JEL C32, C53, C58, C81

BUILDINGS ELECTRICITY CONSUMPTION MODELING USING ARTIFICIAL INTELLIGENCE SYSTEMS

SHCHETININ EUGENE (ORCID 0000-0003-3651-7629)[1]

[1]Financial University under the Government of the Russian Federation

Abstract. Intelligent technologies of energy saving and energy efficiency are a modern large-scale global trend in the development business. The demand for smart buildings is growing not only in the world, but also in Russia, especially in the market of construction and operation of large shopping, entertainment and business centers. In this paper the gradient boosting algorithm is used to simulate energy consumption. On its base the method of modeling the daily energy consumption profile is proposed and a numerical algorithm is developed. Data on the energy consumption of 380 commercial buildings were used to assess its efficiency. The computer experiments showed that the use of the gradient boosting model improved the prediction accuracy in more than 80 percent of cases compared to the куп algorithms..

Keywords: smart buildings, artificial intelligence systems, gradient boosting, neural nets.

INTRODUCTION

The development of intelligent networks in manufacturing, finance and services creates new opportunities for the development and application of effective methods of machine learning and data analysis. Smart technologies for collecting, recording and monitoring data on energy consumption create a huge amount of data of different nature. These data can be used for optimal network management, improving the accuracy of the forecasting load, detection of abnormal effects of power supply (peak load conditions), the formation of flexible price tariffs for different groups of consumers [1, 2, 3]. One of the most important issues in this area is to predict the power load consumption as accurately as possible. Consumption has rather complicated stochastic structure which is difficult for modeling and prediction. Nevertheless, when different methods of aggregation are applied to the group of consumers having similar statistical characteristics of time series of power consumption, it is possible to count on considerable progress in the solution of objectives. However, in order to improve the accuracy, our goal is to investigate the possible benefit of combining clustering procedures with forecasting methods. We have used several known approaches such as Holt-Winters forecasting, ARIMA model, Support Vector Regression and some others. The cluster analysis can also be beneficial for finding the patterns in data [3, 4].

MACHINE LEARNING ALGORITHMS APPLICATIONS FOR ENERGY CONSUMPTION MODELING

The problems of application of clustering methods to the time series of electricity consumption are mainly in high dimension and high noise level of the data, which can be solved with the use of machine learning methods. In the paper the comparative analysis of two approaches of classification of consumers is carried out: on the basis of clustering and without it (aggregation). Our cluster approach has three steps: the first step is to normalize the data and calculate the energy consumption model for each consumer. In the future, the study uses four different models based on the representation of time series, which serve as inputs to the clustering method. The second stage consists of calculating the optimal number of clusters for the given time series representation and the selected data learning window. The third stage is clustering and aggregation of consumption within clusters. For each

DOI: 10.33278/SAE-2018.eng.167-170

cluster, the forecast model is trained and the forecast for the next period is run. Then the forecasts are aggregated and compared with the real consumption data. Next, we construct a forecast for day-ahead for the received representations of the clusters using the above-described prediction methods.

Modeling of the energy consumption time series

Time series X is an ordered sequence of n real variables

$$X = (x_1, x_2, \dots, x_n), x_i \in R \quad (1)$$

The main reason for time series presentation using is a significant decrease in the dimension of the analyzed data, respectively, reducing the required memory and reducing the computational complexity. Four different model-based representation methods were chosen: (a) Robust Linear Model (RLM), (b) Generalized Additive Model (GAM), (c) Holt-Winters Exponential Smoothing, and (d) Kalman filter. The first presentation is based on a robust linear model (RLM). Like other regression methods, it is aimed at modeling the dependent variable by independent variables

$$x_i = \beta_1 u_{i1} + \beta_2 u_{i2} + \dots + \beta_s u_{is} + \varepsilon_i \quad (2)$$

where i = 1, ..., n, x_i– is energy consumption, $\beta_0, \beta_1, \dots \beta_s$ are the regression coefficients, u_{i1}, ..., $u_{is} u_{is}$ are the binary variables, ε_i is a white noise. Extensions for regression model (2) are generalized additive models (GAM) [7]

$$E(x_i) = = \beta_0 + \sum_{l=1}^{K} \beta_i f_s(u_{is}), \quad (3)$$

where f_i are B-splines [14]. Model parameters (3) can be evaluated by the method of gradient descent.

Cluster analysis for energy consumption in smart grids

Usually, utility divided their customers in industrial, commercial and residential sectors based on some fixed information like voltage level, nominal demand etc. Based on this approach a set of customer class load profiles were defined and each user was assigned to one of these classes. However, this is still a fundamental problem, and the procedures for dealing with customers segmentation need to be revised greatly. Firstly, the consumption data of customers, those who have installed smart meters, are now accessible. Secondly, the time period of measurement is not restricted and usage information for some successive years is available. These two factors affect the dimensionality of data which is not comparable with previously used data sets. Finally, as the data is continuously recorded, it can have possible applications for real-time operation and management of power systems. All of these factors emphasize the use of new clustering methods for electricity consumption characterization.

For classification consumers into groups (clusters), we used the centroid based clustering method K-means [9]. K-means is a method based on the mutual distances of objects, measured by Euclidean distance. The advantage over conventional K-means is based on carefully seeding of initial centroids, which improves the speed and accuracy of clustering. Before applying the K-means algorithm the optimal number of clusters k must be determined. For each representation of a data set, we have determined the optimal number of clusters to k using the internal validation rate Davies-Bouldin index [10,11]. The optimal number of clusters ranged from 10 to 15. It works as follows. Let d(x) denote the shortest Euclidean distance from a data point x to the closest centroid we have already chosen. Choose an initial centroid K_1 uniformly at random from X. Choose the next center $K_i = \hat{x} \in T K_i = \hat{x} \in T$, selecting with probability $d(\hat{x})^2 / \sum_{x \in X} d(x)^2$. $d(\hat{x})^2 / \sum_{x \in X} d(x)^2$. Repeat previous step until we have chosen a total of K centers. Each object from data set is connected with a centroid that is closest to it. New centroids are then calculated. Last two steps are repeated until classification to clusters no longer changes. Euclidian distance measure is one of the best measures for comparison of time series of electricity load because of its stronger dependence on time. In each iteration of a batch processing, we have automatically determined the optimal number of clusters to K using the internal validation rate Davies-Bouldin index.

Energy consumption time series forecasting

We used three methods to improve forecasting energy consumption time series: Support Vector Regression (SVR), a method based on a combination of STL decomposition, Holt-Winters exponential smoothing and ARIMA model [8]. Seasonal decomposition of time series based on LOESS regression is a method, which decomposes seasonal time series into three parts: trend, seasonal component and remainder (noise) [11]. For the final three time series any of the forecast methods is used separately, in our case either Holt- Winters exponential smoothing or ARIMA model. The original idea of random decision forests was created by Ho [4]. Also Random Forest (RF) algorithm is suitable for classification and regression. The method constructs the large number of decision trees at training time. Its output is the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Extreme Gradient Boosting (XGB) is an efficient and scalable implementation of gradient boosting framework by Friedman [8]. Bagging (Bagg) predictors generate multiple versions of predictors and use them for determination an aggregated predictor. The aggregation is an average of all predictors. The bagging method gives substantial gains in accuracy, but the vital element is the instability of the prediction method. In the case that perturbing the learning set has significant influence on the constructed predictor, the bagging can improve accuracy.

COMPUTER EXPERIMENTS FOR CUSTOMER ENERGY CONSUMPTION

We performed several computer experiments to evaluate the profit of using clustering procedures on four time series representation methods for one day ahead forecast. Our testing data set contains measurements from customers of Central Russia Region [12,13]. Each forecasting method was evaluated on 5 datasets; 4 datasets are clustered with different representation methods (Median, HW, GAM, RLM) and aggregated electric load consumption (Sum). The following conclusions can be derived from our computations. Optimized clustering of consumers significantly improves accuracy of forecast with forecasting methods SVR, Bagging, XGB. Despite this, clustering with STL+ARIMA, RF, R-Tree does not really improve accuracy of forecast. Three robust representation methods of time series Median, GAM and RLM performed best among all representations, while HW was the worst in most of the cases, because robust representations are stable and less fluctuate. The best result of all cases achieved by Bagging with optimized clustering using GAM representation with mean daily MAPE error under 3,17% [14].

CONCLUSION

Improving the accuracy of forecasts of electricity consumption is a key area in the development of intelligent energy grids. To implement this problem, we used machine learning methods, namely cluster analysis. The main purpose of this paper is to show that the application of the clustering procedure of consumers to the representation of time series of energy consumption can improve the accuracy of their forecasts for energy consumption. Robust linear model, generalized additive model, exponential smoothing and median linear filter were used as such representations. In this paper we applied a modified K-means++ algorithm to more accurately select centroids and the Davis-Boldin index to evaluate clustering results. Numerical experiments have shown that the methods of forecasting such as LOESS+ARIMA, SVR, RF, Bagging considered in the paper are more effective for improving forecast accuracy if used together with clustering. Prediction methods performed the best reliable representations of RLM, GAM, and median filter. The most accurate prediction result is obtained by Bagging with the GAM presentation. Among the perspective applications of clustering for smart grids are benefits for tariff design, compilation smart demand response programs, improvement of load forecast, classifying new or non-metered customers and other tasks.

REFERENCES

1. *Haben S., Singleton C., Grindrod P.* Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data // IEEE Transactions on Smart Grid. 2015. V.PP. no.99. P.1-9.

2. *Chicco G., Napoli R. and Piglione F.* Comparisons Among Clustering Techniques for Electricity Customer Classiffcation // IEEE Trans. Power Sys., 2013, vol. 21. P.933-940.

3. *Gelling C. W.* The Smart Grid: Enabling energy effciency and demand response. The Fairmont Press Inc., 2009.

4. *Aghabozorgi S., Shirkhorshidi A.*, et al, Time-series clustering: A decade review, Information Systems, 2015, vol. 53, P. 16-38.

5. *Shahzadeh A., Khosravi A., Nahavandi S.* Improving load forecast accuracy by clustering consumers using smart meter data, International Joint Conference on Neural Networks (IJCNN), 2015, P. 1-7.

6. *Andersen A.* Modern Methods for Robust Regression. SAGE Publications, Inc, 2008.

7. *Wijaya T.K., Vasirani M.* et al, Cluster-based aggregate forecasting for residential electricity demand using smart meter data, in 2015 IEEE International Conference on Big Data. IEEE, oct 2015, pp. 879-887.

8. *Wood S.* Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC, 2006.

9. *Hyndman R.J., Koehler A.B., Snyder R.D. and Grose S.* A state space framework for automatic forecasting using exponential smoothing methods, International Journal of Forecasting, 2002, vol. 18, no. 3, P. 439-454.

10. *Arthur D. and Vassilvitskii S.* K-means++: The Advantages of Careful Seeding, in SODA '07 Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms, 2007, P. 1027-1035.

11. *Hong W.C.* Intelligent Energy Demand Forecasting, London: Springer Verlag, 2013.

12. *Taylor J.W.* Short-term electricity demand forecasting using double seasonal exponential smoothing, Journal of Operational Research Society, 2003, vol. 54, P. 799-805.

13. *Shchetinin Eu.Yu.* Cluster-based energy consumption forecasting in smart grids, Springer Communications in Computer and Information Science (CCIS), 919, 46-656. Springer, Berlin, 2018.

14. *Shchetinin Eu.Yu., Lyubin P.G.* Fast two-dimensional smoothing with discrete cosine transform, Springer Communications in Computer and Information Science (CCIS), Springer: Berlin, 2016, 678, P. 646-656.