Renewable and Sustainable Energy Reviews 91 (2018) 59–69

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Compression of smart meter big data: A survey

Lulu Wen[a], Kaile Zhou[a,b,*], Shanlin Yang[a,b], Lanlan Li[a]

[a] School of Management, Hefei University of Technology, Hefei 230009, China
[b] Key Laboratory of Process Optimization and Intelligent Decision-making of Ministry of Education, Hefei University of Technology, Hefei 230009, China

ARTICLE INFO

Keywords:
Smart grid
Smart meter
Energy big data
Data compression

ABSTRACT

In recent years, the smart grid has attracted wide attention from around the world. Large scale data are collected by sensors and measurement devices in a smart grid. Smart meters can record fine-grained information about electricity consumption in near real-time, thus forming the smart meter big data. Smart meter big data has provided new opportunities for electric load forecasting, anomaly detection, and demand side management. However, the high-dimensional and massive smart meter big data not only creates great pressure on data transmission lines, but also incur enormous storage costs on data centres. Therefore, to reduce the transmission pressure and storage overhead, improve data mining efficiency, and thus fulfil the potential of smart meter big data. This study presents a comprehensive study on the compression techniques for smart meter big data. The development of smart grids and the characteristics and application challenges of electric power big data are first introduced, followed by analysis of the characteristics and benefits of smart meter big data. Finally, this study focuses on the potential data compression methods for smart meter big data, and discusses the evaluation methods for smart meter big data compression.

1. Introduction

Electricity is an important form of secondary energy, which plays a vital role in modern economic and social development. The demand for industrial electricity grows rapidly with the development of an economy. Furthermore, residential electricity consumption is also increasing significantly with the improvement of living standards [1–3]. The electric power industry has become a fundamental industry and important public utility supporting the development of national economies [4,5]. At present, the recovery and steady growth of the world economy has led to rapid increases in the demand for energy [6]. Meanwhile, the world is facing serious resource shortages, environmental pollution, the greenhouse effect and other difficult challenges. Therefore, it is urgent to develop cleaner energy production and smart energy systems to alleviate these problems [7].

Information and communication technologies (ICTs) have been widely used in electric power systems, making power systems more digital, more intelligent, efficient, and robust [8]. The smart grid is expected to become a general trend in the development of electric power system [9–11]. As typical representatives of the new generation of ICTs, cloud computing, big data analytics and internet of things related technologies have widely penetrated the construction and development of smart grids [12]. Therefore, the data generated by a smart grid increases dramatically, and the forms of data are increasingly complex [13,14]. The data generated in power generation, power system operation, and electricity consumption combine to form electric power big data [15,16]. Electric power big data has great potential to support the optimization of electric power systems and various management decisions. However, it also places great pressure on data transmission lines and increases storage costs. In addition, the data are not all valuable; the existence of redundant data also obviously influences the efficiency of big data analysis. Smart meter big data, including voltage, current, and electricity consumption data, is an important component of electric power big data [17]. This data can provide valuable knowledge for marketing strategies development and demand side management (DSM) of power companies. For example, analysis of electricity consumption data can support electric load forecasting, anomaly detection, and demand response program development [18–20]. Ideally, power generation and system operation can be optimized in near real time, electricity demand can be predicted precisely, electricity can be dispatched in timely fashion, electricity consumption patterns can be discovered accurately, and more effective pricing mechanisms can be developed. However, smart meters record electricity consumption details of consumers in near real time and transmit data to data centres frequently [21]. For example, suppose there are 100 million smart meters in a smart grid and each record consist of five KB. If the data are collected every 15 min, the total amount of data will reach 2920 TB per year. Efficient data compression

* Corresponding author at: School of Management, Hefei University of Technology, Hefei 230009, China.
E-mail address: zhoukaile@hfut.edu.cn (K. Zhou).

https://doi.org/10.1016/j.rser.2018.03.088
Received 20 January 2017; Received in revised form 29 March 2018; Accepted 30 March 2018
Available online 05 April 2018
1364-0321/

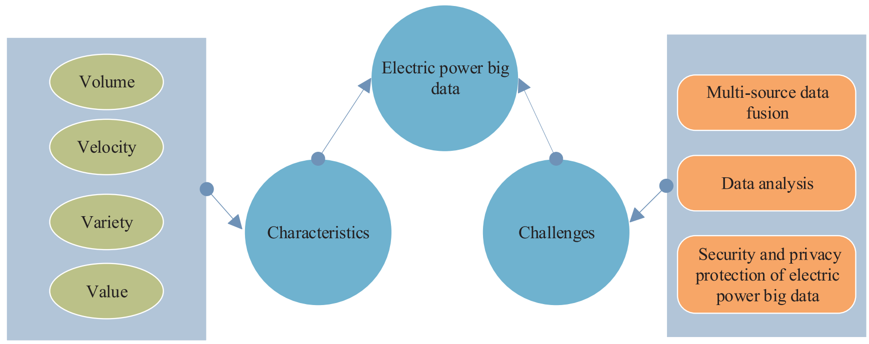

Fig. 1. Characteristics and application challenges of electric power big data.

methods can relieve transmission pressure, reduce storage overhead, and enhance data analysis efficiency. Thus, a comprehensive study on the compression of smart meter big data would be beneficial at this time. Currently, compression techniques involving smart meter big data have been an important research area for smart grids [22,23].

There have been some research works on electric power big data [24–26]. Zhou et al. [24] provided a review of big data driven smart energy management, and the authors presented a system architecture and its related industrial energy management tools. Tu et al. [25] focused on electric power big data and summarized the latest applications leveraged by big data technology in smart grids. Diamantoulakis et al. [26] presented a detailed review of big data analytics for dynamic energy management in smart grids. However, specific research on smart meter big data is rarely reported. To the best of our knowledge, this is the first comprehensive review on the compression of smart meter big data. The objective of this study is to highlight the key research issues in smart meter big data compression methods, including lossy and lossless compression. This paper aims to provide a better understanding of the full potential and to improve data analysis efficiency of smart meter big data, as well as to reduce the transmission pressure and the storage overhead.

The remainder of the paper is organized as follows. The next section introduces the development of smart grids, as well as the structure and characteristics of electric power big data. Section 3 analyses the characteristics and benefits of smart meter big data. In Section 4, it summarizes various popular data compression methods for smart meter big data and discuss corresponding research issues. The evaluation methods for smart meter big data compression are discussed in Section 5. Finally, Section 6 presents the conclusions.

2. Smart grid and electric power big data

2.1. Smart grid

In the 21st Century, the problems of energy shortage and environmental pollution have become more serious, while the demand for electricity has increased significantly year by year [27]. Traditional electric power systems can only convert one-third of the energy they consume to electricity without recycling waste heat, and lose 8% of total electricity during transmission, whereas only 20% of total electricity capacity is used to meet peak demand [28]. Furthermore, other problems are inherent in traditional electric power systems, such as the lack of automatic analysis and the low level of visualization [29]. With the increasing need to allow access to external energy sources such as wind, solar, and other clean energy systems, a variety of adverse effects also emerge, affecting the security and reliability of power system operation [30,31].

A smart grid utilizes advanced communication technology, computer technology and related control technologies to establish an intelligent system for power generation, transmission, substation control, distribution, and consumption [10,32,33]. As a modern infrastructure, the smart grid can find solutions for power system failures with high efficiency, thus reducing the workforce and providing sustainable, reliable, and safe electricity to consumers [34]. It can also integrate renewable and alternative energy resources using automation control technology and ICTs [35–37]. Development of the smart grid has been regarded as a major energy strategy to achieve industrial progress and social sustainable development among many countries of the world.

2.2. Electric power big data

A smart grid is characterized by the bi-directional mobility of electricity and information, which aims to establish highly automated and widely distributed energy exchange networks [38]. As the future direction of the electric power system development, the smart grid is constantly being advanced and developed. At the same time, huge amounts of data are generated by smart meters, power distribution automated devices, digital protection devices, and other intelligent devices in the smart grid, thus forming the electric power big data [24,39]. Specifically, electric power big data is mainly composed of user description data, user behaviour data, electric power system internal data, and business systems related data [40]. User description data refers to the characteristic data of the household population, their living environment, etc. User behaviour data includes the data of customer service systems, demand response systems, etc. The internal data of the electric power system contains electricity consumption data, power production system data, etc. Business system data includes meteorological data, geographic information data, building data, etc. [41].

The general characteristics and application challenges of electric power big data are shown in Fig. 1.

2.2.1. Characteristics of electric power big data

(1) Volume. Large scale power production and consumption data are collected by thousands of advanced measurement devices in near real-time. Additionally, very large amounts of external data, such as temperature, weather, and geographic information data; are also integrated into electric power big data.
(2) Velocity. An electric power system places extremely high demands on data processing speed to ensure real-time electricity supply-demand balance and instantaneity. In addition, rapid data processing is very important for the identification and recovery of power system failures, short-term load forecasting, and other decisions. Therefore, large scale power production and consumption data must be processed at a very high speed [42].
(3) Variety. The types of electric power big data are highly complex and data dimensions are very high. The data not only includes voltage, current, frequency, and power of each measurement node, but also includes a large number of power production data, business data and user data. To further complicate matters, the variety of

unstructured data is also increasing, such as network logs, weather images and geographic location information.

(4) Value. The knowledge discovery of electric power big data is of great significance in supporting power production and distribution, failures recovery, and energy investment. Moreover, the analysis of electric power big data contributes to provide personalized energy efficiency services to customers and to create new business models [43].

2.2.2. Challenges of applying electric power big data

(1) Challenges in multi-source data fusion. Electric power big data are not just the data generated by the intelligent sensing devices, monitoring devices, and communication devices. As solar, wind, and other clean energy sources access a power grid, large amounts of temperature data, humidity data, weather data, and geographic information data are also collected by intelligent devices. Currently, the significance of these data in supporting electric load forecasting, anomaly detection, and demand side management has been recognized. However, it will be a major challenge to achieve multi-source data fusion for effective application of electric power big data [44].

(2) Challenges in data analysis. Millions of advanced sensors and meters will be further deployed. It is inevitable that massive amounts of data will be generated [45]. While it is significant to extract valuable knowledge from these data for grid optimization, load balancing, business planning, and other purposes [46]; new problems and challenges have emerged in the management of data sets and data analysis [47].

(3) Challenges in security and privacy protection. Weak security and privacy protection of electric power big data will inhibit the application of smart grid solutions. In addition, to make the grid more intelligent and efficient, the scope and application of information technology in smart grids is expected to expand [48–50], which will also create new security problems. Therefore, the security and privacy protection of electric power big data is becoming increasingly important to ensure grid security and reliability.

3. Smart meter big data

With the development of smart grids, a large number of smart meters have been deployed. Power companies can obtain a large amount of real-time data through smart meters. Smart meter big data includes voltage, current, power consumption, and other important parameters [51–53].

3.1. Characteristics of smart meter big data

At present, power companies have accumulated a large amount of smart meter data, which cause difficulties in data processing and analysis. It is necessary to take full consideration of the characteristics of smart meter big data to fulfil its potential. With the development of smart grid technology and information technology, the current smart meter big data has the following primary characteristics.

(1) High dimensionality of data. Smart meters can record users'

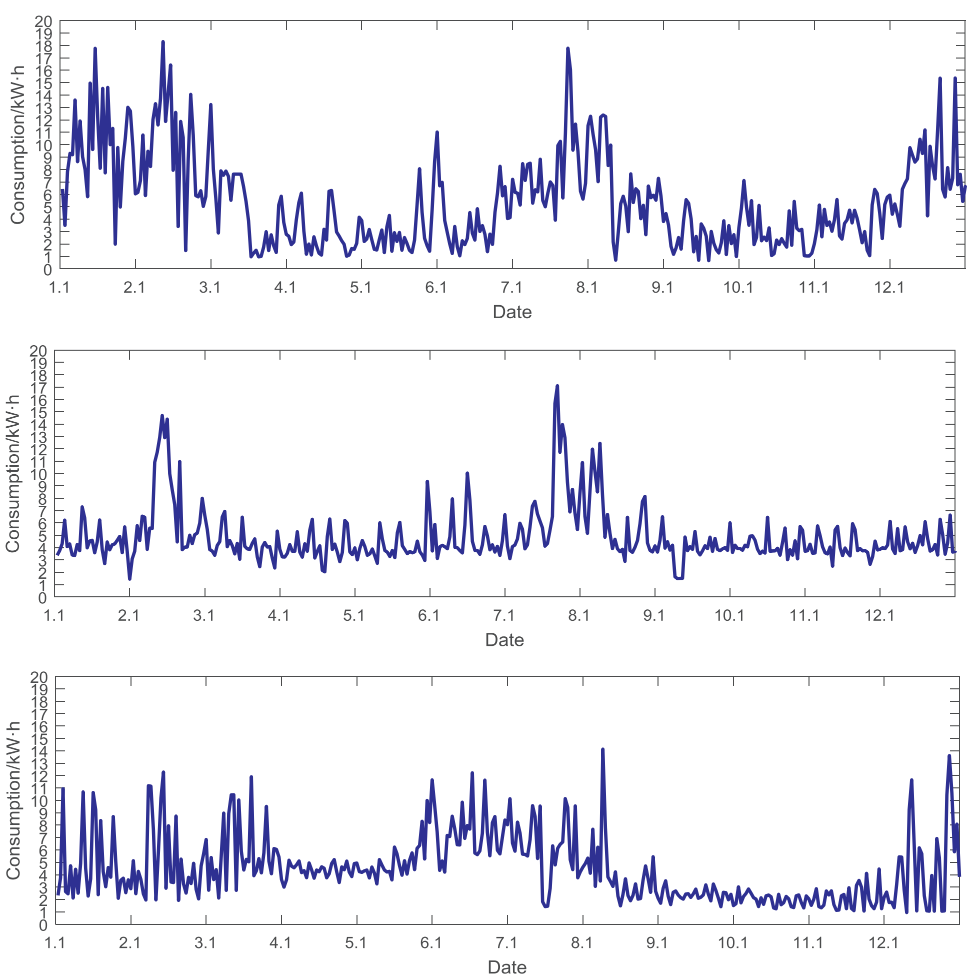

Fig. 2. Yearly electricity consumption profiles of three residential users in a city of China.

electricity consumption information frequently, including current, voltage, power, etc. Advanced smart meters can collect smart meter data every 5–60 min. In 2010, the largest electricity and gas supplier in the United States, PECO, provided a technical solution which can collect smart meter data once a minute [50], and in the future, power companies will receive more fine-grained smart meter data more frequently [54]. The era of smart meter big data is coming and will bring many opportunities and challenges.
Fig. 2 shows the yearly electricity consumption profiles of three residential users in a city of China.
In Fig. 2, the dimension of power consumption profiles is very high.

(2) Large volume of data. Smart meters continue to be deployed with the development of smart grids. They collect users' electricity consumption information with high frequency, and the data volume grows rapidly. Power companies have to establish special data centres to store these high-dimensional smart meter data [55,56]. According to current statistics, the total amount of annual load profile data of 1.658 million households in Ireland produces 216 GB. Comparatively, the population of China is much larger than Ireland, and more smart meters will be deployed. It is roughly estimated that China's annual load profile data will reach 29 TB [56].
(3) High speed of data acquisition and transmission. ICTs have developed rapidly and have been widely applied in smart grids. Intelligent devices and smart systems are widely used in smart grids, and are basically equipped with intelligent processing programs, as well as can collect data in near real-time. They can also significantly improve data transmission capabilities [55–57].
(4) High speed of data analysis and processing. The analysis and processing of smart meter big data contribute to predict users' short-term load demands, determine the location and causes of power failures, implement DSM, and so on. At the same time, it places high demands on the speed of analysis and processing of smart meter big data to maintain the stability and reliability of the electric power system.

Valuable information can be mined from smart meter big data through big data based modelling, and innovative services can be provided to electricity customers, power supply companies, and society in general [57,58]. It is estimated that if the data utilization rate increases by 10%, a power company can increase profits by 20–49% [59]. However, the large scale of smart meter data has placed great pressure on data transmission lines and creates large storage costs to data centres. Therefore, it is essential to compress smart meter big data through models and methods involving mathematics, statistics, and data analysis.

3.2. Benefits of smart meter big data

Smart meter big data implies valuable information that can be used to optimize energy efficiency, maintain stability and reliability of a smart grid, and improve the quality of energy services [54]. For example, it has been used in electric load forecasting, anomaly detection of electric power system, DSM, and so on.

Fig. 3 shows the benefits of smart meter big data.

3.2.1. Electric load forecasting

Electric load forecasting provides the support for power generation planning and development planning of an electric power system. Precise electric load forecasting is of great significance for the economic, safe, and reliable operation of power system [60,61]. Electric load forecasting can be divided into long-term load forecasting, medium-term load forecasting, short-term load forecasting and ultra-short-term load forecasting, corresponding to annual, monthly, single-day, and hourly load forecasting [52,62].

Several methods have been used for electric load forecasting, such as regression analysis, exponential smoothing, and weighted iteration, and other improved algorithms including adaptive prediction and stochastic time series. Neural networks and genetic analysis have also been widely applied to electric load forecasting. Wi et al. [63] proposed a method that combined fuzzy polynomial regression with weather feature selection and adjustment for electric load forecasting. The method improves the accuracy and effectiveness of short-term load forecasting. Kling et al. [64] employed a neural network to predict the current electric load using historical data from smart meters. The forecasting model can be implemented by energy suppliers and distributed system operators to develop optimal power distribution plans. Amjady et al. [65] presented a hybrid forecasting method with wavelet transform and a neural evolutionary model to forecast electric load.

A large amount of valid data accumulated by smart meters provides the foundation for electric load forecasting. However, it is very difficult to use high-dimensional smart meter data for electric load forecasting directly. Furthermore, consumer grouping is the core of load profiling, but there exists some problems in clustering such high-dimensional data [66], so it is necessary to compress smart meter big data to reduce dimensionality.

3.2.2. Anomaly detection of electric power system

Anomalies are commonly referred to as faults in an electric power system. Additionally, some non-technical electricity losses also fall within the range of anomalies of electric power systems, such as illegal power usage and electricity theft [67,68]. How to detect anomalies is critical for power companies to maintain reliability, safety, and stability of their systems, and it is also of great importance for ensuring the safety of users' electricity consumption information and in reducing economic losses [69]. Smart meters generally have built-in intelligent sensors that can send fault information to the fault management system when power failures occur. The geographic information, climate information, and point of failures information can be used to further identify the cause and locations of failures. Interactions between smart meters and operators can also help to determine whether multiple grid faults exist simultaneously. Currently, neural networks, deep learning, and other methods have been used to detect anomalies with smart meter big data.

Power companies can quickly identify failures in an electric power

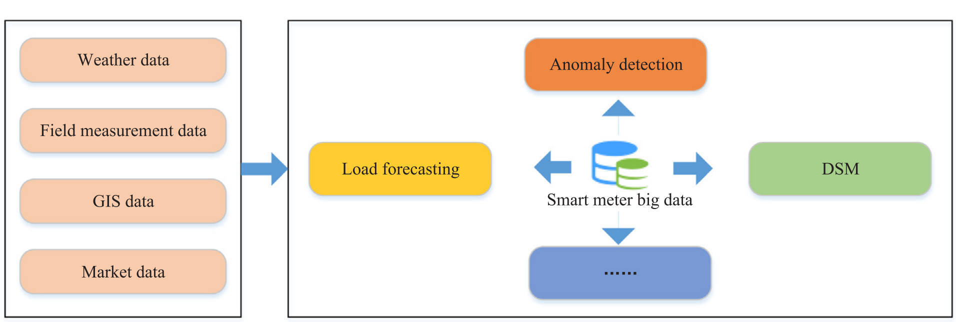

Fig. 3. Benefits of smart meter big data.

system using real-time smart meter data and reduce the interruptions caused by a power failure to a few minutes, depending on the type of failure [70]. Furthermore, many smart meters can detect electricity theft behaviours relatively quickly by detecting the operating behaviour based on tampering of meter box, wiring changes, and meter software updating [52].

Anomaly detection in smart grids mainly depends on the data collected by smart meters. Owing to the rapid communication capability of smart meters, power companies can detect electric anomalies quickly and take relevant measures to avoid unnecessary economic loss, as well as maintain the stability and reliability of the system.

3.2.3. DSM

DSM encourages users to consume less electricity at peak hours and consume more electricity at other times through policy measures, thereby improving the efficiency of power systems by optimizing power consumption [71,72]. It ensures the stability and reliability of an electric power system and suppresses short-term increases in electricity prices [73,74].

The current research on DSM focuses on different pricing strategies and different incentives [75]. The price-based strategies include time-of-use (TOU) price, spot price and critical peak price. TOU price charges different prices for electricity consumption at different times of the day. Electricity price increases with the load increase. DSM encourages users to consume electricity during low load periods, thus effectively reducing the peak loads of the grid. In 1997, Schweppe first proposed the spot pricing strategy [76]. Spot pricing is considered as the optimal pricing strategy for power systems. Power companies adjust electricity prices in real-time according to the power supply and demand situation to stimulate users to consume electricity during low load periods. It can reduce peak load and achieve the purpose of peak load shifting. If applied fairly, it maintains the stability of the grid while saving electricity expenditures for users [77,78]. Critical peak pricing is based on TOU tariffs plus a spike rate and reflects short-term market costs [79]. As for incentive strategies, users are rewarded when they comply with the measures of power companies. The forms of incentives are different, for example, Zhong et al. [80] proposed using discount coupons as incentives. Related to this type of DSM strategy, smart meter data has been widely applied to estimate demand responses of consumers [81].

DSM can guide and encourage electricity consumers to take the initiative to change their electricity consumption patterns. It contributes to the stability and reliability of the electric power system. The premise for power companies implementing DSM is that they can predict users' demands for electricity at a specific time, and smart meter big data is an important resource for electric power load forecasting.

3.2.4. Other aspects

Smart meters can collect users' electricity consumption data, communicate with other devices in the smart grid, and transmit relevant data to power companies in real-time, and smart meter data can reveal behavioural patterns of users. For example, when electricity consumption is at a low level, it can be inferred that there is no one at home. When electricity consumption fluctuates during night hours, it suggests that the family may have a new baby (e.g., breastfeeding behaviour). With the development of science and technology, smart meters can even sense which device consumes electricity in the home at any time. However, the very detailed information about electricity consumption may become a problem if the information is obtained by third parties, such as advertising companies and insurance companies, as customer privacy may be violated [80].

4. Data compression methods for smart meter big data

Smart meters collect user's electricity consumption information in near real-time and send massive amounts of data to power companies. Through the processing and analysis of smart meter big data, power companies can predict electric load of users precisely, detect and process abnormalities and failures in timely fashion, and implement more flexible DSM. However, smart meter big data places high burden on transmission lines and data storage centres [82]. Meanwhile, the high-dimensional nature of smart meter data represents difficulties for data analysis and processing, so it is necessary to compress the smart meter big data.

The research on compression methods for smart meter big data is extensive. There are three classification criteria: (i) Whether compression occurs in real-time; (ii) Whether the compression process involves static data compression or dynamic data compression; (iii) Whether lossy compression or lossless compression methods can be applied depending on whether the method can restore the original information after compression.

In this paper, compression methods of smart meter big data are divided into lossy compression methods and lossless compression methods. This classification is based on the criterion of whether the compressed data can be restored to the original data [83,84]. Lossless compression can completely restore all original information, but lossy compression will suffer some information loss in the recovery of data. In general, lossy compression has better compression effects than lossless compression, if the trade-off between compression ratio and information loss is acceptable when selecting a specific compression method [85,86].

In today's smart grid, several data compression methods have been applied in data transmission, storage, analysis, and mining. Generally, lossless compression methods are used to compress data for transmission and storage, and lossy compression methods are used to improve the efficiency of data analysis and mining. This study proposed a compression system for smart meter data to relieve data transmission line burden and reduce storage overhead in Fig. 4.

The following discusses several common methods of smart meter big data compression.

4.1. Lossy compression

Lossy compression compresses data under the condition of sacrificing some part of the information, but ideally retains the most valuable information of the original data. In this paper, several common lossy

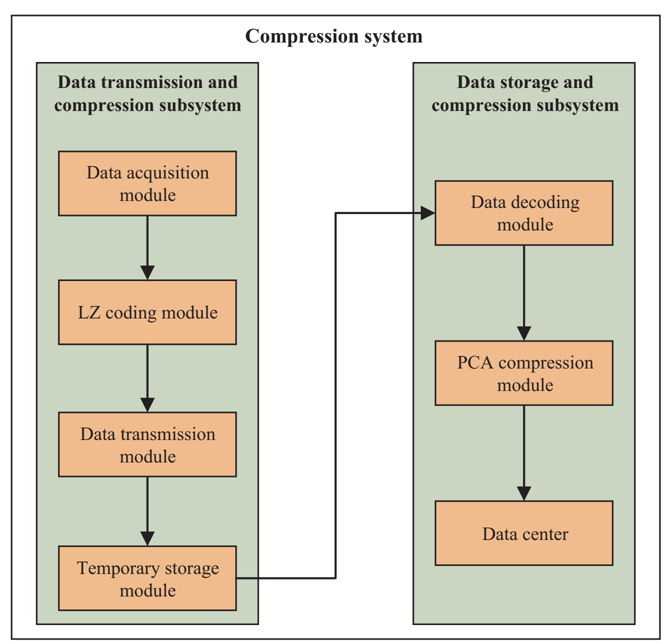

Fig. 4. Compression system for smart meter data to relieve data transmission line burden and reduce storage overhead.

compression methods are discussed, such as wavelet transform (WT), symbolic aggregation approximation (SAX), principal component analysis (PCA), singular value decomposition (SVD), linear regression based dimension reduction, and sparse coding (SC).

4.1.1. WT

Wavelets were developed by Morlet and Grossman in the early 1980's as a perfect combination of Fourier analysis, functional analysis, harmonic functions, numerical analysis, etc. They can be used to localize time-frequency data and analyse signals by stretching and translation [87]. WT, as a new transform analysis method, inherits and extends the idea of local change of the short-time Fourier transform. At the same time, it solves a variability problem in that window size does not change with frequency, so it is an effective tool for time-frequency analysis and processing of signals [88]. Generally, WT can be divided into discrete wavelet transform (DWT) and continuous wavelet transform (CWT) categories. DWT is usually used for signal coding and CWT is often employed to analyse signals. DWT has also been widely used in signal and image compression. The signal is represented by a wavelet basis with fast attenuation of finite length, and the input signal is matched by stretching and translation. The basic formula of DWT is as follows.

$$(W_\varphi f)(m,n) = < f, \varphi_{m,n} > = \alpha|^{-\frac{m}{2}} \int_R f(x)\varphi\left(\frac{x-nb}{\alpha^m}\right)dx \tag{1}$$

Ning et al. [89] decomposed the time series into proportional coefficients and wavelet coefficients by WT-based multi resolution analysis (MRA). Then trivial data points were removed to achieve data compression. The inverse multi-resolution transform can reconstruct the time series perfectly. WT can also effectively suppress sine-wave and white noise. If the algorithm is embedded in the monitoring equipment of the smart grid, the data can be compressed before sending it to a data centre. In [90–92], WT is used to compress power quality data and power system disturbance. The method not only reduces the burden on data communication lines in a smart grid, but can find any disturbance or fault in remote relay protectors and acts as a fault recorder. However, the compression ratio requires further improvement. Khan et al. [93] presented a wavelet packet decomposition method for data de-noising and compression. The authors extended the wavelet decomposition tree into a complete binary tree. By searching for the best binary tree from a certain number of discrete wavelet packets, the method can provide a better signal representation than simple wavelet decomposition, and achieve a higher compression ratio. However, the time complexity of the method is high.

WT-based data compression methods involve decomposing time series into proportional coefficients and wavelet coefficients, and then trivial data points are deleted. The objective of the method is to determine the best wavelet coefficients and optimal decomposition scale. In this case, a 0-tree wavelet transform can compress data according to different compression ratios. Wavelet packet decomposition can achieve good noise reduction and compression, as well as guarantee the integrity of information. Finally, most research has focused on one-dimensional signal compression in smart grids and the effect is no better than image compression. Overall, the compression methods based on 0-tree WT and wavelet packet can achieve better compression performance than the simple wavelet transform method.

4.1.2. SAX

SAX is a powerful method for data dimension reduction and compression, especially for dealing with time series data with small lower-bound Euclidean distance [66,94]. It discretizes time series data into symbolic strings. SAX involves two steps: First, the time series data are represented by piecewise aggregation approximation (PAA), and then the data processed by PAA is represented as a discrete string. PAA is used to replace discrete amplitude values with the average value of the amplitude over a period of time, which can be expressed using formula (2).

$$\overline{Y_i} = \frac{1}{T_i - T_{i-1}} \sum_{j=T_{i-1}+1}^{T_i} Y_j \tag{2}$$

Here, j is the index value of load data, i is the index value of the converted load data, T_i is the ith time breakpoint, and $\overline{Y_i}$ is the average value of the load in the ith time period [66,95]. Then the amplitude is divided into several intervals, and each interval is represented with different characters. That is to say, different $\overline{Y_i}$ correspond to different characters [96].

SAX can greatly reduce data dimensions and has been widely applied in data compression of smart meter big data. Notaristefano et al. [97] presented an improved algorithm for classical SAX that divides the time domain irregularly. The approach allows for denser partitioning when high-load variations occur. In [66], the authors divided the time domain through in-depth analysis of TOU tariffs and user behaviour, where the amplitude interval division is determined by the (statistical) fractile of data. This method achieves a better compression effect, but the method may ignore some data points that possess important information. Wijaya et al. [98] transformed smart meter data into symbols by means of SAX. The authors not only compressed the data, but also overcame the difficulties of massive data analysis. At the same time, some analysis algorithms used for character have been applied to smart meter data to expand data analysis and mining methods.

In general, SAX can compress real-time data effectively, especially massive and high-dimensional data, such as smart meter big data. As a lossy compression method, SAX achieves high compression ratios. The crux of SAX is the partition of time domain and amplitude interval, which influences the data compression. SAX is more efficient compared with PCA and Sammon Mapping and can be embedded in smart meters. However, the data recoverability of SAX is poor and some important information may be lost after compression.

4.1.3. PCA

PCA has been widely used in statistical analysis. It is a multivariate statistical method used to investigate the correlations among multiple variables and study how to reveal the internal structures of a number of variables through a few principal components. To retain as much information of the original variables as possible, a small number of principal components are derived from the original variables and they are not related to each other. Specifically, the principle of PCA is transforming a set of variables into a set of linearly independent variables through orthogonal transforms. The variables after transforming are called principal components.

PCA is often regarded as a linear dimensionality reduction method, which uses as few variables as possible to maintain the original data characteristics [66]. This lossy compression method has been widely applied in data compression in smart grids [66,96,99–101]. Mehra et al. [96] presented a method for smart grid data compression using PCA and iterative PCA. By comparing the experimental results under system steady state and system fault conditions, the authors found that the method can eliminate Gaussian noise in signals and effectively compresses steady-state and transient signals. Das and Rao also employed PCA to compress steady-state data in an electric power system successfully [100]. There are many other papers that use PCA to extract the most valuable features of data from load profiles [99,101,102].

The most important challenge in data compression using PCA is how to determine the number of principal components, which has great impact on the compression performance and the amount of information loss after compression. PCA has been widely used in power consumers clustering. It is generally employed to reduce the dimension of smart meter big data and obtain fine-grained results, and is of great significance for improving data mining efficiency. The advantage of PCA is that it can effectively eliminate noise while achieving data compression. With the growth of smart meter big data, power companies need faster data analysis capabilities. PCA provides an effective means to reduce data dimensions, promotes data mining efficiency, and should benefit the management of smart grids

4.1.4. SVD

SVD is an important matrix decomposition method based on linear algebra and is an extension of normalized matrix unitary diagonalization in matrix analysis. It is very popular in signal processing, statistics etc. Assuming Z is an $m \times n$ matrix, all of the elements belong to a field M, and M is either a real number field or a complex number field, then the singular value decomposition can be described by formula (3).

$$Z = X \sum Y^* \tag{3}$$

Here, X is an $m \times m$ unitary matrix; $\sum$ is a $m \times n$ positive definite diagonal matrix, and Y^* is the conjugate transpose of Y and is also an $n \times n$ unitary matrix. The diagonal elements of $\sum$ are called singular values [103,104]. Then the singular values of a given data set are arranged according to their importance. Data dimensions are reduced by discarding unimportant feature vectors according to certain criteria.

SVD has been successfully applied in image compression and other data compression fields [105–109]. In [105], an SVD based lossy compression method was proposed. The method was performed on the data obtained from different measurement devices at different time points in a smart grid. It is simple to use and has high data compression efficiency. Thus, they developed a new method to compress electric power big data.

SVD has been used in image compression widely, but application in smart meter big data compression is still in its infancy. However, the decomposition is simple, efficient, and is able to achieve an acceptable compression effect. After data reconstruction, the loss of information is very low. If combined with other methods, SVD may achieve higher compression ratios. These advantages will promote its research in the field of smart meter big data compression.

4.1.5. Dimension reduction method based on linear regression

Linear regression uses dummy variables to estimate the effect of explanatory variables of different quantitative levels on the explained variables. In economics, there is a linear relationship between explained variables and explanatory variables before the value of an explanatory variable T reaches a certain level T'. When the value of the explanatory variable reaches or exceeds T', the relationship with the explained variable will change. At this point, if we know the inflection point of T, we can use dummy variables to estimate the slope of each section of the line [110].

A piecewise linear interpolation approach was proposed in [111]. First, N_0 data points are read, and the least squares method is used to obtain vectors β_1 andβ_0. Next, the mean square error and unbiased standard deviation are identified. Then N_0data points are read again to determine whether the current line is still valid through noise estimation. If Y_i is not within the $(\pm m)$ range of results predicted by formula (4)

$$y = \beta_0 + \beta_1 x, \tag{4}$$

then the least squares estimation is carried out again until the fitting of all data points is completed. When the “energy error” is 0.001%, the method achieves a compression ratio of 100:1. Considering that the accuracy of typical hardware standards is within 0.5%, the method can maintain the integrity of the information and is a lossless compression to a certain extent. Eichinger et al. [112] applied a piecewise linear regression and learned different levels of polynomial regression functions to compress smart meter big data. The technique ensures that the user-defined maximum deviations from the original values do not exceed any point in the compression time series. There are some related works that have focused on the selection of regression functions. For example, a constant function with an approximate fixed-length interval was used as the regression function in [113] Furthermore, a variable time-interval constant function was used to achieve better compression [114,115]. The selection of regression function affects compression effect, so it is advisable to evaluate different regression functions for different data sources to achieve the best compression effect.

Linear regression is a traditional statistical method and its development is mature. Its application in smart meter data compression is extensive. Because of the high dimensionality of smart meter big data, we usually adopt piecewise linear regression to complete data compression. The choice of regression function is very important when smart meter big data is compressed, and both the variability and the length of the time interval have major impact on compression effect.

4.1.6. SC

SC is an artificial neural network method and is characteristic of spatial locality, directivity, and band-pass of the frequency domain. As a data compression and feature extraction technology, SC is widely employed in image processing and semantic recognition. The sparsity of smart meter big data provides the possibility for data compression using SC [85].

Wang et al. [85] provided a smart meter data compression method based on sparse and redundant representation. The authors compressed the data from a single smart meter and extracted a variety of partial usage patterns (PUPS) from load profiles. A number of training and optimization processes were applied to determine the redundant partial usage patterns, and each load profile was decomposed into linear combinations with very few PUPS components. Assuming x_i is a load profile denoted as $x_i = [x_{i,1}, x_{i,2}, \cdots x_{i,n}]^T$, it represents a set of k elementary vectors where each vector is a PUP.

$$x_i = \sum_{k=1}^{k} a_{i,k} d_k \tag{5}$$

In (5), d_k is the kth PUP, $a_{i,k}$ is a coefficient vector of the k PUPS, and these k PUPS form a redundant dictionary $D(k > n)$. Given the dictionary $D(k > n)$, searching for sparse vectors is called encoding. The linear combination of the base vectors is called the reconstruction of the load profile. Yu et al. [86] used SC to fit the load profile, and achieved good compression of smart meter big data. Then, the compressed data was used to predict a user's electricity demand. Similarly, to compress data in a smart grid and enhance the data transmission speed, Rui et al. [116] employed SC to process the data.

SC is an effective pattern extraction technology, and has been used in user's load profiles fitting extensively. With this method, users’ load profiles can be broken down into partial use patterns. Compared with PCA and DWT, SC can achieve a higher compression ratio with lower information loss after compression. The performance of SC is determined by the sparse coefficient, if the appropriate sparse coefficient cannot be found, the recovery of compressed data will be significantly degraded.

4.2. Lossless compression

Lossless compression is a compression method that can preserve all the information of the original data. In the morden smart grid, lossless compression methods are usually employed to relieve pressure on the communication lines. This paper focuses on two common lossless compression algorithms, namely Huffman coding and the Lempel-Ziv (LZ) algorithm.

4.2.1. Huffman coding

Huffman coding is a coding method involving variable word length, which constructs a codeword with different prefix and the shortest average header length based on the probability of occurrence of certain characters. It is sometimes referred to as the optimum coding. The main processes of Huffman coding are as follows. First, given a binary tree initial set $T = \{t_1, t_2, \cdots t_m\}$, each binary tree has only one root node with weight w_i, and its left and right sub-trees are empty. Next, the two trees with smallest weight root nodes in T are selected as the left and right sub-trees of the newly constructed binary tree. The weight of the root node in the new binary tree is the sum of the root nodes weights of the left and right sub-trees. Then the two trees are deleted from T and the new binary tree is added to the set T in ascending order. Finally, the second and third steps are repeated until there is only one binary tree in set T [117–119].

Zeinali et al. [120] discussed an adaptive form of Huffman coding. Only one pass is needed to construct a Huffman tree. A novel compression method was proposed for Wireless Sensor Network data based on the principle of adaptive Huffman coding [121,122]. It encodes the elementary characters in the difference value. However, Huffman coding has limits regarding encoding and decoding efficiency as the length of the code is variable.

As a traditional compression algorithm, it has an edge on compressing smart meter big data due to of its ability for perfect data recovery. However, the compression ratio is not very high and the algorithm efficiency is low compared with other compression algorithms. With the creation and development of LZ algorithms, Huffman coding is gradually being replaced in smart meter big data compression.

4.2.2. Lempel-Ziv (LZ) algorithm

LZ is a compression method invented by Lempel and Ziv based on the table search algorithm. The basic principle of the LZ algorithm is to create a compiled table based on the characters extracted from the original text file data. An index of each character in the compiled table is used to replace the corresponding character in the original text file data [123]. Suppose that the data symbol set $S = \{s_1, s_{2,} \cdots s_m\}$ has m symbols and the input data symbol sequence is $U = \{u_1, u_2, \cdots u_n\}$. Encoding divides the sequence into different segments. First, one symbol is taken as the first segment, and then continue to segment (a subsequent segments may contain several symbols). If the symbols of the next segment are the same as previous segments, a following symbol is added to form a new segment together that is different from previous segments. All the segments are used to construct a dictionary. When the dictionary reaches a certain size, the process of segmentation should be checked to determine whether there is a duplicate phrase in the dictionary. If there is a repetition, a new symbol will be added to one of the duplicate phrases in the dictionary until the source sequence ends. The resulting code words consists of a segment number and a symbol [124,125].

With further research, some improved compression algorithms based on LZ have been proposed. Zeinali et al. [120] used the LZW algorithm to replace a repetitive string with a single code word in the dictionary and assign a new reference number to the dictionary for the new codeword. The algorithm allocates new code words for subsequent characters, and then the compression processing of smart meter big data is performed. A tiny Lempel Ziv Markov Chain algorithm was also introduced in [122], where a 128-byte constraint history window was used for the dictionary encoding step and range encoding was completely ignored. Kraus et al. [126] proposed a method to compress smart grid data and compared it with LZMA. The authors found that the compression ratio of LZMA is limited to a certain extent.

It is vital to choose the appropriate total length of the dictionary list in LZ algorithms. With increasing data sequence, the efficiency of LZ will increase. As a lossless compression method, LZ can save all the information of the original data and can be used to compress some key data. The algorithm has been widely applied in measurement equipment to relieve transmission pressure on communication lines.

5. Compression results evaluation of smart meters big data

There are essentially three evaluations pertaining to compression of smart meter big data: (i) Whether the compressed data can be accurately restored to original data; (ii) Whether the required storage space ratio before and after compression is large; (iii) Whether the compression algorithm is simple to apply. It would also be better to achieve real-time compression and decompression.

The common lossy compression methods of smart meter big data include WT, SAX, PCA, SVD, linear regression, and sparse coding. LZ encoding and Huffman coding algorithms are two commonly used lossless compression methods. Generally, lossy compression methods have higher compression ratios than lossless compression methods. However, lossy compression methods have a certain amount information loss, and the loss in each method is different. It is important to measure the loss of information and compression ratio under different requirements. For example, for smart meter data transmission to a data centre, it would be better to use lossless compression methods to ensure the integrity of electricity consumption data. In the mining of smart meter big data, lossy compression methods can accelerate data mining efficiency. Furthermore, different compression methods have different advantages and disadvantages, and they are suitable for different applications. For example, lossless compression methods are generally employed to compress data before they are transmitted to data centres, thus relieving the pressure on data transmission lines and reducing the cost of storage. Lossy compression methods have been widely applied to alleviate the data mining issues in the smart meter big data environment.

Table 1 lists the advantages and disadvantages of some data compression methods.

Compression rate generally refers to the size of compressed data divided by the size of data before compression. For example, if the size of a 100 M file changes to 90 M after compression, the compression rate is $90/100 \times 100\% = 90\%$. The lower the compression rate the better it is, but the lower compression rate will cause longer compression and decompression time. For the compression of smart meter big data, the compression rate refers to the ratio of the amount of data before compression to the amount of data after compression. For example, if the original data volume is 1000 GB, compressed data size is 10 GB, the compression ratio is 100. Table 2 lists the compression ratios of several data compression methods.

In Table 2, the lossy compression methods mainly achieve higher compression ratios than the lossless compression methods.

There exists another evaluation standard of compression results, which is the running time of the compression algorithm. Running time is very important for compressing data in real-time and for further data analysis. The embedded 0-tree wavelet algorithm is faster than the simple wavelet transform, and it is considered highly suitable for real-time data compression, especially for smart meter big data [72]. Because studies regarding running time of smart meter big data compression are few, it is difficult to provide a detailed comparison among different compression methods. Generally, the running times of lossless compression algorithms are longer than those of lossy compression algorithms.

The selection of a data compression method may depend on its intended use, involving trade-offs among compression rate, information restorability, and algorithm efficiency.

6. Conclusions

With the continued development of smart grids, the number of smart meters are increasing and smart meter big data is growing rapidly. It is very important to compress the data to relieve transmission pressure on communication lines and reduce the storage overhead of data centres, as well as to enhance the efficiency of data mining. This paper analyses the characteristics and applications of smart meter big data, discusses various compression methods that may be applicable to smart meter big data, and discusses evaluation considerations (e.g., compression ratio, algorithm speed).

The compression methods applicable to smart meter big data can be

Table 1
Advantages and disadvantages of some data compression methods.

Methods	Data sources	Advantages	Disadvantages	References
WT	Phasor data Power quality data	• It can compress disturbance signals • It can depress sinusoidal and white noise • It can find the disturbances and faults	• Difficult in selection of order and scale of WT	[89–92]
WPD	PMU data	• Better signal representation • Satisfied features preservation and redundancy removal	• Difficult in selection of a suitable W function • High time complexity	[92,93]
SAX	Smart meter data	• Fast data processing • It can extract features and make data discrete	• Poor data recovery	[66,98]
PCA	Phasor data Smart meter data Steady-state data	• It can effectively eliminate Gaussian noise • It can effectively compress steady-state and transient signals • Low information loss • Flexible control of compression ratio	• It may lose some pivotal data points	[93,100–102]
SVD	Metering data from substations	• Simple operation and high efficiency • Accurate reconstruction of original data	• High time complexity	[105]
Linear Regression	Smart meter data	• High compression ratio	• High information loss • Difficult in selection of a proper regression function	[113–115]
SC	Smart meter data	• It can compress data and extract patterns effectively	• Difficult in selection of a proper sparse coefficient	[85,86,116]
Huffman coding	Smart grid data	• Complete data recovery	• Very time-consuming	[120–122]
LZ	smart grid or power quality data	• Simple encoding and decoding	• Low compression ratio compared with lossy compression	[120,126]

Table 2
Compression ratios of several data compression methods.

Group	Category	Methods	Compression ratio	References
Lossy Compression	Wavelet Transform	Daubechies wavelet	6:1 to 3:1	[90]
		2Daubechies wavelet	5.4:1	[89]
	Wavelet Packet	WPD	8.98:1	[93]
		WPT and LZW	10:1	[92]
	SVD	SVD	27.8:1	[105]
Lossless Compression	LZW	LZW Chain	7.8:1	[105]
		LZMH	5:1 to19:1	[122]

categorized as lossy compression methods and lossless compression methods. Lossy compression methods cannot restore the original data completely after compression, but can typically retain the most valuable information. In contrast, lossless compression methods have the ability to restore data completely. However, lossless methods are often not very efficient and achieve lower compression ratios than lossy compression methods. At present, lossy compression methods have been widely used in data mining and analysis in the environment of electric power big data.

Although there are many studies on the compression of smart meter big data, some problems remain. Especially in practical applications, there are major challenges to be addressed: With the development of smart grids, power companies demand higher requirements for real-time data compression, so it is important to achieve an acceptable balance between efficiency and the loss ratio in the selection of compression methods. Currently, there is no perfect fool-proof system available to evaluate the ideal compression effect of algorithm(s) to process smart meter big data.

Acknowledgement

This work is supported by the National Natural Science Foundation of China under Grant Nos. 71501056, 71521001 and 71503065, Anhui Science and Technology Major Project under Grant No. 17030901024, China Postdoctoral Science Foundation under Grant No. 2017M612072, and Anhui Provincial Natural Science Foundation Program under Grant No. 1608085QG165.

References

[1] Jansson PM, Udo VE. The role of the electric power industry in global sustainable development. In: Proceedings of IEEE international conference on systems, man and cybernetics; 2003, p.1729–36.
[2] Lai TM, To WM, Lo WC, Choy YS, Lam KH. The causal relationship between electricity consumption and economic growth ina gaming and tourism center: the case of Macao SAR, the People's Republic of China. Energy 2011;36:1134–42.
[3] Aytac D, Guran MC. The relationship between electricity consumption, electricity price and economic growth in Turkey: 1984–2007. Argum Oeconomica 2011;27:101–23.
[4] Tang CF. A re-examination of the relationship between electricity consumption and economic growth in Malaysia. Energy Policy 2008;36:3077–85.
[5] Voropai NI. Investments and development of electric power industry in market environment. In: Proceedings of international conference on power system technology, Powercon; 2002, p. 32–5.
[6] Sharma SS. The relationship between energy and economic growth: empirical evidence from 66 countries. Appl Energy 2010;87:3565–74.
[7] Beard LM, Cardell JB, Dobson I, Galvan F. Key technical challenges for the electric power industry and climate change. IEEE Trans Energy Convers 2010;25:465–73.
[8] Kabalci Y. A survey on smart metering and smart grid communication. Renew Sustain Energy Rev 2016;57:302–18.
[9] Amin SM, Wollenberg BF. Toward a smart grid: power delivery for the 21st century. IEEE Power Energy Mag 2005;3:34–41.
[10] Fang X, Misra S, Xue G, Yang D. Smart grid — the new and improved power grid: a survey. IEEE Commun Surv Tutor 2012;14:944–80.
[11] Irfan M, Iqbal J, Iqbal A, Iqbal Z, Riaz RA, Mehmood A. Opportunities and challenges in control of smart grids–Pakistani perspective. Renew Sustain Energy Rev 2017;71:652–74.
[12] Leiva J, Palacios A, Aguado JA. Smart metering trends, implications and necessities: a policy review. Renew Sustain Energy Rev 2016;55:227–33.
[13] Taylor JW, McSharry PE. Short-term load forecasting methods: an evaluation based on european data. IEEE Trans Power Syst 2007;22(4):2213–9.
[14] Zheng H, Jin N, Ji C, Xiong Z, Li K. Analysis technology and typical scenario application of electricity big data of power consumers. Power Syst Technol 2015;39(11):3147–52.
[15] Saha B, Srivastava D. Data quality: The other face of Big Data. In: Proceedings of IEEE international conference on data engineering. IEEE; 2014. p.19-46.
[16] Blake R, Mangiameli P. The effects and interactions of data quality and problem complexity on classification. J Data Inf Qual 2011;2:160–75.
[17] Schleich J, Faure C, Klobasa M. Persistence of the effects of providing feedback alongside smart metering devices on household electricity demand. Energy Policy 2017;107:225–33.
[18] Weiss M, Helfenstein A, Mattern F. et al. Leveraging smart meter data to recognize home appliances. In: Proceedings of IEEE international conference on pervasive computing and communications 2012. IEEE; 2012. p.190-7.
[19] Buchmann E, Hm K, Burghardt T, Kessler S. Re-identification of smart meter data. Personal Ubiquitous Comput 2013;17:653–62.
[20] Liu X, Nielsen PS. A hybrid ICT-solution for smart meter data analytics. Energy 2016;115:1710–22.
[21] Al-Wakeel A, Wu J, Jenkins N. State estimation of medium voltage distribution networks using smart meter measurements. Appl Energy 2016;184:207–18.
[22] Liu Y, Liang J. Current status and challenges of big data processing technology for

smart grid. Innov Appl Sci Technol 2015:184.
[23] Song Y, Zhou G, Zhu Y. Current status and challenges of big data processing. Technol Smart Grid Power Grid Technol 2013;37:927–35.
[24] Zhou K, Fu C, Yang S. Big data driven smart energy management: from big data to big insights. Renew Sustain Energy Rev 2016;56:215–25.
[25] Tu C, He X, Shuai Z, Jiang F. Big data issues in smart grid–A review. Renew Sustain Energy Rev 2017;79:1099–107.
[26] Diamantoulakis PD, Kapinas VM, Karagiannidis GK. Big data analytics for dynamic energy management in smart grids. Big Data Res 2015;2:94–101.
[27] Zainudin A, Arshaduddin A. Development of portable solar electricity generating system. Universiti Teknologi Petronas; 2014.
[28] Gungor VC, Sahin D, Kocak T, Ergut S, Buccella C, Cecati C, et al. Smart grid technologies: communication technologies and standards. IEEE Trans Ind Inform 2011;7:529–39.
[29] Pingkuo L, Zhongfu T. How to develop distributed generation in China: in the context of the reformation of electric power system. Renew Sustain Energy Rev 2016;66:10–26.
[30] Zhou K, Yang S, Shen C, Ding S, Sun C. Energy conservation and emission reduction of China's electric power industry. Renew Sustain Energy Rev 2015;45:10–9.
[31] Zhang D, Yao L, Ma W. Development strategy of Chinese and foreign smart grid. J China Electro Eng 2013;33(31):2–14.
[32] Farhangi H. The path of the smart grid. IEEE Power Energy Mag 2010;8:18–28.
[33] Sun X, Wang L, Wang X. Development of Low-Carbon Electricity System Based on Smart Grid. In: Proceedings of international conference on control and automation; 2015. p. 25-8.
[34] Bayindir R, Colak I, Fulli G, Demirtas K. Smart grid technologies and applications. Renew Sustain Energy Rev 2016;66:499–516.
[35] Zhou K, Yang S, Chen Z, Ding S. Optimal load distribution model of microgrid in the smart grid environment. Renew Sustain Energy Rev 2014;35:304–10.
[36] Chen SY, Song SF, Lan-Xin LI, Shen J. Survey on smart grid technology. Power Syst Technol 2009;33:1–7.
[37] Brown RE. Impact of Smart Grid on distribution system design. In: Proceedings of power and energy society general meeting on conversion and delivery of electrical energy in the 21st century; 2008. p. 1–4.
[38] Galli S, Scaglione A, Wang Z. For the grid and through the grid: the role of power line communications in the smart grid. Proc IEEE 2011;99:998–1027.
[39] Nie Y, Yang W, Huang L, Lu X. A privacy preserving scheme for change increment of electricity big data base on symmetric cryptography. In: Proceedings of IEEE international conference on cloud computing and intelligence systems; 2014. p. 487-90.
[40] Wang X, Shi Y, Zhang J, Bo L, Cheng C. Computation services and applications of electricity big data based on Hadoop. Power Syst Technol 2015;39(11):3128–33.
[41] Zhan J, Huang J, Niu L, Peng X, Deng D, Cheng S. Study of the key technologies of electric power big data and its application prospects in smart grid. In: Proceedings of power and energy engineering conference (APPEEC), IEEE PES Asia-Pacific; 2014. p. 1-4.
[42] Huang Y, Zhou X. Knowledge model for electric power big data based on ontology and semantic web. J Power Energy Syst CSEE 2015;1:19–27.
[43] Peng X, Deng D, Cheng S, Wen J, Zhaohui LI, Lin N. Key technologies of electric power big data and its application prospects in smart grid. J China Electromech Eng 2015;35:503–11.
[44] Ukil A, Zivanovic R. Automated analysis of power systems disturbance records: smart grid big data perspective. In: Proceedings of IEEE innovative smart grid technologies Asia - ISGT ASIA; 2014. p. 126-31.
[45] Koutitas G, Tassiulas L. Smart grid technologies for future radio and data center networks. IEEE Commun Mag 2014;52:120–8.
[46] Niyato D, Wang P. Cooperative transmission for meter data collection in smart grid. IEEE Commun Mag 2012;50:90–7.
[47] McLoughlin F, Duffy A, Conlon M. A clustering approach to domestic electricity load profile characterisation using smart metering data. Appl Energy 2015;141:190–9.
[48] Kim YJ, Thottan M, Kolesnikov V, Lee W. A secure decentralized data-centric information infrastructure for smart grid. IEEE Commun Mag 2010;48:58–65.
[49] Murrill BJ, Liu EC, Thompson RM, Ii. Smart meter data: privacy and cybersecurity. congressional research service reports. Congressional Research Service, Library of Congress. [Avaliable from < http://marylandsmartmeterawareness.org/wp-content/uploads/2012/07/Congress-Research-Serv ice-SM-privacy-and-cybersecurity.pdf >].
[50] Molina-Markham A, Shenoy P, Fu K. et al. Private memoirs of a smart meter. In: Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building. ACM; 2010. p. 61-6.
[51] Benzi F, Anglani N, Bassi E, Frosini L. Electricity smart meters interfacing the households. IEEE Trans Ind Electron 2011;58:4487–94.
[52] Xu D, Luan W, Wang P, Zhang Z. Data analysis method and application of smart meter. Power Supply Demand 2015;32:25–30.
[53] Reddy K, Kumar M, Mallick T, Sharon H, Lokeswaran S. A review of integration, control, communication and metering (ICCM) of renewable energy based smart grid. Renew Sustain Energy Rev 2014;38:180–92.
[54] Guo Z, Wang ZJ, Kashani A. Home appliance load modeling from aggregated smart meter data. IEEE Trans Power Syst 2015;30:254–62.
[55] Madnick S, Zhu H. Improving data quality through effective use of data semantics. Ssrn Electron J 2004;59:460–75.
[56] Tong X, Kang C, Xia Q. Smart metering load data compression based on load feature identification. IEEE Trans Smart Grid 2016;7:2414–22.
[57] Wang Y, Gao Y, Jin F. Research on intelligent meter data analysis and application. Inform Technol 2015;2:64–8.
[58] Viegas JL, Vieira SM, Melício R, Mendes V, Sousa JM. Classification of new electricity customers based on surveys and smart metering data. Energy 2016;107:804–17.
[59] Zhang L, Wang X, Li X. Operation and maintenance technology of electric energy information acquisition system. China Electric Power Press; 2010.
[60] Zhou K, Yang S, Shen C. A review of electric load classification in smart grid environment. Renew Sustain Energy Rev 2013;24:103–10.
[61] Zhou K, Yang S. Understanding household energy consumption behavior: the contribution of energy big data analytics. Renew Sustain Energy Rev 2016;56:810–9.
[62] Shao Z, Yang S, Gao F. Density prediction and dimensionality reduction of mid-term electricity demand in China: a new semiparametric-based additive model. Energy Convers Manag 2014;87:439–54.
[63] Wi YM, Joo SK, Song KB. Holiday load forecasting using fuzzy polynomial regression with weather feature selection and adjustment. IEEE Trans Power Syst 2012;27:596–603.
[64] Asare-Bediako B, Kling WL, Ribeiro PF. Day-ahead residential load forecasting with artificial neural networks using smart meter data. In: Proceedings of PowerTech (POWERTECH) IEEE Grenoble; 2013. p.1-6.
[65] Amjady N, Keynia F. Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. Energy 2009;34:46–57.
[66] Wang Y, Chen Q, Kang C, Xia Q. Clustering of electricity consumption behavior dynamics toward big data applications. IEEE Trans Smart Grid 2016;7:2437–47.
[67] Nishiya K, Hasegawa J, Koike T. Dynamic state estimation including anomaly detection and identification for power systems. IEE Proceedings C (Generation, Transmission and Distribution): IET; 1982. p. 192-198.
[68] Ahmad T. Non-technical loss analysis and prevention using smart meters. Renew Sustain Energy Rev 2017;72:573–89.
[69] Coutinho MP, Lambert-Torres G, Silva LEBD, Martins HG, Lazarek H, Neto JC. Anomaly detection in power system control center critical infrastructures using rough classification algorithm. In: Proceedings of IEEE international conference on digital ecosystems and technologies. IEEE; 2009. p. 733–8.
[70] Depuru SSSR, Wang L, Devabhaktuni V. Smart meters for power grid: challenges, issues, advantages and status. Renew Sustain Energy Rev 2011;15:1–7.
[71] Siano P. Demand response and smart grids—a survey. Renew Sustain Energy Rev 2014;30:461–78.
[72] Yu M, Hong SH. Supply–demand balancing for power management in smart grid: a Stackelberg game approach. Appl Energy 2016;164:702–10.
[73] Zhou K, Yang S. Demand side management in China: the context of China's power industry reform. Renew Sustain Energy Rev 2015;47:954–65.
[74] Vardakas JS, Zorba N, Verikoukis CV. Performance evaluation of power demand scheduling scenarios in a smart grid environment. Appl Energy 2015;142:164–78.
[75] Every J, Li L, Dorrell DG. Leveraging smart meter data for economic optimization of residential photovoltaics under existing tariff structures and incentive schemes. Appl Energy 2017;201:158–73.
[76] Schweppe FC, Caraminis MC, Tabors RO, Bohn RE. Spot pricing of electricity. 144. Kluwer International; 1997. p. 399–405.
[77] Singh JG, Singh SN, Srivastava SC. An approach for optimal placement of static VAR compensators based on reactive power spot price. IEEE Trans Power Syst 2007;22:2021–9.
[78] Davison M, Anderson CL, Marcus B, Anderson K. Development of a hybrid model for electrical power spot prices. IEEE Trans Power Syst 2002;17:257–64.
[79] Zhang P, Wu X, Wang X, Bi S. Short-term load forecasting based on big data technologies. J Power Energy Syst CSEE 2015;1:59–67.
[80] Zhong H, Xie L, Xia Q. Coupon incentive-based demand response (CIDR) in smart grid. In: Proceedings of power and energy society general meeting; 2012. p. 1–6.
[81] Dyson ME, Borgeson SD, Tabone MD, Callaway DS. Using smart meter data to estimate demand response potential, with application to solar energy integration. Energy Policy 2014;73:607–19.
[82] Beckel C, Sadamori L, Staake T, Santini S. Revealing household characteristics from smart meter data. Energy 2014;78:397–410.
[83] Tcheou MP, Lovisolo L, Ribeiro MV, Silva EABD. The compression of electric signal waveforms for smart grids: state of the art and future trends. IEEE Trans Smart Grid 2014;5:291–302.
[84] Sun Y, Cui C, Lu J, Wang Q. Data compression and reconstruction of smart grid customers based on compressed sensing theory. Int J Electr Power Energy Syst 2016;83:21–5.
[85] Wang Y, Chen Q, Kang C, Xia Q, Luo M. Sparse and redundant representation-based smart meter data compression and pattern extraction. IEEE Trans Power Syst 2017;32(3):2142–51.
[86] Yu CN, Mirowski P, Ho TK. A Sparse coding approach to household electricity demand forecasting in smart grids. IEEE Trans Smart Grid 2016:1–11.
[87] Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 1989;11:674–93.
[88] Antonini M, Barlaud M, Mathieu P, Daubechies I. Image coding using wavelet transform. IEEE Trans Image Process A Publ IEEE Signal Process Soc 1992;1:205–20.
[89] Ning J, Wang J, Gao W, Liu C. A wavelet-based data compression technique for smart grid. IEEE Trans Smart Grid 2010;2:212–8.
[90] Santoso S, Powers EJ, Grady WM. Power quality disturbance data compression using wavelet transform methods. IEEE Trans Power Deliv 1997;12:1250–7.
[91] Brindha S, Sundararajan D. Power quality monitoring and compression using the discrete wavelet transform. In: Proceedings of international conference on advanced computing and communication systems; 2013. p. 1–6.
[92] Littler TB, Morrow DJ. Wavelets for the analysis and compression of power system

disturbances. IEEE Trans Power Deliv 1999;14:358–64.
[93] Khan J, Bhuiyan S, Murphy G, Williams J PMU data analysis in smart grid using WPD. In: Proceedings of the 40th annual conference of the IEEE on industrial electronics society, IECON 2014; 2014. p.1–5.
[94] Lin J, Keogh E, Lonardi S, Chiu B. A symbolic representation of time series, with implications for streaming algorithms. In: Proceedings of ACM SIGMOD workshop on research issues in data mining and knowledge discovery. ACM; 2003. p. 2–11.
[95] Sevcech J, Bielikova M. Symbolic time series representation for stream data processing. in: IEEE international workshop on real time data stream analytics; 2015. p. 217-22.
[96] Mehra R, Bhatt N, Kazi F, Singh NM. Analysis of PCA based compression and denoising of smart grid data under normal and fault conditions. in: 2013 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT). IEEE.p.1-6; 2013.
[97] Notaristefano A, Chicco G, Piglione F. Data size reduction with symbolic aggregate approximation for electrical load pattern grouping. Iet Gener Transm Distrib 2013;7:108–17.
[98] Wijaya TK, Eberle J, Aberer K. Symbolic representation of smart meter data. In: Proceedings of the joint EDBT/ICDT 2013 workshops. ACM; 2013. p. 242–8.
[99] Koivisto M, Heine P, Mellin I, Lehtonen M. Clustering of connection points and load modeling in distribution systems. IEEE Trans Power Syst 2013;28:1255–65.
[100] Das S, Rao PSN. Principal component analysis based compression scheme for power system steady state operational data. Innovative Smart Grid Technologies – India (ISGT India). 2011 IEEE PES; 2011. p. 95–100.
[101] Zakaria Z, Lo KL. Two-stage fuzzy clustering approach for load profiling. In: Proceedings of universities power engineering conference; 2009. p. 1-5.
[102] Chicco G, Napoli R, Piglione F. Comparisons among clustering techniques for electricity customer classification. IEEE Trans Power Syst 2006;21:933–40.
[103] Golub GH, Reinsch C. Singular value decomposition and least squares solutions. Numer Math 1970;14:403–20.
[104] Klema V, Laub AJ. The singular value decomposition: its computation and some applications. IEEE Trans Autom Control 1980;25:164–76.
[105] De Souza JCS, Assis TML, Pal BC. Data compression in smart distribution systems via singular value decomposition. IEEE Trans Smart Grid 2017;8(1):275–84.
[106] Wei JJ, Chang CJ, Chou NK, Jan GJ. ECG data compression using truncated singular value decomposition. IEEE Trans Inf Technol Biomed 2001;5:290–9.
[107] Ayubi J, Mehdi-Rezaei. Lossy color image compression based on singular value decomposition and GNU GZIP. Adv Comput Sci: Int J 2014;3:16–21.
[108] Rufai AM, Anbarjafari G, Demirel H. Lossy medical image compression using Huffman coding and singular value decomposition. In: Proceedings of signal processing and communications applications conference; 2013. p. 1-4.
[109] Sadek RA. SVD based image processing applications: state of the art, contributions and research challenges. Int J Adv Comput Sci Appl 2012;3(7).
[110] Blu T, Thevenaz P, Unser M. Linear interpolation revitalized. IEEE Trans Image Process Publ IEEE Signal Process Soc 2004;13:710–9.
[111] Clements AF, McCulloch MD, Nixon KJ. Low-loss, high-compression of energy profiles. In Proceedings of IEEE international conference on renewable energy research and applications (ICRERA); 2015. p. 949–53.
[112] Eichinger F, Efros P, Karnouskos S, Böhm K. A time-series compression technique and its application to the smart grid. VLDB J 2015;24:193–218.
[113] Yi BK, Faloutsos C. Fast time sequence indexing for arbitrary Lp norms. In: Proceedings of the 26th international conference on very large data bases; 2000. p. 385–94.
[114] Lazaridis I, Mehrotra S. Capturing sensor-generated time series with quality guarantees. In: Proceedings of the 19th IEEE international conference on data engineering; 2003. p. 429-/40.
[115] Keogh E, Chakrabarti K, Pazzani M, Mehrotra S. Locally adaptive dimensionality reduction for indexing large time series databases. ACM Trans Database Syst 2002;27:188–228.
[116] Rui P, Lucani DE, Phulpin Y, Nistor M. Network coding protocols for smart grid communications. IEEE Trans Smart Grid 2014;5:1523–31.
[117] Bledsoe RE. Data communication with modified Huffman coding. Google Patents; 1987.
[118] Knuth DE. Dynamic Huffman coding. J Algorithms 1985;6:163–80.
[119] Kavousianos X, Kalligeros E, Nikolos D. Optimal selective Huffman coding for test-data compression. IEEE Trans Comput 2007;56:1146–52.
[120] Zeinali M, Thompson JS. Impact of compression and aggregation in wireless networks on smart meter data. In: Proceedings of IEEE international workshop on signal processing advances in wireless communications; 2016.
[121] Mo Y, Qiu Y, Liu J, Ling Y. A data compression algorithm based on adaptive Huffman code for wireless sensor networks. In: Proceedings of the Fourth International Conference on Intelligent Computation Technology and Automation. IEEE; 2011. p.3-6.
[122] Ringwelski M, Renner C, Reinhardt A, Weigel A. The Hitchhiker's guide to choosing the compression algorithm for your smart meter data. In: Proceedings of the IEEE international energy conference and exhibition IEEE; 2012. p.935-940.
[123] Welch TA. A technique for high-performance data compression. Computer 1984;17:8–19.
[124] Wolff FG, Papachristou C. Multiscan-based test compression and hardware decompression using LZ77. In: IEEE test international conference 2002 proceedings; 2002. 331-339.
[125] Kreft S, Navarro G. Self-indexing based on LZ77. In: Proceedings of Conference on Combinatorial Pattern Matching; 2011. p. 41-54.
[126] Kraus J, Štěpán P, Kukačka L. Optimal data compression techniques for smart grid and power quality trend data. in: Proceedings of IEEE international conference on harmonics and quality of power; 2012. p.707–12.