

基于地理相关性的电力物联网数据 DCS 算法研究与应用

马成林 奚波 李静坤 徐森 孙开宁 杨龙

(国网新疆电力公司运营监测(控)中心,新疆 乌鲁木齐 830002)

摘要: 压缩感知是近几年来发展起来的理论,首先对压缩感知理论做了简单的介绍,重点阐述了 WSN 中基于地理相关性分布式压缩感知理论。在电力信息采集系统中,信息量比较大,所需传感器节点数目繁多,数据的传输量也很大,对数据进行压缩是减少数据传输量的有效途径,根据实际应用情况,选用了基于地理相关性分布式压缩感知理论对网络层数据进行压缩,并取得了理想的实验结果。

关键词: 物联网; DCS; 压缩; 分布式

Abstract: Compressive sensing is the theory developed in recent years. Firstly the compressive sensing theory is introduced briefly, and the emphasis is on distributed compressive sensing theory based on geographical correlation in wireless sensor network (WSN). In electrical power information acquisition system, the amount of information is large, the number of sensor nodes is various, and the amount of data transmission is massive, so compressing the data is an effective way to reduce the amount of data transmission. According to the actual application, the distributed compressive sensing theory based on geographical correlation theory is selected to compress the data of network layers, which obtains an ideal result.

Key words: internet of thing (IoT); DCS; compressed; distributed

中图分类号: TP393 文献标志码: A 文章编号: 1003-6954(2016)06-0054-05

DOI:10.16527/j.cnki.cn51-1315/tm.2016.06.013

0 引言

物联网技术是智能电网的重要支撑技术,智能电网和物联网的深度融合,将大大促进电网智能化产业发展。在电力采集系统中物联网技术大多依赖于无线传感器网络(wireless sensor network, WSN)^[1-3]。目前国家电网公司已完成输变电状态监测系统、智能变电站、配电自动化、统一视频及用电信息采集等。电力物联网的一个特点是以数据为中心,数据处理是其中一个关键问题,在很大程度上决定了网络的性能。在电力物联网中,一般信息感知环境区域较大,节点采集及处理的信息量比较大;所需传感器节点数目繁多,如何依据实际地形有效部署节点,并保证信息的全面性是个面临的问题;同时数据的冗余容易造成网络堵塞^[4-6];所以对数据进行融合是不可或缺的。压缩感知理论首先是由 Candes、Rom-Beg 提出,于 2006 年首先构建了压缩传感理论的框架,正式提出压缩传感(compressed sensing, CS)理论,信号的稀疏、测量矩阵的设计以

及信号重构算法的设计是压缩感知理论中很重要的理论部分。

1 算法模型

标准的 CS 理论主要基于单节点的内部数据结构设计,利用单节点感知数据的相关性来降低该节点测量数据的维度。在无线传感器网络中,传感器节点数量众多。对于空间密集分布的节点,相邻节点的数据之间必然存在很大的空间相关性。利用这一特性,自然联想到对各个节点协同合作形成测量值进行组合,利用节点之间数据的相关性来降低整体数据测量的维数,会得到更大的压缩效果。基于此,分布式压缩传感理论(distributed compressed sensing, DCS)应运而生。DCS 理论指出,如果多个信号都能够同一个基上稀疏表达,而且这些信号之间又是具有相关性的,那么每一个信号在编码端都可以利用另一个与该稀疏基不相关的基(例如一个随机矩阵)来对感知数据信息进行观测和编码,从而得到远远少于信号原始长度的编码。

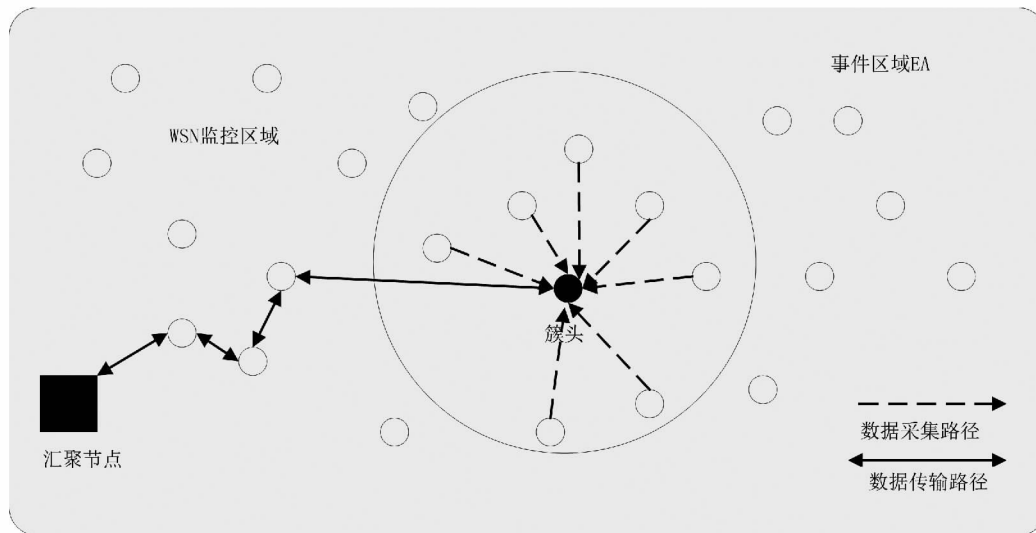


图1 基于地理相关性的DCS示意图

在物联网中通过 Sink 节点来搜集各个传感器节点的感知数据,并对监控区域内的事件源 V 进行估计,使得估计结果的失真度能够满足电力物联网应用的要求。图1给出了 WSN 中基于地理相关性的 DCS 示意图。

在物联网中的节点是密集分布的,在 N 之间以及 N 和 N 之间在空间上存在着不同程度的相关性。因此,在满足失真的要求下,通过确定事件区域 EA 的范围,同时利用 WSN 节点间感知数据的空间相关性对数据进行压缩以及重构,在 WSN 中具有非常重要的意义。

假设 WSN 中事件源 S 所处的位置为 $(0, \rho)$, 分布在事件区域 EA 中的节点 $n(x, y)$ 、 $n(x_r, y_r)$ 坐标分别为 (x, y) 、 (x_r, y_r) , 其感知数据信息是 $N(x, y)$ 、 $N(x_r, y_r)$, 定义变差函数为

$$\gamma(x) = [N(x, y) - N(x_r, y_r)]^2 / 2 \quad (1)$$

式中, $(x - x_r)^2 + (y - y_r)^2 = r^2$ 。若变差函数 $\gamma(x)$ 的值越小,那么感知数据之间的相关性就会越强。在极坐标中可以定义 $n(r, \theta)$ 的节点感知数据信息为 $N(r, \theta)$, 在 WSN 监控区域内事件源 S 所触发的事件区域 EA 范围内,节点 $n(0, \rho)$ 的感知数据信息和周围其他节点的感知数据信息之间有如下的相关性:

$$N(0, \rho) = I_{(U=\tau)} + I_{(U=H)} \int \int_{(\theta, r)} (S_{(r, \theta)} + Z_{\theta}) \times \delta(R=r) \delta(\Theta=\theta | R=r) \quad (2)$$

式中, $U=T$ 表示感知数据信息 $N(0, \rho)$ 是通过随机变量 Y 来获得的,概率为 β ; $U=H$ 表示感知数据信息 $N(0, \rho)$ 是通过相邻的节点 $n(r, \theta)$ 的感知数据信息 $N(r, \theta)$ 的值得到的,概率为 $1 - \beta$ 。

随机变量 Y 和 Z 的概率密度函数分别为 $f_Y(y)$ 和 $f_Z(z)$, 并且随机变量 Y 和随机变量 Z 都与感知数据信息 $N(r, \theta)$ 相互独立。随机变量 Y 反应的是相邻节点间感知数据无相关性的情况;随机变量 Z 反应的是存在空间相关性的情况下,WSN 相邻节点之间感知数据信息的差异。其中变量 Z 符合 $Z \sim (0, \delta_z^2)$, 即:

$$f_Z(z) = \frac{1}{\sqrt{\pi}\delta} e^{-\frac{z^2}{\delta^2}} \quad (3)$$

均方差 δ_z 以根据节点的历史数据统计获得。

采用变差函数来对 WSN 的监控区域内数据场的空间相关性进行分析,在应用所要求的误差门限范围内,计算得到事件区域 EA 的分布范围,EA 范围内的节点以 $n_i (i=1, 2, \dots, N)$ 组成一个簇,并从簇中选出节点 $n_h (h \in \{1, 2, \dots, N\})$ 作为簇首,簇首负责收集 EA 范围内各个节点的感知数据信息。

$$X_i: X_i = S_i + N_i (i=1, 2, \dots, N)$$

式中, S_i 是节点 n_i 的感知数据信息; N_i 为观测噪声。定义 X_i 为同一时刻各个传感器节点的感知数据信息,同时定义向量 $X_N = (X_1, X_2, \dots, X_N)^T \in R^N$ 表示事件区域 EA 中 N 个节点的感知数据信息。由于 WSN 的空间相关性,数据 X_N 在小波基 ψ 上呈现出

k 稀疏性,因此感知信息数据 X_N 、小波基 ψ 和变换系数向量 θ 关系符合式(4)。

$$X_N = \psi\theta \quad (4)$$

式中 稀疏系数向量为

$$\theta = (\lambda_0 \hat{r}_0 \cdots \hat{r}_{J-1})^T, \|\lambda_0\|_0 + \sum_{i=0}^{J-1} \|\hat{r}_i\| = k$$

簇首节点获得感知数据向量 X_N 之后,运行 DCS 编码算法,将稀疏度为 k 的感知数据向量 X_N 随机投影到一个观测矩阵 Φ 上,从而得到 M ($M = ck < N, 2 \leq c \leq 4$) 个观测数据向量 $Y_M = (Y_1, Y_2, \dots, Y_M)^T \in R^M$,从而可以得出式(5)。

$$Y_M = \Phi X_N = \Phi\psi\theta = \Theta\theta \quad (5)$$

式中,观测矩阵 $\Phi = (\varphi_1 \varphi_2 \cdots \varphi_N)$, $\varphi_i \in R^M$ 且观测矩阵 Φ 与小波基 ψ 不相关,将矩阵 $\Theta = \Phi\psi$ 看作是 DCS 的测量矩阵。

簇首节点在得到观测数据向量 Y_M 后,以无限多跳的方式将 Y_M 传送到 Sink 节点,由于簇内各个节点的感知数据向量在小波基上具有稀疏性,而且小波基矩阵和随机观测矩阵之间不相干, Sink 节点便可运行 DCS 的解码算法:

$$\min \|\theta\|_1, \text{subject to } Y_M = \Phi\psi\theta \quad (6)$$

通过求解上式的稀疏最优解 θ^* 便可以重构各个节点的感知数据

$$\hat{X}_N = (\hat{x}_1 \hat{x}_2 \cdots \hat{x}_n)^T = \Phi\theta^* \quad (7)$$

WSN 中基于地理相关性的分布式压缩感知模型如图 2。

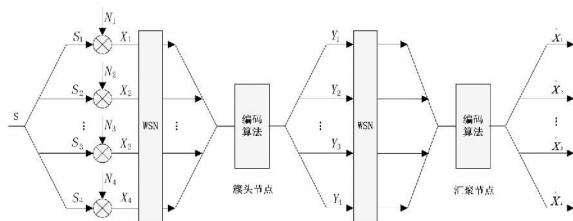


图 2 WSN 基于地理相关性模型

事件区域 EA 范围内的簇首节点对感知数据向量 $X_N \in R^N$ 进行编码,生成观测数据向量 $Y_M \in R^M$,并将 Y_M 以无线多跳的方式传递给汇聚节点。由于 WSN 中节点感知数据的空间相关性,感知数据向量 X_N 呈现 k 稀疏性, M 一般小于 N 。在实际应用中,簇首节点与汇聚节点一般会间隔较远,因此经过压缩算法后,为整个 WSN 网络节省大量的传输能量。

2 基于地理位置相关性的 DCS 算法

首先,需要得到事件区域 EA 的分布范围。假设事件源 S 所处的位置处有一个虚拟的节点为 $n(0, \rho)$,该节点的信息数据为 $S(0, \rho)$,在事件源 S 所触发的事件区域的边界节点 $n(r, \theta)$ 感知的信息数据 $S(r, \theta)$ 符合条件 $|S(r, \theta) - S(0, \rho)| \leq \mu$, μ 是误差门限,它用来表示处于不同位置的节点的感知数据信息与事件源 S 的差异, r 代表事件区域 EA 的分布半径。

根据式(1)可得:

$$\begin{aligned} \gamma(r) &= \frac{1}{2} E [(S(r, \theta) - S(0, \rho))^2] = \frac{1}{2} E [Z^2] \\ &= \frac{1}{2} \int_{-\mu}^{\mu} Z^2 \frac{1}{\sqrt{2\pi\sigma_z}} e^{-\frac{z^2}{2\sigma_z^2}} dz = \frac{1}{2} \sigma_z^2 \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma_z}\right) - \\ &\quad \frac{1}{\sqrt{2\pi\sigma_z}} e^{-\frac{\mu^2}{2\sigma_z^2}} = \psi(\sigma_z, \mu) \end{aligned} \quad (8)$$

式中 $\psi(\sigma_z, \mu)$ 是 σ_z 和 μ 的函数。

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} e^{-t^2} dt$$

通常情况下, $\gamma(r)$ 有很多种估计模型,这里采用估计模型取 $\gamma(x) = c(1 - e^{-\lambda x^2})$,参数 c 是数据之间相关性的强弱程度的指标,参数 λ 反映的是数据相关性随着距离变化的快慢程度,这两个参数取决于监控区域内数据场的空间相关特性。

将 $r(x) = c(1 - e^{-\lambda x^2})$ 代入式(7)得到:

$$c(1 - e^{-\lambda r^2}) = \psi(\sigma_z, \mu) \Rightarrow r = \left[\frac{1}{\lambda} \left(\frac{c}{c - \psi(\sigma_z, \mu)} \right) \right]^{\frac{1}{2}} \quad (9)$$

在各向同性统计过程中,由节点的历史数据统计规律,根据式(9)可以求得在不同的误差门限 μ 下事件区域 EA 的分布半径 r 。因此 Sink 节点就只需要采集事件区域 EA 中各个节点的感知信息数据,从而大大减少了数据处理的能量开销以及执行时间。

其次,要得到基于提升小波变换的小波基矩阵 ψ 。假设感知数据向量 X_N ,对应的是事件区域 EA 中的节点集合 $E_j: \{n_1, n_2, \dots, n_N\}$ 在同一时刻的感知数据,令:

$$\lambda_j = X_N (\lambda_j \in R^N)$$

由于各个节点的地理位置各不相同,根据分布式小波变换,将节点集合 $E_j: \{n_1, n_2, \dots, n_N\}$ 裂为两个集合:偶数集 E_{j-1} 和奇数集 O_{j-1} 。它们所对应的感知信息数据集合分别是 $\lambda_{j,E}$ 和 $\lambda_{j,O}$,用 T 来表示提升小波变换,可得:

$$\begin{cases} T(\lambda_j) = (\lambda_{j-1}, \gamma_{j-1}) \\ \gamma_{j-1} = \lambda_{j,O} - P(\lambda_{j,E}) \\ \lambda_{j-1} = \lambda_{j,E} + U(\gamma_{j-1}) \end{cases} \quad (10)$$

式中: γ_{j-1} 代表小波系数集合; λ_{j-1} 代表尺度系数集合; P 代表线性预测算子; U 代表线性更新算子。由于 WSN 各节点的感知数据信息在空间上具有相关性,因此根据 $\lambda_{j,E}$ 能够比较准确地预测 $\lambda_{j,O}$,而 γ_{j-1} 只包含比较少的信息量。

对 $\lambda_i (i \in \{1, 2, \dots, J\})$ 进行 J 次递归提升小波变换,可得:

$$T^J(\lambda_j) = \{\lambda_0, \gamma_0, \gamma_1, \dots, \gamma_{j-1}\} \quad (11)$$

式中,小波系数向量集合以 $\gamma_i (i \in \{1, 2, \dots, J-1\})$ 中含有许多非常小的元素,对 γ_i 中小于门限值的小波系数进行清零后得到具有稀疏结构的集合 $\hat{\gamma}_i$,利用 $\hat{\gamma}_i$ 对原始信号 λ_j 进行精确重建,令 $T^{-J}(\cdot)$ 为提升小波逆变换。由于预测算子和更新算子都属于线性运算,因此 $T^J(\cdot)$ 和 $T^{-J}(\cdot)$ 都属于线性变换,可得:

$$T^{-J}(\lambda_0, \hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_{j-1}) = \hat{\gamma} \approx \lambda_j \Leftrightarrow \psi \theta \approx \lambda_j = X_N \quad (12)$$

式中 $\theta = (\lambda_0, \hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_{j-1})$, $\|\lambda_0\|_0 + \sum_{i=0}^{j-1} \|\hat{\gamma}_i\| = k < N$,因此只要得到了事件区域 EA 中各个节点的拓扑结构,就可以求出感知数据向量 X_N 的小波变换基矩阵 ψ ,且 X_N 在 ψ 上是 k 稀疏的。

综上所述,WSN 中基于地理相关性的分布式压缩感知编码以及解码算法可以表示如下:

1) 根据 $T^J(\lambda_j) = \{\lambda_0, \gamma_0, \gamma_1, \dots, \gamma_{j-1}\}$, 汇聚节点计算事件区域 EA 的分布范围,通过组播路由的方式激活事件区域 EA 中的节点 $n_i (i = 1, 2, \dots, N)$ 组成簇,并选出节点 $n_h (h \in \{1, 2, \dots, N\})$ 作为簇首, Sink 传递给簇首 n_h 随机节点 $S_M: \{s_1, s_2, \dots, s_M\}$ 。

2) 簇首节点 n_h 生成观测矩阵 $\Phi = R(s_M, \tau_N)$,

其中 $R(\cdot)$ 为伪随机数发生函数, $\tau_N: \{\tau_1, \tau_2, \dots, \tau_N\}$ τ_i 是 n_i 的编号。

3) 簇首节点与簇内各节点进行通信,从而获得同一时刻 WSN 事件区域 EA 范围内所有节点的感知数据向量 $X_N = (X_1, X_2, \dots, X_N)^T$,然后运用基于小波变换的分布式压缩感知编码算法得到测量值向量 $Y_M = \Phi X_N$,通过压缩编码将 N 维的感知数据向量 $X_N = (X_1, X_2, \dots, X_N)^T$ 换成 M 维的观测数据向量 $Y_M = (Y_1, Y_2, \dots, Y_M)^T$,其中 $M = ck$, $2 \leq c \leq 4$,并且 $M < N$ 。

4) 簇首节点将编码后的数据 Y_M 通过无线多跳的方式传递到汇聚节点。

5) Sink 节点生成与簇首节点相同的观测矩阵 $\Phi = R(s_M, \tau_N)$,然后根据网络的拓扑结构得到小波基 ψ 变换矩阵,运行分布式压缩感知解码算法:根据 $Y_M = \Phi \psi \theta$,采用 OMP 重构算法得到的节点的原始感知数据的稀疏小波表示形式为 θ^* ,然后根据 $\psi \theta^* = \hat{X}_N$ 精确重构感知数据集合。

3 实验及结果

3.1 数据融合实例

前面介绍了分簇有限自组网的组网方式,即按照区域进行分簇,各簇首之间采用自组网结构,簇内节点采用星型、树状或链状结构。假设对 4 个小区内的抄表数据进行查询,首先发送查询请求,采用树形路由的方式,传感器位置、通信路径器位置以及通信路径如图 3 所示。

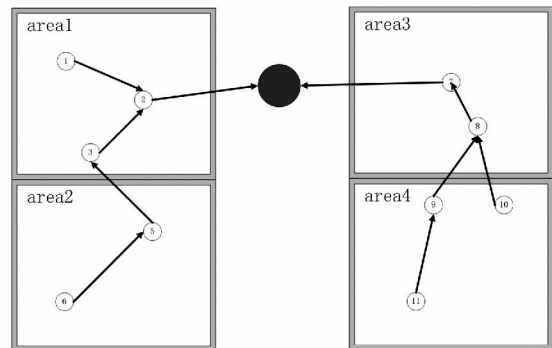


图 3 数据融合实例

图 3 中各个传感器节点都准备好了一份数据,并以 (Filed, Soil moisture) 的形式表示。需要进行以下操作:

1) 各个传感器节点检查是否符合上传要求,以决定是否要参与发送;

2) 各个节点接收到其他节点发送来的数据,进行本地运算,运算结果继续向上游节点提交;

3) 中间节点如果在一定时间内没有收到邻居节点发送来的数据,则默认自邻居节点以下,没有节点提交数据请求。

按照上面的操作,上述过程总共在网内传送了 5 份数据。如图 3 所示,Node 7 传送了 2 份数据,Node 8 传送了 2 份数据,Node 9 传送了 1 份数据。假如不用任何数据融合手段,各个节点单独发送数据到汇聚节点,并由汇聚节点集中计算结果,网络的数据传送量为 25 份,远大于 5 份,相应的节点能耗必将显著增加。

3.2 仿真实验及结果

仿真的侧重点在于研究并分析 WSN 中基于时间相关性的 DCS 编码解码算法中的重构误差与观测数据的个数之间的关系。

仿真采用 Matlab 为工具,WSN 中的空间相关性数据源采用二维高斯分布来模拟,WSN 中节点随机地均匀分布在 $60\text{ m} \times 60\text{ m}$ 的事件区域范围内。研究并分析重构误差与观测数据个数 M 之间的关系。在系统中,假设 WSN 中所有的传感器节点在同一时刻产生感知数据并将其发送到编码端的簇首节点,簇首节点利用 DCS 编码算法对接收到的感知数据进行压缩编码,并将生成的观测数据通过无线多跳的方式传送到解码端,汇聚节点利用 DCS 解码算法对 WSN 事件区域中各个节点的感知数据进行精确重构。

重构误差与抄表数据个数之间的关系如图 4 所示,假设感知值数目为 100。

由图 4 可以看出,当观测值 M 增加时,重构误差越来越小,适合多观测值数据采集系统。

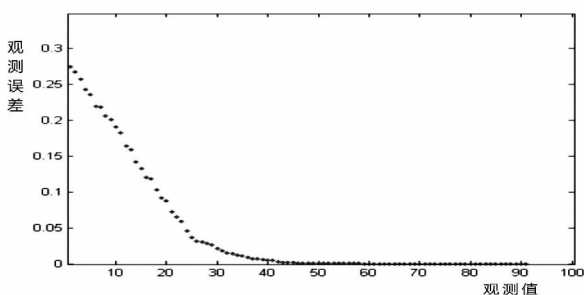


图 4 重构误差和观测个数之间的关系

由图 4 还可以看出,观测值 M 的个数与重构误差之间是呈现反比的关系。假设感知值数目为 100 时,当观测值的个数 M 为 20 时,重构误差为 8%,当

观测值的个数 M 大于 30 时,重构误差很小,几乎接近于 0,可以精确地实现数据压缩后的重构,从而可以降低节点信息发送次数,降低功耗,有助于 WSN 节点的大规模应用。

4 结 论

在对数据进行压缩时,重构误差是一个很重要的参考指标。通过仿真实验的方法,得出了重构误差与观测值 M 之间的关系。由实验可以得到,假设感知值数目为 100 时,当观测值的个数 M 为 20 时,重构误差为 8%,当观测值的个数 M 大于 30 时,重构误差很小,几乎接近于 0,可精确实现对压缩数据的重构,适合 WSN 中数据群的压缩。通过压缩算法,可在很大程度上减少数据的传送量,扫清了网络拥堵的障碍,节省了网络的能耗。在电力 WSN 中,传感器节点的数据存在着较强的空间相关性,正是考虑到这一因素,采用了基于空间相关性的分布压缩感知理论,并对重构误差与观测值 M 之间的关系进行了仿真。仿真结果表明,这一数据算法能满足大规模电力 WSN 应用需求,实现对数据群的压缩。

参考文献

- [1] 荆孟春,王继业,程志华,等. 电力物联网传感器信息模型研究与应用[J]. 电网技术, 2014, 38(2): 532 - 537.
- [2] 陈鹏. 电力物联网应用软件开发[D]. 西安: 西安工程大学, 2011.
- [3] 李娜,陈晰,吴帆,等. 面向智能电网的物联网信息聚合技术[J]. 信息通信技术, 2010, 4(2): 21 - 28.
- [4] 王琼. 基于物联网的电力设备信息管理研究与实现[D]. 保定: 华北电力大学, 2013.
- [5] 王金峰. 电力物联网的数据采集系统与网络传输[J]. 物联网技术, 2013, 3(11): 40 - 42.
- [6] 李福童. 物联网感知层实时数据处理技术研究[D]. 沈阳: 辽宁大学, 2014.
- [7] 王金峰,王进. 电力物联网的数据采集系统及其网络[J]. 科技风, 2012(18): 59.
- [8] 肖颖,周靖. 电力物联网环境下一种有效的云数据安全策略[J]. 信息技术与信息化, 2015(4): 54 - 55.

作者简介:

马成林(1981),工程师、硕士,主要研究方向为电力大数据。

(收稿日期: 2016 - 06 - 26)