Renewable and Sustainable Energy Reviews 60 (2016) 615–630

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

A knowledge discovery in databases approach for industrial microgrid planning

Carlos Gamarra[a], Josep M. Guerrero[b,*], Eduardo Montero[a]

[a] *Department of Electromechanical Engineering, University of Burgos, Spain*
[b] *Department of Energy Technology, Aalborg University, Denmark*

ARTICLE INFO

Article history:
Received 26 April 2015
Received in revised form
2 January 2016
Accepted 13 January 2016
Available online 8 February 2016

Keywords:
Microgrid planning
Knowledge discovery in databases
Energy Management Systems
Data Mining
Machine Learning
Sustainability

ABSTRACT

The progressive application of Information and Communication Technologies to industrial processes has increased the amount of data gathered by manufacturing companies during last decades. Nowadays some standardized management systems, such as ISO 50.001 and ISO 14.001, exploit these data in order to minimize the environmental impact of manufacturing processes. At the same time, microgrid architectures are progressively being developed, proving to be suitable for supplying energy to continuous and intensive consumptions, such as manufacturing processes.

In the merge of these two tendencies, industrial microgrid development could be considered a step forward towards more sustainable manufacturing processes if planning engineers are capable to design a power supply system, not only focused on historical demand data, but also on manufacturing and environmental data. The challenge is to develop a more sustainable and proactive microgrid which allows identifying, designing and developing energy efficiency strategies at supply, management and energy use levels.

In this context, the expansion of *Internet of things* and *Knowledge Discovery in Databases* techniques will drive changes in current microgrid planning processes. In this paper, technical literature is reviewed and this innovative approach to microgrid planning is introduced.

Contents

Abbreviations: (IoT), Internet of things; (MG), Microgrid; (KDD), Knowledge discovery in databases; (DM), Data mining; (DW), Data warehousing; (ICTs), Information and communication technologies; (ML), Machine learning; (MES), Manufacturing Execution System; (ENMS), Environmental Management System; (EMS), Energy Management System; (ERP), Enterprise Resource Planning system; (IoE), Internet of Energy; (DER), Distributed Energy Resources; (CERTS), Consortium for Electric Reliability Technology Solutions; (LMP), Locational Marginal Pricing; (M2M), Machine-to-Machine; (H2M), Human-to-Machine

* Correspondence to: Department of Energy Technology, Aalborg University, Microgrids Research Programme, Pontoppidanstraede 101, Room 79, 9220 Aalborg East, Denmark. Tel.: +45 9940 9726.
E-mail address: joz@et.aau.dk (J.M. Guerrero).
URL: http://www.microgrids.et.aau.dk (J.M. Guerrero).

http://dx.doi.org/10.1016/j.rser.2016.01.091
1364-0321/

1. Introduction

During 2008, the world's companies processed 63 terabytes of information annually on average and the world servers processed 12 gigabytes of information daily for the average worker (about 3 terabytes of information per worker per year) [1]. For sure Internet has changed the amount of data available for companies. Following this technological evolution (towards data acquisition, transmission and storage) new concepts have appeared around computer-based science in business environments. *Internet of things* (IoT) is perhaps one of the trending topics in this field nowadays. Many authors have approached it since this term arose in 1999. For example F. Mattern and C. Floerkemeier affirm in [2] that IoT *represents a vision in which the Internet extends into the real world embracing everyday objects. Physical items are no longer disconnected from the virtual world, but can be controlled remotely and can act as physical access points to Internet services.*

Hence, it can be expected that the progressive connection of everyday objects to internet will be used to *remotely determine their state so that information systems can collect up-to-date information on physical objects and processes* [2]. Also *devices should be able to communicate each other, and to develop a certain level of intelligence.* The IoT vision is grounded in the steady advances in electronics, communications and information technologies. Due to their diminishing size [3], falling price and declining energy consumption, processors, communications modules and other electronic devices are being increasingly integrated into everyday objects. Main objectives of the integration of this kind of devices are data gathering, measuring and communication. Perera et al. identify smart grid, smart homes and smart industries between main contributors to smart products sales market by 2016 [4].

As J. Short et al. point out in [1], there exist some differences between two related concepts: *data* and *information*. Since *data are collections of numbers, characters, images or other outputs from devices that represent physical quantities as artificial signals intended to convey meaning,* they define information as *a subset of data,* considering data as *the lowest level of abstraction from which information and knowledge are derived*. During the period from 1986–2007, general-purpose computing capacity grew at an annual rate of 58%, and the world's capacity for bidirectional telecommunication grew at 28% per year, closely followed by the increase in globally stored information (23%) [5].

KDD is essentially the process of discovering useful knowledge from a collection of data. A. Berstein et al. also define KDD as *the result of an exploratory process involving the application of various algorithmic procedures for manipulating data, building models from data, and manipulating the models* [6]. The exponential grow of the amount of data in many systems, no longer allows the manual search of underlying patterns, as it used to be. *The main objective of KDD is to extract high-level knowledge from these low-level information, or in other words, to automatically process large quantities of raw data, identify the most significant and meaningful patterns, and present these as knowledge appropriate for achieving the user goals* [7]. Relationship between KDD, IoT and Data Mining (DM) is described in an accurate way by N. Ramakrishan in [8]:

- IoT collects data from different sources, which may contain data for the IoT itself.
- KDD, when applied to IoT, will convert the data collected by IoT into useful information that can then be converted into knowledge.
- DM is responsible for extracting patterns or generating models from the output of the data processing step and then feeding them into the decision-making step, which takes care of transforming its input into useful knowledge.

There are critical steps along a KDD process. Yoong and Kerschberg assert in [9] that knowledge discovery critically depends on how well a database is characterized and how consistently the existing and discovered knowledge is evolved. The step definition of the KDD process can also have a strong impact on the final results of mining. For example, not all the attributes of the data are useful for mining. The consequence is that DM algorithms may have a hard time to find useful information if the selected attributes cannot fully represent the characteristics of the data [8]. DM is described by Fayad et al. in [10] as a step in the KDD process that consists of *applying data analysis and discovering algorithms that produce a particular enumeration of patterns (or models) over the data.* But every DM process requires a previous data processing step, also defined by Fayad et al. as *data warehousing* (DW). DW refers to *collecting and cleaning transactional data to make them available for online analysis and decision support*. DW helps set the stage for KDD in two important ways: data cleaning and data access.

Typical KDD process includes five general stages: selection, pre-processing, transformation, data mining and evaluation. But, instead of being based in the same principles, different authors propose different KDD processes. Fayad et al. [10] define KDD as an *iterative and interactive* process based in nine steps such as:

- *Developing an understanding of the application domain and the relevant prior knowledge and identifying the goal of the KDD process from the customer's viewpoint.*
- *Creating a target data set.*
- *Data cleaning and pre-processing.*
- *Data reduction and projection*
- *Matching the goals of the KDD process to a particular data mining method.*
- *Exploratory analysis and model and hypothesis selection.*
- *Data mining.*
- *Interpreting mined patterns.*
- *Acting on the discovered knowledge.*

M. Last et al. introduce a specific time series databases KDD process [11] based on seven stages: data pre-processing, feature extraction, transformation, dimensionality reduction, prediction and rule extraction. A review on time series DM techniques is also presented by Fu [12]. Between mining tasks he highlights pattern discovery and clustering but also classification, rule discovery, summarization and other recent research directions. Finally, a deep review of 13 different KDD process models is presented by Kurgan and Musilek in [13]. Analyzing these KDD methodologies, data preparation can be considered the foundation, while *DM can be considered as the pillar of KDD*. The existence of similar, but at the same time different KDD methodologies makes sense since:

- KDD techniques have not been widely applied to manufacturing processes, neither standardized yet.
- KDD techniques are based on optimization problems between different alternatives, under different constraints and towards different goals depending not only on the characteristics of the manufacturing process, but also on environmental, social and legal conditions: there are some aspects in a KDD-based approach that might not be standardized.

On the basis of KDD, a growing body of emerging applications is changing the landscape of business decision support [14] such as: risk analysis, targeted marketing, customer retention [15], portfolio management and brand loyalty [16]. *Traditional DM approaches have proven to be efficient on modeling variables of interest*, so that these variables may be forecasted in future scenarios, and effective decisions taken based on that forecast. DM technologies are reviewed, described and classified [8] into