A Contextual Anomaly Detection Framework for Energy Smart Meter Data Stream

Liu, Xiufeng; Lai, Zhichen; Wang, Xin; Huang, Lizhen; Nielsen, Per Sieverts

Published in:
Proceedings of the International Conference on Neural Information Processing

Link to article, DOI:
10.1007/978-3-030-63823-8_83

Publication date:
2020

Document Version
Peer reviewed version

Citation (APA):
Liu, X., Lai, Z., Wang, X., Huang, L., & Nielsen, P. S. (2020). A Contextual Anomaly Detection Framework for Energy Smart Meter Data Stream. In *Proceedings of the International Conference on Neural Information Processing* (pp. 733-742). Springer. https://doi.org/10.1007/978-3-030-63823-8_83

A Contextual Anomaly Detection Framework for Energy Smart Meter Data Stream

Xiufeng Liu[1](✉), Zhichen Lai[2](✉), Xin Wang[3], Lizhen Huang[4], and Per Sieverts Nielsen[1]

[1] Technical University of Denmark, 2800 Lyngby, Denmark
xiuli@dtu.dk
[2] Sichuan University, Chengdu, China
ryanlai.cs@gmail.com
[3] Southwest Jiaotong University, Chengdu, China
[4] Norwegian University of Science and Technology, Trondheim, Norway

Abstract. Monitoring abnormal energy consumption is helpful for demand-side management. This paper proposes a framework for contextual anomaly detection (CAD) for residential energy consumption. This framework uses a sliding window approach and prediction-based detection method, along with the use of a concept drift method to identify the unusual energy consumption in different contextual environments. The anomalies are determined by a statistical method with a given threshold value. The paper evaluates the framework comprehensively using a real-world data set, compares with other methods and demonstrates the effectiveness and superiority.

Keywords: Contextual · Anomaly detection · Data stream · Concept drift

1 Introduction

According to the EU energy in figures [3], households account for 27% of total energy consumption and 16% carbon emissions. Several studies [8,10,11] have shown that thermal renovation often does not deliver the expected energy savings. This is because household energy consumption is largely dependent on household characteristics and occupant behaviour. Therefore, an effective approach is to monitor user energy consumption, in particular, to detect and diagnose abnormal consumption. This now become possible with the use of smart meters. By analysing smart meter data, utilities can then provide personalised services and energy-efficiency suggestions.

Anomaly detection refers to the process of identifying unusual observations that do not meet expectations, also known as outlier detection [7]. Contextual factors often play an important role in influencing energy consumption, which must be taken into account when detecting anomalies. We call it contextual

H. Yang et al. (Eds.): ICONIP 2020, CCIS 1333, pp. 733–742, 2020.
https://doi.org/10.1007/978-3-030-63823-8_83

anomaly detection in this paper. For example, the increase in electricity consumption in summer may not need to be considered as anomalies if we consider the effects of weather temperature. This is because people usually use more energy in summer due to air conditioning for cooling. However, if the temperature is mild and no air conditioning is used, the high consumption should be detected as an anomaly. This is also true in other contexts, such as changes in consumption patterns, thermal renovation of buildings or changes in occupants, which may cause the general consumption pattern shifts. Therefore, contextual anomaly detection consists of identifying individual observations or patterns that differ from the masses in the same context. In this paper, the consumption trend or pattern changes due to variations in the contextual environment are considered as *concept drift*, which means that the overall pattern or trend may change abruptly or gradually. The anomalies will be detected accordingly in each context.

Depending on how the streaming data are processed, the methods can be divided into batch and stream data anomaly detection. This paper focuses on the latter. Compared to the detection of anomalies in batch data, the detection of anomalies in data streams consists of additional challenges due to their complex nature and data stream evolving. These challenges include high velocity, the presence of noisy data, and concept drift. The model trained on static data may not be suitable for detection on a data stream because the time series may drift, e.g., the data distribution changes. Therefore, the detection models should be retrained if a concept drift occurs. To remedy this, our earlier work [7] updates the detection models iteratively using a lambda system and uses the latest model to detect anomalies of a data stream. This approach runs in a cluster environment and requires a large amount of computing capacity. It is therefore not an ideal solution for stream data mining in an environment with limited computing resources such as IoT devices.

In this paper, we implement a contextual anomaly detection framework, called *CAD*. It integrates with a concept drift detector to identify each context and trains the detection model when a concept drift occurs. It can be integrated into a building energy management system or a building information management platform (BIM) to enable automation of energy performance monitoring or design of building energy systems. This paper makes the following contributions: (1) We propose an algorithm of detecting contextual anomalies for high-frequent energy consumption data streams; (2) We propose an unsupervised concept drift detection method for contextual detection; (3) We evaluate the proposed framework comprehensively, and compare with other approaches using a real-world energy consumption data set.

2 The Framework

The proposed framework for contextual anomaly detection is presented in Fig. 1. This framework detects anomalies on the streaming data from smart meters. First, it uses online concept drift detection to identify different contexts representing the change of energy consumption environment. The reason is that context-dependent factors are often difficult to determine. Then, the framework uses an algorithm to predict short-term energy consumption values. In the end, the predicted value is compared with the actual value by calculating the distance which is used to determine an anomaly or not according to a given threshold value. The anomaly can be used for notification purposes or control signal in smart energy systems.

Figure 2 shows an illustrating example using synthetic data to explain when and how an anomaly is detected. In this time series, two concept drifts are first identified, and the prediction model is retrained when a concept drift is determined. Thus, the concept drifts are used to identify the context shifts and the consumption values are predicted using the retrained model in each context (see the red line). Anomaly scores are calculated based on prediction errors, and those with anomaly scores above a threshold value are classified as anomalies. The concept drift detection, short-term prediction, and anomaly classifier are described in greater detail in the following.

2.1 Context Shift Detection

The concept drift detection can be formalised as follows. For a data stream, $\mathcal{D} = \{(X_1, y_1), (X_2, y_2), \ldots, (X_t, y_t), \ldots\}$, where X is the vector of features and y is the class, the concept drift can be defined as the distribution shift detection, i.e., $p(X_t, y_t) \neq p(X_{t-1}, y_{t-1})$, where p represents the probability of concept drift. The entire data stream is divided into the segments in which the drifts of the underlying distributions are detected accordingly. We perform the model update only when a drift is detected. The algorithm is called *CAD-D3*, derived from the Discriminative Drift Detector (D3) [5] (see Algorithm 1). It detects a drift by comparing the divergence of data statistics between two segments in a time window.

Figure 3 shows the architecture. A fixed-length time window, W, slides on a time series with two segments partitioned by a ratio of p, w and pw, used to hold old and new data respectively. The separability of the two segments is determined by an AUC value ranging from 0.5 to 1. 1 means that the two classes can be perfectly separated and 0.5 means that they overlap completely. With a given threshold τ, if the AUC value is greater than or equal to τ, the old data will be dropped and the new data will be moved into it (the lower left); otherwise, the pw elements at the tail of the sliding window will be dropped and the remaining data in the sliding window will be shifted to the left (see Algorithm 1 for details). Therefore, this method can identify different contexts of energy consumption.

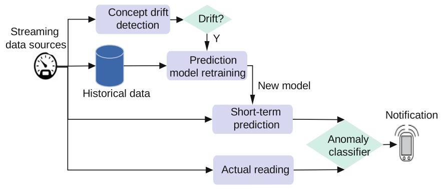

Fig. 1. Architecture overview

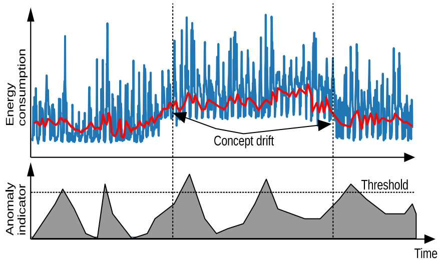

Fig. 2. Contextual anomaly detection overview (Color figure online)

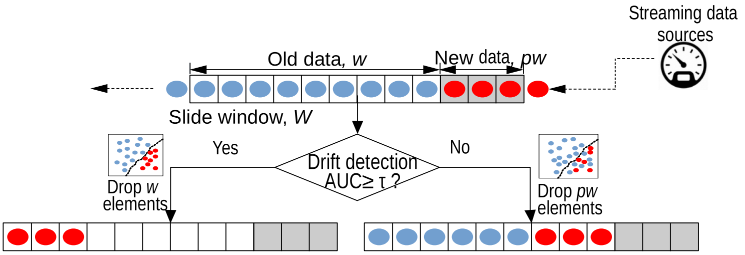

Fig. 3. Concept drift detection architecture of CAD-D3

2.2 Consumption Prediction

We now present how to predict short-term energy consumption. The energy consumption pattern of residential households typically has a periodicity characteristics, e.g., with low consumption during the night and high consumption during the day with morning and evening peaks. The yearly pattern has a seasonality (or seasonal periodicity), e.g., with higher consumption in both summer and winter, due to electrical cooling and heating. The fluctuations with a long range of time are considered as contextual shifts in this paper, e.g., weather temperature changes. We choose the LSTM (Long Short-Term Memory) model which has shown good performance in time series prediction [12]. LSTM is a deep learning method with numerous hidden layers, which has a gate structure and a memory cell to manage the temporal correlation of time series. A LSTM model can be seen as the function $\mathcal{F}(\cdot)$ that maps a sequence of past observations to an output as the predicted value. It is defined as: $X_t' = \mathcal{F}(X_{t-1}, X_{t-2}, \cdots, X_{t-24})$, where X_t' is the predicted value, and $X_{t-1}, X_{t-2}, \cdots, X_{t-24}$ are the past values, representing the last 24 h energy consumption that constitute a daily periodicity.

2.3 Aggregated for Contextual Anomaly Detection

We now aggregate the above concept drift detection and the LSTM prediction model to detect contextual anomalies for energy consumption time series. The aggregated model is presented in Algorithm 1. The input includes an individual consumption time series $\mathcal{TS}$, the period d ($d = 24$ as the default for daily periodicity), the size of window holding old data w, the window ratio for holding new data p, and the drift threshold τ. The algorithm is for online concept drift detection, $\mathcal{C}$ and anomaly detection $\mathcal{AD}$, as well as retraining the prediction model, $\mathcal{AP}$.

This algorithm first divides the elements in the sliding window W into multiple groups according to the period d (line 16), then extracts the group statistics as the features to train the concept drift detection model, including mean, standard deviation, and range. The drift discriminative classifier is a logistic regression model trained by the extracted features (lines 21–22), then a concept drift is detected (lines 23–24). The group statistical characteristics is used as the features for training the classifier. The AUC is then used as the measure to determinate the separability between the old and new data. This is done by introducing a slack variable s, which labels old data as 1 and new data as 0 (line 19–20). If the AUC is greater than or equal to the user-specified drift threshold value τ, a concept drift is determined. In the end, the auto-regressive predictor $\mathcal{AP}$ (i.e., the LSTM model) is retrained to predict energy consumption in the next context after the concept drift (lines 30–31).

The following describes the anomaly classifier $\mathcal{AD}$. For the input X_t at the time t, the corresponding prediction error $err_t = |X_t - X_t'|$ is calculated (line 8). The error represents the prediction capability of the generated model. Instead of directly thresholding the error to identify an anomaly, the error distribution is used as the indirect metric for calculating anomaly score. At the time t with N_t data points, the errors are modeled as a rolling normal distribution, $\mathcal{N}(\mu_t, \sigma_t^2)$, where:

$$\mu_t = \frac{\sum_{i=1}^{N_t} err_i}{N_t}, \quad \sigma_t^2 = \frac{\sum_{i=1}^{N_t} (err_i - \mu_t)^2}{N_t}, \quad S(err) = 1 - \frac{1}{\sigma_t\sqrt{2\pi}} e^{-\frac{(err-\mu_t)^2}{2\sigma_t^2}} \tag{1}$$

If the anomaly score, S, is greater than or equal to a user-defined threshold value, ϵ, i.e. $S(err) \geq \epsilon$, the corresponding data point is classified as an anomaly.

3 Evaluation

The experiments were conducted on an HP laptop, configured with an Intel(R) Core(TM) i7-8700 processor (3.20 GHz, 6 Cores), 16 GB RAM, and a Western Digital Hard driver (1TB, 6 Gb/s, 64 MB Cache and 7200 RPM), running Windows 10 Professional edition. The LSTM network is implemented with Keras 2.3.1 [2].

The Irish Commission for Energy Regulation (CER) data set [6] is used, with time series of 4,182 households from July 2009 to December 2010, a resolution of 30 min. We made the following pre-processing: aggregate to an hourly resolution and concatenate multiple time series to obtain a longer time series. The concatenations of different time series with statistical significance in values can be seen as a context shift, as well as the changes due to weather temperatures.

Algorithm 1: Contextual anomaly detection model

```
function CAD($TimeSeries\ \mathcal{TS}$, $WindowSize\ w$, $Period\ d$, $NewDataRatio\ p$,
  $DriftThreshold\ \tau$)
    Initialize window $W$ where $|W| = w(1 + p)$;
    Discriminative classifier $\mathcal{C}$;
    Auto-regressive predictor $\mathcal{AP}$;
    Statistical anomaly detector $\mathcal{AD}$;
    while $\mathcal{TS}$ generates a new sample $X$ do
        $X' = AP(W[|W| - d, |W|])$;
        $err = |X - X'|$;
        if $AD(err)$ alarms then
            $Anomaly = True$;
        else
            $Anomaly = False$;
        end
        if $W$ is not full then
            $W \leftarrow W \cup X$;
        else
            $G = \{G_1, ..., G_{|W|/d}\}$ ← Split $W$ into $|W|/d$ groups, each of which $G_i$ has $d$
              elements, $i \in [1, |W|/d]$;
            $F = \{F_1, ..., F_{|W|/d}\}$ ← Calculate mean, standard deviation and range as the
              classification feature $F_i$ of $G_i$;
            $S$ is vector of $s = \{0, 1\}$, $|S| = |W|/d$, label for old data (0) and new data (1);
            $s = 0$ for old data $F[1, (w/d)]$;
            $s = 1$ for new data $F[(w/d) + 1, |W|/d]$;
            $(F_{train}, S_{train}), (F_{test}, S_{test})$ ← Split classification feature $F$ into train and
              test set with $S$;
            Train the discriminative classifier $C(F_{train}, S_{train})$ on the train set;
            $AUC(F_{test}, S_{test})$ ← Test $C(F_{test}, S_{test})$ on the test set;
            if $AUC(F_{test}, S_{test}) \geq \tau$ then
                $Drift = True$;
                Drop $w$ elements from the tail of $W$;
            else
                $Drift = False$;
                Drop $wp$ elements from the tail of $W$;
            end
        end
        if $Drift$ is $True$ then
            Retrain $\mathcal{AP}$ model;
        end
    end
end
```

The following subsections will first investigate the optimal hyperparameters for the CAD model, then evaluate the concept drift model and the influence of the model update frequency on the prediction accuracy, and finally evaluate the anomaly detection.

3.1 Hyperparameter Investigation for the CAD Model

The hyperparameters include the LSTM units of the time series predictor, the sliding window size and the AUC threshold of the concept drift detector. The following studies will be based on the performance evaluated by mean square

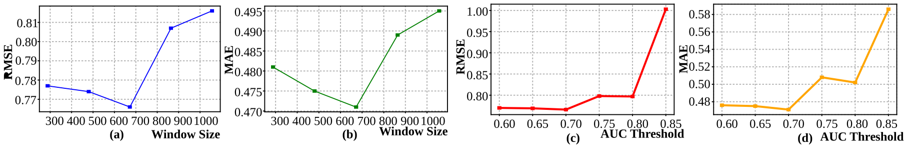

Fig. 4. Impact of sliding window size and AUC threshold on prediction

Table 1. Different LSTM units in CAD

No. LSTM units	RMSE	MAE
5	0.963	0.557
10	0.842	0.512
25	0.766	0.471
50	0.792	0.496

Table 2. Concept drift on prediction

Model	RMSE	MAE
DDM	1.117	0.694
ADWIN	1.101	0.674
D3	0.826	0.558
CAD-D3	0.766	0.471

error (RMSE) and mean absolute error (MAE).

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(X_i' - X_i)^2}{N}}, \qquad MAE = \frac{\sum_{i=1}^{N} |X_i' - X_i|}{N} \tag{2}$$

where X' is the predicted value, X is the observation (actual reading) and N is the number of observations.

We first investigate the optimal number of LSTM units by varying the number of LSTM units. The corresponding RMSE and MAE values are in Table 1, which shows that the CAD model performs the best when the number of units is set to 25.

We next investigate the optimal sliding window size W by varying from 288 (corresponding to the hourly readings of 6 days) to 1,056 (corresponding to the hourly readings of up to 44 days). The results are shown in Fig. 4a and b, in which the lines both have a "V" shape, with the lowest value at $W = 672$ (corresponding to the hourly readings of 28 days). We then investigate the AUC threshold of the concept drift detector, and use a fixed window size $W = 672$, but vary the threshold value from 0.6 to 0.85. The results are shown in Fig. 4c and d, showing that the AUC threshold of $\tau = 0.7$ has the best result. Therefore, the window size $W = 672$ and the concept drift threshold $\tau = 0.7$ are used as the default parameters.

3.2 Comparison with Other Drift Detection Models

We now compare CAD-D3 with other baselines, including the original D3, and DDM [1] and ADWIN [4]. The reason for choosing DDM and ADWIN is that they are the most commonly used concept drift detection algorithms for data stream. The results are presented in Table 2, which shows how different concept drift detectors affect prediction accuracy. D3 and CAD-D3 outperform DDM and ADWIN according to the RMSE and MAE values, and CAD-D3 performs slightly better than D3. It is because CAD-D3 improves the D3 by taking into account the statistical features in its classifier, including mean, standard deviation and range of daily consumption data. This results in greater accuracy, but also allows more updates to be triggered for predictions.

3.3 Evaluation of Model Update Frequency

In this study, we compare the model without update (static model), the model with regular updates (regular model) and the model update triggered by concept drift (CAD model). We randomly select a time series for the experiment. Figure 5a–c show the results of the three models, while Fig. 5d shows the original time series for better visual comparison. The results show that the static model predicts electricity consumption with a relatively fixed pattern because it was trained by the early data and is not able to adapt well to data variations in different contextual environments. In Fig. 5b, the regular model shows a similar result to the static model in Fig. 5a. Note that the prediction accuracy is related to the freshness of the model used. Here we specify the update frequency for every 4,000 h for the regular model, which does not obtain the same prediction accuracy as the CAD model. In contrast, the CAD model is more adaptable. The model is updated according to the contextual environment variation, which maintains the freshness of the model.

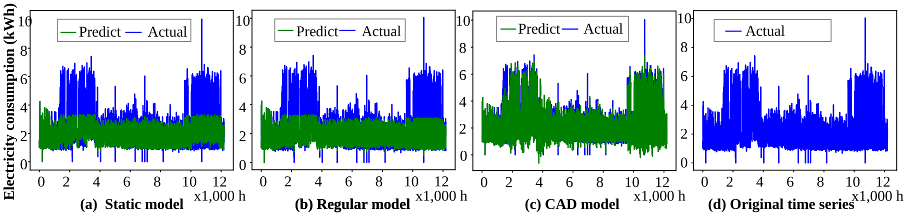

Fig. 5. Prediction by the models with different update frequency

We calculate the corresponding RMSE and MAE values in Table 3. The static model has the highest RMSE and MAE values, followed by the regular model. However, we can see that the regular model has improved the RMSE value by 27.3% and the MAE value by 26.9% compared to the static model. In this case, the CAD model gives the best result, i.e., it can achieve good prediction accuracy in contextual environments.

Table 3. Impact by update frequency

Model	RMSE	MAE
Static model	1.203	0.722
Regular model	0.918	0.528
CAD model	0.766	0.471

Table 4. Comparison between models

Model	Precision	Recall	F1-score
Tukey BP	0.544	0.831	0.564
Static model	0.557	0.671	0.537
Regular model	0.580	0.793	0.585
CAD model	0.584	0.844	0.636

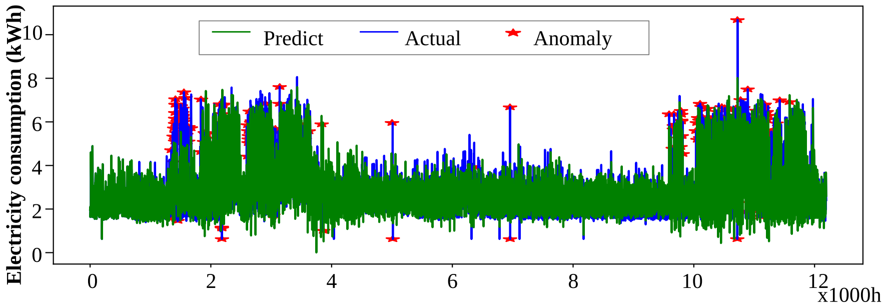

Fig. 6. A showcase of anomaly detection using the CAD model (Color figure online)

3.4 Evaluation of Anomaly Detection

Since there are no ground-truth anomalies, we label the data manually, aided by a time series visualisation tool and a Python program. Again, we compare the CAD model with the static model and the regular model. We also introduce *Tukey Boxplot (BP)* [9] as an additional baseline to evaluate the proposed framework.

Figure 6 is a showcase of anomaly detection by the CAD model, with detected anomalies marked with a red star. Note that the ϵ anomaly detection threshold is set to 0.6. Table 4 summarises the performance of anomaly detection based on the metrics including precision, recall and F1 score. According to the F1 score, the static model has the lowest accuracy in anomaly detection. It is noteworthy that there is a big difference between the static model and the other models in terms of recall value. The accuracy of the regular model is better than the static model, but its recall value is lower than the Tukey BP. The CAD model outperforms others in all metrics, which means that our model outperforms others in detecting contextual anomalies. The others, like the static model, have not taken into account contextual shifts because their models are fixed, which makes it less accurate. Although the regular model performs better than the static model, blind updates may not help to obtain better accuracy because a concept drift can occur exactly after the model is updated. Because Tukey BP analyses the overall statistical characteristics without considering the periodicity and drift of the data, its performance is not superior. In contrast, the contextual CAD model is well suited for detecting anomalies for time series with periodicity and drifts, such as energy consumption data, as it can maintain model freshness.

4 Conclusions and Future Work

This paper proposed a framework for online contextual anomaly detection for the data stream from smart meters. Concept drift was used to signal the refreshing of the prediction model. In this paper, the errors between the actual and predicted values were modeled into a rolling normal distribution and used to calculate the score for identifying anomalies. We have conducted extensive experiments to evaluate the framework, and the results verified the superiority of the proposed CAD model over other baselines, being better able to adapt to changes in contextual conditions and environment, but also offering the best performance. For future work, we plan to support the detection of different types of anomalies such as energy theft and leakage.

Acknowledgement. This research was supported by the CITIES project (No. 1035-00027B) and the HEAT 4.0 project (No. 8090-00046B) funded by IFD.

References

1. Bifet, A., Gavalda, R.: Learning from time-changing data with adaptive windowing. In: Proceedings of the 2007 SIAM International Conference on Data Mining, pp. 443–448. SIAM (2007)
2. Chollet, F., et al.: Keras. https://github.com/keras-team/kerasas. Accessed 10 Sept 2020
3. EU Energy in Figures: Statistical pocketbook 2018. European Union (2018)
4. Gama, J., Medas, P., Castillo, G., Rodrigues, P.: Learning with drift detection. In: Bazzan, A.L.C., Labidi, S. (eds.) SBIA 2004. LNCS (LNAI), vol. 3171, pp. 286–295. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-28645-5_29
5. Gözüaçık, Ö., Büyükçakır, A., Bonab, H., Can, F.: Unsupervised concept drift detection with a discriminative classifier. In: Proceedings of CIKM, pp. 2365–2368 (2019)
6. ISSDA: Irish social science data archive. http://www.ucd.ie/issda/data/commissionforenergyregulationceras. Accessed 10 Sept 2020
7. Liu, X., Nielsen, P.S.: Scalable prediction-based online anomaly detection for smart meter data. Inf. Syst. **77**, 34–47 (2018)
8. Lundström, L., Wallin, F.: Heat demand profiles of energy conservation measures in buildings and their impact on a district heating system. Appl. Energy **161**, 290–299 (2016)
9. Mosteller, F., Tukey, J.W., et al.: Data Analysis and Regression: A Second Course in Statistics. Addison-Wesley Publishing Company, Boston (1977)
10. Passer, A., Ouellet-Plamondon, C., Kenneally, P., John, V., Habert, G.: The impact of future scenarios on building refurbishment strategies towards plus energy buildings. Energy Build. **124**, 153–163 (2016)
11. Risholt, B., Time, B., Hestnes, A.G.: Sustainability assessment of nearly zero energy renovation of dwellings based on energy, economy and home quality indicators. Energy Build. **60**, 217–224 (2013)
12. Siami-Namini, S., Tavakoli, N., Namin, A.S.: A comparison of ARIMA and LSTM in forecasting time series. In: Proceedings of ICMLA, pp. 1394–1401 (2018)