面向电力大数据分布式云存储的性能分析与优化

在电力调度系统中，电力数据以大数据散布的形式分布在云存储系统中，由于云存储节点和数据分布在城市电力调度环境具有非均匀自组织特性，制约电力大数据的调度效率，造成云存储冗余，需要进行性能优化。传统的电力大数据分布式云存储和数据调度算法采用能量谱密度函数特征提取算法，由于能量谱密度函数多模盲均衡特性，造成云存储系统中的干扰数据较多，电力数据调度效率不好。提出一种基于宽频带数据特征压缩的电力系统大数据分布式云存储算法。构建面向电力系统大数据调度的分布式云存储系统模型，采用匹配滤波器检测方法进行大数据特征预处理，构建分布式数据聚集树实现电力大数据聚集调度，通过宽频带数据特征压缩对数据特征进行信息降维融合，减少云存储数据冗余，提高电力数据调度效益。仿真结果表明，采用该算法进行电力调度系统的数据云存储设计，降低了电力数据的聚集和调度延时，提高数据传输的性能，电力系统的抗干扰性能较好，展示了较好的应用价值。

王曙霞／湖北工程学院计算机与信息科学学院
陈茂辉／孝感供电公司

王曙霞／副教授

关键词/Keywords
大数据·
电力系统·
分布式·
调度·

在电力电气研究领域，通常的电力系统划分为发电、输电和配电三大部分。电力调度系统是电力系统中从降压配电变电站（高压配电变电站）出口到用户端，实现电力分配和电力调度的智能管理系统，城市电力调度系统作为一种最重要的基础性设施，在人们的生产和生活中发挥重要的作用。在电力配电运输生产和建设中，电力调度系统利用各种通信方式进行各种信息传送和处理。这些数据信息以大数据的形式寄存在电力调度云存储系统中，可见，在电力配电和调度系统中，对云存储系统的大数据调度和通信设计是关键，需要进行电力配电通信网络管理，提高大数据的存储和调度性能。然而，在电力调度系统中，电力数据以大数据散布的形式分布在云存储系统中，由于云存储节点和数据分布在城市电力调度环境具有非均匀自组织特性，制约电力大数据的调度效率，造成云存储冗余，需要进行性能优化。因此，研究电力大数据的分布式优化云存储技术具有重要的价值和意义，得到了广泛关注[1]。

传统方法中，对电力调度大数据的云存储调度技术主要有线性规划电力大数据网格资源云存储调度算法、网络空间的资源调度挖掘算法和基于神经网络自组织映射的大数据的云存储分布式设计方法等。其中，现有的电力大数据的云存储开源系统主要有Lessfs和Open Dedup系统，其建立在应用层之上，进行电网组网数据调度，然而传统的云存储模型存在数据调度冗余[2]，计算复杂等问题。对此，相关文献进行了算法改进，其中B. Hunt、J. Marin和P. T. Stone等人提出了C/S负载均衡大型电力系统的云存储优化调度和数据融合技术，采用决策树ID3算法对电力系统的重复大数据进行删除和融合，提高数据的存储性能[3]；文献[4]提出一种基于水平阵列信号处理的大型电力数据的云存储数据布局整合方法，实现对电力数据的优化调度，但算法对数据聚集特征具有不可预测性，导致调度延时。文献[5]采用能量谱密度函数特征提取算法，由于能量谱密度函数多模盲均衡特性，造成云存储系统中的干扰数据较多，电力数据调度效率不好。

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

针对上述问题，本文提出一种基于宽频带数据特征压缩的电力系统大数据分布式云存储算法。论文首先构建面向电力系统大数据调度的分布式云存储系统模型，采用匹配滤波器检测方法进行大数据特征预处理，构建分布式数据聚集树实现电力大数据聚集调度，通过宽频带数据特征压缩对数据特征进行信息降维融合，减少云存储数据冗余，提高电力数据调度效益。最后通过仿真实验进行性能验证，展示了本文算法在实现电力调度大数据的优化云存储方面的卓越性能，得出了有效性结论。

电力系统大数据调度的分布式云存储系统模型构建及数据特征分析

1. 电力调度分布式云存储系统模型构建

为了提高整个电力系统的调度控制性能，需要进行云存储分布式设计，构建电力调度分布式云存储系统模型，整个电力调度控制网络模型用一个连通的无向图$G=(V, E)$表示，所有节点都有相同的传输半径r，WSN网络的边$(u, v)\in E$。电力调度模型中，设$A\subset V$，$B\subset V$，且$A\cap B=\varphi$，每个传输调度集$S_i(i=1, 2, \cdots, L)$满足以下条件

$$S_i\cap S_j=\varphi,\quad \forall i\neq j \tag{1}$$

$$\bigcup_{i=1}^{L}S_i=V-\{Sink\} \tag{2}$$

电力调度分布式云存储系统根据分码的结构形成一个包含n个向量的向量组$\boldsymbol{A}=\{\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_n\}$。其中，$\boldsymbol{A}_j$表示包含$M$个已调节数据向量，为了保证在任何一次数据传输过程中，不产生干扰冲突，需要对电力配电数据进行完整性检测[6]。传统方法通过混合映射建立两种域间链接可以解决权限隐蔽提升和角色映射冲突问题[7]，但是却提高了管理难度。本文采用联合特征信息增益提取方法，构建多路复用大型电力调度云动态数据采集模型。本文假设被处理云采集数据是可分类的，引入了一个云存储管理因子$\beta\subseteq(0, 0.5)$。在这个管理因子的限定范围内，实现大数据的信息融合，以及大数据调度标签生成阶段（TagBlock），采集数据集合S。当$X\subseteq U$，$R\subseteq A$条件成立时，电力调度大数据信息系统状态响应函数表达式为

$$h(t)=\sum_i a_i(t)\mathrm{e}^{\mathrm{j}\theta_i(t)}\delta(t-iT_S) \tag{3}$$

式中，a_i为电力调度大数据信息系统的调度个数；t为调度状态的响应时间。

根据通信理论中频分复用理论，得到大数据频率与行为关系，常用多普勒效用表示，多普勒频移用公式表示为

$$y_b=y_0\frac{w}{z}\cos\varphi \tag{4}$$

式中，y_b为电力调度通信接收端检测到的发射频率变化量；y_0为电力调度通信基站发射端的载波频率；z为传输功率。

通过上述设计，进行电力调度分布式云存储系统模型构建，为下一步进行电力调度数据分布式云存储设计提供模型基础。

2. 电力调度分布式大数据聚集树的构造和特征分析

在上述构建面向电力系统大数据调度的分布式云存储系统模型的基础上，采用匹配滤波器检测方法进行大数据特征预处理，引入C4.5决策树模型进行冗余数据住特征建模，如图1所示。

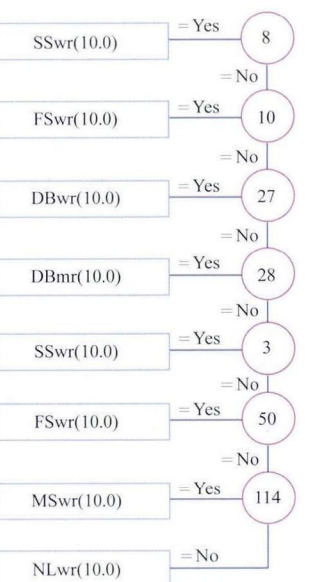


图1　C4.5决策树模型进行云存储冗余数据住特征建模

图1中，云存储冗余数据伴随跟踪属性取值{0，1}映射到决策树中表现为映射值{No，Yes}，C4.5决策树下主特征设置特征聚类类型参量值为

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

164类，提取的主特征类型为19类，进行云存数据的特征降维和压缩[8-9]，得到数据调度的时间减弱函数为

$$f(k)=\begin{cases} f(k-1)-\dfrac{1}{n} & 1\leqslant k<n \\ 1 & k=n \end{cases} \tag{5}$$

由此得到云存储系统中主特征量有效行为概率计算公式为

$$D''_{i,j}(t_{n+1})=\frac{D'_{i,j}(t_{n+1})+f(n)D'_{i,j}(t_n)}{2} \tag{6}$$

式中，t_{n+1}时刻和t_n时刻相差一个更新周期。此时，电力调度系统相当于一个多层节点的非线性映射器，其重组规律定义为

$$u(t)=K_{\mathrm{p}}[e(t)+\frac{1}{T_{\mathrm{I}}}\int_0^t e(t)\mathrm{d}t+T_{\mathrm{D}}\frac{\mathrm{d}e(t)}{\mathrm{d}t}] \tag{7}$$

式中，K_{p}为电力调度通信的效率；e为调度系数。

采用匹配滤波器检测方法进行大数据特征预处理构建匹配滤波器，滤波器的系统函数为

$$H_{\mathrm{B}}(z)=\frac{(1+\sin\theta_2)}{\cos\theta_2}\cdot \frac{\cos\theta_1(k)\cos\theta_2 z^{-1}}{1+\sin\theta_1(k)(1+\sin\theta_2)z^{-1}+\sin\theta_2 z^{-2}}G(z) \tag{8}$$

其中

$$G(z)=\frac{1-\sin\theta_2}{2}\cdot \frac{1-z^{-2}}{1+\sin\theta_1(k)(1+\sin\theta_2)z^{-1}+\sin\theta_2 z^{-2}} \tag{9}$$

在上述匹配滤波设计中，$x(k)$为观测序列，它是由信号和背景噪声共同构成的；$y(k)$为匹配器的输出，由此实现电力调度分布式大数据的特征分析和提取，为进行优化云存储及数据调度提供准确的数据基础。

宽频带数据特征压缩的提出及大数据分布式云存储算法实现

1. 宽频带数据特征压缩处理

在上述模型设计和特征预处理的基础上，进行电力系统大数据的云存储优化设计，通过上述分析可见，传统的电力大数据分布式云存储和数据调度算法采用能量谱密度函数特征提取算法。由于能量谱密度函数多模盲均衡特性，造成云存储系统中的干扰数据较多，电力数据调度效率不好。对此，本文提出一种基于宽频带数据特征压缩的电力系统大数据分布式云存储算法。首先给出宽频带数据特征压缩处理算法的关键技术，采用匹配滤波器检测方法进行大数据特征预处理的基础上，以特征预处理结果为输出向量，此时特征具有高维性，在云存储中形成存储空间冗余。本文为了降低存储冗余，降低计算开销，通过宽频带数据特征压缩对数据特征进行信息降维融合，进行宽带压缩频谱的联合特征识别，其检验统计量为

$$y=\sum_{k=1}^{N}|x(k)|^2 \tag{10}$$

式中，N为云存储节点的采样点数。通过云储存实现资源调度，大量数据不断传输调度过程中产生异常数据，此时局部性交叉项信息链表示为

$$CUB(i)=\frac{CBT(i)}{T}\times CB \tag{11}$$

式中，CB为传输调度中产生的异常数据个数；T为传输调度产生异常数据的时间。

大数据分布式频谱感知的聚类中心为

$$f(x)=\beta\mathrm{e}^{-\beta x} \tag{12}$$

式中，β为频谱感知系数。

采用竞争集来描述一个节点在传送数据时，节点u的竞争集的定义如下

$$CS(u)=N(p(u))\cup(\cup_{v\in N(u)\setminus Ch(u)}CH(v))\setminus\{u,p(u)\} \tag{13}$$

两个云$C_1(Ex_1, En_1, He_1)$，$C_2(Ex_2, En_2, He_2)$通过云发生器各产生出n个云滴分区区域，构建模糊隶属函数，多源节点中形成新的映射

$$\boldsymbol{x}_n=[x(0),x(1),\cdots,x(N-1)]^{\mathrm{T}} \tag{14}$$

把文件块和文件块的标签信息关联到S－Table上。其中$h(\cdot)$为同态哈希函数，设计宽频带数据特征压缩能量检测器，如图2所示。此时，节点发送给数据融合中心的概率为

$$\begin{aligned} p_d&=\int_0^d f(x)\mathrm{d}x \\ &=\int_0^d \beta\mathrm{e}^{-\beta x}\mathrm{d}x \\ &=1-\mathrm{e}^{-\beta d} \end{aligned} \tag{15}$$

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

式中，β为数据融合中心的数据集；x为节点个数。

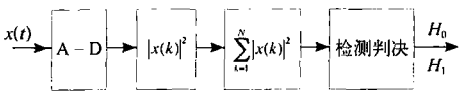


图2　宽频带数据特征压缩能量检测器

基于上述二元假设模型，构建的检验统计量y服从渐进的正态分布，通过能量检测和判决，实现大数据的云存储状态特征压缩。

2. 电力大数据优化分布式云存储实现

在上述宽频带数据特征压缩能量检测设计的基础上，通过构建分布式数据聚集树实现电力大数据聚集调度，减少云存储数据冗余，提高电力数据调度效益。通过特征压缩，得到电力大数据的分布式云粗糙带宽压缩优化特征识别的虚警概率和检测概率分别为

$$P_F=\sum_{j=k}^{N}\sum_{\sum u_i=j}\prod_{i=1}^{N}(P_{fi})^{u_i}(1-P_{fi})^{1-u_i} \tag{16}$$

$$P_D=\sum_{j=k}^{N}\sum_{\sum u_i=j}\prod_{i=1}^{N}(P_{di})^{u_i}(1-P_{di})^{1-u_i} \tag{17}$$

式中，N为电力大数据虚警的统计总和；P_{fi}为接收信号的时段数；P_{di}为电力大数据的配置参数。

对电力调度系统中的大数据设定脉冲压缩方式为LFM和二次调频信号表示，那么可以将云存储中的大数据特征核函数写成

$$k_z^g(t,\tau)=\prod_{k=-q/2}^{q/2}[z(t+c_k\tau)]^{b_k} \tag{18}$$

式中，t为云存储所需的时间；z为云存储中的数据集。

基于动态频谱接入的认知技术，得到大数据的差异化特征识别数学模型为

$$y_i=f(x_1,x_2,\cdots,x_m) \tag{19}$$

对于两个标量时间序列y_1和y_2，其联合概率密函数为$f(y_1, y_2)$，计算大数据簇内的灰度相关特征，头节点增加无线功率来发送信标信息，得到$C_i\subseteq V$，$1\leqslant i\leqslant k$，将电力调度系统的大数据库的灰度相关特征切分为若干数据块Chunk，由此实现了大数据分布式云存储改进。

仿真实验与结果分析

为了测试本文设计的电力系统大数据云存储模型在电力数据调度和监测中的优越性能进行了仿真实验。仿真实验模拟了1个客户端和5个服务器节点，仿真实验的硬件环境为Intel Core3－530 1 G内存，操作系统为Windows 7，实验数据集采用KDD2012电力调度系统大数据库，云存储系统的构建采用Ubuntu 12.10进行系统设计，基于Hadoop平台进行C/S云平台设计。

在电力调度云存储系统中，包含了大量的云数据，根据城市调度数据的需求和实际应用环境，电力系统中的大数据格式分别为DOC、TXT、PPT、VMDK、EXE和PDF等。数据量从100 MB到1 GB，以100 MB为单位线性增长。在上述仿真环境和参数设计的基础上，首先进行电力调度大数据采集，得到数据采集的时域和频域波形如图3所示。

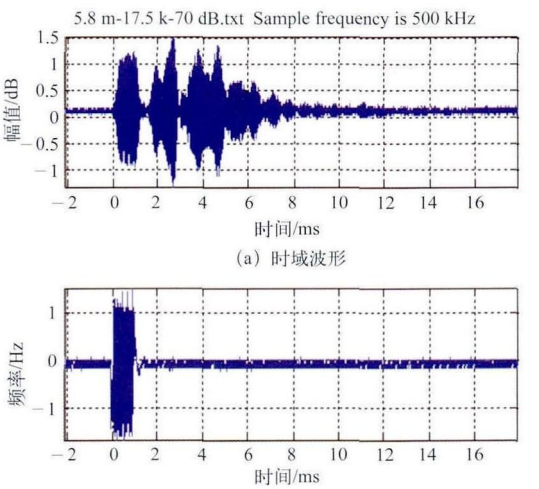
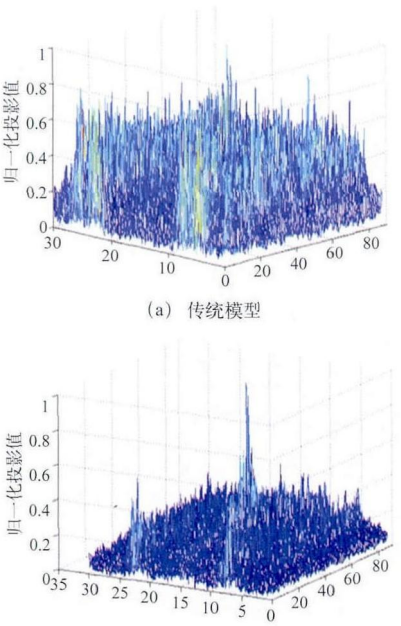


图3　电力数据采样时域和频域波形

图3中采样的数据信息样本作为大数据云存储系统调度的信息流研究样本，其采样频率为500 kHz，采样数据为TXT数据。以此为基础，进行大数据分布式云存储设计，采用本文模型和传统的云存储模型进行对比得到相同的数据样本在云存储系统的存储空间分布如图4所示。

分析上述结果可见，采用传统算法具有较大的数据冗余，影响电力数据的调度性能，需要进行云存储特征降维。通过本文设计的云存储模型进行信息降维融合，减少云存储数据冗余，提高电力数据调度效益。最后以不同算法进行电力大

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net



(a) 传统模型

(b) 本文模型

图4 云存储空间状态分布对比

数据调度聚集延时为测试指标，得到不同算法的聚集延时仿真结果如图5所示。分析可见，采用本文算法，降低了电力数据的聚集和调度延时，提高数据传输的吞吐量，电力系统的抗干扰性能较好，展示了较好的应用价值。

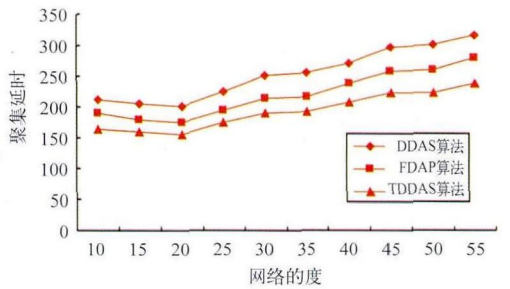


图5 不同算法的聚集延时

结束语

在电力配电运输生产和建设中，电力调度系统利用各种通信方式进行各种信息传送和处理。电力数据以大数据散布的形式分布在云存储系统中，由于云存储节点和数据分布在城市电力调度环境具有非均匀自组织特性，制约电力大数据的调度效率，造成云存储冗余，需要进行性能优化。提出一种基于宽频带数据特征压缩的电力系统大数据分布式云存储算法。构建面向电力系统大数据调度的分布式云存储系统模型，采用匹配滤波器检测方法进行大数据特征预处理，构建分布式数据聚集树实现电力大数据聚集调度，实现算法改进。仿真实验表明，采用本文设计的电力大数据云存储模型进行信息降维融合，减少云存储数据冗余，提高电力数据调度效益，展示了较好的应用前景。

参考文献

[1] 谢平．存储系统重复数据删除技术研究综述 [J]．计算机科学，2014，41 (1)：22-30.

[2] Miorandi D，Sicari S，Pellegrini F D，et al. Internet of things：vision，applications and research challenges [J]. Ad Hoc Networks，2012，10 (7)：1497-1516.

[3] Chen L，Brian K，Jamie E. Theoretical characterization of nonlinear clipping effects in IM/DD optical oFDM systems [J]. IEEE Transactions on Communications，2012，60 (8)：2304-2312.

[4] 朱继萍，戴君波．中长期电力负荷预测相关影响因素优化选择 [J]．计算机仿真，2008，25 (5)：225-228.

[5] 彭洲红，马国强．嵌入式电力监控系统的研究与实现 [J]．嵌入式电力监控系统的研究与实现，2005，18 (4)：12-13.

[6] 李红升．基于粒子群算法改进电力信息系统的安全研究 [J]．科技通报，2013，29 (4)：158-161.

[7] 张文亮，刘壮志，王明俊．智能电网的研究进展及发展趋势 [J]．电网技术，2009，33 (13)：1-11.

[8] 王晓波，樊纪元．电力调度中心统一数据平台的设计 [J]．电力系统自动化，2006，30 (22)：89-92.

[9] 蒋海波，王晓京，范明钰，等．基于水平纠删码的云存储数据布局方法 [J]．四川大学学报：工程科学版，2013，45 (2)：103-109.

(收稿日期：2014－02－10) EA

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net