



无线传感器网络数据自适应稀疏变换^{*}

陈 暄

(浙江工业职业技术学院 绍兴 312000)

摘 要:针对无线传感器网络节点位置移动及传输干扰等因素可导致数据稀疏结构改变的问题,提出一种基于字典学习的无线传感器网络自适应稀疏变换方法。根据网络数据稀疏结构变化,自适应构建最优稀疏变换基,同时在字典学习问题中引入无线传感器网络数据稀疏基的可压缩约束,以满足无线传感器网络中大规模数据处理特点及稀疏变换的高实时性要求。理论分析和仿真结果表明,所提算法可有效提高无线传感器网络数据稀疏变换算法的顽健性,同时具有良好的实时性。

关键词:压缩感知;字典学习;无线传感器网络;稀疏变换;可压缩

doi: 10.3969/j.issn.1000-0801.2013.12.010

Adaptive Sparse Transform for Wireless Sensor Network Data

Chen Xuan

(Zhejiang Industry Polytechnic College, Shaoxing 312000, China)

Abstract: Aiming at the change of sparse structure introduced by mobility of the wireless sensor network (WSN) nodes and noise in data transmission, an adaptive sparse transform method based on dictionary learning (DL) for WSN data was proposed. The optimum sparse basis can be adaptively constructed according to the change of sparse structure, and the compressibility of WSN data basis was introduced to DL to satisfy the real time requirement for large-scale data processing. Analysis and experimental results demonstrate that the proposed algorithm can significantly improve the robustness and the real time performance of WSN data sparse transform.

Key words: compressed sensing, dictionary learning, wireless sensor network, sparse transform, compressibility

1 引言

无线传感器网络(wireless sensor network, WSN)节点数量多,采集和传输数据量大,但无线传感器网络节点体积微小,节点能量受限,因此降低节点能耗、延长网络生命周期是无线传感器网络技术发展和应用面临的重要挑战^[1,2]。作为近年来发展起来的全新采样理论——压缩感知(compressed sensing, CS)^[3~5],可利用无线传感器网络数据的冗余性进行数据压缩,减少冗余信息传输,为降

低网络节点能耗提供了理想解决方案。

压缩感知理论中,对信号准确重构的前提是采用适当稀疏变换得到原始信号在变换域中的稀疏表示。信号稀疏表示的稀疏性不仅决定了重构准确度,更对重构所需的测量数,即准确获取网络数据所需传输的数据量具有直接影响。由于小波变换、离散余弦变换等较为成熟的稀疏变换方法的存在,现有的网络数据压缩感知方法中,大多直接使用或借鉴、改进现有方法对网络数据进行稀疏变换。然而,由于无线传感器网络拓扑结构及工作节点位置变化、

^{*} 国家自然科学基金资助项目(No.61273072)

数据传输过程中易受干扰等原因,网络数据的稀疏结构将发生改变,从而对传统使用固定稀疏基进行的稀疏变换算法性能造成严重影响。

字典学习(dictionary learning, DL)^[6-8]是稀疏表示领域中一种通过训练样值得到字典进而对信号进行稀疏表示的新兴理论。相对传统的稀疏变换方法,字典学习理论框架下的稀疏变换基可由即时信号样值训练得到,从而使稀疏变换过程充分考虑了信号的稀疏结构及其变化,可得到最佳的信号稀疏表示。上述优点使得字典学习尤其适用于稀疏结构易发生变动的无线传感器网络数据。然而,由于无线传感器网络数据量巨大,而现有的字典学习算法^[9-11]大多未考虑所得字典结构对数据稀疏变换运算量的影响,难以满足无线传感器网络数据稀疏变换的实时性要求。

针对上述问题,本文提出一种基于压缩字典学习的无线传感器网络数据自适应稀疏变换方法。针对无线传感器网络数据特点,在字典学习问题中引入可压缩约束以降低稀疏基构造算法的复杂度。由于稀疏基由即时信号样值训练得到,所得稀疏基与网络数据的稀疏结构更为吻合。与以往的网络数据稀疏变换方法相比,本文算法可在网络数据稀疏性变化时保证稀疏变换效果,从而进一步降低网络数据压缩感知所需的测量数。

2 问题描述

假设无线传感器网络中的传感器节点个数为 J , 每个节点在数据采集过程进行 n 次采样, 单节点数据长度为 n 。单个节点采集数据作为矩阵的列, 则网络数据可表示为:

$$X_{2D} = \begin{pmatrix} x_{11} & \cdots & x_{1J} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nJ} \end{pmatrix}$$

为便于表述, 将各节点数据依次衔接, 网络数据可写为向量形式:

$$X = \text{vec}(X_{2D}) = (x_{11}, \cdots, x_{1n}, x_{12}, \cdots, x_{n2}, x_{1J}, \cdots, x_{nJ})^T = (x(1), x(2), \cdots, x(N))^T \quad (1)$$

其中, $X \in R^N$, $N = J \times n$ 。

由于无线传感器节点采集的环境数据(如大气温度、湿度、光照强度等)在时间维度上并不具有稀疏性, 因此对其进行压缩感知测量前, 需要采用适当的稀疏变换方法求得网络数据在变换域中的稀疏表示。用 Ψ 表示稀疏基, 则数据的稀疏变换过程可写为:

$$X = \Psi\theta = \sum_{i=1}^N \Psi\theta_i = \sum_{k=1}^K \Psi_k \theta_k, \theta_k \neq 0 \quad (2)$$

变换所得 N 维稀疏系数向量 θ 中包含 K ($K \ll N$) 个非零元素。记节点 j ($j \in 1, 2, \cdots, J$) 产生的 $m \times n$ ($m \ll n$) 维测量矩阵为 Φ_j 。将各节点所得测量值写为向量形式为:

$$Y = (y_{11}, \cdots, y_{1m}, y_{12}, \cdots, y_{n2}, y_{1J}, \cdots, y_{nJ})^T = (y(1), y(2), \cdots, y(M))^T \quad (3)$$

则网络数据测量过程可由如式(4)所示矩阵向量形式表示:

$$Y = \Phi X = \Phi \Psi \theta \quad (4)$$

其中, $Y \in R^M$, $M = J \times m$, 测量矩阵为:

$$\Phi = \begin{pmatrix} \Phi_1 & 0 & \cdots & 0 \\ 0 & \Phi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi_J \end{pmatrix} \in R^{M \times N}$$

为实现无线传感器网络数据恢复, 采用压缩感知重构算法即可在已知测量值和测量矩阵条件下对稀疏系数向量进行重构。压缩感知理论框架下, 网络数据重构准确度直接决定于网络数据稀疏向量的稀疏性, 现有网络数据稀疏变换方法, 如离散余弦变换(discrete cosine transform)、小波变换(wavelet transform)等, 未考虑因节点移动或数据传输干扰造成的网络数据稀疏结构改变, 而导致系数向量稀疏性不佳。

3 网络数据自适应稀疏基构造算法

由于传统稀疏字典学习算法复杂度较高, 无法满足无线传感器网络稀疏变换的实时性要求。为此, 将无线传感器网络数据字典的可压缩性约束引入字典学习问题, 在此基础上进行算法步骤描述。

3.1 可压缩字典学习

给定训练样值 $\{x_{Tr_i}\} (i \in \{1, 2, \cdots, L\})$, 字典学习问题可转化为以字典及系数向量为自变量的目标函数最小化问题, 即:

$$\arg\min_{\theta} (\|X_{Tr} - \Psi\theta\|^2) + \lambda J_p(\theta) \quad (5)$$

其中, X_{Tr} 表示由训练样值 x_{Tr_i} 为第 i 列组成的训练样值矩阵, Ψ 和 $\theta = \{\theta_{ij}\}$ 分别表示稀疏基(字典)和系数矩阵,

稀疏度量 $J_p(\theta) = \sum_{i,j} |\theta_{ij}|^p$ 。为使表述清晰直观, 首先对信号的可压缩性进行定义^[12]。若信号 s 元素满足式(6), 则称 s 为可压缩信号。

$$|s|_{(k)} \leq c_k k^{-r} \quad (6)$$

其中, $|s|_{(k)}$ 表示信号 s 中幅值最大的第 k 个元素, $r \geq 1$,



c_r 为常数。类似地,若矩阵 S 元素满足式(6),则称 S 为可压缩矩阵。使用 S_k 表示矩阵 S 中仅保留幅值最大的 K 个元素,其他元素置零得到的矩阵,则 S_k 可作为矩阵 S 的最优稀疏近似,近似误差满足:

$$\|S - S_k\|_F \leq c_r K^{-r+1/2} \quad (7)$$

利用上述可压缩性定义,若字典 $d \times N$ 维字典 Ψ 可由 $d \times M$ 维满秩矩阵 V 表示为:

$$\Psi = SV \quad (8)$$

则字典 Ψ 称为可压缩字典,其中 $M \times N$ 维可压缩矩阵 S 称为字典生成矩阵。

结合式(5)所示的字典学习问题和可压缩字典定义,无线传感器网络数据稀疏基构造可转化为如下问题:在给定 $d \times L$ 维训练样值矩阵 X_{tr} 和 $d \times M$ 维满秩矩阵 V 基础上,对生成矩阵 S 进行估计,即通过求解以 (Θ, S) 为自变量的目标函数 $v(\Theta, S)$ 的极小值得到稀疏基:

$$\underset{s}{\operatorname{argmin}} \{ \min_{\Theta} v(\Theta, S) : v(\Theta, S) = \|VS\Theta - X_{\text{tr}}\|^2 + \lambda J_{\rho}(\Theta) + \gamma J_q(s) \} \quad (9)$$

其中 $\rho \in \{p, q\} \leq 1$ 。正则化参数 λ, γ 分别对变换系数和生成矩阵的稀疏度进行加权。为便于表述,不妨假设 $p=q=1$,稀疏度量使用 1 范数。下面对式(9)求解算法步骤进行推导和描述。

3.2 算法描述

为求解目标函数 $v(\Theta, S)$ 最小值,可在每次迭代时分别针对 Θ 或 S 进行,同时固定另外一个变量。用 $(\Theta^{[0]}, S^{[0]})$ 表示迭代初始值,算法通过 $\Theta^{[n]} \rightarrow \Theta^{[n+1]}$ 或 $S^{[n]} \rightarrow S^{[n+1]}$ 的解优化使目标函数值下降。在进行算法步骤描述前,首先需明确变量更新方程。

当变量 S 固定时,记目标函数为 $v_s(\Theta)$,则此时强函数(majorizing function)可写为:

$$\mu_s(\Theta, \Theta^{[n]}) = \operatorname{tr}\{c_r c_s \Theta^T \Theta - 2\Theta^T (S^T V^T (X_{\text{tr}} - VS\Theta^{[n]})) + c_r c_s \Theta^{[n]}\} + \lambda J_1(\Theta) + c \quad (10)$$

注意到 μ_s 为不可微凸函数,因此矩阵 Θ 在 μ_s 极小值的次梯度方向上,可得:

$$O \in \partial \mu_s(\Theta^{[n+1]}, \Theta^{[n]}) \quad (11)$$

由式(10)可得:

$$\partial \mu_s(\Theta, \Theta^{[n]}) = 2c_r c_s \Theta - 2(S^T V^T (X_{\text{tr}} - VS\Theta^{[n]}) + c_r c_s \Theta^{[n]}) + \lambda \partial J_1(\Theta) \quad (12)$$

将式(11)代入有:

$$\Theta^{[n+1]} = S_{\lambda/2} \left[\frac{1}{c_r c_s} (S^T V^T (X_{\text{tr}} - VS\Theta^{[n]}) + c_r c_s \Theta^{[n]}) \right] \quad (13)$$

其中, S_{α} 为软收缩算子^[13],对矩阵 $A = \{a_{ij}\}$ 有:

$$S_{\alpha}(A) = \begin{cases} a_{ij} - \operatorname{sign}(a_{ij}) \cdot \alpha/2, & \alpha/2 < |a_{ij}| \\ 0, & \text{其他} \end{cases} \quad (14)$$

同理可得,当变量 Θ 固定时,变量 S 更新方程为:

$$S^{[n+1]} = S_{\lambda/2} \left[\frac{1}{c_r c_s} (V^T (X_{\text{tr}} - VS\Theta^{[n]}) \Theta^T + c_r c_s S^{[n]}) \right] \quad (15)$$

其中 S_{α} 由式(14)给出。

结合以上分析,算法可按如下步骤描述。

初始化 $c_r, c_s > \|V^T V\|, K_{\Theta}, K_S$ 取自然数, $t=0, n=0$ 。对 $t=0, 1, 2, \dots, T$ 的每一个取值,重复步骤 1 和步骤 2。

步骤 1 $c_s > \|S^T S\|, \Theta^{[0]} = \Theta_t$ 。对 $n=0, 1, 2, \dots, K_{\Theta}-1$, 重复步骤 1.1。

步骤 1.1 $\Theta^{[n+1]} = S_{\lambda/2} \left[\frac{1}{c_r c_s} (V^T (X_{\text{tr}} - VS\Theta^{[n]}) \Theta^T + c_r c_s S^{[n]}) \right]$ 。循

环终止: $\Theta_{t+1} = \Theta^{[K_{\Theta}]}$ 。

步骤 2 $c_{\lambda} > \|\Theta \Theta^T\|, \Theta^{[0]} = \Theta_t$ 。对 $n=0, 1, 2, \dots, K_{\Theta}-1$, 重复步骤 2.1。

步骤 2.1 $S^{[n+1]} = S_{\lambda/2} \left[\frac{1}{c_r c_s} (V^T (X_{\text{tr}} - VS\Theta^{[n]}) \Theta^T + c_r c_s S^{[n]}) \right]$ 。循

环终止: $S_{t+1} = S^{[K_S]}$ 。

输出 S_T

由上述算法描述进一步分析算法所得可压缩字典 Ψ 进行向量相乘的计算复杂度。所得字典 Ψ 与任意向量的相乘运算可分解为两步:首先与字典生成矩阵 S 相乘,而后与满秩矩阵 V 相乘。注意到 S 为可压缩矩阵,由式(7)可知相乘过程等价于与最优稀疏近似 S_k 相乘,算法复杂度与 K 同数量级;在矩阵 V 为傅里叶矩阵或小波变换矩阵等具有快速矩阵向量相乘性质时,相乘运算复杂度与 $M \ln N$ 同数量级。实际无线传感器网络应用中通常满足 $K \leq M \ln N$, 则与 $d \times N$ 维字典 Ψ 相乘的总运算复杂度为 $N \ln N$, 远小于普通字典相乘计算复杂度 $d \cdot N$ 。

4 仿真与结果分析

采用 MATLAB 软件对本文算法进行仿真。为验证在节点数据不同稀疏结构下的算法性能,选取如图 1 所示的两组温度数据,数据长度 $N=500$,其中一组数据相关性较强,另外一组受到干扰因素影响,部分相关性被破坏。使用本文算法、K-SVD 算法构造的稀疏基及传统 DCT 分别对网络数据进行稀疏变换。选用元素服从零均值高斯分布的随机矩阵进行压缩感知测量,随后使用压缩采样匹配追踪(compressive sampling matching pursuit, CoSaMP)算法^[14]进行数据重构。

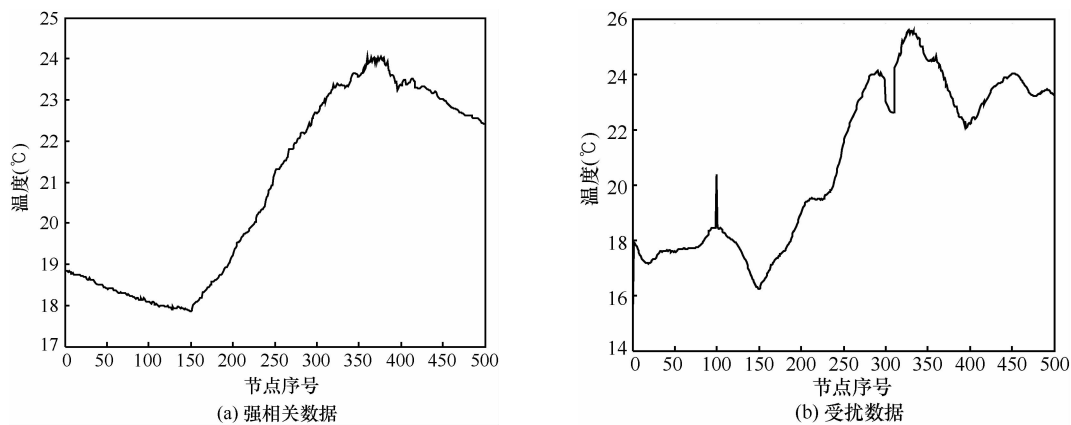


图 1 原始温度数据

实验中变换测量数,使得测量数与数据长度之比(M/N)在 0.1~0.8 依次取值,定义重构相对误差 $\eta=\frac{\|\hat{x}-x\|_2}{\|x\|_2} =$

$$\frac{\sqrt{\sum_{i=0}^{N-1} (\hat{x}_i - x_i)^2}}{\sqrt{\sum_{i=0}^{N-1} (x_i)^2}}$$
。每种测量数取值重复实验 100 次,计算平均

重构误差。若定义重构失败率 $P_{er}=\frac{num(\max(\hat{x}-x)>x\cdot0.1)}{num_{total}}$,统计成功重构次数并计算重构失败率,结果如图 2 和图 3 所示。可知,在数据相关性较强时,本文算法与经典算法相近,而在数据相关性被破坏时,本文算法得到的稀疏系数可获得更为准确的重构结果。

实验中同时使用不同字典学习算法所得的稀疏基对上述网络数据进行变换,进行测量、重构后经过反变换得到原始温度数据的重构值。统计整个重构过程的运算时间(实验计算机使用 Intel 双核 Q9500 型双核 CPU,主频 2.83 GHz),由表 1 所示重构时间可以看出,经过本文算法变换的网络

数据重构运算时间小于其他算法。由于使用压缩感知测量矩阵及重构算法,可认为由稀疏变换时间导致了上述运算时间差异。上述结果验证了本文算法的实时性。

表 1 运算时间对比

	DCT 运算 时间(s)	K-SVD 运算 时间(s)	本文算法运算 时间(s)
相关数据	1.573	2.677	1.032
受扰数据	1.641	3.323	1.049

5 结束语

本文利用字典学习思想提出了一种面向无线传感器网络数据的稀疏变换方法。与以往算法不同,本文算法可在网络节点相关性变化或节点数据受到干扰时自适应构建最优稀疏基,从而有效避免了传统稀疏变换方法中固定稀疏基与网络数据可能发生的稀疏结构变化间的矛盾。实验表明,本文算法可在网络数据相关性变化时获得稀疏性更强的变换系数,进而提高无线传感器网络数据压缩感知重构的准确度。

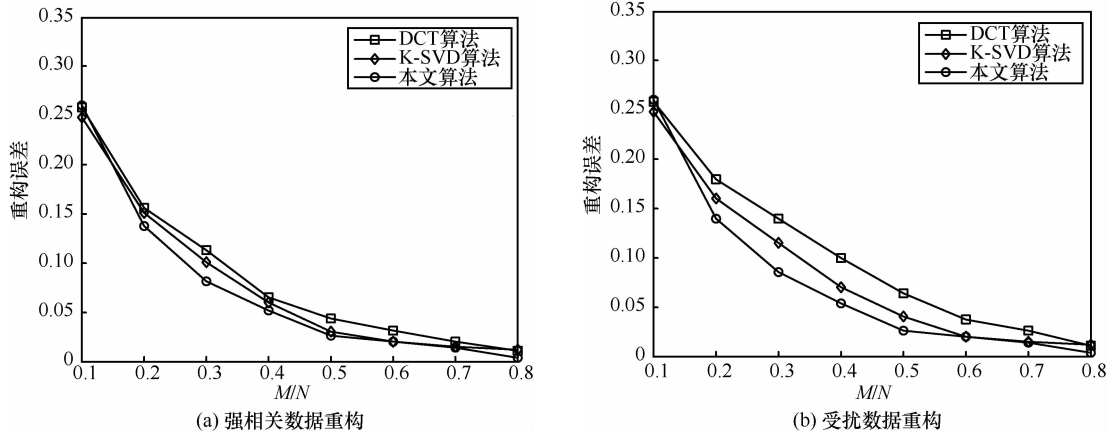


图 2 重构误差对比

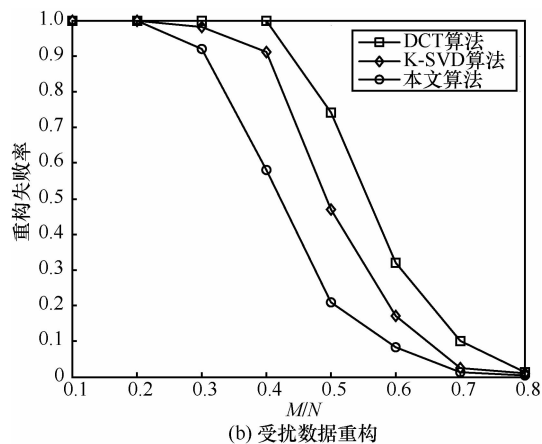
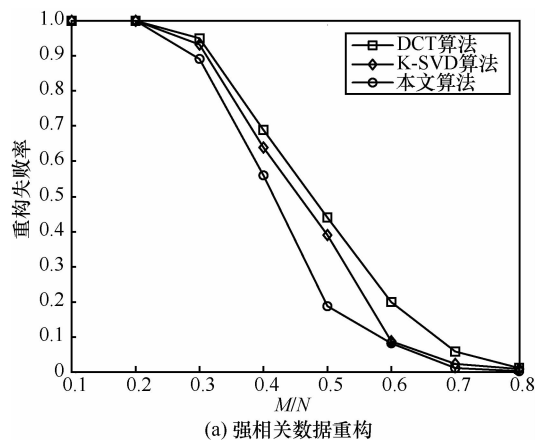


图3 重构概率对比

参考文献

- 1 Akyildiz I F, Su W, Sankarasubramaniam Y. A survey on sensor networks. IEEE Communications Magazine, 2002, 40(8)
- 2 李建中, 高宏. 无线传感器网络的研究进展. 计算机研究与发展, 2008, 45(1)
- 3 Candes E, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory, 2006, 52(4)
- 4 焦李成, 杨淑媛, 刘芳等. 压缩感知回顾与展望. 电子学报, 2011, 20(7)
- 5 戴海琼, 付长军, 季向阳. 压缩感知研究. 计算机学报, 2011, 54(3)
- 6 Tomic I, Frossard P. Dictionary learning. IEEE Transactions on Signal Processing, 2011, 28(2)
- 7 Aharon M, Elad E, Bruckstein A M. K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing, 2006, 54(11)
- 8 Kreutz-Delgado K, Murray J F, Rao B D, et al. Dictionary learning algorithms for sparse representation. Neural Computation, 2003, 15(2)
- 9 Skretting K, Engan K. Recursive least squares dictionary learning algorithm. IEEE Transactions on Signal Processing, 2010, 58(4)
- 10 Yaghoobi M, Blumensath T, Davies M. Dictionary learning for sparse approximations with the majorization method. IEEE Transactions on Signal Processing, 2009, 57(6)
- 11 Gowreesunker B V, Tewfik A H. Learning sparse representation using iterative subspace identification. IEEE Transactions on Signal Processing, 2010, 58(6)
- 12 Candes E, Romberg J, Tao T. Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics, 2006, 59(8)
- 13 Donoho D L, Johnstone J M. Ideal spatial adaptation by wavelet shrinkage. Biometrika, 1994, 81(3)
- 14 Needell D, Tropp J. CoSaMP: iterative signal recovery from incomplete and inaccurate samples. Applied and Computational Harmonic Analysis, 2009, 26(3)

[作者简介]



陈暄, 男, 浙江工业职业技术学院讲师、系统分析师、信息系统项目管理师, 主要研究方向为软件方法学、无线传感器网络。

(收稿日期: 2013-06-15)

· 简讯 ·

2013年10月份通信业经济运行情况

2013年10月份, 我国通信业总体运行平稳。电信业务总量和业务收入继续保持稳定增长。10月份, 全国电信业务总量完成1 183.5亿元(按照2010年不变单价计算), 同比增长7.7%; 电信主营业务收入完成1 004.6亿元, 同比增长8.2%。1-10月, 全国电信业务总量累计完成11 592.7亿元, 连续两个月保持7.8%的同比增长; 电信主营业务收入累计完成9 713.3亿元, 同比增长8.6%, 与上月持平。

(来源: 工业和信息化部运行监测协调局)