可以综合利用更多的资源进行同步处理。对并行化相关的算法，其并行化时可选择的模式主要有三类，其一为数据间并行，后两种则为属性间、节点间的并行方式，其各有一定的适用范围。运用决策树算法，基于 MapReduce 并行创建 CART 树形成随机森林模型，建立气象业务预测以及各领域服务功能模块。

基于随机森林算法的特征和 MapReduce 性能，考虑到研究要求，采用下述的并行化策略：对原始数据的随机采样的方式选取，建立相关决策树模型过程中，随机筛选对应的属性，对应的样本集保持独立。因而建立的单棵决策树间不存在相关性。这种模型在进行并行化操作时，主要应用 MapReduce 编程实现，相关的工作流程可划分为，设置决策树、确定出森林模型、结果投票。在 MapReduce 中并行创建 CART 树，构建随机森林模型并行化过程如图 1 所示。

具体分析图 1 进而可发现，这种模型建立过程中主要是并行化单棵树的建立过程。首先基于 Map 方法切片处理 HDFS 中数据，接着通过 Bagging 随机抽样，将抽取后的数据划分为训练集和属性集，然后分布到集群的各节点上，在此基础上利用并行化 Map 任务进行分析，而得到其中各

属性的 Gini 值，将所得结果输出到 Reduce 中，利用 Reduce 功能进行处理，而得到最优的属性、节点的编号，这样既可以获得所需要的单棵决策树，对树的参数进行输出后，对多棵 CART 树进行一定的集群映射而得到随机森林。在一定投票基础上汇总多棵树的结果，在此处理时可调用 Map、Reduce 函数。

在气象大数据云平台下，在完成收集气象相关数据处理后，创建相应的预报预警。此时将相关样本集经由 MapReduce 并行创建 CART 树形成随机森林，训练模型，根据相关气象数据进行预测天气预测或分类，并由控制器将新生成的气象情况以及其它相关附加信息等一并传输至各个功能模块中，使得气象大数据云平台中应用层功能模块能够实现动态数据更新，更好的为农业、交通、旅游等领域服务。

2 气象大数据云平台的设计

2.1 Hadoop 生态系统架构

Hadoop 是谷歌公司研发的高性能云计算平台，在分布式计算领域被广泛应用。HDFS 是 Hadoop 的分布式文件系统，也是 Hadoop 系统的最重要组成单元，HDFS 集群中包含了大量的服务器，这样可以通过并行模式来高效的处理海量气象数据。在 MapReduce 模型下，程序可以高效快速的处理大量集群上的海量数据。

Hadoop 云平台在不断发展改进基础上其性能水平明显提升，功能组件也日益丰富，满足各领域的分析要求。其中的模块主要包括固定模块以及特定功能模块，根据相关资料可知其中的核心模块主要如下。

1) Hadoop Common: 为其中的基础模块，可为其他单

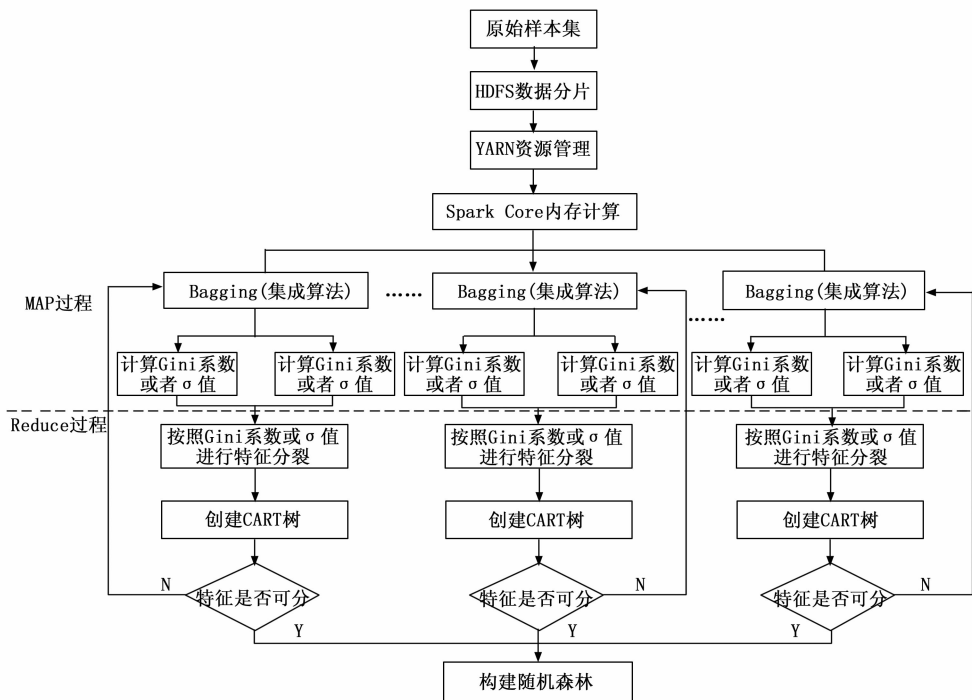


图 1 并行创建 CART 树构建随机森林模型

元的功能实现提供支持, 这种模块中设置了很多公用 API 接口, 与 HTTP 认证功能密切相关。

2) HDFS: 主要是对海量数据进行分布式存储和处理, 在数据处理时依据一次写入多次读取相关原则, 此外还可提供流式数据访问。

3) YARN: 为一种全新框架, 发展初期的 MapReduce 功能繁杂, 对系统运行可靠性产生很不利影响, 因而此后分布的版本中独立的处理数据和资源, 这样就通过 YARN 模块进行资源管理, 且可实现一定调度功能

4) MapReduce: 为一种高性能的分布式计算框架, 在进行编程过程中不需要考虑到数据存储相关因素, 从而大幅度简化了编程过程。主要是编写 Map ()、Reduce () 方法就可满足要求。此种平台的系统结构如图 2。

Hadoop 中的组件主要如下。

1) ZooKeeper: 主要是提供集群相关的协调服务, 主要是改进 YARN 形成的, 在应用过程中可实现的功能主要包括集群管理、分布式锁等, 可据此实现特定的功能。

2) Flume: 主要的作用是进行日志采集和传输, 对应的采集模式有两种, 分别为在单服务器和集群上采集, 对获得的日志进行汇总, 然后在一定统一格式基础上发送到 HDFS 中。

3) Spark: 为一种高性能的计算系统, 其功能强大, 可实现机器学习、图形计算相关的功能。此外也有很强的支持性, 对 Java、SQL 等相关的语言都支持。可以运行在 EC2、YARN 框架之上而满足应用要求。

4) HBase: 这种模块主要是基于“Bigtable”思想建立的, 可以看作为一种特殊的分布式数据库, 可以为集群的协调提供支持。

5) Hive: 在其支持下, 用户可对 HDFS 上的数据通过

SQL 相关的语言进行读写和查询处理, 对降低编程难度有重要意义, 使得用户对 Hadoop 上手更容易。

2.2 MapReduce 任务流程

MapReduce 在进行数据处理时的操作主要包括 Map、Shuffle、Reduce 阶段, 如图 3。以下对这三个阶段的工作内容和特征进行说明:

1) Map 阶段。在此阶段处理过程中, 应用程序基于输入的数据进行分片处理, 建立起一定量的 Map 任务, 接着对分片数据依据相关的规则映射形成一定量键值对, 然后通过 Map () 方法处理这些键值对, 处理后的结果保持为 <key, value>的形式, 且进行输出。

2) Shuffle 阶段。上一阶段处理后的结果为无规则的, 然后通过 Shuffle () 方法进一步转换处理, 形成规则的数据输出。在此环节处理时主要是合并同大小的键值对, 进行排序后形成相应的键值对列表。

3) Reduce 阶段。上一环节处理后的数据通过 Reduce () 方法处理, 将上述的列表进行汇总, 且基于特定的函数输出。

在实际处理时, 用户只需要编写出 Map () 和 Reduce () 方法, 在运行过程中程序可自动的调用这些方法进行操作, 从而满足特定的功能要求。

2.3 气象大数据云平台集群部署结构

气象大数据云平台的构建基于 Hadoop 集群、Impala 集群和 GBase 集群构建。其中大数据云平台的数据层中全部与气象密切相关的数据进行适当的分类处理后, 形成一定量的数据块, 这些数据可基于特征和来源不同进行划分, 而分为互联网数据、结构化数据、气象外部数据等。Impala 集群在运行过程中可接收结构化和基础数据, 且调用函数来实现数据处理功能。Hadoop 集群可对非结构化数据进行

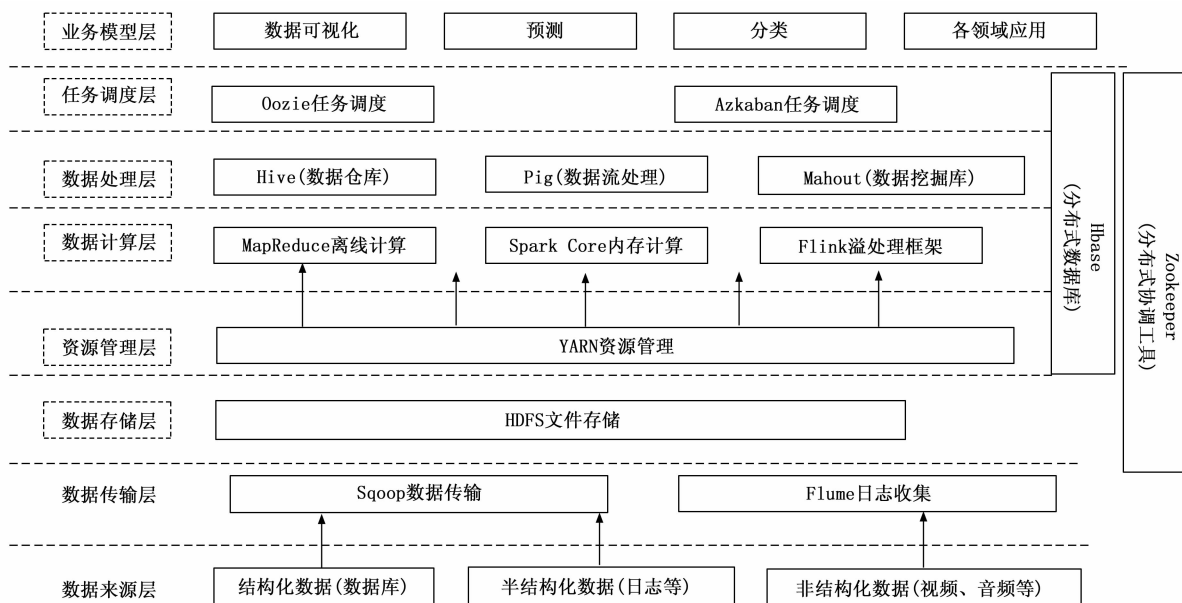


图 2 Hadoop 生态系统架构

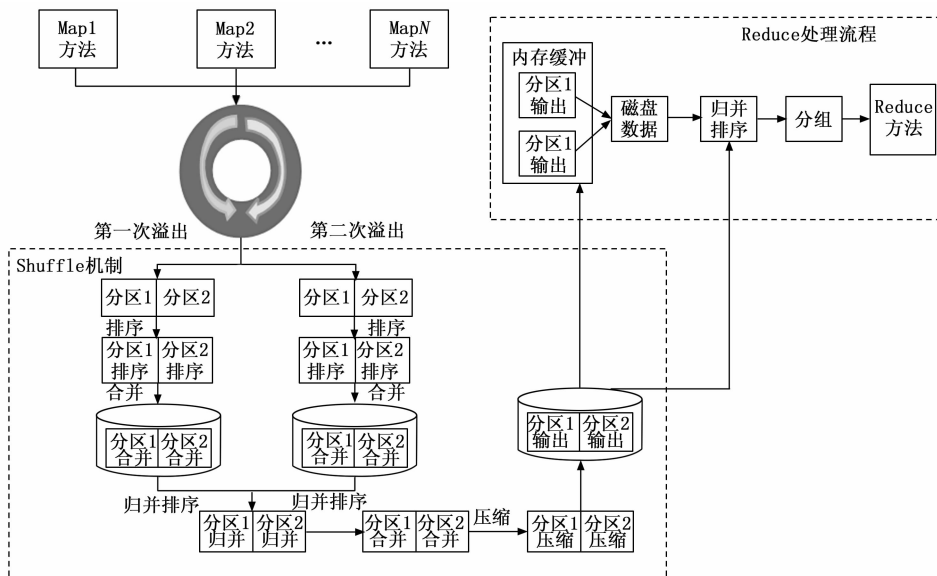


图 3 MapReduce 任务流程

存储,其中包含了 HDFS, HBase 相关的单元。GBase 集群在也应用过程中可对相关专题应用数据进行存储,系统上层设置 Solr 集群,这样在运行过程中可充分利用 Solr 的集中式配置、容错相关功能,满足高索引和访问相关的数据应用需求。Solr 集群可对索引数据进行存储,其上层设置了气象数据统一服务接口,在运行过程中大数据云平台,气象管理人员可通过这一接口来获取不同类型的数据。此外,还可以在后台上部署各种格式的数据,对相关数据进行统一集中的展示,为其后的处理提供支持。

2.4 气象大数据云平台总体架构

在以上基础上建立起大数据可视化平台,在运行过程中其可以实现复杂的管理功能,如调度管理、元数据管理、分布式 ETL 管理、质量管理,此外也可以实现数据共享功能。为系统的运行维护以及功能实现提供支持。总体架构分为基础设施层、数据处理与管理层、应用层,气象大数据云平台总体架构如图 4。在运行过程中为确保数据高效处理,这种平台在应用时可以全流程管理各类型的气象数据信息,且可以方便的进行交换、质控,表现出较高的应用性能优势。

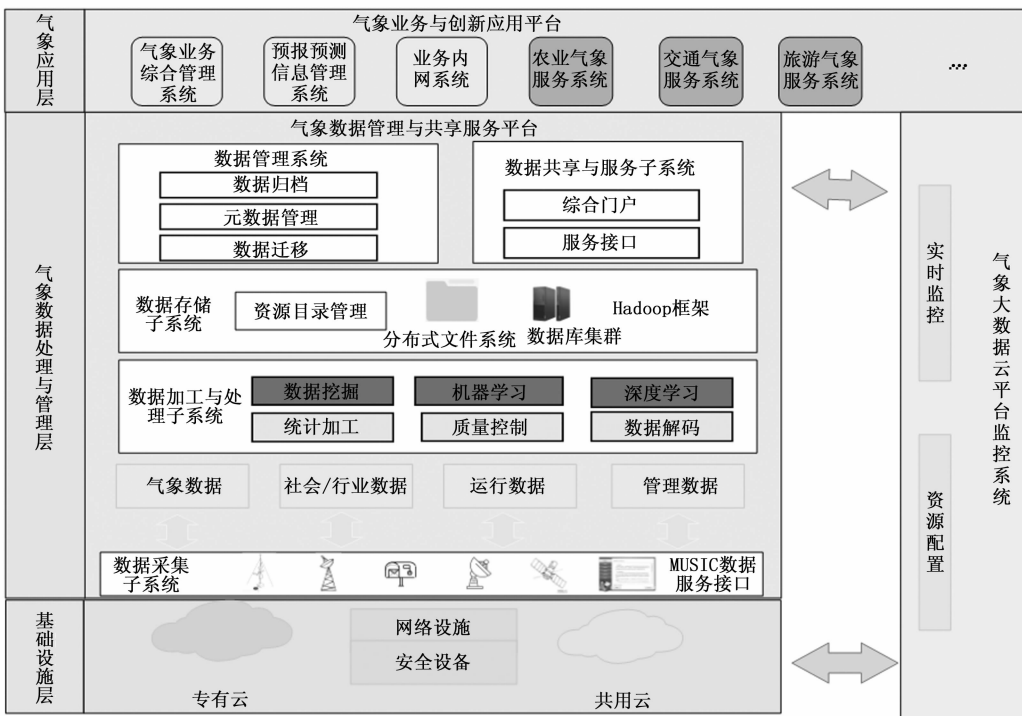


图 4 气象大数据云平台总体架构

表 1 平台性能测试

数据类型	响应时间(毫秒)							
	CIMISS 系统				气象大数据云平台			决策树挖掘算法
	统计	质控	解码	加权平均(10:1:1)	Map	Reduce	小计	数据挖掘、机器学习、深度学习
气象数据	820	930	1050	848	75	127	202	137
行业数据	1200	110	850	1080	88	139	227	159
运行/管理数据	945	—	—	945	—	—	350	256

大数据云平台层对全部的气象数据可基于接口机集群高效处理, 进行一定转换后得到各种类型的数据产品, 且对不同类型的数据进行分类管理和存储; 对采集的原始数据通过分布式 ETL 工具抽取、清洗操作后, 接着基于设定的仓库模型进行加载处理。对这种架构而言, 大数据云平台有重要的意义, 也是其中的最重要单元, 能够提供气象业务综合管理、天气预测预报、农业气象服务、交通气象服务、旅游气象服务等功能。

3 平台性能测试与分析

全国综合气象信息共享系统 (CIMISS) 是气象部门数据收集、存储、加工和共享的业务支撑系统, 通过数据统一服务接口整合 CIMISS 与本地数据服务, 提供数据访问服务和应用服务。对比测试平台的性能, 将相同大小的任务分别在 CIMISS、气象大数据云平台及基于 CART 气象大数据云平台上运行, 响应时间如表 1 所示。其中, 在 CIMISS 中, 根据统计、质控、解码应用频率, 按照 10:1:1 频次做加权平均; 气象大数据云平台按照 Map 和 Reduce 任务统计响应时间, 考虑 HDFS 分布式特性, 对不同大小数据处理性能也有不同, 通过 Map 操作、设置 reducer 参数、任务大小提高平台执行效率, 其中 Map 和 Reduce 任务个数通过多次设置优化执行时间; 采用了决策树算法的气象大数据云平台, 自动选择数据挖掘、机器学习、深度学习来处理提交的任务。

结果可以看到, 气象大数据云平台对数据统计、质控、解码等处理性能均有较大提升, 平均提升 3.6 倍。采用决策树挖掘算法后, 平台性能可以提升到 1.4 倍。

4 结束语

当今已经进入大数据时代, 气象数据规模迅速的增长, 在此背景下如何高效的利用海量气象数据并充分发挥其资源价值, 就成为当前气象研究的热点。Hadoop 平台是内网分布式在运行时不会产生明显的延迟, 且简化了系统结构。气象大数据云平台的整体解决方案, 对收集气象相关数据进行处理, 将相关样本集经由 MapReduce 并行创建 CART 树形成随机森林, 训练模型, 使得气象大数据云平台中应用层功能模块能够实现动态数据更新, 更好地为其他领域服务。Hadoop 平台表现出明显的异构和集群化特征, 未来很有必要研究异构下的数据集处理方法, 更好的满足异构条件下的数据处理要求, 同时也为并行算法的广泛应用提供支持。在气象信息管理的实际应用中, 许多数据是复杂异构的, 有很多非结构化数据, 下一步可以对非结构化气

象数据挖掘方法展开研究。

参考文献:

- [1] 杜建华, 王立俊, 王双双. 基于 CIMISS 的海南气象信息服务网设计与实现 [J]. 计算机测量与控制, 2022, 30 (3): 244-250.
- [2] 胡争光, 薛峰, 于连庆. 海量气象数据计算处理及可视化在决策气象服务移动平台上的应用 [J]. 气象科技, 2020, 48 (5): 615-621.
- [3] 杨永毅. 基于分布式架构的气象数据网设计 [J]. 信息技术, 2020, 44 (10): 6.
- [4] 周雪莹, 李芬, 李显凤, 等. 气象水文信息实时共享系统的设计与实现 [J]. 计算机技术与发展, 2020, 30 (7): 194-198.
- [5] 彭显忠, 王谦, 元昌安, 等. 数据挖掘技术在气象预报研究中的应用 [J]. 干旱气象, 2015, 33 (1): 19-27.
- [6] 刘钊, 谷星月. 形成开放融合协同数字气象业务生态 [N]. 中国气象报, 2022-01-14 (1).
- [7] 曾乐, 孙超, 张来恩, 等. 基于大数据技术的气象业务监视数据采集处理 [J]. 计算机仿真, 2021, 38 (7): 181-188.
- [8] 朱雪峰, 高美美, 侯康. 网络环境下气象数据自动查询系统设计 [J]. 计算机测量与控制, 2017, 25 (8): 261-264.
- [9] 刘骥超, 叶钊, 谢寒生. 云计算环境下气象大数据的应用研究 [J]. 计算机技术与发展, 2019, 29 (5): 168-171.
- [10] BILGIN T T, ZAMURCU A Y. A Data Mining Application on Air Temperature Database [C] // International Conference on Advances in Information Systems. Springer-Verlag, 2004: 68-76.
- [11] REIKARD G. Combining frequency and time domain models to forecast space weather [J]. Advances in Space Research, 2013, 52 (4): 622-632.
- [12] 邱声春. 数据挖掘和数据融合技术在天气预报和气象服务中的应用研究 [J]. 山西气象, 2007 (2): 34-36.
- [13] NUI Z F, ZHU Y A, JIANG L X. Distance Metric Learning Approach for Weather Data Mining [J]. Bdiot 2018: Proceedings of the 2018 2nd International Conference on Big Data and Internet of Things, 2018, 4 (10): 33-36.
- [14] FAROUK M, ALI M, ASKLANY S A, et al. Data Mining Algorithms for Weather Forecast Phenomena: Comparative Study [J]. International Journal of Computer Science and Network Security, 2019, 19 (9): 76-81.
- [15] 王丹妮, 包世泰, 王春林, 等. 基于气象数据挖掘的广东省农业高温灾害预测研究 [J]. 华南农业大学学报, 2015 (2): 106-112.
- [16] 陈凯, 王立松. 一种新的加权最近邻算法的降水预报试验

[J]. 计算机仿真, 2014, 31 (6): 325-328.

[17] 柯秀文. 基于二型模糊数据挖掘的船用气象数据网络应用 [J]. 舰船科学技术, 2017, 39 (10): 164-166.

[18] 胡冬梅. 基于时空数据挖掘的城市空气污染分析与预测 [D]. 北京: 清华大学, 2018.

[19] 王春政, 胡明华, 杨磊, 等. 基于 Agent 模型的机场网络延误预测 [J]. 航空学报, 2021, 42 (7): 452-465.

(上接第 115 页)

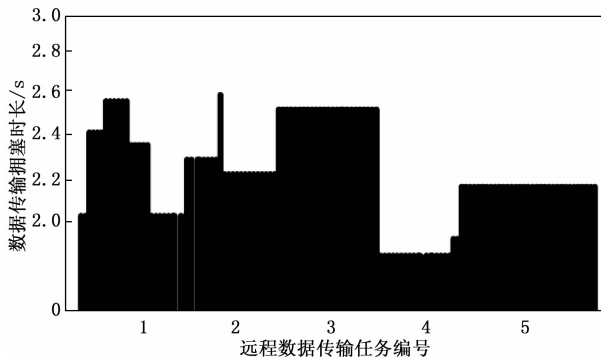


图 7 控制系统下远程数据传输拥塞总时长统计

控制功能指标的测试结果均低于预设值, 即优化设计系统的控制功能满足设计要求。另外通过式 (8) 的计算得出系统运行性能的测试结果, 如图 8 所示。

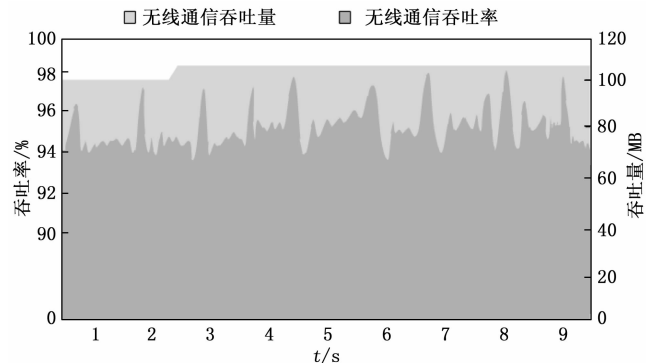


图 8 通信远程数据传输控制系统运行性能测试结果

从图 8 中可以看出, 无线通信远程数据传输控制系统吞吐率的最小值为 91.5%, 由此可见设计控制系统的运行性能满足设计与应用要求。

4 结束语

从系统测试结果中可以看出, 通过 STM32 芯片的嵌入式应用, 有效提高了无线通信远程数据传输控制系统的控制功能与运行性能, 对于无线通信技术而言具有较高的应用价值。然而实验中设置的传输样本数量较少, 且未对数据传输延迟进行测试, 因此得出的系统测试结果存在一定的局限性, 因此在控制系统投入应用之前, 还需要进一步补充相关的测试数据。

参考文献:

[1] 赵佳, 李永亮, 曹磊. 基于 STM32 的无线视频传输控制

[20] 张晨阳, 马志强, 刘利民, 等. Hadoop 下基于粗糙集与贝叶斯的气象数据挖掘研究 [J]. 计算机应用与软件, 2015, 32 (4): 72-76, 90.

[21] 王昊, 师卫, 李欢. Hadoop 下基于贝叶斯网络的气象数据挖掘研究 [J]. 电子器件, 2016, 39 (4): 841-846.

[22] 黄飞龙, 谭晗凌, 霍亚. 物联网与自动气象站深度融合方法研究 [J]. 计算机测量与控制, 2022, 30 (3): 217-221.

[1] 系统设计 [J]. 电子制作, 2021 (13): 85-86, 68.

[2] 陈远港, 杨明山, 王梦灏, 等. 基于直流载波通信的 LAM-OST 控制系统设计 [J]. 测控技术, 2021, 40 (10): 79-84.

[3] 周言, 程仁杰, 周晓亮. 工业控制系统数据传输明密文分流处理技术的研究与应用 [J]. 中国石油和化工标准与质量, 2020, 40 (11): 233-234.

[4] 王波, 肖寒, 张捷. 基于 ARM 的数据安全传输系统的软件设计 [J]. 工业控制计算机, 2021, 34 (2): 14-16, 20.

[5] 马鑫龙, 艾冬生, 孙发鱼, 等. 基于 MQTT 的遥测数据无线传输 [J]. 探测与控制学报, 2020, 42 (2): 61-64.

[6] 宋大伟, 马凤娟. 基于流量延时调度机制的大数据传输控制算法研究 [J]. 宁夏师范学院学报, 2020, 41 (10): 47-52, 60.

[7] 苏成志, 侯代国, 原泽栋. 车载云数据传输拥塞优化控制 [J]. 制造业自动化, 2020, 42 (5): 92-94.

[8] 陈瑞军, 孟伟君, 孟飞, 等. 基于 PC+控制卡的机器人手势容错控制系统设计 [J]. 计算机测量与控制, 2020, 28 (7): 127-131, 147.

[9] 赵健. 大数据通信仪表的自动化传输控制模型研究 [J]. 制造业自动化, 2021, 43 (8): 104-108.

[10] 杨帆, 车现法, 陈波, 等. 云湛高速公路防雷监测系统数据传输控制通信协议 [J]. 广东公路交通, 2020, 46 (3): 59-62.

[11] 张静. 基于 Struts2 的医疗病案数据异地传输控制系统设计 [J]. 机械设计与制造工程, 2019, 48 (7): 52-55.

[12] 张大伟, 叶蓓蓓. 面向冗余数据消除的多维异质网络数据传输控制方法 [J]. 中国电子科学研究院学报, 2019, 14 (5): 519-523, 552.

[13] 张海波. 基于 RFID 的通信数据安全传输系统设计 [J]. 数字通信世界, 2021 (12): 97-99.

[14] 徐良. 异构环境下无线传感大数据跨域传输安全控制系统设计 [J]. 计算机测量与控制, 2020, 28 (12): 117-121.

[15] 龙恒. 一种基于 UDP 的可靠数据传输协议 [J]. 计算机时代, 2020 (4): 33-37.

[16] 吴婷. 基于稳定调度重传机制的 WSN 数据传输算法研究 [J]. 系统仿真技术, 2020, 16 (1): 40-45.

[17] 金鑫, 杨阳, 李振兴, 等. 基于模糊 L₁(1/2) 正则化的无线传感网络拥塞控制研究 [J]. 长春理工大学学报 (自然科学版), 2019, 42 (4): 66-70.

[18] 常梦磊, 罗述翔, 李幸睿, 等. 低时延传输的 ERDQN 数据调度算法 [J]. 哈尔滨工业大学学报, 2021, 53 (8): 132-136.

[19] 孙浩凯, 刘博, 佟世继. 基于双模通信的路权控制系统设计 [J]. 物联网技术, 2022, 12 (1): 90-94, 98.

[20] 张一宁. 通信传输线路的质量控制及优化策略研究 [J]. 中国设备工程, 2022 (1): 145-146.