Future Generation Computer Systems 63 (2016) 131–147

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

A hybrid integrated architecture for energy consumption prediction

Alejandro Maté [a], Jesús Peral [a,*], Antonio Ferrández [a], David Gil [b], Juan Trujillo [a]

[a] Department of Software and Computing Systems, University of Alicante, Spain

[b] Department of Computing Technology and Data Processing, University of Alicante, Spain

HIGHLIGHTS

- Energy consumption predictions based on data mining and supported by external data.
- Heterogeneous data are combined through DW and Information Extraction (IE).
- Multidimensional model integrates information extracted from Social Networks and IE.
- The scenario: consumption prediction is modified with external unstructured data.

ARTICLE INFO

Article history:
Received 12 June 2015
Received in revised form
19 March 2016
Accepted 21 March 2016
Available online 31 March 2016

Keywords:
Data mining
Energy consumption
Information Extraction
Big data
Decision trees
Social networks

ABSTRACT

Irresponsible and negligent use of natural resources in the last five decades has made it an important priority to adopt more intelligent ways of managing existing resources, especially the ones related to energy. The main objective of this paper is to explore the opportunities of integrating internal data already stored in Data Warehouses together with external Big Data to improve energy consumption predictions. This paper presents a study in which we propose an architecture that makes use of already stored energy data and external unstructured information to improve knowledge acquisition and allow managers to make better decisions. This external knowledge is represented by a torrent of information that, in many cases, is hidden across heterogeneous and unstructured data sources, which are recuperated by an Information Extraction system. Alternatively, it is present in social networks expressed as user opinions. Furthermore, our approach applies data mining techniques to exploit the already integrated data. Our approach has been applied to a real case study and shows promising results. The experiments carried out in this work are twofold: (i) using and comparing diverse Artificial Intelligence methods, and (ii) validating our approach with data sources integration.

1. Introduction

Energy resources have been used irresponsibly and negligently in the past five decades. Therefore, the European Union [1–4] and other international organizations [5,6]are promoting initiatives to foster responsible and efficient actions to ensure the sustainability of cities.

Several initiatives have highlighted how Information and Communication Technologies (ICTs) can be used to achieve cities' climate targets by lowering energy use and greenhouse gas (GHG) emissions from other sectors [7,8]. Some of these initiatives include proposals such as dematerialization and demobilization, as well as comprehensive concepts for smart logistics and smart cities [9]. Hilty [10] describes how ICT can be seen as an enabling technology for improving or substituting processes in other sectors. According to the report 2020 [11] of the Climate Group and McKinsey which focused on the potential for reducing GHG emissions in six different sectors: power, transportation, agriculture, building, manufacturing and consumer and services), ICT is a key player in the battle against climate change and offers the possibility of 7.8 Gt reduction of CO_2 emission in 2020. In addition, the European Commission is stressing the importance of ICT for energy reduction and sustainability, and invests in research in this area [12]. Therefore, a number of research programmes combining ICT with energy have been developed.

According to the Climate Group, a smart city is a city that uses data, information and communication technologies strategically to provide efficient services to citizens, monitors policy outcomes, manages and optimizes existing infrastructure, employs cross-sector collaboration and enables new business models [11].

* Corresponding author.
E-mail address: jperal@dlsi.ua.es (J. Peral).

http://dx.doi.org/10.1016/j.future.2016.03.020
0167-739X/

Smart cities generate increasingly huge amounts of data which traditional data processing applications are incapable of dealing with. This is the reason why Big Data appears associated to smart cities, and represents a hot research topic [13,14]. The overall aim of this paper is to explore the opportunities of using ICT as an enabling technology to predict energy consumption in cities.

A more efficient use of energy is to minimize the energy loss caused by inaccuracies in energy prediction. Energy prediction already makes use of large volumes of data in order to make more accurate estimations. However, improving energy prediction further requires introducing streaming and highly heterogeneous information into the estimation process. While the technologies for processing this kind of information already exist, they have not yet been adapted for improving energy consumption. Therefore, more specifically, the main aim of this paper is to explore a novel architecture for predicting energy consumption in cities, based on the combination of internal data already stored in Data Warehouses with external data gathered from Big Data sources.

This paper presents a study in which we propose a hybrid architecture that is able to integrate on the fly highly heterogeneous data sources, enabling the use of current energy information combined with contextual data which are integrated with a variety of information sources or Big Data. The benefit is to improve knowledge acquisition for better decision making. The objective of this paper is to make some predictions about energy consumption based on energy data mining and supported by knowledge that provides a torrent of information in many cases hidden across heterogeneous and unstructured data sources.

The remainder of this paper is structured as follows: in the next section, the related work is reviewed; thereafter, we design, create and implement an architecture which includes an integrated model and uses data mining to access Big Data in combination with other data sources. An explanation of the architecture follows in the next section with a case scenario. Subsequently, the analysis of the results section evaluates our proposal. Finally, we include a discussion of the advantages of the model and the difficulties related to its implementation as well as the further areas of research in this field.

2. Related work

In the following subsection, the main ICT solutions for smart cities are overviewed. In the second one, the previous work on energy consumption prediction is summarized. In the third one, given that our approach uses ontologies, the main ontologies in the domain of energy are outlined. Finally, the contributions of our proposal are enumerated.

2.1. ICT solutions for smart cities

Although very often there is no explicit connection between smart and sustainable cities, it is obvious that ICT plays an essential role for supporting the transition to more sustainable cities, not only regarding the management of urban systems but also offering more support for sustainable urban lifestyles. Mitchell [9] has defined five main opportunities for how ICT can contribute to the reduction of energy use in cities. Four of these have direct effects and one has indirect effects on the reduction of energy use. For this purpose he makes use of opportunities. The first one is labeled as dematerialization. Here, physical products or services are converted to digital ones (we can imagine how the CDs are now streamed music and the bank offices are online banking services mainly). According to Hilty [10] software represents the immaterial resources and the services provided represent the value that could become the pattern for a discontinued economy. The second opportunity is demobilization, where everything that has been digitalized can be transported via the telecom network instead of being physically transported. We are now aware how transport and travel are totally or partially replaced by telecommunications. The third opportunity is mass customization where less resource use is accomplished through intelligent adaptation, personalization and demand management. The fourth opportunity, intelligent operation, involves more resource-efficient operations of, for instance, water, energy and transport systems. The fifth and indirect opportunity is soft transformation where the existing physical infrastructure is transformed because of new opportunities presented by the information paradigm. These principles can be applied to product design, architecture, urban design and planning at regional, national and global levels [9]. The Smart 2020 report [15] identifies ICT solutions by combining the abatement potential of ICT (called change levers) with economic end-use sectors. Somewhat similar to the opportunities put forward by Mitchell, the change levers are (1) digitalization and de-materialization, (2) data collection and communication, (3) systems integration and (4) process, activity and functional optimization. With the aim of mapping out ICT solutions which have the potential to offer beneficial environmental effects, [10] used a combination of economic sectors and environmental indicators to compile a list of ICT solutions for sustainable development: e-business, virtual mobility (teleworking, teleshopping, virtual meetings), virtual goods (services partially replacing material goods), ICT in waste management, intelligent transportation systems, ICT in energy supply, ICT in facilities management and ICT in production process management. In addition to this, The Climate Group [11] proposes a comprehensive list of possible ICT solutions that can be implemented, as well as setting out metrics in order to understand which solutions could be implemented to reach a specific city's goals.

2.2. Antecedents on energy consumption prediction

Within these ICT solutions, we are focusing on energy supply, specifically on the prediction of energy consumption in cities. In [16], a brief overview of over 100 years of energy forecasting practices is reported, which concludes that the electric power industry needs forecasts of supply, demand and price, so called energy forecasts, to plan and operate the grid. While many other industries have some form of inventory to store and buffer their products and services, those of the electric power industry, electricity, cannot be massively stored using today's technologies. As a result, electricity has to be generated and delivered as soon as it is produced. In other words, the utilities have to balance the supply and demand every moment.

The storage limitation and societal necessity of electricity lead to several interesting features of energy forecasting, such as the complex seasonal patterns, 24/7 data collection across the grid, and the need to be extremely accurate. In [17], a review and categorization of electric load forecasting techniques is presented. These techniques are classified there into nine categories: (1) multiple regression, (2) exponential smoothing, (3) iterative reweighted least-squares, (4) adaptive load forecasting, (5) stochastic time series, (6) ARMAX models based on genetic algorithms, (7) fuzzy logic, (8) neural networks and (9) expert systems. The state-of-the-art on energy consumption prediction techniques depends on the type of consumption at hand. Ideally, having all the information about the system would yield highly accurate results. However, it is the case that most often all information is not available, and thus, approximate models must be built.

The research line about Energy Informatics (EI) is related to this issue, which means to increase the efficiency of energy