IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 20, 2018, accepted December 30, 2018, date of publication January 7, 2019, date of current version January 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2891315

Drift-Aware Methodology for Anomaly Detection in Smart Grid

GIUSEPPE FENZA, (Member, IEEE), MARIACRISTINA GALLO, AND VINCENZO LOIA, (Senior Member, IEEE)
Dipartimento di Scienze Aziendali-Management and Innovation Systems, University of Salerno, 84084 Fisciano, Italy
Corresponding author: Giuseppe Fenza (gfenza@unisa.it)

ABSTRACT Energy efficiency and sustainability are important factors to address in the context of smart cities. In this sense, smart metering and nonintrusive load monitoring play a crucial role in fighting energy thefts and for optimizing the energy consumption of the home, building, city, and so forth. The estimated number of smart meters will exceed 800 million by 2020. By providing near real-time data about power consumption, smart meters can be used to analyze electricity usage trends and to point out anomalies guaranteeing companies' safety and avoiding energy wastes. In literature, there are many proposals approaching the problem of anomaly detection. Most of them are limited because they lack context and time awareness and the false positive rate is affected by the change in consumer habits. This research work focuses on the need to define anomaly detection method capable of facing the concept drift, for instance, family structure changes; a house becomes a second residence, and so forth. The proposed methodology adopts long short term memory network in order to profile and forecast the consumers' behavior based on their recent past consumptions. The continuous monitoring of the consumption prediction errors allows us to distinguish between possible anomalies and changes (drifts) in normal behavior that correspond to different error motifs. The experimental results demonstrate the suitability of the proposed framework by pointing out an anomaly in a near real-time after a training period of one week.

INDEX TERMS Anomaly detection, concept drift, machine learning, smart grid, time series analysis.

I. INTRODUCTION

A. CONTEXT

Recent studies in the area of smart grids highlight important issues. The Northeast Group, LLC by studying 125 countries around the world, realizes that utility companies lose about \$96 billion every year due to *non-technical losses* also known as NTLs (i.e., electricity theft, fraud, billing errors, and other lost revenue) [1]. It is also estimated that the global energy consumption will increase really much for 2040 [1] impacting the economy and the planet.

Moving toward the smart cities and the smart energy [2], energy efficiency and sustainability are important factors to address. In this sense, smart metering and nonintrusive load monitoring play a crucial role for fighting energy thefts and for optimizing the energy consumption of home, building, city, and so forth. The estimated number of smart meters will exceed 800 million by 2020. By providing near real-time data about power consumption, smart meters can be used to analyze electricity usage trends and to point out anomalies guaranteeing companies' safety and avoiding energy wastes. In order to address these challenges, it is required to apply emerging results in the area of Artificial Intelligence for enabling advanced monitoring capabilities. Among others, anomaly detection [3] is an important issue to address companies' revenues and safety by implementing continuous monitoring of consumers' activities. On the other hand, by recognizing, for instance abnormal energy consumption growth for a specific consumer, it should be considered as a way for an early alerting that induces a reduction of energy waste in buildings. The outcome will be the smoothing of the impact of the energy consumption growth on the economy and the planet. This paper introduces a novel method for identifying anomalies by collecting and analyzing data coming from these smart meters. The research stresses the importance of profiling electricity consumptions in order to estimate nearly real-time eventually fraud, concept-drift or anomalies in energy consumption. The proposed approach uses machine learning in order to forecast changes in energy

[1] IEA. World energy outlook 2017 executive summary.

consumption and consequently alerting the expert about anomalies or changes of habits. In this sense, the research contribution can actually address the challenge of energy efficiency and sustainability that are important issues in the context of smart cities. Indeed, the proposed methodology relies on (i) a clustering analysis applied to historical time series for identifying energy consumption profiles, (ii) an LSTM model trained on the curves of the clustering centroids in order to forecast future individual consumptions with respect to the most common profiles, and (iii) the analysis of the forecasting error motif for identifying potential anomalies. The idea is applying Machine Learning techniques in order to model consumers' behavior based on their consumptions along the time, and identify possible anomalies in such behaviors. However, for real world time series what is defined as a "normal behavior" can change along the time making start to the *concept drift*. In fact, when we talk about stream processing, we must consider the evolution of the data along the time, and due to the non-stationary characteristics of streaming data, it is important to continuously adapt the decision model to concept drifts. In the energy environment, a concept drift could be represented by events such as the following: (i) a family decrease (e.g., it passes from 5 members to 3 members because one son gets married and another one goes to study outside); (ii) sometimes the family moves out to another house (e.g., a cottage or a beach house); (iii) one person retires; etc. Despite to the existing approaches, this framework also considers concept drift as the possibility to both avoiding false positives and recognizing new behaviors in a focused re-train of the neural network.

B. PROBLEM DEFINITION

The problem consists in identifying anomalies in consumers' behaviors through data stream analysis, considering circumstances in which hypothetical anomaly is due to a changing into costumer profile (i.e., concept drift) rather than fraud, theft, or losses. In Machine Learning, concept drift refers to an unexpected changing in the statistical properties of the predicted variable [4] that could cause a substantial decrease in the prediction performances.

C. PROPOSED SOLUTION

Since the literature reports very high performance by regression algorithms for prediction of energy consumption, and with the availability for the advanced meters of more online data, the objective of this paper is to define a noise tolerant concept drift and time-aware system able to distinguish an anomaly from a user profile's change. The idea consists of training a Long Short Term Memory (LSTM) network by means of some different consumers' profiles extracted by applying a clustering algorithm on their consuming (i.e., 4 clusters experimentally obtained). The obtained model is exploited for predicting next consumption at each instant. Successively, by comparing the prediction against the real consumption we calculate a prediction error which is compared with the errors did during the last week. More precisely, we consider the standard deviation (σ) of the prediction errors made in the last week and fix -2σ and 2σ as a lower and upper bound, respectively. So, if the standard deviation of the last 24 hours is not in this range, an anomaly is pointed out.

D. CONTRIBUTIONS

The main contributions of the proposed research are:

- A concept drift aware method able to distinguish between normal and anomalous consuming profiles that also discriminates between profile changes (i.e., concept drift) and real anomalies, with a *near real-time* anomaly detection.
- A context-aware system for the anomaly detection: the regression model makes prediction considering both the last 24 hours consumptions and context information such as the month of the year, the day of the week and so on.
- Pre-processing dataset by means of a clustering allows to extract the most representative profiles to use during the model training, and training a unique neural network makes the framework more efficient.
- Concept drift recognizing capacity avoids the need to re-train the model after behavior changes because the system automatically recognize them. As opposite, using of the system by a human expert helps to establish when a new model training is needed (e.g., it is emerging a new profile never seen before).
- Training the model by cluster centroids makes the anomaly detection heuristics independent from the model and its predictive performances.

E. EXPERIMENTAL RESULTS

The proposed framework has been tested on the ElectricityLoadDiagrams20112014 [2] dataset. Outcomes confirm the suitability of the combination between the LSTM neural network and the proposed anomaly detection algorithm with promising results evaluated in terms of Precision and Recall performed on the number of corrected detected anomalies.

F. OUTLINES

The paper is structured as follows: Section II analyzes the state of the art in terms of anomaly detection and concept drift aware systems. Section III and IV present, respectively, the proposed framework and its evaluation. Finally, Section V concludes the paper.

II. RELATED WORKS

In the last few years, the interest about anomaly detection in the area of energy consumption has attracted much interest, and the proposed solutions are numerous. In [5], it is presented a methodology for non-technical loss (NTL) detection that exploits smart meter data together with auxiliary databases with contextual information. Chen *et al.* [6]

[2]https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014

propose the use of fractional-order self-synchronization error-based Fuzzy Petri nets (FPNs) to detect nontechnical losses and outage events in micro-distribution systems. A rough set based approach is presented in [7]. Ford *et al.* [8] demonstrate the effectiveness of an artificial neural network as a technique for the modeling of consumers' energy utilization and the identification of anomalies. In [9], two linear regression-based algorithms aiming to (i) model consumers' energy consumption, and (ii) evaluate potential energy theft caused by meter tampering, are presented. The work in [10] presents a theft detector based on the combination between meter audit logs of physical and cyber events with consumption data. Another approach based on the combination of text mining, neural networks, and statistical techniques for the detection of NTLs is presented in [11]. Finally, some other approaches aiming to reduce NTLs, propose lower level solutions such as a GSM-based Energy Recharge system [12], or a state estimation based approach for distribution transformer load estimation [13].

Concerning concept drift, in the literature, its applications regard different types of task: monitoring and control, information management, and diagnostics. Together with different target applications, concept drift can be applied to different data types: sensor streams, time-stamped documents, relational data tables and so on [14]. Ogundele *et al.* [15] described the importance of capturing time drifting patterns in user preferences. They presented two leading recommender techniques: factor modeling and item-item neighborhood modeling. The time feature is considered crucial also in POI [16] and service [17] recommendation through the evaluation of a time-aware user similarity; and in mining customer preference in physical stores [18]. Sun and Dong [19] adopt clustering and time impact factor matrix to predict user interest drifts through a linear regression.

In terms of combination between two presented issues (i.e., anomaly detection and concept drift), a very recent work [20] presents a time series anomaly detection system based on a Recurrent Neural Networks (RNNs) which is updated from time to time after an anomaly detection. Since the approach is not thought for the energy domain, it does not fit our goal because of the risk that a fraud seen as an anomaly becomes a network parameter and influences negatively subsequent predictions. So, what emerges is the lack of systems able to recognize anomalies without the influencing of concept drift events.

III. CONCEPT DRIFT AWARE ANOMALY DETECTION FRAMEWORK

The proposed solution consists of a classification technique based on a regression model exploiting sensor data coming from the smart grid. This model aims to help an expert in the identification of anomalies in energy consumptions through the monitoring of the differences between the predicted and real consumptions. In particular, the framework consists of two main stages: (i) a regression model preparation phase, and (ii) an anomaly detection phase. Following, we will define theoretical background about the proposed methodology and, then, detail it.

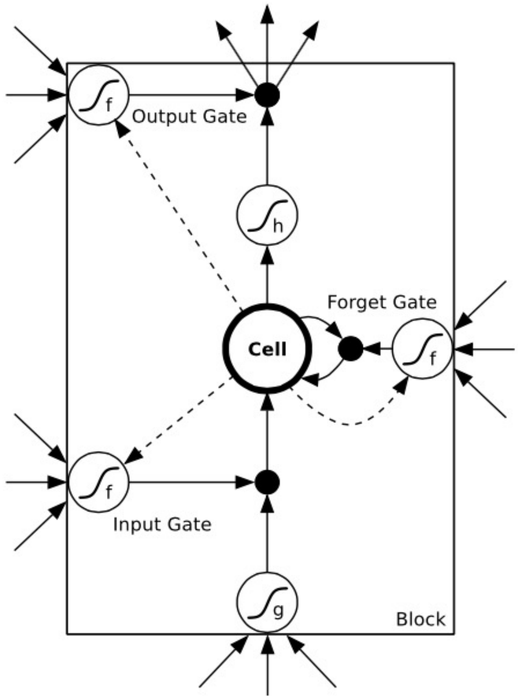

FIGURE 1. LSTM memory block [23].

A. K-MEANS CLUSTERING

The K-means algorithm [21] is considered the most popular and simplest clustering algorithm. Starting from a given number of clusters K, it arranges the input objects in K clusters based on their attributes through a set of iterations. At each iteration, the algorithm chooses a centroid for each cluster and links each object to the nearest centroid after the evaluation of a distance between them. The algorithm ends when the last iteration does not make any change into the location of the centroids.

More formally, let $X = \{x_1, \ldots, x_n\}$ be the set of observations to be clustered into a set of K clusters, $C = \{C_1, \ldots, C_K\}$ whose centroids are $\{c_1, \ldots, c_k\}$. The algorithm aims to minimize the sum of the squared distances of each object (x_i) with respect to its centroid (c_k):

$$\min \sum_{k=1}^{K} \sum_{x_i \in C_k} || \, x_i - c_k \, ||^2 \tag{1}$$

where $\sum_{x_i \in C_k} || \, x_i - c_k \, ||^2$ is the Euclidean distance between the centroid and all observations in C_k. Each centroid is evaluated as:

$$c_k = \frac{\sum_{x_i \in C_k} x_i}{n_k} \tag{2}$$

where n_k is the number of observations belonging to the cluster C_k.

The algorithm works as follows:

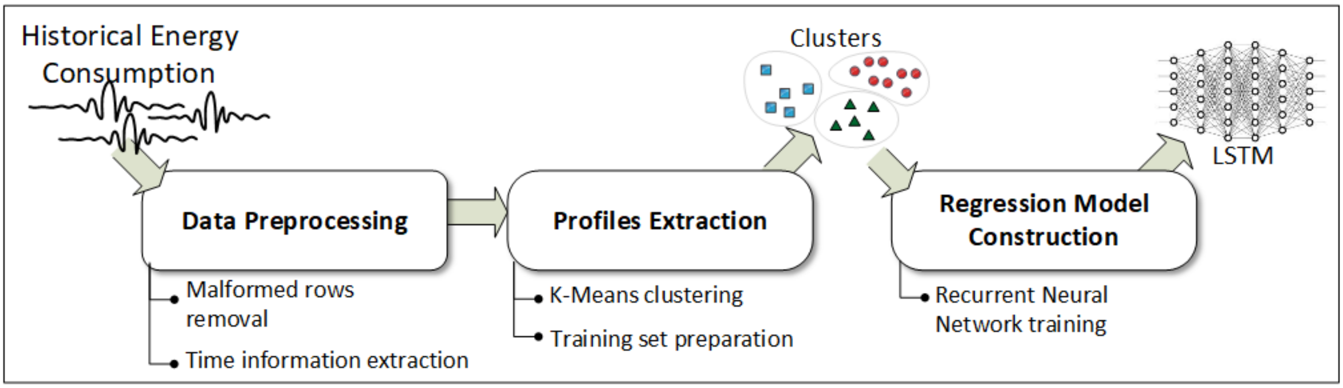

FIGURE 2. Process of LSTM model training for energy consumption forecasting.

1) Select K observations as the initial set of centroids (generally in a random fashion or by heuristics).
2) Assign each observation to the cluster having the closest centroid.
3) Recalculate the centroids when all observations are arranged.
4) Repeat Step 2 and 3 until centroids do not change.

B. LONG SHORT TERM MEMORY NETWORK

Long Short Term Memory [22] is a particular type of Recurrent Neural Networks (RNN). These last have been designed in order to understand the sequential nature of the data by using past history. Technically, this is achieved The architecture of LSTM consists of one input layer, one hidden (i.e. LSTM) layer, one output layer, and some memory blocks. Each memory block has multiple cells which have recurrent connections among them, an input gate (i), a forget gate (f) and an output gate (o). The input gate learns information to store in the memory. The forget gate learns the length of stored information, and the output gate learns when the stored information should be used. An example of single memory block is shown in Fig. 1.

The purpose of a LSTM network consists of identifying a correlation between an input sequence $x = (x_1, \ldots, x_T)$ and an output sequence $y = (y_1, \ldots, y_T)$. It is done by calculating the network unit activations by processing following equations iteratively [24]:

$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_i) \tag{3}$$
$$f_t = \sigma(W_{fx}x_t + W_{mf}m_{t-1} + W_{cf}c_{t-1} + b_f) \tag{4}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{cm}m_{t-1} + b_c) \tag{5}$$
$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}c_t + b_o) \tag{6}$$
$$m_t = o_t \odot h(c_t) \tag{7}$$
$$y_t = \phi(W_{ym}m_t + b_y) \tag{8}$$

where the W terms denote weight matrices (e.g. W_{ix} is the matrix of weights from the input gate to the input), the b terms denote bias vectors, σ is the logistic Sigmoid function, m is the cell output activation vector which have the same size of i, f, o, and c (the cell activation vectors). $\odot$ is the element-wise product of the vectors, g and h are the cell output and input activations that use hyperbolic tangent activation functions, ϕ is the softmax output activation function.

C. REGRESSION MODEL PREPARATION

The regression model has been implemented by means of a deep neural network. In particular, we have chosen a Long Short Term Memory (LSTM) network due to its capacity to learn long-term dependencies. Its preparation mostly consists of the design and training of the neural network whose process is depicted in Fig. 2. Starting from a dataset of historical time series about energy consumptions, we make a data pre-processing in order to remove malformed rows, and extract needed additional information (e.g., timeslice, previous average consumption, etc.). Cleaned data becomes input for a clustering operation which aims to aggregate similar consumption profiles and to guide the creation of a training set. Finally, the network is trained by means of cluster centroids.

The structure of the network consists in a first visible layer with 1 input, a hidden layer with 80 LSTM neurons, and an output layer that makes a single value prediction. The number of lag observations for each input is relative to the last 24 hours (i.e., 96, considering an observation every 15 minutes). The default sigmoid activation function is used for the LSTM blocks, and the training of the network is done for 90 epochs.

The deep neural network gives as input features information as the energy consumption, and the time aspects. In particular, the input vector for training is composed by:

- the day of the week;
- the day of the month;
- the month of the year;
- the timeslice (extracted through a partition of the day in 6 slices);
- an average consumption about the previous hour;
- an average consumption about the last but one hour;

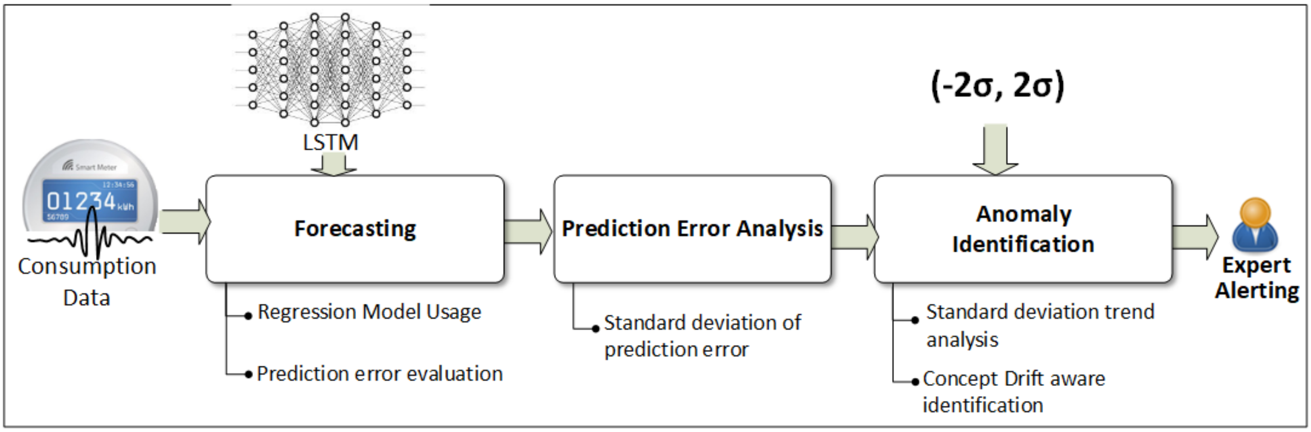

FIGURE 3. Process of anomaly detection through LSTM model usage.

```
1. Training the network until instant t;
2. Take the real consumption at t+1: CR_{t+1};
3. Forecast the next value: CP_{t+1};
4. Evaluate the prediction error: δ_{t+1} = (CR_{t+1} − CP_{t+1})^2;
5. Evaluate σ(Δ_k) Δ_k = {δ_{t+1}, δ_t, δ_{t−1}, ..., δ_{t−94}}^3;
6. Check if: σ(Δ_k) ∈ (−2σ, 2σ).
```

Listing 1. Anomaly Detection Algorithm.

and the *real consumption* value as a label, which becomes the prediction value in the test phase.

The extracted LSTM model is unique for all consumption behaviors.

D. ANOMALY DETECTION

The core of the framework is the anomaly detection process depicted in Fig. 3. It analyzes the prediction errors achieved by the neural network trained in the previous phase in order to distinguish between normal or anomalous behaviors. The general idea is evaluating the error trend along the time by means of its standard deviation. Unlike the most of works at the state of the art, we do not limit the analysis to the estimation of the absolute error between predicted and real consumption at a specific time instant, but analyzing its trend. We define a range between -2σ and 2σ, where σ is the standard deviation of the prediction errors made in the week precedes the actual instant. A value of standard deviation of the error prediction at the actual instant that does not lay in such range, is an indication of possible anomaly. Considering values about the last week allow us to distinguish anomaly from customers profile changing. The idea is that the assumption of anomaly must be verified by a continuous abnormal activity that does not reflect any of the profiles identified during the learning phase. Summing up, given the range $(-2\sigma, 2\sigma)$, at each timestamp t, the proposed anomaly detection algorithm makes the process exposed in the Listing 1 and in the flowchart in Fig. 4.

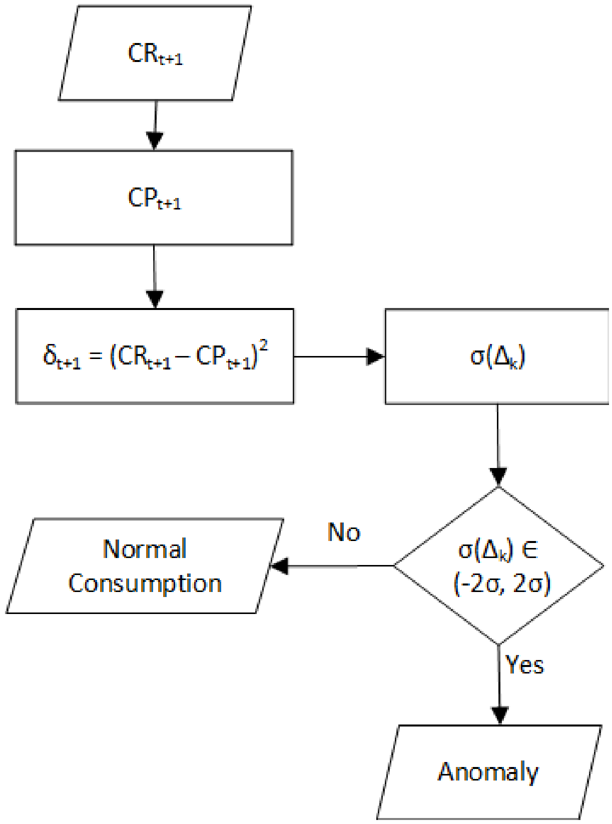

FIGURE 4. Anomaly Detection Algorithm.

More formally, at each instant t, let be $S = \{s_1, s_2, \ldots, s_7\}$ the set of standard deviations evaluated on 24 hours for each day of the week that precedes t, we define the anomaly as follows:

$$\forall s \in S\ s_t > 2 \times s \bigcup \forall s \in S\ s_t < -2 \times s \tag{9}$$

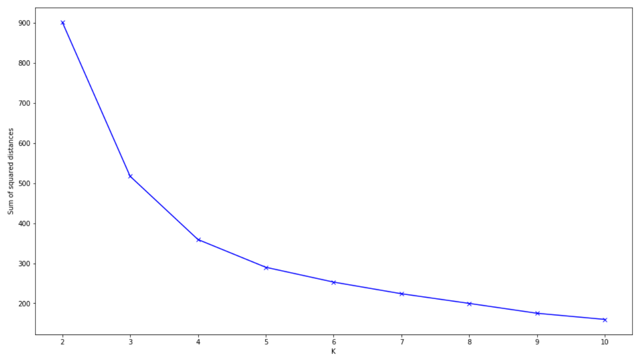

FIGURE 5. Elbow Method for optimal *k*.

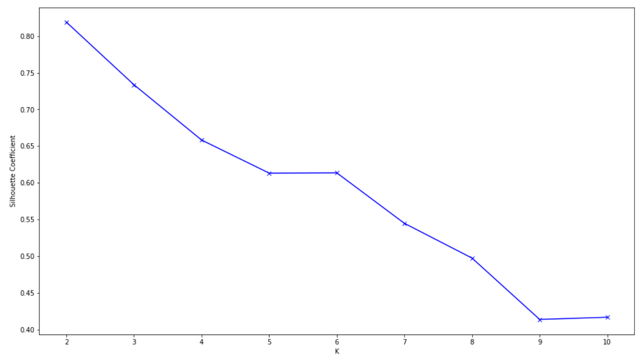

FIGURE 6. Silhouette Coefficient for optimal *k*.

where s_t is the standard deviation evaluated in the 24 hours that immediately precede t.

The choose to set the range basing on the standard deviation was inspired by the empirical rule of normal distribution from which the 95% of values lie within roughly in range $(-2\sigma, 2\sigma)$, and most influenced from [25] and [26].

IV. EXPERIMENTATION

The proposed framework consists of Python scripts that:

1) Generating clusters of data in order to determine the most representative profiles to use during the subsequent training. The adopted clustering algorithm was the K-means, it was applied by means of Scikit-learn library,[3] after an outliers removal.
2) Constructing and training a LSTM neural network by means of Keras library.[4] The regression model creation was made after a clustering of values of consumption of all the users in one year, in order to extract a significant sample of consuming.
3) Identifying anomalies in the test series.

For identifying the number of the clusters in order to execute K-means algorithm, we adopted two different heuristics: the Elbow Method and the Silhouette Coefficient [27]. The first one, starting with $K = 2$, and keep increasing it in each step by 1, evaluates the clustering and its goodness by means of the sum of squares distance inside each cluster. More precisely, if we plot the average sum of squares distance inside of each cluster (i.e., W_k), and the number of the clusters, we can see that the first clusters will add much information (explain a lot of variance), and for certain K the graph begins to flatten significantly. This point is named "elbow"and is the value we are looking for (see Fig.5). W_k is calculated as follows:

$$W_k = \sum_{r=1}^{k} \frac{1}{n_r} D_r \tag{10}$$

where k is the number of the clusters, n_r is the number of points in cluster r and D_r is the sum of distances between all points in a cluster, evaluated as follows:

$$D_r = \sum_{i=1}^{n_r-1} \sum_{j=i}^{n_r} \|d_i - d_j\| \tag{11}$$

The second method, the Silhouette Coefficient, measures the difference between the within-cluster tightness and separation from the rest. In particular, the silhouette width $s(o)$ for object $o \in O$ is defined as:

$$s(o) = \frac{b(o) - a(o)}{\max(a(o), b(o))} \tag{12}$$

where $a(o)$ is the average distance between o and all other objects of the cluster to which o belongs, and $b(o)$ is the minimum of the average distances between o and all the objects in each other cluster. Values of silhouette width range between -1 and 1. If all the silhouette width values are close to 1, it means that the set O is well clustered. A clustering can be characterized by the average silhouette width S of individual objects. The largest average silhouette width, over different K, indicates the best number of clusters. In our case, obtained results for $k = 2 \ldots 10$ are in Fig. 6.

Since the returned values of Elbow method and Silhouette Coefficient seem not to correspond, we also applied a hierarchical agglomerative clustering (with Euclidean distance) that does not require number of clusters as input. Hierarchical clustering extracts 3 clusters for representing the overall set of consumers. So, we decided to execute the k-means clustering by varying K in the range $[2, 5]$, and trained our LSTM model with resulting centroids in four different experimentations. Thus, we evaluated the performance as described in Section IV-B for each resulting model. As shown in Table 1, the best performance are obtained with $K = 4$

TABLE 1. Performance about LSTM model trained with different number of centroids.

K	Precision	Recall
2	71%	82.5%
3	75%	86%
4	**78%**	**88%**
5	77%	86%

[3]http://scikit-learn.org/stable/

[4]https://keras.io/

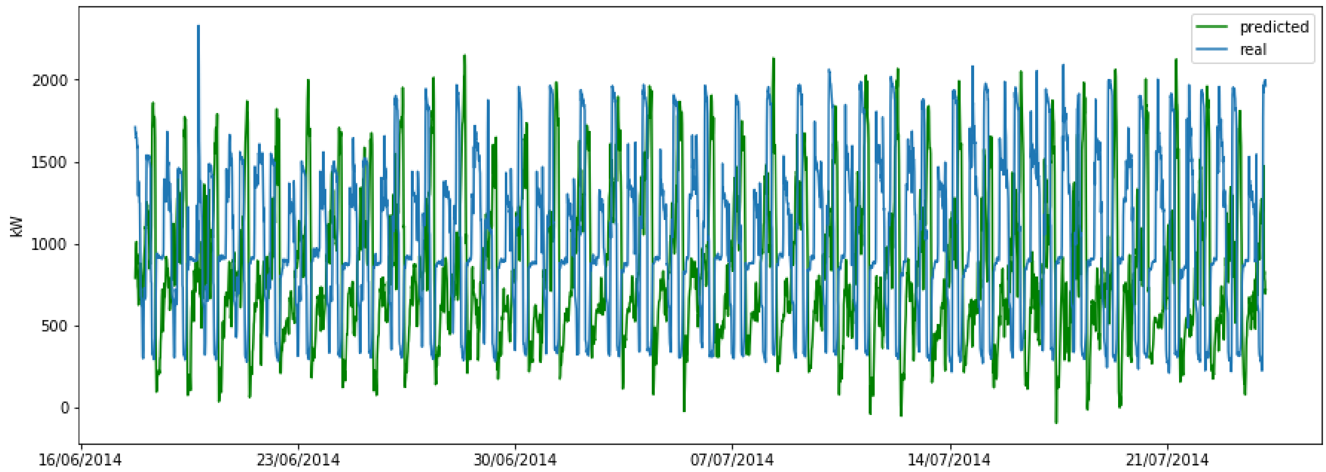

FIGURE 7. Prediction example of a normal consumption.

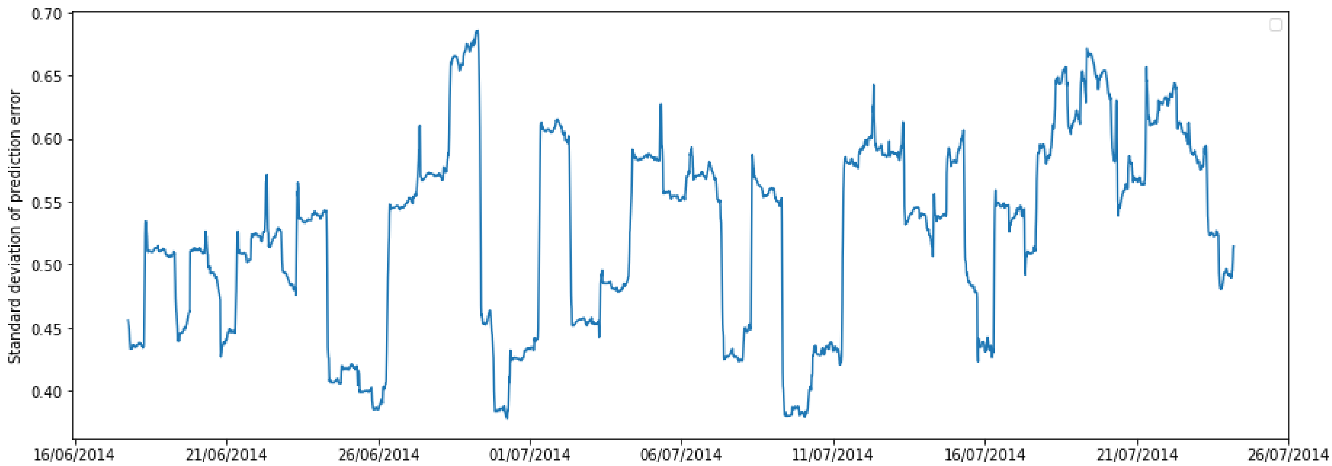

FIGURE 8. Example of standard deviation trend of prediction errors for a normal behavior.

A. DATASET

The dataset used for the validation of the proposed framework is *ElectricityLoadDiagrams20112014*[5] publically available on UCI repository. It is composed by the consumptions about 370 users from 2011 to 2014 with an observation every 15 minutes. Involved users are clients with different economic activities such as offices, factories, hotels, restaurants, and so on.[6] A further dataset description is available in [28].

A pre-processing of data removed empty rows, and extracted needed additional information (e.g., timeslice, previous average consumption, etc.).

After an extraction of 4 clusters of users' consumptions during the year 2014, the training of the model was conducted on the extracted centroids in the period between January and November. The evaluation of the model was conducted on the consumptions of the same consumers in December of the same year. The test set, instead, consisted of a set of 300 users. Concerning the anomaly detection, we selecting, inside the dataset some anomalous profiles (i.e., time series which presents some irregularities) for a total of 14% of the overall test set, and create some profile changes (i.e. concept drifts) by combining consumptions belonging to multiple clusters for a total of 13% of the overall test set. In this case, we are assuming the hypothesis in which a customer change his tendencies passing from a cluster to another (e.g., a hotel launches services of an health center).

[5]http://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014

[6]This justifies very high consumption values shown into the following diagrams.

B. EVALUATION

The evaluation of the system consists of two main phases: (i) performances analysis about predictions made by the LSTM model; (ii) algorithm's capability to anomaly and concept drift detection. For the first stage, we adopt the

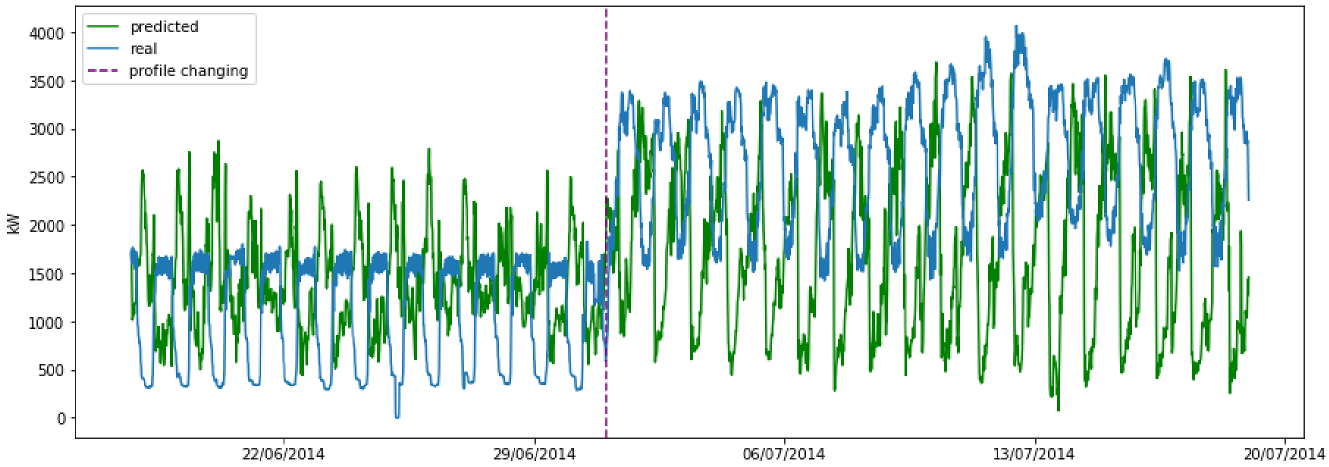

FIGURE 9. Prediction example during a concept drift.

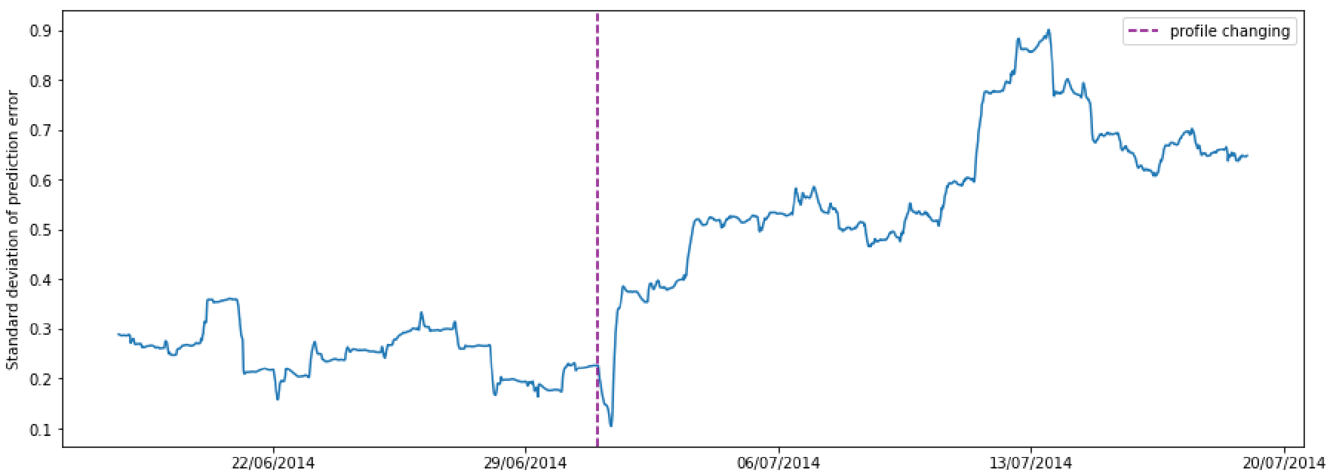

FIGURE 10. Standard deviation trend of the prediction error during a concept drift.

Root Mean Squared Error (RMSE) metrics, a widely used statistical method to calculate the difference between real and forecasted values, and defined as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y'_t - y_t)^2}{n}} \tag{13}$$

where y'_t is the prediction value, y_t is the real consumption value in the test set, and n is the number of instances.

The validation phase carried out a max RMSE of 0.15 kW and a minimum RMSE of 0.001 kW. However, objectives of present work are beyond the validity proof of an LSTM model (and, more generally, recurrent neural networks) for time series prediction. It is already considered a fact as shown in a variety of works also in the smart grid domain [29] and [22]. The main goal of this paper is demonstrating the feasibility of an LSTM-based system in recognizing anomalies in the time series and distinguish them from concept drift events. Furthermore, we are demonstrating that the proposed heuristics is independent from the model and its precision. So, we adopt the classical pattern recognition metrics *Precision* and *Recall*. They are defined as following:

$$Precision = \frac{|A_{gold} \bigcap A_{ret}|}{|A_{ret}|} \tag{14}$$

$$Recall = \frac{|A_{gold} \bigcap A_{ret}|}{|A_{gold}|} \tag{15}$$

where $A_{gold} = \{a_{gold_1}, a_{gold_2}, \ldots, a_{gold_m}\}$ and $A_{ret} = \{a_1, a_2, \ldots, a_n\}$ are, respectively, the gold set of anomalies existing in the test set, and the result set of anomalies retrieved by the system.

The test set execution carried out values of Precision and Recall equal to 78% and 88%, respectively.

C. RESULTS

In this section, we will present some example of execution about the anomaly detection algorithm. Fig. 7 shows an

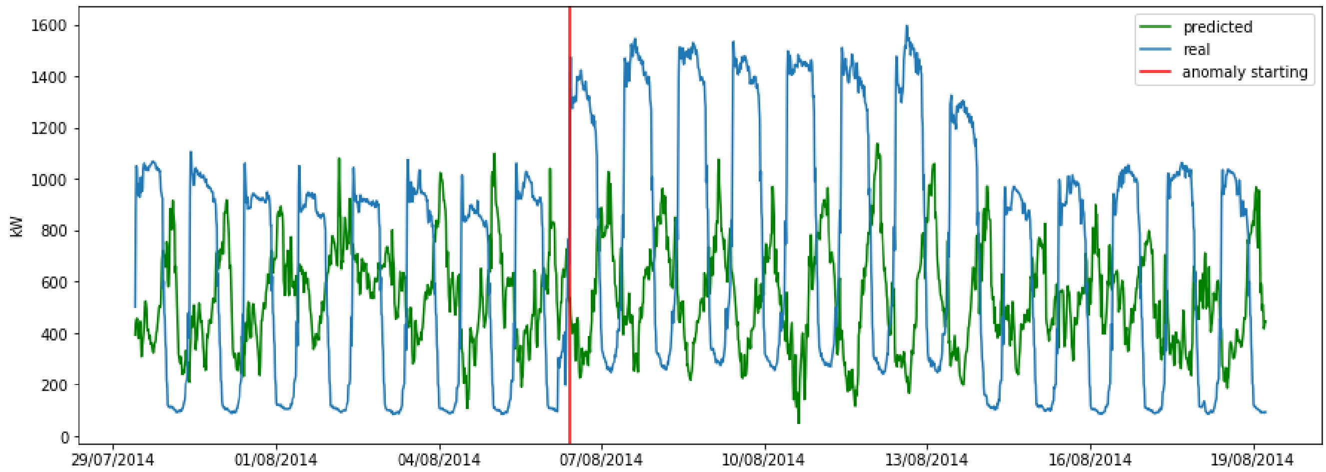

FIGURE 11. Prediction example of an anomalous event.

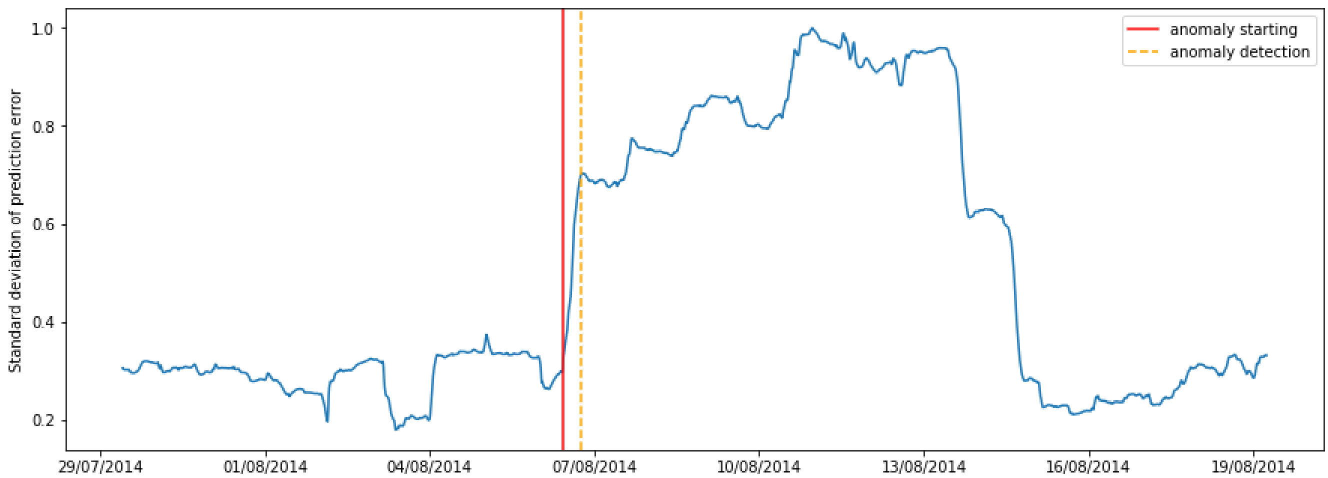

FIGURE 12. Standard deviation trend of anomalous consumption recognized by the system.

example of prediction about a normal consumption. It is evident that, beyond instant error values, the predictions follow the real trend of the series. This is also evident in the graph about the standard deviation of the prediction error in Fig. 8.

Fig. 9 shows a profile that changes along the time by illustrating real and predicted curves, that is an example of concept drift. In addition, in Fig. 9 the dashed vertical line points out when the change starts. Fig. 10 shows the standard deviation at intervals of 24 hours of the difference between real and predicted time series shown in Fig. 9. In particular, Fig. 10 points out that the error trend evaluated when the profile change happens is almost normal, and after a first period also the predictions will be re-adapted to the novel consumer's behavior. This allows us to not raise any anomaly and happens because we trained the LSTM model for recognizing the most representative behaviors altogether.

Fig. 11 and Fig. 13 show the predicted time series against the real consumption series about two anomalous curves, in Fig. 12 and Fig. 14, respective standard deviation of prediction errors are depicted. It is evident that the system recognizes anomalies some hours later the start of them. It happens that the model tries to predict the consumptions as in the past predictions, since it is guided from acquired knowledge and actual context. So, while the real consumptions grow, the predicted ones remain the same and consequently, the prediction error increases. In this case, the standard deviation of the error in the last 24 hours exceeds the 2σ threshold and the anomaly is pointed out.

In order to verify the real efficacy of the proposed algorithm, we made additional tests on other anomaly types. In particular, while the anomalies shown previously consist of unexpected increases of the consumptions, the anomalies shown in Fig. 15 and Fig. 17 consist of new different trends

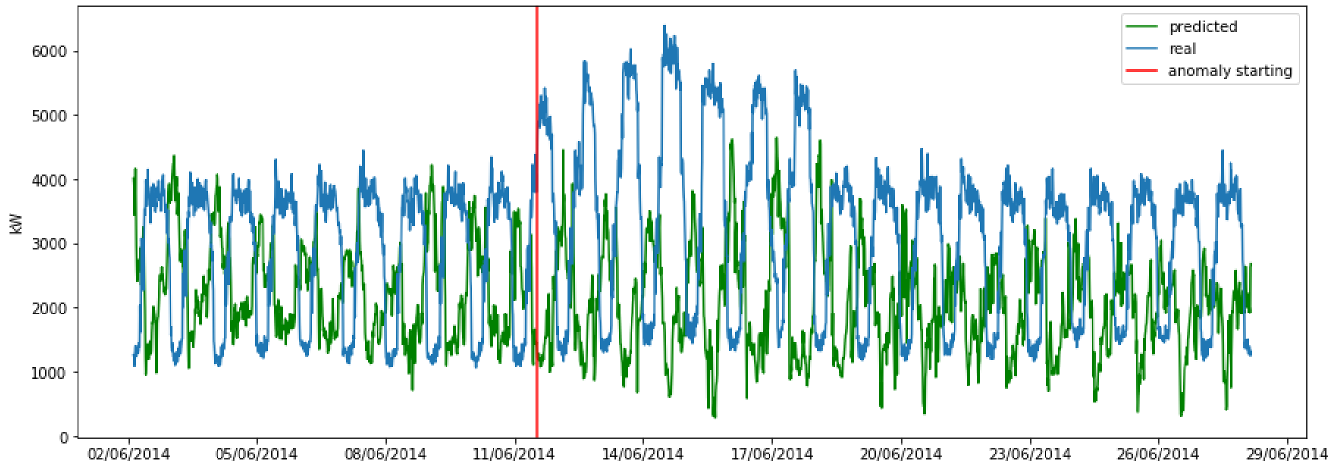

FIGURE 13. Prediction example of an anomalous event.

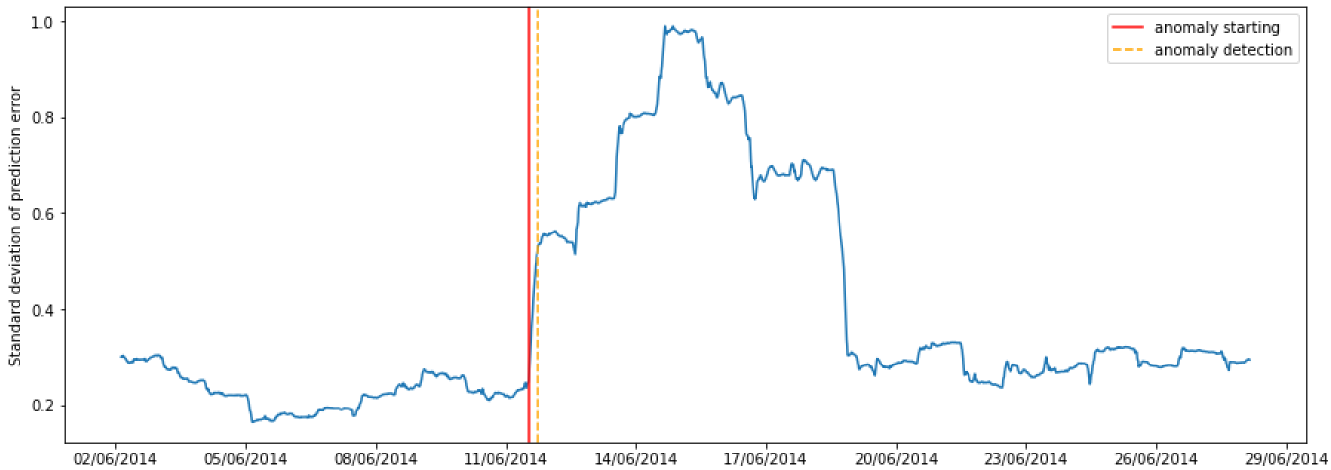

FIGURE 14. Standard deviation trend of anomalous consumption recognized by the system.

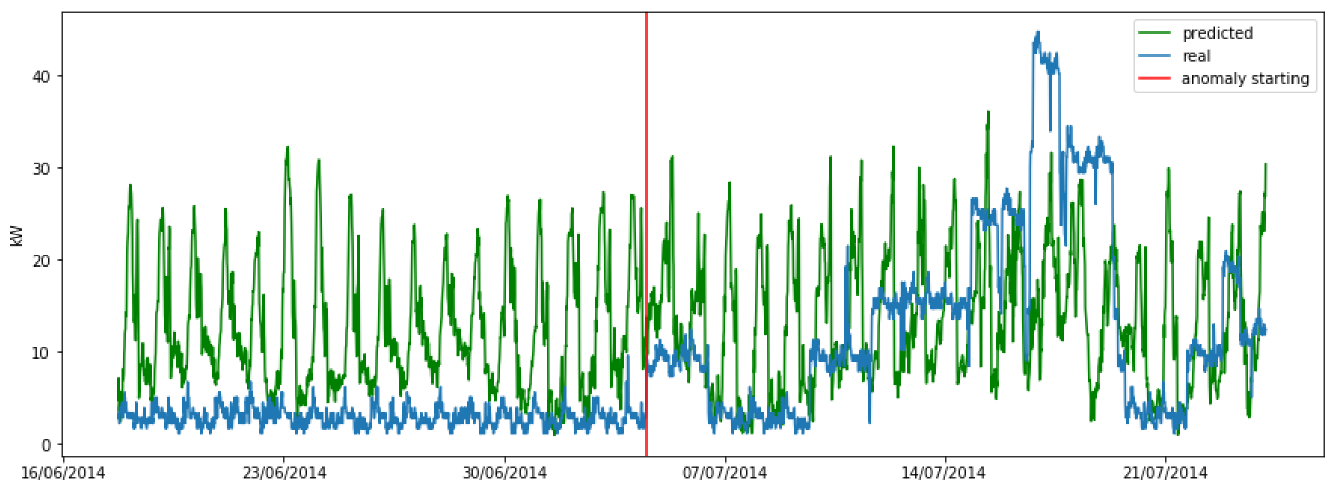

FIGURE 15. Prediction example of an anomalous event.

in the consumption time series. In all the cases, as also visible in Fig. 16, and Fig. 18 the delay of the anomaly detection is of few tens of observations corresponding to few hours. This delay may be reduced by using more curves in the training.

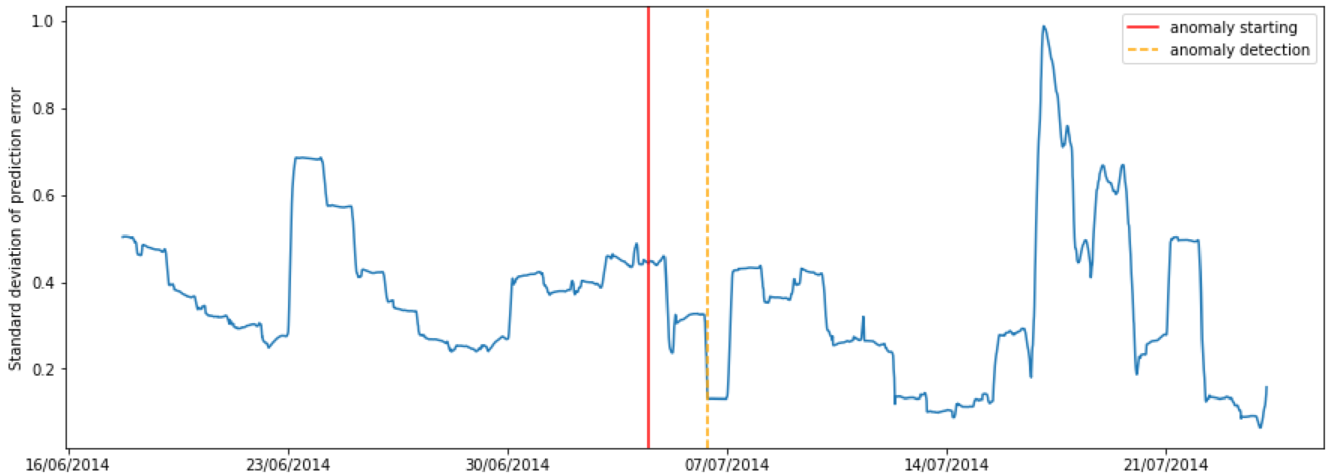

FIGURE 16. Standard deviation trend of anomalous consumption recognized by the system.

FIGURE 17. Prediction example of an anomalous event.

As shown in all the previous examples, the choice of training a unique model by means of centroids resulting from the clustering application, make the system more robust in terms of pattern recognition when such patterns have been known during the training itself. In this sense, with the purpose of demonstrating the ability of the model to distinguish known patterns, concept drifts, and anomalies, we have collected some consumptions in different periods of time for a specific consumer. In particular, in Fig. 19 we show the consumptions about the first user labeled as anomaly (see Fig. 11) in the six month that precede the anomalous period while, in the Fig. 20, we show consumption about the same period but in the previous year (i.e., 2013).[7] This suggests that the model is able to keep trace about the consumption motifs for all the users. However, what could happen is that after an unknown period of time, the LSTM model becomes obsolete and needs to be retrained with updated information.

[7]Let us note that consumptions about the year 2013 are not included in our test-set.

D. DISCUSSION

An emerged limitation of the proposed framework regards the inability of recognizing anomalies in the first week of the system execution. In fact, as expressed in Section III-D, the anomaly detection algorithm compares the actual prediction error trend with previous one needing observations about a whole week to understand the consumer profile. In fact, the anomaly is highlighted when the actual standard deviation (i.e., of the last 24 hours) about the prediction error double exceeds the standard deviation of all day of the previous week or, analogously, is lower than its half. So, the prerequisite is one week of observation. Furthermore, the practical realization of this type of system is conditioned to the availability of streaming data about different user's profiles, constraints about privacy, and computational capabilities.

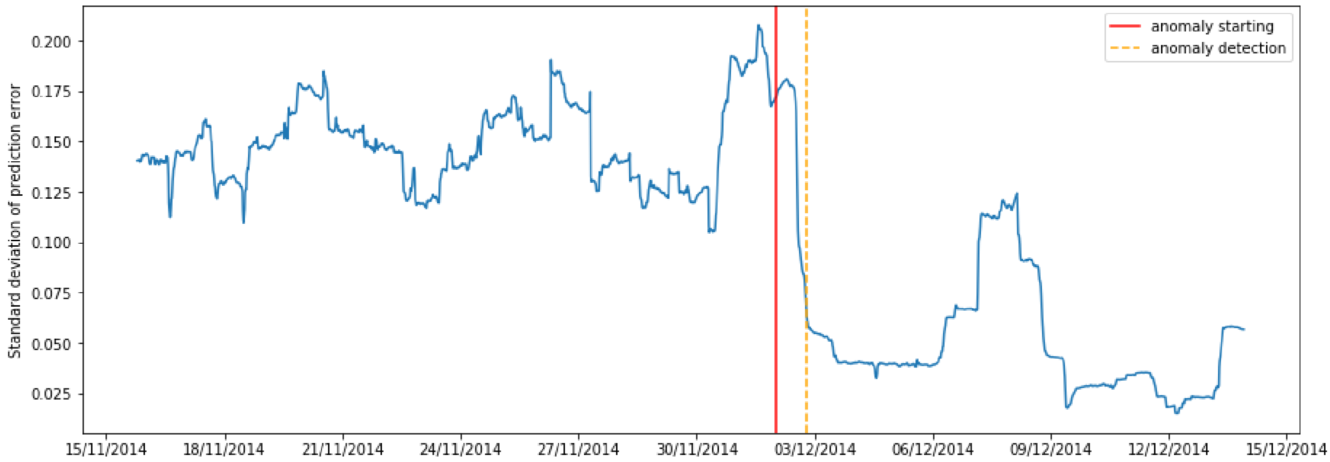

FIGURE 18. Standard deviation trend of anomalous consumption recognized by the system.

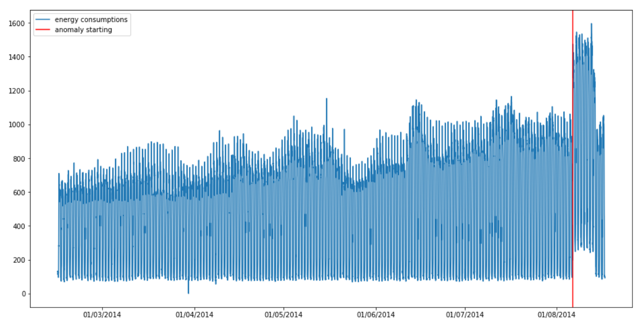

FIGURE 19. Consumptions six months before the anomaly detection shown in Fig.11.

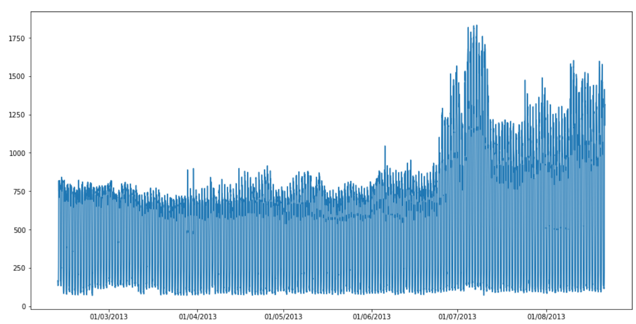

FIGURE 20. Consumptions one year before the anomaly detection shown in Fig.11.

V. CONCLUSION AND FUTURE WORKS

This work consists of a framework based on a Recurrent Neural Network trained in order to forecast energy consumption in a smart grid. Starting from the prediction error carried out by the network, the framework tries to recognize anomalies in the consumptions distinguishing them from concept drifts (i.e., changes in the customers' practices). The experimentation demonstrates that training a unique model for different profiles allows to recognize known trends although the consumption changes and highlighting possible anomalies. So, it confirms the suitability of the LSTM model for the proposed aim.

The future of the proposed method will move in two directions, that are: 1) studying the motif of the prediction errors will allow us to classify the habits change enabling other kinds of services in terms of consumption management and anomalous consumptions detection; 2) improving the performance of the proposed approach by reducing the delay between the anomaly and its detection, at the moment it is of few hours (corresponding to few tens of observations, that happens one each 15 mins).

REFERENCES

[1] Northeast Group LLC. Electricity Theft and Non-Technical Losses: Global Markets, Solutions, and Vendors. Brochure, 2017. [Online]. Available: http://www.northeast-group.com/reports/Brochure-ElectricityTheft&Non-TechnicalLosses-NortheastGroup.pdf

[2] N. Mohamed, J. Al-Jaroodi, I. Jawhar, S. Lazarova-Molnar, and S. Mahmoud, "Smartcityware: A service-oriented middleware for cloud and fog enabled smart city services," *IEEE Access*, vol. 5, pp. 17576–17588, 2017.

[3] M. Zanetti, E. Jamhour, M. Pellenz, M. Penna, V. Zambenedetti, and I. Chueiri, "A tunable fraud detection system for advanced metering infrastructure using short-lived patterns," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 830–840, Jan. 2017.

[4] A. L'Heureux, K. Grolinger, H. F. Elyamany, and M. A. M. Capretz, "Machine learning with big data: Challenges and approaches," *IEEE Access*, vol. 5, no. 5, pp. 777–797, 2017.

[5] M.-M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, and A. Gómez-Expósito, "Detection of non-technical losses using smart meter data and supervised learning," *IEEE Trans. Smart Grid*, to be published, doi: 10.1109/TSG.2018.2807925.

[6] S.-J. Chen, T.-S. Zhan, C.-H. Huang, J.-L. Chen, and C.-H. Lin, "Nontechnical loss and outage detection using fractional-order self-synchronization error-based fuzzy Petri nets in micro-distribution systems," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 411–420, Jan. 2015.

[7] J. V. Spirić, S. S. Stanković, M. B. Dočić, and T. D. Popović, "Using the rough set theory to detect fraud committed by electricity customers," *Int. J. Elect. Power Energy Syst.*, vol. 62, pp. 727–734, Nov. 2014.

[8] V. Ford, A. Siraj, and W. Eberle, "Smart grid energy fraud detection using artificial neural networks," in *Proc. IEEE Symp. Comput. Intell. Appl. Smart Grid (CIASG)*, Dec. 2014, pp. 1–6,

[9] S.-C. Yip, K. Wong, W.-P. Hew, M.-T. Gan, R. C.-W. Phan, and S.-W. Tan, ''Detection of energy theft and defective smart meters in smart grids using linear regression,'' *Int. J. Elect. Power Energy Syst.*, vol. 91, pp. 230–240, Oct. 2017.

[10] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, ''A multi-sensor energy theft detection framework for advanced metering infrastructures,'' *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1319–1330, Jul. 2013.

[11] J. I. Guerrero, C. León, I. Monedero, F. Biscarri, and J. Biscarri, ''Improving knowledge-based systems with statistical techniques, text mining, and neural networks for non-technical loss detection,'' *Knowl.-Based Syst.*, vol. 71, pp. 376–388, Nov. 2014.

[12] B. O. Omijeh and G. I. Ighalo, ''Modeling of GSM-based energy recharge scheme for prepaid meter,'' *IOSR J. Elect. Electron. Eng.*, vol. 4, no. 1, pp. 46–53, 2013.

[13] S.-C. Huang, Y.-L. Lo, and C.-N. Lu, ''Non-technical loss detection using state estimation and analysis of variance,'' *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2959–2966, Aug. 2013.

[14] I. Žliobaitė, M. Pechenizkiy, and J. Gama, ''An overview of concept drift applications,'' in *Big Data Analysis: New Algorithms for a New Society*. Cham, Switzerland: Springer, 2016, pp. 91–114.

[15] T. J. Ogundele, C.-Y. Chow, and J.-D. Zhang, ''Socast*: Personalized event recommendations for event-based social networks: A multi-criteria decision making approach,'' *IEEE Access*, vol. 6, pp. 27579–27592, 2018.

[16] H. Zheng, Y. Zhou, N. Liang, X. Xiao, A. K. Sangaiah, and C. Zhao, ''Exploiting user mobility for time-aware poi recommendation in social networks,'' *IEEE Access*, to be published, doi: 10.1109/ACCESS.2017.2764074.

[17] L. Qi, X. Xu, W. Dou, J. Yu, Z. Zhou, and X. Zhang, ''Time-aware IoE service recommendation on sparse data,'' *Mobile Inf. Syst.*, vol. 2016, Nov. 2016, Art. no. 4397061.

[18] Y. Chen, Z. Zheng, S. Chen, L. Sun, and D. Chen, ''Mining customer preference in physical stores from interaction behavior,'' *IEEE Access*, vol. 5, pp. 17436–17449, 2017.

[19] B. Sun and L. Dong, ''Dynamic model adaptive to user interest drift based on cluster and nearest neighbors,'' *IEEE Access*, vol. 5, pp. 1682–1691, 2017.

[20] S. Saurav *et al.*, ''Online anomaly detection with concept drift adaptation using recurrent neural networks,'' in *Proc. ACM India Joint Int. Conf. Data Sci. Manage. Data*, 2018, pp. 78–87.

[21] A. K. Jain, ''Data clustering: 50 years beyond K-means,'' *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.

[22] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, ''Long short term memory networks for anomaly detection in time series,'' in *Proc. Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn.*, 2015, p. 89.

[23] W. De Mulder, S. Bethard, and M.-F. Moens, ''A survey on the application of recurrent neural networks to statistical language modeling,'' *Comput. Speech Lang.*, vol. 30, no. 1, pp. 61–98, 2015.

[24] M. B. Kamal, G. J. Mendis, and J. Wei, ''Intelligent soft computing-based security control for energy management architecture of hybrid emergency power system for more-electric aircrafts,'' *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 4, pp. 806–816, Aug. 2018.

[25] M. Brown, C. Barrington-Leigh, and Z. Brown, ''Kernel regression for real-time building energy analysis,'' *J. Building Perform. Simul.*, vol. 5, no. 4, pp. 263–276, 2012.

[26] J.-S. Chou and A. S. Telaga, ''Real-time detection of anomalous power consumption,'' *Renew. Sustain. Energy Rev.*, vol. 33, pp. 400–411, May 2014.

[27] T. M. Kodinariya and P. R. Makwana, ''Review on determining number of cluster in k-means clustering,'' *Int. J. Advance Res. Comput. Sci. Manage. Stud.*, vol. 1, no. 6, pp. 90–95, 2013.

[28] F. Rodrigues and A. Trindade, ''Load forecasting through functional clustering and ensemble learning,'' *Knowl. Inf. Syst.*, vol. 57, no. 1, pp. 229–244, 2018.

[29] S. Masum, Y. Liu, and J. Chiverton, ''Multi-step time series forecasting of electric load using machine learning models,'' in *Proc. Int. Conf. Artif. Intell. Soft Comput.* Cham, Switzerland: Springer, 2018.

GIUSEPPE FENZA received the Ph.D. degree in computer science from the University of Salerno, Italy, in 2009. Since 2009, he has been collaborating on several research initiatives mainly focused on knowledge extraction from unstructured resources defining intelligent systems based on the combination of techniques from soft computing and semantic web, areas in which he has many publications. He is currently an Assistant Professor in computer science with the Department of Management and Innovation Systems, University of Salerno. He has been deeply involved in several EU and Italian research and development projects on ICT and, in particular, on situation awareness, service discovery, enterprise information management, and e-commerce. He has published extensively about fuzzy decision making, ontology elicitation, situation and context awareness, and semantic information retrieval. He is currently working in the fields of big data, social media analytics, and web intelligence by proposing novel methods, for instance, to support microblog summarization, time-aware information retrieval, and recommendations extraction. He serves as an Associate Editor for international journals, such as *Neurocomputing*, the *International Journal of Grid and Utility Computing*, and the *International Journal of Engineering Business Management*.

MARIACRISTINA GALLO received the master's degree in computer science from the University of Salerno, Italy, in 2009, where she is currently pursuing the Ph.D. degree in big data management. From 2009 to 2017, she collaborates on several research initiatives and projects mainly focused on computational intelligence, data mining, and ontology e-learning semantic information retrieval in different domains, such as health, e-commerce, and enterprise. Recently, she has worked in the fields of social media analytics and semantic web to study users' interests, characteristics of their posts, and potential cyclic nature of both of them.

VINCENZO LOIA (SM'08) received the master's degree in computer science from the University of Salerno, Italy, in 1984, and the Ph.D. degree in computer science from the University of Paris VI, France, in 1989. Since 1989, he has been a Faculty Member with the University of Salerno, where he is currently the Head of the Department of Management and Innovation Systems. He was the Principal Investigator in a number of industrial research and development projects and in academic research projects. He has authored over 350 original research papers in international journals, book chapters, and international conference proceedings. He has edited four research books about agent technology, Internet, and soft computing methodologies. His current research interests include merging soft computing and agent technology to design technologically complex environments, particularly in web intelligence applications. He is the Co-Editor-in-Chief of *Soft Computing*, and the Editor-in-Chief of the *Journal of Ambient Intelligence and Humanized Computing*, both published from Springer-Verlag. He serves as an Associate Editor for several international journals, such as the IEEE Transactions on Industrial Informatics, the IEEE Transactions on Systems, Man, and Cybernetics: Systems, the IEEE Transactions on Fuzzy Systems, and the IEEE Transactions on Cognitive and Developmental Systems. He holds several roles in the IEEE Society in particular for Computational Intelligence Society (the Chair of the Emergent Technologies Technical Committee, the IEEE CIS European Representative, and the Vice Chair of the Intelligent System Applications Technical Committee).