Bernard, an energy intelligent system for raising residential users awareness

Marta Zorrilla[a], Álvaro Ibrain[b]

[a] Department of Computer Science and Electronics, University of Cantabria, Spain
[b] Computer Science Degree, University of Cantabria, Spain

ABSTRACT

Energy efficiency is still a hot topic today. Coming roughly the 25% of the energy consumption in EU from the residential sector, very few cheap and simple tools to promote energy efficiency in home users have been de- veloped. The purpose of this paper is to present Bernard, a concept proof designed for filling this gap. This aims that householders become aware of their energy habits and have useful information that help them to redirect their consumption pattern. To achieve these goals, Bernard offers, through a mobile application, the home energy consumption monitoring in real time, the energy price forecast for the next hour and the appliances which are switched on, among others. Furthermore, it is important to highlight that the system has been de- signed with the premises of being cheap, non-intrusive, reliable and easily scalable, in order that utilities can gradually deploy and provide it to their customers, gaining at the same time valuable information for decision making and improving its corporate social image. Therefore, the adopted solution is based on a real time streaming data architecture suitable for handling huge volumes of data and applying predictive techniques on a cloud-computing environment. The paper provides a detailed description of the system and experimental results evaluating the performance of the predictive modules built. As case study, REFIT and REDD datasets were used.

KEYWORDS

Energy awareness Smart consumption, Real time monitoring, Cloud computing, Deep learning

1. Introduction

Environmentally responsible energy use is still a world challenge today. The effect of the growth of population in developing countries, the increase in the demand for energy, the depletion of fossil energy sources and the still reduced use of renewable energies, as claimed Ugursal (Ugursal, 2014), will likely have a profound impact on the socio-economic development of the world in the coming decades. Therefore, it is highly recommended that governments and, in particular, the energy sector establish policies aimed at changing consumption habits, in addition to promoting awareness campaigns addressed to involve everyone from producer to final consumer. This paper presents Bernard, a cheap, non-intrusive and easily scalable system addressed to the last element of the chain, the residential sector since this sector represented 25.4% of final energy consumption in EU in 2016 (Eurostat, 2018) and, to the best of our knowledge, there is a shortage of systems for this purpose.

In the last years, some investments focused partially on empowering residential consumers to manage their energy usage more actively and efficiently have been developed, but, in general, the home users are the ones who have to buy the devices and the expected savings are achieved in the very long term. Furthermore, as Bhati, Hansen, and Chan (2017) pointed out, the average consumer is not worried about his energy efficiency, he is not interested in improving it if that does not lead to get a reward (Bhati et al., 2017). Even more, though they are strongly motivated, they refuse to use energy monitors and smart plugs, due to the complexity of installing and understanding their output (Piccolo et al., 2016).

Therefore, our proposal is based on the premise that the final product must be inexpensive, non-intrusive, easy-to-install and understandable, pursuing in this way that users adopt the system, become aware and act accordingly.

Bernard, thus, was conceived as a smart system designed for utilities to develop and offer their residential customers a product that guides them to consume energy more efficiently. This system, from the electrical power readings supplied by a single sensor as suggested in 1992 by Hart (1992), provides the users with information about their current and historical consumption in real time, and offers recommendations about when it is most beneficial for them to consume, impacting as little as possible on their habits.

The huge volume of data that must be managed (reading of thousands of power curves from clients) and the rate of speed at which data should be processed so that the recommendations sent to users are effective (received at quasi-real time), led to adopt a big data solution. Furthermore, a big data architecture meets also the requirements of being easily deployed, upgraded and scalable (O'Donovan, Leahy,

E-mail addresses: Marta.zorrilla@unican.es (M. Zorrilla), alvaro.ibrain@alumnos.unican.es (Á. Ibrain).

Bruton, & O'Sullivan, 2015), necessary condition so that utilities can gradually provide this service. Moreover, these can take advantage of ingested data, which suitably processed and analysed, allow them to make better decisions.

In short, the goal of this paper is to describe Bernard, a smart system addressed to both, residential users to become aware of their consumption pattern and utilities to gain valuable information about their clients' consumption patterns and use them to decision making as well as to increase their corporate social responsibility. The design decisions adopted for the construction of each module as well as the cutting-edge technologies used for its implementation are also valuable contributions for its application in other industrial scenarios (towards Industry 4.0 (Scoop, 2016)).

This paper is organized as follows. Section 2 relates scientific works that address aims, technologies and approaches similar to our proposal. Section 3 describes the system requirements and architecture and details the main tasks entrusted to each module. Section 4 details the design and operating mode of each component of the system, evaluating its viability and performance. Section 5 discusses the results achieved and compares these with others found in the literature. Section 6 states the findings and managerial insights. Finally, Section 7 summarizes the main contributions of this work and comments the lines of future research.

2. Related work

This section is organised in two folds: a first subsection where some works in the field of home energy saving systems are referenced and a second subsection that relates big data technologies for the building of real time applications today.

2.1. Home energy management systems

Many home energy management systems (HEMSs) have been proposed and developed over time with the aim of improving the energy consumption efficiency of residential users. HEMSs address optimal consumption and production schedules by considering multiple objectives such as energy costs, environmental concerns, load profiles, and consumer comfort (Beaudin & Zareipour, 2015).

To name a few systems with a similar goal as Bernard, but with a different approach in relation to their development and deployment are listed in what follows: Ltd. (2018c) is an energy monitoring system which alerts the householders about exclusively solar generation and consumption; Ltd. (2018a) enables the user to monitor, program and control an ecosystem of devices and appliances in the home, but this requires the installation of sensors and their corresponding setting up; Plogg Network Controller (Ltd., 2018b) is a commercial smart meter and plug providing real-time and accurate energy reading but this cannot scale or be extended with another functionality since it only uses a sensor to measure power in the home power loop; the same happens with (Buono, 2015); finally, *Sense* (Sense, 2018) is a product that tells the user the appliances which are turned on but, as AlertMe, can be only utilised by domotic homes.

Other initiatives for raising people awareness in energy saving not related with smart meters have been proposed. This is the case of EnergyUse (Piccolo et al., 2016) an online social and collaborative platform to discover, share, and discuss tips for conserving energy, or the in-progress European project, eTEACHER (de Estudios de Materiales y Control de Obra et al., 2017), which aims at developing a set of tools for encouraging and enabling a behavior change of building users in order to save energy and optimise indoor environment quality.

To the best of our knowledge, a system as Bernard has never been proposed because, until relatively recently, there was neither technology nor infrastructure that could process, transform, analyse and store huge quantities of data coming from different sources, applying parallel processing and real-time analytics.

2.2. Big data technologies for building real time applications

Real time analytics is the discipline that analyse data as soon as it becomes available in the system. It aims at providing insights for making better decisions quickly. This is today an imperative activity for organizations and companies that want to advance towards digital transformation, as it is the case of the industrial sector.

Technologies of this arena are focused on high availability, performance and scalability, i.e., they are designed to cope with and perform well under an increased or expanding workload. In general, this kind of applications are built following a kappa architecture (Wingerath, Gessert, & Friedrich, 2016), which comprises three essential modules: a streaming data pipeline, a stream processor that reads data from the pipeline and performs a certain task and whose result is forwarded to a serving layer which might be an analytics web GUI or a database where a materialised view is maintained.

Tools such as *Cosmos, Kafka* or *EventHub* can be used as event logging system (pipeline), whereas *Spark, Storm, Flink* or *Azure Stream analytics* fall in the category of stream processors. As data-stores, any data management system which supports high query performance can be considered. Some examples are Cassandra, HBASE or MemSQL.

Generally, a streaming ecosystem also includes mining modules addressed to characterise big data sets, recommend or predict values or facts. For instance, Luo et al. (2017) recommend energy saving appliances to users based on the analysis of their energy consumption patterns and Li, Ding, Zhao, Yi, and Zhang (2017) performed a study with the aim of utilizing one popular deep learning approach, the SAE method, to improve the predicted results of building energy consumptions. Tools such as Spark MLlib, Google's Tensorflow, Theano, PyTorch or Keras are frequently used in this field.

3. System requirements and architecture

The main goal of this home smart system is to offer householders valuable information that allows them to know their energy consumption patterns as well as strategies and suggestions that help them to redirect these patterns towards a more efficient and sustainable consumption. In particular, the functional requirements to be satisfied were:

- Identify home consumption patterns and, based on these, predict the household energy consumption for a given day.
- Predict the KWh price with, at least, one hour in advance to inform householders about when it is cheaper to switch devices on.
- Detect and recognize the equipment that are on in the house in real-time.
- Offer an easy-to-use graphic interface that displays consumption data in real-time, the energy price forecast for the next hour, the electronic devices turned on, the outdoors temperature and tips to increase efficiency.

With the aim of facilitating the development, deployment and future upgrades, Bernard was conceived under a modular architecture. This follows the paradigm of the third generation platforms proposed by the Industrial Internet Consortium (IIRA, 2017) for the building of applications for the Industry 4.0 (Scoop, 2016), which rely on an data-oriented architecture (Data-as-a-Service) deployed in cloud environments (Kleppmann, 2017). This paradigm leads to select and utilize software that can be deployed in a distributed environment, scalable and suitable for real time data processing (see Section 2.2).

In particular, the technologies chosen for Bernard implementation were Apache Kafka, as logging system and Apache Spark, as stream processor mainly because they are open source projects, their performance is very high (Chintapalli, Sanket, & Derek Dagit, 2001) and have a large and active community of developers and supporters. Regarding data-stores, MemSQL, although it is not free, is a high-performance, in-

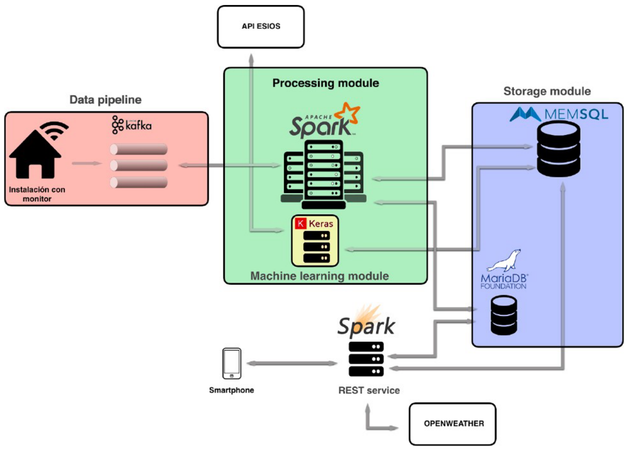

Fig. 1. Bernard architecture.

memory database that combines the horizontal scalability of distributed systems with the familiarity of SQL language. Regarding the machine learning components, Tensorflow (with Keras) was chosen due to its great user community and the support that it receives from Google, an enterprise which has been pushing forward the state of the art in this area in the last decade.

Fig. 1 depicts the six modules that comprise Bernand:

- **Data pipeline module**: This module is composed of a low energy computer (sensor) that monitors the house electric load and sends the readings to a low latency and high availability Kafka cluster to be processed by a Spark job. This is also responsible for sending messages that notify the switching on and off of appliances in the house and the request of identification of a new device plugged-in (see Section 4.4).
- **Processing module**: This module runs on top of a Spark cluster and is responsible for the execution of the following jobs:
  1. **Storage of readings**: a Spark job reads real-time data coming from Kafka cluster and saves it into a high performance database (MemSQL).
  2. **Data Aggregation**: Hourly, a Spark job cleans and aggregates the received data and, next stores processed data in MemSQL to be later used by the predictive module.
  3. **Data Backup**: A daily process stores the household consumption in a secondary database (MariaDB) with the aim of serving as backup and answering non-real-time user queries.
  4. **Prediction of home consumption**: a Spark job that estimates the home consumption for each hour of the current day and stores the forecast into MemSQL to be shown to user if requested.
  5. **Identification of appliance on**: this task is performed by a phyton programme that reads the power signal received and passes it through appliance disaggregation neural networks to extract and store the signature of each device turned-on in that home.
- **Machine learning module**: This includes two artificial neural networks built with *Keras* over *Tensorflow* for:
  1. **Electricity price prediction**: This component regularly gets data from the Spanish Electrical Grid, processes it, makes a prediction the KWh price and stores it in MemSQL, so that this can be later shown to the users.
  2. **Appliance disaggregation**: This artifact analyses the home power load readings, extracts the activations of the appliances and stores the isolated signal into MemSQL. This allows users to know the appliances that are switched on in real time.
- **Storage module**: This hosts two database management systems: MemSQL and MariaDB. The first one is a high performance in-memory based database focused on real-time analysis. It supports the queries which require very low latency. MariaDB is an open source relational engine that answers the least frequent and lowest cost queries requested by the user. It also works as a secondary data storage.
- **REST service**: This module is responsible for serving the information gathered and processed by the whole system to the user mobile application.
- **Smartphone**: Mobile application which offers the functionalities for the householder.

The details of the design and implementation of each module are described in the next section.

4. Modules's description

This section describes the criteria of design and schemes of implementation followed in the building of the main components of the system.

During the development phase, two publicly available datasets were used. The first one, *REFIT*, contains low frequency-sampled aggregated power readings of 21 UK houses along a year. The second one, *REDD*, contains one-second-sampled power readings of 6 houses during 6 months. *REDD* is though for appliance disaggregation tasks whereas *REFIT* is more general-purpose.

4.1. Home consumption forecast

This component aims at predicting the consumption curve of the current day. As it is well-known, the home consumption pattern is affected by several factors such as the number of occupants, ages, gender, home construction year, the day of the year, the weather, and so on. A house thus does not have an only consumption pattern during all the year (Jenkins, Patidar, & Simpson, 2014), but its consumption varies according to features like the occupants' routines or bank holidays, as can be observed in Fig. 2 which displays four consumption profiles for the same home on days selected randomly.

The predictive component was built from REFIT data set because this has data sampled each 8 s and a higher number of houses to be characterised. First, a Spark program was written to preprocess data, eliminate missing data and outliers and calculate the average consumption per hour. Next, a feature vector was defined. This included the following home metadata: the building year, the number of male and female occupants, the mean age of occupants and the average consumption per each hour of the day (see Eq. (1)). Then, a new dataset with this information was created in order to develop this component.

$$\overrightarrow{C_d} \quad (build\_year, mean\_age, num\_male, num\_female, p_0, \ldots, p_{23}) \qquad (1)$$

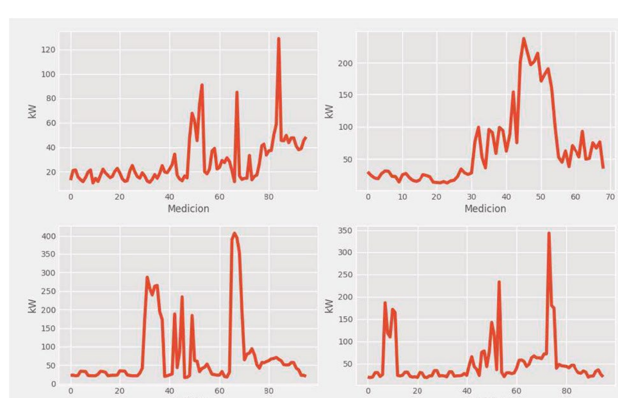

Fig. 2. Example of four consumption profiles for the same house on days selected randomly.

Two strategies were studied. The first one followed a clustering-based approach. In particular, a model was built by applying a *kmeans* algorithm with optimum k centroids $= 30$ on the previous dataset, where the value of k was established by applying the elbow method (Tibshirani, 2001). In this approach, the centroid of the nearest cluster is selected as the most probable consumption pattern.

In order to compute the distance between real vector and centroids, an estimation of the components of the real vector corresponding to the next hours had to be carried out. To solve this issue, the formula in Eq. (2) was utilised. This gives a probability to a vector $\overrightarrow{C_d}$ (partially complete) of belonging to a cluster (C_i) given the day of the week d.

$$P\left(\overrightarrow{C_d} \in C_i \middle| d, H\right) = \frac{1}{\alpha|\sigma| + \coth(\mu(d, H) + \tau)} \tag{2}$$

where:

- α: Calibrator. Must be $\ll$.
- σ: Euclidean distance between $\overrightarrow{C_d}$ and $\overrightarrow{C_i}$, but only with known components.
- H: Historical data with information about which cluster was assigned to what house each day.
- μ **(d, H)**: Function which returns the number of times that a house was assigned to a cluster when the weekday and its historic are provided.
- τ: Certainty degree of the calculation. It's equal to $\frac{|\overrightarrow{C_d}|}{|\overrightarrow{C_i}|}$. The more components are known, the higher probability is reached.

To validate the accuracy of the forecast, we calculated the difference in hours between the predicted maximum consumption and the real one. This approximation performed well in most of the cases. However, on days in which the house presented a radically different consumption pattern, the difference was not acceptable (about 7 h).

Therefore, another alternative was searched. This was based on discretizing the power vector values by assigning each one to a label of the following ones: *Low, Medium-Low, Medium, Medium-High* or *High* and performing the comparison of the partial known vector with the load patterns of the house corresponding to the two previous weeks instead of using the clusters. The evaluation method used in this case was changed. We only considered that the prediction was right if the predicted maximum consumption was the same as the real one for that day. This approximation had a very low error rate, only , and thus the method was considered satisfactory. Nevertheless, we do not discard that the first method outperforms in a real environment in which a higher number of consumption profiles and metadata will be available.

4.2. Price prediction forecast

This job is responsible for predicting the price of the kWh for the next hour in the power pool. To this end, the whole set of several electric market indicators from ESIOS (Spanish electrical grid API) was downloaded. After studying the indicators that this includes (using correlation techniques among others), the following ones were selected for training the neural network.

- **PVPC**: Price of the energy for users who have less than 10 kW contracted power.
- **SPOT market price**: Price of kWh in a market called SPOT. In this market, utilities trade with their surpluses of energy.
- **Total energy**: Total energy in the grid.
- **Actual demand, sum of generation:** Total energy destined to supply the demand of a certain moment.
- **Actual demand**: Energy demanded to the grid in a certain moment.

Several neural networks were designed and trained with Keras library over Tensorflow framework. The best results were achieved with a LSTM model (Long short-term memory) with 3 hidden layers of 128 neurons, using Adam algorithm as optimizer and Mean Average Error as cost function. We chose LSTM units due to their "longer-term memory" feature that is very suitable for dealing with time-series data.

Next, the network architecture is shown.

- Input layer (4 units)
- LSTM layer (128 units)
- LSTM layer (128 units, dropout=0.2)
- LSTM layer (128 units, dropout=0.2)
- Output layer (1 unit, activation=selu)

As can be observed, a dropout of **0.2** was set for the regularization across two layers. The training of the network was performed with the 80% of dataset, leaving a **20%** for validation. The training error was very low as is shown in Figs. 3 and 4.

An example of a prediction out of the training/validation set is shown in the Fig. 5 where prediction is depicted in orange colour and the current price is displayed in blue colour. As it can be observed, the mean average error is really low, **3** €/MWh (note that the consumption scale is kWh, ***10 3**).

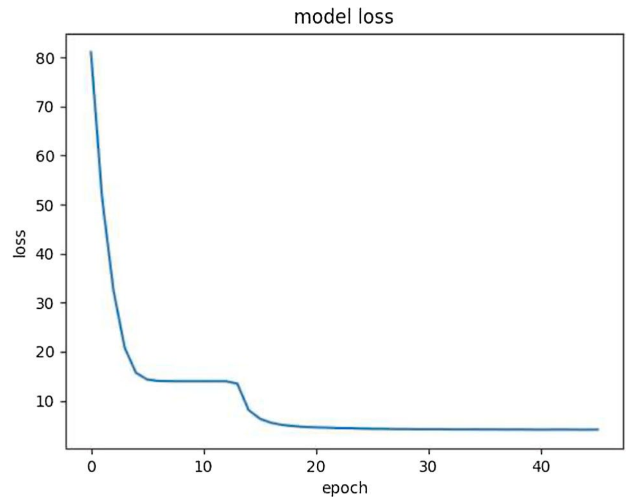

Fig. 3. Training set loss.

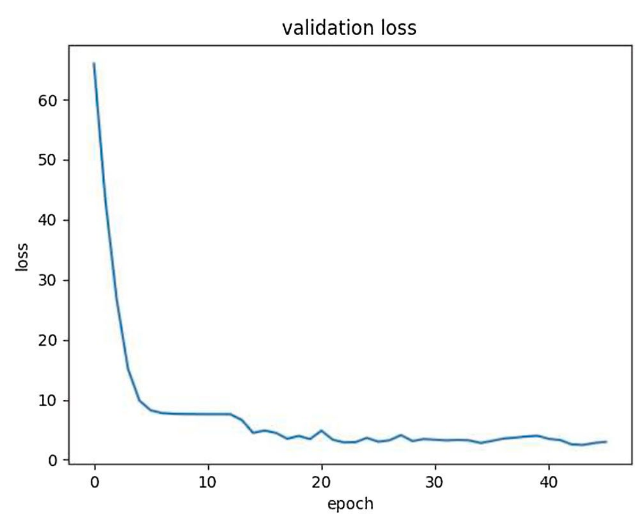

Fig. 4. Validation set loss.

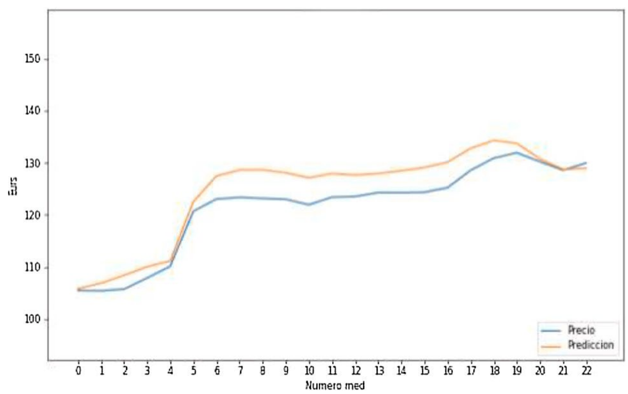

Fig. 5. Result of prediction with real data.

4.3. Appliance disaggregation

This task aims at the disaggregation of each home appliance as a consequence of the fact that Bernard receives only the global power consumed in the house (design requirement in order to be a cheap and non-intrusive system (see Section 4.4)). The resulting models are used to show homeowners the appliances switched on.

This deep learning module was built with REDD dataset which contains power measures of the appliances of 6 houses, along with the total power readings sampled each 3 s. The approximation followed to achieve a successful isolation of an appliance from the aggregated power signal was the use of Denoise Autoencoders neural networks since they act as a filter. These are updated in the backpropagation phase in such a way that, in each epoch, their ability to extract the features improves.

In this prototype, the goal is focused on extracting the signal of a fridge. Several configurations were tested based on (Kelly & Knottenbelt, 2015). Finally, the architecture of the neural network that achieved the best accuracy is the following:

- 1-Dim. Convolutional layer with 8 filters of 4x1 kernels, same padding and linear activation (0.2 dropout probability).
  Fully connected layer with 256 units and Linear activation.
- Fully connected layer with 128 units and Linear activation.
- Fully connected layer with 256 units and Linear activation.
- Fully connected layer with 256 units and ReLU activation.
- Fully connected layer with 256 units and Linear activation.
- Fully connected layer with 120 units and Linear activation (0.2 dropout probability).
- 1-Dim. Convolutional layer with 1 filter with 4x1 size, same padding and linear activation.
- Output layer (120 units).

The network was trained with the Adam optimizer and the *Mean Average Error* was selected as loss function. The aggregated power signal of the two first houses of the data set was used as training set and the third house was utilised as validation set. The signal was divided into vectors of 120 components (input layer). This number was chosen because of the sampling rate of the readings in the dataset is every 3 s and the need of establishing a period of time, in this case 6 min, that guarantees that the appliance was fully activated completely (e.g. the washing-machine, the vitroceramic cooker or the fridge).

The results of this training can be observed in Figs. 6 and 7.

After this training process the mean average error in the validation set was 35 W. The results achieved are depicted in Fig. 8 (Left). As can be observed, the extraction tends to add noise on the "active" part of the signal. This was fixed by programming a function to soften this effect. The final result can be seen in Fig. 8 (Right) where the orange signal is the one extracted and softened and the blue is the disaggregated one.

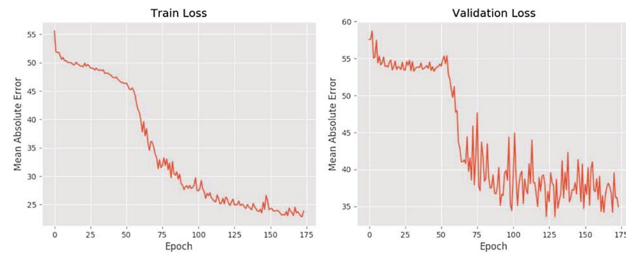

Fig. 6. (Left) Training set loss on house 1; (Right) Validation set loss on house 1.

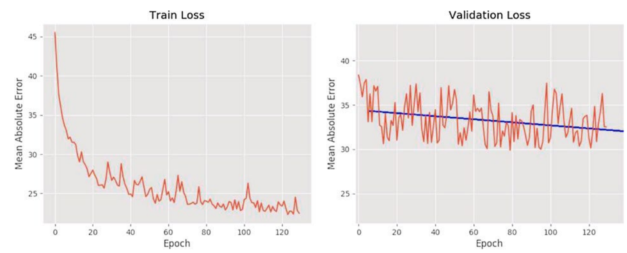

Fig. 7. (Left) Training set loss on house 2; (Right) Validation set loss on house 2.

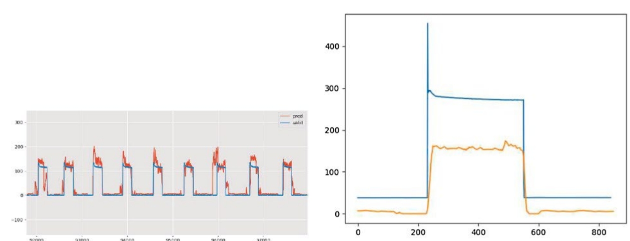

Fig. 8. (Left) Random interval of house 3 disaggregated; (Right) Softened disaggregated vector.

Following this approach, a neural network per appliance type (i.e. fridges, televisions, computers, microwaves, etc.) should be built. Although this seems naive, this provides a decoupling between the system and the evolution of electrical appliances over time and facilitates the upgrade of each appliance detector as more consumption curves become available.

4.4. Sensor logic

One of the goals of Bernard is that this was the least intrusive as possible, that means, avoiding the installation of any extra circuit in the home electric distribution system. That is the reason why we proposed the use of current clamps as shown in Fig. 9. These are connected to a low energy computer such as a RaspberryPi which is the responsible for

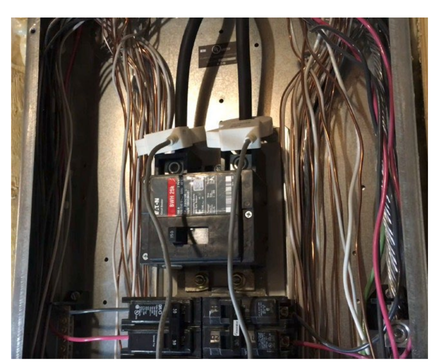

Fig. 9. Current clamps in the main distribution panel.

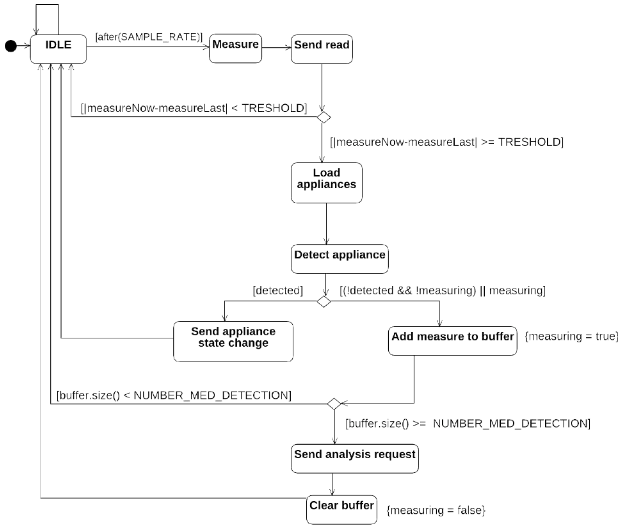

Fig. 10. Activity diagram of the sensor.

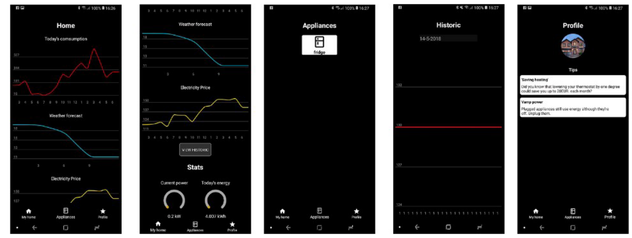

Fig. 11. APP screenshots.

reading the consumed power.

The sensor logic carries out three tasks:

1. **Sending power value**: Every 3 s, the sensor sends a message to a Kafka topic with the value of the power in the house.
2. **Sending the state changes of appliances**: Every change detected in an appliance is notified to Bernard through a Kafka topic.
3. **Sending disaggregation requests**: If a change occurs and it does not correspond to a known appliance, the sensor records the aggregated power and sends the signal to a Kafka topic.

It is important to highlight that this minicomputer stores the signature of each appliance which has been previously identified in the house by means of the already explained deep learning module and stored in the MemSQL database, in such way that every time a sudden shift in the aggregated power signal occurs, the program compares the jump in the power signal with all the stored signatures. If one matches, the appliance has been recognized, if not, a request of disaggregation is sent.

The activity diagram of the sensor logic is displayed in Fig. 10.

4.5. Android mobile application

With the aim of raising the users' awareness in the responsible and efficient use of energy, an Android app using our REST service was developed. This app shows the householders all the information available in the system in order that they can analyse it, reflect and change their consumption mode towards a more sustainable one. In short, they can observe data about their daily and historical energy consumption, the temperature forecast (read from OpenWeather API), their historical and forecasted electricity price for each day, their real-time power consumption, the active appliances and some tips for improving energy efficiency.

The app design is very simple, self-contained and easy-to-use for everyone. Fig. 11 shows some screenshots of the APP.

5. Discussion

Bernard is a proof of concept built from public datasets, therefore the results achieved by our predictive modules will be discussed in comparison with the ones published in other research works that used the same or other datasets with similar features and goals.

We have developed three smart components in this work: the first one aimed at predicting the home power consumption per hour of the current day; the second one addressed to forecast the price of kWh for the next hour in the power pool; and, finally, a third component for identifying each appliance switched-on from the analysis of the home consumed power signal.

Regarding the prediction of residential electricity consumption, we must first point out that this is a difficult problem to solve using modern deep learning techniques because these require a large amount of data to train the models to guarantee that their answer is general enough to new and unknown instances (He & Chai, 2016). Due to the absence of large datasets, techniques such as "Moving average" (Veit, Goebel, Tidke, Doblander, & Jacobsen, 2014) have been proposed to estimate the home power consumption with results of up to 146% deviation for the consumption prediction within a 24 h horizon. In order to reduce this error margin, we decided to simplify this task, predicting only the maximum daily consumption in a house one day in advance. Bernard is capable of predicting the electrical power demand of the user for each hour of the current day with an accuracy of **99.66%**. This provides a competitive advantage to the utilities, since they can determine the purchase of energy that it requires in quasi-real-time. Nevertheless, with larger data sets, it would be interesting to try approaches based on neural networks as proposed by Shabani and Zavalani (2017) in order to know, not only when a house will consume more, but also how much energy it will require in a whole day.

Concerning the electricity price forecasting module built in this work, we can say that this is highly competitive, since the validation error achieved is very low (**0.003** €/kWh), almost negligible in terms of household consumption. This result is noticeably better than the ones obtained with classic statistical methods (see Li, Lawarree, & Liu, 2010) whose RMSE are higher than **0.005**/kWh or those achieved with more traditional machine learning approaches (Gianfreda, Ravazzolo, & Rossini, 2018), such as RF or SVR, whose RMSE ranges between **0.006**/kWh and **0.008**/hWh. However, deep learning strategies such as the presented here or described in Zhu, Lu, Dai, Liu, and Wang (2018) clearly outperform the other approaches with RMSE near **0.004**/kWh or lower.

Finally, the module responsible for extracting the electrical signal of a fridge from the global consumption reading obtained a mean error of 35 W, which is relatively low taking into account that the consumption of this appliance is much greater (about 300 W). In literature, we found two papers (Kelly & Knottenbelt, 2015; He & Chai, 2016), that both use a LSTM Based Deep Learning Model (auto-encoders) as we do with results slightly better. In Kelly and Knottenbelt (2015), the network architecture is pretty similar to ours but it uses a fewer number of connected layers achieving a MAE of 26 W. On the other hand, the network proposed by He and Chai (2016) presents several more complex network configurations. These connect several branches of neurons with two dense layers obtaining a MAE of 3.46 W. Strictly speaking, our proposal is less competitive but it is simpler, which means less training time and lower computational cost.

6. Findings and managerial insights

Raising awareness among residential users about the responsible use of energy and motivating changes in their behaviour and habits concerning this regard is the main Bernard's goal. This is based on the visualization of useful information such as the power consumption data in real time, the forecasted cost of the KW/h in the next hour and the identification of appliances that are switched on. The developed solution, designed for being deployed by electric utilities, allows us to state that:

- Machine learning techniques offer a good performance for the resolution of complex prediction problems in the energy arena (Mosavi et al., April 2019).
- The kappa architecture provides the needed scalability to gradually deploy the system according to resources available and the existing demand (householders interested in this service); furthermore, this makes the upgrade and improvement of the deployed components as well as the adding of new services easier thanks to their uncoupled nature.
- It is feasible to deploy a low-cost and non-intrusive solution that allows the digitization of the value chain for electric companies.

On the other hand, the experience gained during this development allows us to state that big data technology is very recent and there is still a lack of methodologies, open data and standards that guide and help companies to develop these solutions and to build proofs of concept that demonstrate their viability. Furthermore, there is a large number of development tools available and these are in constant growth and change which makes the choice of the most appropriate tool for each purpose difficult. Another challenge comes from the management and processing of huge quantities of data that require the hiring of resources on the cloud, that, despite being inexpensive, companies cannot have the same computational power in the development environment as in the production one, which makes testing and deployment phases more complex. Finally, regarding the building of predictive modules, we find that there is a shortage of home consumption public datasets which limits the developing of solutions as well as the need of advance in the research of algorithms addressed to build models that learn and can adapt themselves while reading data streams (still under research Shabani & Zavalani, 2017).

Big data solutions are complex and there is no enough experience or workforce for their development and deployment (Jabbour, de Sousa Jabbour, Sarkis, & Filho, 2019). Bernard is a prototype that fills this gap and demonstrates that a smart and scalable system can be built with reasonable guarantees of success. The industry must move towards the digitalization of the entire value chain so that our contribution can help to orientate companies in this challenge.

7. Conclusions

To achieve the desired environmental sustainability, governments and industry must promote and develop training activities, resources and tools that boost environmental awareness. One of the areas where there is still work to be done is in the field of home energy efficiency. This paper describes Bernard, a proof of concept of a smart real-time system focused on the household energy efficiency improvement. Its main objective is to make residents aware of their consumption patterns by offering them information that allows them to make a more responsible use of energy. The system meets the requirements of being cheap and non-intrusive for residential users and presents advantages for both end users and distribution companies. The former can observe their consumption patterns in real time, know the hours in which the energy is cheaper or what appliances are switched on and so, try to change their routines in order to reduce their bill; the latter can gain valuable information about consumers' behaviour and use it to bid in the electric pool and adjust their purchase to demand, to offer their clients new personalized services or products (e.g. for customer fidelization) and to send pieces of advice to them, improving their brand image.

Bernard, in order to deal with the large volume of data while achieving a scalable solution, has been implemented using cutting-edge technologies under a kappa architecture. It must be kept in mind that the system presented here is still in a very early stage and must be deployed in houses and further developed using real world data.

The future steps will be directed to recognize more appliance types and build more accurate and varied consumption patterns as well as to train and test deep learning modules to improve the accuracy of predictive modules. It is also among our objectives to apply these technologies to the industrial sector where their potential benefits would be broader.

Acknowledgements

The research leading to these results has received partial funding from Spanish Government under grant TIN2017-86520-C3-3-R B.

References

Beaudin, M., & Zareipour, H. (2015). Home energy management systems: A review of modelling and complexity. *Renewable and Sustainable Energy Reviews, 45*, 318–335. http://www.sciencedirect.com/science/article/pii/S1364032115000568.

Bhati, A., Hansen, M., & Chan, C. M. (2017). Energy conservation through smart homes in a smart city: A lesson for Singapore households. *Energy Policy, 104*, 230–239. http://www.sciencedirect.com/science/article/pii/S0301421517300393.

Buono, P. (2015). *A low cost system for home energy consumption awareness*. <http://ceur-ws.org/Vol-1528/paper9.pdf>, [Online; Accessed on 04/22/2018].

de Estudios de Materiales y Control de Obra et al., C. (2017). eteacher: End-users tools to empower and raise awareness of behavioural change towards energy efficiency. <https://cordis.europa.eu/project/rcn/211927_en.html>, [H2020-EU.3.3.1. - Reducing energy consumption and carbon footpint by smart and sustainable use. Online; Last access 16-09-2018].

Eurostat (2018). *Energy consumption in households*. <http://ec.europa.eu/eurostat/statistics-explained/index.php/Energy_consumption_in_households>, [Online; Last access 22-04-2018].

Gianfreda, A., Ravazzolo, F., & Rossini, L. (2018). *Comparing the forecasting performances of linear models for electricity prices with high res penetration.*

Hart, G. W. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE, 80*(12), 1870–1891.

He, W., & Chai, Y. (2016). *An empirical study on energy disaggregation via deep learning*.

He, W., & Chai, Y. (2016). An empirical study on energy disaggregation via deep learning. In *2016 2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2016)*. Atlantis Press. <https://doi.org/aiie-16.2016.77>.

i Scoop (2016). *Industry 4.0: the fourth industrial revolution - guide to industrie 4.0.* <https://www.i-scoop.eu/industry-4-0/>, [Online; Last access 16-09-2018].

IIRA (2017). *Industrial internet reference architecture v1.8*. <http://www.iiconsortium.org/IIRA.htm>, [Online; Last access 16-09-2018].

Jabbour, C. J. C., de Sousa Jabbour, A. B. L., Sarkis, J., & Filho, M. G. (2019). Unlocking the circular economy through new business models based on large-scale data: An integrative framework and research agenda. *Technological Forecasting and Social Change, 144*, 546–552. https://app.dimensions.ai/details/publication/pub.1091893748.

Jenkins, D., Patidar, S., & Simpson, S. (2014). Synthesising electrical demand profiles for uk dwellings. *Energy and Buildings, 76*, 605–614. http://www.sciencedirect.com/science/article/pii/S0378778814002321.

Kelly, J., & Knottenbelt, W. (2015). Neural NILM. *Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments - BuildSys*ACM Presshttps://doi.org/10.1145/2821650.2821672.

Kleppmann, M. (2017). *Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems.* O'Reilly Media.

Li, C., Ding, Z., Zhao, D., Yi, J., & Zhang, G. (2017). Building energy consumption prediction: An extreme deep learning approach. *Energies, 10*, 1–20.

Li, G., Lawarree, J., & Liu, C.-C. (2010). *State-of-the-art of electricity price forecasting in a grid environment.*

Ltd., A. (2018a). *Smart meter*. < https://www.alertme.com/how-we-do-it/products-and-services/smart-data/ >, [Online; Last access 16-09-2018].

Ltd., B. (2018b). *The plogg smart energy meter*. < http://www.bytesnap.co.uk/about-us/experience/case-studies/the-plogg-smart-energy-meter/ > , [Online; Last access 16-09-2018].

Ltd., E. (2018c). *Wattson energy monitors from energeno-display solar pv energy generation and consumption*. < http://www.diykyoto.com/uk/wattson > , [Online; Last access 16-09-2018].

Luo, F., Ranzi, G., Kong, W., Dong, Z. Y., Wang, S., & Zhao, J. (2017). Non-intrusive energy saving appliance recommender system for smart grid residential users. *IET Generation, Transmission Distribution, 11*(7), 1786–1793.

Mosavi, A., Salimi, M., Ardabili, S. F., Rabczuk, T., Shamshirband, S., & Varkonyi-Koczy, A. (2019). State of the art of machine learning models in energy systems, a systematic review. *Energies, 12*(7), 1301–1345. https://eprints.qut.edu.au/128290/.

O'Donovan, P., Leahy, K., Bruton, K., & O'Sullivan, D. T. J. (Nov 2015). An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities. *Journal of Big Data, 2*(1), 25. https://doi.org/10.1186/s40537-015-0034-z.

Piccolo, L.S.G., & Alani, H. (2016). Strategies and tools to raise energy awareness collectively. In: *Behave 2016 - 4th European conference on behaviour and energy efficiency*. < http://oro.open.ac.uk/47088/ > .

Chintapalli, Sanket, & Derek Dagit, e. a. (2001). Benchmarking streaming computation engines: Storm, flink and spark streaming. In *Proceedings of the first annual workshop on emerging parallel and distributed runtime systems and Middleware. IEEE*.

Sense (2018). *Sense, the smart energy monitor*. <https://sense.com/>, [Online; Last access 16-09-2018].

Shabani, A., & Zavalani, O. (Jul. 2017). Hourly prediction of building energy consumption: An incremental ann approach. *European Journal of Engineering Research and Science, 2*(7), 27–32. https://ejers.org/index.php/ejers/article/view/397.

Tibshirani R., W.G.. H.T. (2001). Estimating the number of clusters in a dataset via the gap statistic. *Journal of Royal Statistical Society B*.

Ugursal, V. I. (2014). Energy consumption, associated questions and some answers. *Applied Energy, 130*, 783–792. http://www.sciencedirect.com/science/article/pii/S030626191300980X.

Veit, A., Goebel, C., Tidke, R., Doblander, C., & Jacobsen, H.-A. (2014). *Household electricity demand forecasting – Benchmarking state-of-the-art methods*.

Wingerath, W., Gessert, F., Friedrich, S., et al. (2016). Real-time stream processing for big data. *Information Technology, 58*(4), 186–194.

Zhu, Y., Lu, S., Dai, R., Liu, G., & Wang, Z. (2018). Power market price forecasting via deep learning. *CoRR* abs/1809.08092. http://arxiv.org/abs/1809.08092.