Advanced Engineering Informatics 30 (2016) 422–448

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Full length article

A data mining based load forecasting strategy for smart electrical grids

Ahmed I. Saleh [a,*], Asmaa H. Rabie [a], Khaled M. Abo-Al-Ez [b]

[a] *Computer Engineering & Systems Department, Faculty of Engineering, Mansoura University, Mansoura, Egypt*
[b] *Electrical Engineering Department, Faculty of Engineering, Mansoura University, Mansoura, Egypt*

A R T I C L E I N F O

Article history:
Received 10 February 2016
Received in revised form 21 April 2016
Accepted 30 May 2016

Keywords:
Smart grids
Load forecasting
Data mining
Outlier rejection
Feature selection

A B S T R A C T

Smart electrical grids, which involve the application of intelligent information and communication technologies, are becoming the core ingredient in the ongoing modernization of the electricity delivery infrastructure. Thanks to data mining and artificial intelligence techniques that allow the accurate forecasting of power, which alleviates many of the cost and operational challenges because, power predictions become more certain. Load forecasting (LF) is a vital process for the electrical system operation and planning as it provides intelligence to energy management. In this paper, a novel LF strategy is proposed by employing data mining techniques. In addition to a novel load estimation, the proposed LF strategy employs new outlier rejection and feature selection methodologies. Outliers are rejected through a Distance Based Outlier Rejection (DBOR) methodology. On the other hand, selecting the effective features is accomplished through a Hybrid technique that combines evidence from two proposed feature selectors. The first is a Genetic Based Feature Selector (GBFS), while the second is a Rough set Base Feature Selector (RBFS). Then, the filtered data is used to give fast and accurate load prediction through a hybrid KN^{3}B predictor, which combines KNN and NB classifiers. Experimental results have proven the effectiveness of the new outlier rejection, feature selection, and load estimation methodologies. Moreover, the proposed LF strategy has been compared against recent LF strategies. It is shown that the proposed LF strategy has a good impact in maximizing system reliability, resilience and stability as it introduces accurate load predictions.

1. Introduction

Smart electrical grids (SEGs) that employ modern IT/communication/control technologies are becoming a global trend nowadays [1]. They depend on an intelligent electricity transmission and distribution networks that employ modern information, communication and control technologies to enhance economy, efficiency, reliability, and security of the electrical grid [2]. SEGs co-ordinate the needs and capabilities of all generators, grid operators, end-users and electricity market stakeholders to operate all parts of the system as efficiently as possible, minimizing costs and environmental impacts while maximizing system reliability, resilience and stability. The potential of SEGs is huge. They could revolutionize the way to generate and use energy, enabling new forms of generation, which brings customers into the heart of the equation with their ability to shift demand and balance the system. To realize that potential, it is need to rethink the roles and responsibilities of all the players in the electrical system. Moreover, there is a critical need to continue investing in smart technologies in the near term and integrate them into existing networks. While many countries have already begun to "smarten" their electrical systems, significant additional investment and planning are required to achieve smarter grids [3].

In the context of SEG, there are several new trends related to information and communication technologies such as smart meters, demand response mechanisms, online customer interactions by PCs/mobile devices using internetworking system, and online billing. The application of these technologies leads to a more reliable power transmission and distribution with dynamic load balancing [1–3]. However, to realize the full potential of SEG, accurate load forecasting (LF) is a must. LF is a vital part for the power system planning and operation to provide intelligence to energy management within smart grids [4]. In addition, it provides a correct decision to other subsystems with in SEG. This leads to substantial savings in operating and maintenance costs, increased reliability of power supply and delivery system, and provide correct decisions for future development [1,4]. The main challenge of LF is the complex demand pattern due to the deregulation of electricity markets [4,5]. It is an important to provide an advanced LF model suitable for stable and mature conditions, as well as dynamic fast growing utilities to achieve fast and accurate forecasting decisions.

Generally, two important issues should be handling carefully before applying the forecasting model as they have strong impact

* Corresponding author.
E-mail address: aisaleh@yahoo.com (A.I. Saleh).

http://dx.doi.org/10.1016/j.aei.2016.05.005
1474-0346/

on the model performance, which are; (i) feature selection, and (ii) outlier rejection. Electrical loads are affected by a variety of features such as time factors, weather conditions, class of customers, special events, population, economic indicators, and electricity price, etc. [6]. Several prediction algorithms don't perform well with large amounts of features, while many ineffective features may be presented in electrical load data [7]. Feature selection has a good impact in improving the model performance as well as providing faster decisions as it minimizes the considered features to the only effective ones [7]. On the other hand, training of the forecasting model with raw data that usually contains outliers certainly degrades the system's performance as the decisions are affected by those rare data whose behavior is very exceptional.

Outlier detection is an outstanding data mining task referred to as outlier mining [8]. The major objectives of outliers' rejection are to improve model performance as well as the accuracy during training phase. In Electrical power system, outliers are represented as bad data which have an unexpected effect on the loads, for example the existence of a football game in the winter. This event causes the load demand of the electrical power is higher than the normal loads in winter. Because this event is not repeated, it cannot be relied upon for the learning system. Thus, it is very important to present an advanced outliers' rejection technique for eliminating all bad data from the electrical load data sets before the forecasting process.

There are several techniques that are used to improve the load forecasting accuracy. Researchers have been trying to find a solution for the problem of electricity load forecasting since 1990s [1,2,9]. Those techniques differ in the mathematical formulation and the features used in each formulation. Generally, LF techniques can be classified into; (i) classical load forecasting (CLF) methods and (ii) soft computing based load forecasting (SLF) techniques. CLF techniques such as; regression, multiple-regression, exponential smoothing and Iterative reweighted least-squares technique [10] are more complex in computational operations, with high time penalty, as well as less performance compared to SLF. On the other hand, SLF are based on artificial intelligence (AI) techniques [10] such as; fuzzy logic (FL), neural networks (NNs), evolutionary algorithms (EAs) like genetic algorithms (GAs), Wavelet Networks, expert system methods and support vector machines (SVM) [11]. However, even with soft computing techniques, many of them don't perform well with large number of features, while many irrelevant features may be presented in electrical load data sets. In addition, electrical load data sets usually contain bad data (e.g., outliers) whose behavior is very exceptional. Recently, several LF strategies for SEG have been introduced, which employ data mining techniques, such as Improved ARIMAX, KNN, ANNs model, and K-mean [12–15]. In spite of their effectiveness, recently introduced LF strategies still suffer from overfitting and high time and computational penalties. Moreover, their performance badly affected by the existence of large features as well as unwanted outliers. Therefore, it is important to propose an advanced soft computing technique for an accurate and fast load prediction with high performance.

The originality of this paper is concentrated in introducing a complete LF strategy for ESG. The proposed forecasting strategy is composed of two phases, which are; (i) Data pre-processing (DP) and (ii) Load estimation (LE). During DP, the task is to collect the training (historical) load data, then represent it in a suitable form for the forecasting model. To accomplish such aim, two main processes are performed, which are; outliers rejection and features selection by employing data mining techniques to give a meaningful pattern of this data. Outliers are rejected through a new method called Distance Based Outlier Rejection (DBOR). On the other hand, selecting the effective features is accomplished through a Hybrid methodology that combines evidence from two proposed feature selectors. The first is a Genetic Based Feature Selector (GBFS), while the second is a Rough set Base Feature Selector (RBFS). During the second phase (e.g., LE), the filtered data is used to give fast and accurate load prediction through a hybrid KN^{3}B predictor, which combines KNN and NB classifiers. Each of the proposed methodologies (e.g., outlier rejection, feature selection, and load estimation) have been evaluated through excessive experiments. The overall LF strategy is tested to prove the compatibility of the proposed methodologies. Results have shown that the proposed LF strategy outperforms recent ones as it introduces the best prediction accuracy with the minimum time penalty.

This paper is organized as follows; Section 2 represents a background that shows an overview about the smart electrical grids as well as their main parts. It also illustrates the basic concepts of outlier rejection techniques as well as feature selection methodologies. Section 3 shows the previous efforts about load forecasting. Section 4 focuses on the proposed load forecasting strategy. Section 5 depicts the experimental results. Finally, conclusions are presented in Section 6.

2. Background and basic concepts

Through this section, a quick view of the smart electrical grids will be introduced. Then, the concept of outlier rejection will be explained as rejecting outliers from the input training dataset is one of the main contribution of this paper. Finally, feature selection methodologies will be briefly mentioned.

2.1. Smart electrical grid (SEG)

The architecture and main components of SEG are illustrated in Fig. 1, which are classified into; (i) Low level (basic) components (LLC) and (ii) High level (intelligent) components (HLC) [1–3,16]. LLCs are the basic SEG subsystems, which are; power generation, transmission, and distribution. Control actions, protection and planning are carried out separately at each subsystem of this level.

HLCs, on the other hand, can be considered as the brain of SEG. While LLC are the actuators of SEG, HLCs are used mainly to add intelligence to such actuators to achieve better performance. Like LLCs, HLCs is composed of three sub-systems, which are [1–3];

1. Power load forecasting sub-system, which is used to predict the future load either in short term or long term manners. Decisions of such sub-system are forwarded to both the demand response sub-system to derive the future switching decisions and to the power generation sub-system to indicate the future electrical demands.
2. Demand Response sub-system, which achieves the most planning efficiency of the system resources based on the load forecasting values. Moreover, it controls switching resources to balance the supply and the demand of electricity.
3. Wireless networking sub-system, which is used to exchange both of communication and information between all subsystems in smart grid depending on wireless devices in a secure and a reliable manner.

The load forecasting process is performed for predicting future power consumption by the customers at the electric distribution level. This process is affected by different consumption characteristics of the customers, which depend largely on the operating conditions that effect on the consumption. These conditions are manly the time of load operation during the day (peak hours or off peak hours), the season, the weather conditions, and the holidays which differ considerably from normal week day's consumption. One or more of these factors can affect the consumption pattern of the customer; hence they also affect the forecasted power.