Modelling Electricity Consumption Forecasting Using the Markov Process and Hybrid Features Selection

Hadis Dalkani
Department of Computer Engineering, Liyan Institute of Education, Bushehr, Iran
E-mail: hadis.dalkani76@yahoo.com

Musa Mojarad*
Department of Computer Engineering, Firoozabad Branch, Islamic Azad University, Firoozabad, Iran
E-mail: m.mojarad@iauf.ac.ir

Hassan Arfaeinia
Department of Computer Engineering, Liyan Institute of Education, Bushehr, Iran
E-mail: harfaeinia@gmail.com

Abstract: Given the problem of electrical energy storage, it is critical to predict the amount of load required in order to have a reliable and stable power distribution network. Predicting electricity consumption of subscribers and analyzing their consumption behavior under the influence of various factors and time variables is important. Given the large volume of subscriber consumption data and the effective factors, it is only possible to analyze the data using new information technology tools such as data mining. In this paper, feature selection, clustering and Markov process techniques are used to model and predict the power consumption data of subscribers. First, the selection of a subset of effective features is based on the combined PCA approach and the Firefly algorithm. Subscribers are then clustered based on the features selected by the K-means. Finally, subscriber behavior patterns are modeled to predict consumption using the Markov process on high-risk clusters. This study is simulated based on the data of electricity subscribers in Bushehr-Iran Power Distribution Company. The simulation results show the superiority of the proposed model over other similar algorithms such as LASSO-QRNN and HyFIS. The accuracy of power consumption prediction in the proposed method is about 1% compared to LASSO-QRNN and about 0.5% compared to HyFIS.

Index Terms: Power consumption, Markov process, feature selection, clustering, prediction, power distribution company.

1. Introduction

Due to the low price of energy in Iran, its consumption has increased sharply. The government also pays a lot of money as an energy subsidy to cover the gap between the cost price and the selling price of energy to consumers [1]. Among these, the industrial sector, as one of the common factors among other economic sectors in society, plays a decisive role in energy consumption and efficiency; Because this sector in terms of industrial processes and its products is somehow involved in the activities of other sectors [2]. Prediction is one of the issues that is considered by scientists and paperers because of its applications in the real world. From a scientific point of view, prediction means reducing the error in the result obtained with the available results. Having an accurate estimate of the amount of electricity consumption, as one of the requirements for the development of any country and also the limitations of energy storage, is one of the interesting and of course sensitive issues in the field of applied sciences [3].

Due to the change in attitude and the development of new forecasting techniques, identifying groups of customers who have similar behaviors and patterns to receiving services is one of the most important challenges in the electricity distribution company [4]. To measure the subscriber's behavior, the criterion of their value in the sequence of power consumption is used. This is done by implementing a direct evaluation approach based on subscriber consumption patterns in the past. Today, the increase in electricity consumption, especially in urban areas, has led to the identification of high-consumption customers as well as efforts to identify and encourage low-consumption customers [5]. Among these, the issue of modeling electricity consumption of subscribers and identifying the factors affecting it is

of great importance.

Subscriber consumption management cannot be done without recognizing the patterns in the data associated with them. Because the purpose of data mining is to extract meaningful patterns and connections between big data, the use of data mining techniques can model and extract patterns embedded in subscribers' data [6]. Today, more than ever, different organizations and companies tend to identify customer patterns and behaviors [7]. The large amount of data generated by communication organizations is done in traditional and manual ways that are often not analyzable. This challenge is an important reason for using data mining techniques to analyze data. In fact, data mining enhances our understanding of our customers' future consumption forecasts as well as consumption management [6].

In this regard, feature selection methods are used to search for effective subsets to predict electricity consumption. Many solutions and algorithms have been proposed for the feature selection problem. The problem with some algorithms is their complexity and high computational load, but with the advent of fast computers and large storage resources, this problem is less common today. But on the other hand, the huge data sets in new issues have made it still important to find a fast algorithm for this. In addition, when there is a wide range of data with heterogeneous and scattered properties, the need for more accurate dimensional algorithms seems necessary.

Forecasting electricity consumption plays an important role in the economies of countries. Accurate forecasting of electricity consumption is essential for policymakers to formulate electricity supply strategies. However, limited data and variables usually do not provide enough information to obtain satisfactory forecast accuracy. To solve this problem, we propose a new model of power consumption forecasting using the Markov process, which also uses a combination of clustering and optimization algorithms. To evaluate the efficiency and simulation of the proposed method, the data of electricity subscribers in Bushehr-Iran Power Distribution Company are used.

The remaining of the paper is organized as follows: Section 2 is dedicated to related works. The details of the proposed method for forecasting power consumption are given in Section 3 and the discussion and experiments are given in Section 4. Finally, Section 5 conclusion and future work is given.

2. Literature Review

Numerous papers have been done in the field of identifying high-consumption subscribers as well as forecasting the consumption of subscribers in the electricity distribution company. In this section, some of the latest related paper in this field is analyzed.

In [8], a method was introduced to identify high-consumption electricity subscribers by data mining and optimization of nonlinear functions with information DE algorithm. In this paper, a power consumption pattern recognition system is implemented with the help of data mining tools and using differential evolution algorithm. Finally, three categories are considered for electricity consumption subscribers (normal, low consumption and high consumption).

In [9], electricity consumption management is analyzed using the revenue management approach. In this paper, we have tried to manage the demand and consumption of electricity and use the challenges of this industry by using the revenue management approach. For this purpose, by analyzing the compliance of prerequisites and Features of revenue management with the electricity industry, a conceptual model for implementing revenue management in this industry is presented. In this model, three factors of capacity, customer type, time are considered.

In [10], energy consumption management in smart power distribution network with data mining in advanced metering infrastructure is proposed. For this purpose, subscriber consumption time series are analyzed and using data mining algorithms and pattern recognition, subscribers with similar consumption behavior are placed in a category (cluster). Depending on the prevailing consumption behavior in each of these clusters and price signals at different hours of the day, an optimal consumption pattern can be suggested for each cluster. Also, with almost real-time monitoring provided by intelligent measurement technology, any changes in subscribers' consumption behavior can be detected. In this way, the effect of subscribers of different clusters on reducing energy consumption in the network and the benefit of each of them from reducing costs will be different as a result of modifying the consumption pattern.

In [11], electricity consumption forecasts for commercial and residential buildings have been made using deep recursive neural networks. With advances in sensors and intelligent technologies, there is a need for medium- and long-term forecasts of power consumption in residential and commercial buildings at hourly intervals for support. The main objectives of this paper are: (a) Development and optimization of new models of deep recursive neural network with the aim of predicting medium- and long-term electric charge with one-hour resolution. (B) Analyze the relative performance of the model for different types of power consumption patterns; and (c) Use a deep recursive neural network to perform calculations in a power consumption dataset.

In [12], China's electricity consumption forecast was presented using a new gray forecasting model. To this end, a new optimized gray prediction model, designed by combining a new initial condition and rolling mechanism, is based on the priority of new information. The previous initial conditions have inherent disadvantages of having a fixed structure and poor compatibility with raw data modification. To overcome these shortcomings, the new initial conditions have variable weight coefficients. Here, the parameters are determined using particle swarm optimization algorithm according to different features. In this paper, a new model is used to predict the future of China's total and

industrial electricity consumption from 2015 to 2020.

In [13], a new multivariate gray model is used to predict electricity consumption in Shandong Province in China. Here, the GMC model $(N, 1)$ is optimized by fractional order accumulation by changing the order number and the effectiveness of the first original data pair is proved by the model. The results of practical numerical examples show that this model offers significant forecasting performance compared to the traditional gray forecasting model. The predicted results show that the increase in electricity consumption in Shandong Province will accelerate.

In [14], power consumption forecasting is presented using a new hybrid model. In this paper, several machine learning methods including experimental mode analysis (EMD), support vector regression model (SVR) and particle swarm optimization (PSO) as well as thermal reaction dynamics theory and econometric model (AR-GARCH) for Power consumption bin is combined. The new hybrid model is known as the EMD-SVR-PSO-AR-GARCH. This is a new perspective on electricity consumption and consumer economic behavior. Here, power consumption data from the New South Wales market (NSW, Australia), the developed model for power consumption forecasting is used.

In [15], the LASSO-QRNN algorithm is presented. In this algorithm, power consumption prediction is proposed using a new hybrid model with power consumption probability density based on the LASSO quantum regression neural network. First, important features of external factors affecting the forecast of power consumption are extracted by LASSO regression. Then, the LASSO-QRNN model was developed to predict annual electricity consumption. The results of forecasting electricity consumption in different quantities in the next few years will be evaluated. Prediction accuracy is assessed through empirical analysis of the Guangdong Provincial Database in China and the California Database in the United States.

In [16], the HyFIS algorithm is presented. In this paper, power consumption prediction is performed using a method based on a combined neural fuzzy inference system. The HyFIS method considers two separate strategies, namely a strategy that uses only power consumption as input to the method, and the second strategy uses a combination of power consumption and ambient temperature as input. A case study with respect to the consumption forecast of an office building using this method is also presented. The results show that the second strategy is able to achieve better results.

3. Proposed Method

In this paper, feature selection, clustering and Markov process techniques are used to model and predict the power consumption data of subscribers. Initially, the data of the electricity subscribers is pre-processed in order to better predict the electricity consumption of the subscribers. In the next step, a subset of effective features is selected from the main feature set. This is based on a combined approach with PCA and the Firefly algorithm. In the next step, subscribers are clustered based on the features selected by the K-means algorithm. Subscriber clustering, then, calculates the risk of each cluster based on the actual sample label. In other words, clusters with more high-consumption users are considered high-risk clusters. The next step is to model the behavioral patterns of subscribers to predict consumption using the Markov process on high-risk clusters. This is done based on the subscriber consumption sequence. Finally, high-consumption subscribers are identified based on the predicted results of the Markov model. Here, the results of identifying high-consumption subscribers are evaluated for performance evaluation. Fig. 1 shows the flowchart of the proposed method.

3.1. Data Pre-Processing

In data mining, in order to be able to process the data properly, it is better to convert them in one hand with preprocessing, because the data is often collected from different sources [3]. Pre-processing in data mining consists of different stages. In this paper, three stages of purification, qualitative data conversion to quantitative and normalization are performed. In wiping, we encounter data with unrelated or missing values, which we delete in this paper. In the step of converting qualitative data to quantitative, all qualitative features are converted to quantitative by assigning distinct numeric codes. Finally, the "divide by maximum" method is used for the normalization work, where this method limits the value of all samples for a feature between 0 and 1.

3.2. Effective Features Selection

In order to reduce the size of the data, one of the most common methods is Principal Component Analysis (PCA). On the other hand, evolutionary approaches have many applications for selecting a subset of effective features. In this paper, a combined approach including PCA and Firefly algorithm for selecting/extracting optimal features is proposed.

The main purpose of PCA is to describe the changes in a set of correlated variables $x' = \{x_1, x_2, \ldots, x_q\}$ according to a new set of uncorrelated variables $y' = \{y_1, y_2, \ldots, y_q\}$, each of which is a combination are a line of variables x, is the new variables are considered in descending order of importance [17]. Thus, y_1 calculates the largest possible change in the initial data among all the linear compounds of x. Then y_2 is selected to calculate as much as possible the residual change so that it is unrelated to y_1. The new variables defined by this process, namely $y_1, y_2, \ldots, y_q$, are the main components.
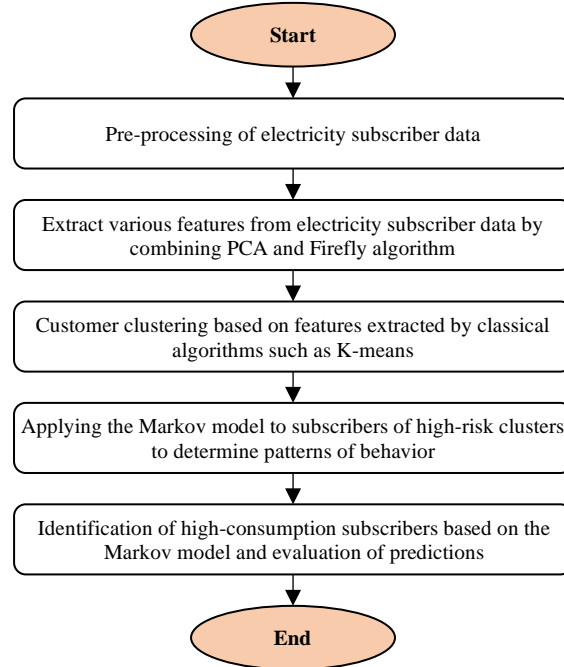
Fig.1. Flowchart of the proposed method

Therefore, the PCA method reduces the dimensions based on the determination of the main components in the data. The principal components are in fact the specific vector of the data covariance matrix. The maximum variance of the data is in the direction in which the eigenvector corresponds to the largest eigenvalue. Similarly, the smaller the eigenvalue, the smaller the variance of the data in the direction of the corresponding eigenvector. The PCA algorithm maps the original $n \times d$ dataset to a structure with a definite principal component q and to a new dataset of size $n \times q$. Here n refers to the number of instances, d to the number of dimensions of the main dataset, and q to the number of principal components. When the next reduction is done with PCA, a subset of the original dataset features is not selected for the new dataset. Because PCA is a feature extraction approach, principal components can be considered as extracted features.

Next, the features extracted from the PCA are combined with the main feature set of the dataset, from which a subset of effective features is selected by the Firefly algorithm. The details of this algorithm are as follows:

- *Solution representation structure:* Each solution is represented as an M-bit binary vector, which for each index 0 indicates no selection and 1 indicates the selection of the corresponding feature. Here, M is the sum of "total number of features" and "number of features extracted from PCA". For example, Fig. 2 shows a solution with $M = 10$ in which features 2, 5, 6 and 9 are selected. Therefore, the other features, 1, 3, 4, 7, 8, and 10, are not used for predictive work.

F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
0	1	0	0	1	1	0	0	1	0

Fig.2. An example of the solution structure in the Firefly algorithm

- *Initial population:* The initial population consists of N_p solutions that are created randomly.
- *Computing of Luciferin:* Luciferin expresses the fitness solution for each cream (solution). Here, the Luciferin of solution i is represented as $\ell(i)$ and is calculated based on a classical classification model such as SVM.
- *Neighborhood definition:* The sensor radius parameter (rs) is used to detect solution i neighbors. Here, rs is defined by the amount of luciferin in relation to Eq. (1).

$$f_j \in NG(f_i), \quad \forall \ell(j) > \ell(i) \tag{1}$$

Where, $NG(f_i)$ is the neighbor of solution i, and neighborhood means a higher amount of luciferin.

In addition, this algorithm is defined to control the motion of the decision radius (rd). Therefore, the neighbors of solution i are corrected according to Eq. (2), where the neighbor is limited to the maximum threshold of th due to the higher amount of luciferin.

$$f_j \in NG(f_i), \quad \forall \, \ell(j) > \ell(i) \; and \; \ell(j) - th \leq \ell(i) \tag{2}$$

- *Probability of movement:* The probability of motion of solution i relative to solution j is calculated as Eq. (3). Based on this probability, a solution is selected as a neighbor to move the solution i.

$$p_{ij} = \frac{\ell(j) - \ell(i)}{\sum_{f_k \in NG(f_i)} \ell(k) - \ell(i)} \quad j = 1,2,\ldots,N_p \tag{3}$$

- *Firefly movement:* Here, Eq. (4) is used to simulate movement to move solution i to solution j.

$$X_i(k) = \begin{cases} X_i(k) & \alpha > Rnd(0,1) \\ X_j(k) & Otherwise \end{cases}, k = 1,2,\ldots,M \tag{4}$$

Where, α is a control parameter to determine the movement step, $X_i(k)$ is the k-th position of the solution i and $X_j(k)$ is the k-th position of the solution j.
- *Luciferin value update:* The luciferin update for solution i is based on Eq. (5).

$$\ell_i(t) = (1 - \rho) \, \ell_i(t - 1) + \gamma \, J(X_i(t)) \tag{5}$$

Where, $\ell_i(t)$ and $\ell_i(t - 1)$ are the new and previous values of luciferin for solution i, respectively. $J(X_i(t))$ is the suitability value of solution i in iteration t. ρ is the control parameter for considering the gradual decrease of luciferin over time and γ is a constant value for the effect of fitness on luciferin.
- *Termination condition:* The steps of the Firefly algorithm are repeated up to a maximum of Max_{Iter}.

3.3. Subscriber clustering

At this stage, customers are clustered based on selected/extracted features. Among the existing clustering algorithms, one of the most common and simple is K-means, which is used in this paper. This algorithm divides a dataset into K clusters, where in this paper K is considered fixed and predetermined.

3.4. Calculate the risk of clusters

In the electricity subscriber database, specify the class label of each subscriber available. This label is 0 or 1 for each subscriber, which indicates its low consumption or high consumption, respectively. Therefore, the method of learning this paper is with the observer. Here, the risk of each cluster is calculated based on the actual label of each sample, as shown in Eq. (6). Finally, the highest risk cluster is used to extract behavioral patterns.

$$Risk(c_k) = \frac{\sum_{u \in c_k, l_u = 1} 1}{\sum_{u \in c_k} 1}, \quad \forall k = 1,2,\ldots,K \tag{6}$$

Where, c_k refers to the k-cluster and K refers to the total number of clusters. $u \in c_k$ the subscribers of the clusters c_k and l_u are the labels of subscriber u.

3.5. Identify patterns with the Markov model

In this paper, the Markov model is used to predict high-consumption electricity subscribers [1]. High-consumption subscribers are selected based on high-risk clusters. To track the behavioral patterns of high-consumption subscribers, their past consumption sequence is used. Here, subscriber consumption is categorized into five groups: "very low", "low", "medium", "high" and "very high" according to Table 1.

Table 1. Subscriber consumption classification

Category	Symbol	Transaction range
Very Low	VL	Less than 500 kWh
Low	L	Between 500 and 800 kWh
Medium	M	Between 800 and 1500 kWh
High	H	Between 1500 and 5000 kWh
Very High	VH	More than 5000 kWh

In fact, this grouping describes customers in "high risk", "medium risk", "normal risk", "low risk" and "no-risk" modes. Markov calculates and predicts the probability of the next consumption according to the subscriber consumption sequence of length k (Markov order). This is done based on the transfer probability matrix. The probability of transfer refers to the number of times the s_i state follows the s_j state. For higher levels, the following states can be presented in

the same way. In Eq. (7) the details of the transfer matrix are provided.

$$TL = \begin{bmatrix} tl_{1,1} & tl_{1,2} & \dots & tl_{1,j} & \dots & tl_{1,S} \\ tl_{2,1} & tl_{2,2} & \dots & tl_{2,j} & \dots & tl_{2,S} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ tl_{i,1} & tl_{i,2} & \dots & tl_{i,j} & \dots & tl_{i,S} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ tl_{S,1} & tl_{S,2} & \dots & tl_{S,j} & \dots & tl_{S,S} \end{bmatrix} \tag{7}$$

The TL matrix describes a Markov chain on a given state space (i.e., S). If the probability of transition from state i to state j at a given time step such as $p_{i,j} = tl_{i,j}$, the TL matrix uses $tl_{i,j}$ as the probability of transition from state i to state j is shown. Here, the total probability of transition from state i to other states is 1. This issue is presented in Eq. (8).

$$\sum_{j=1}^{S} tl_{i,j} = 1 \tag{8}$$

The transfer probability matrix is created when the order of the Markov model is determined. The input $t_{i,j}$ in the matrix is the probability of transfer in the sense of the number of times the state s_j is observed by the activity a_i. Fig. 3 shows an example of calculating the 1st and 2nd order probability matrices for five subscriber usage sequences. These uses are $\langle L, M, VL \rangle$, $\langle L, H, M, VL, VH \rangle$, $\langle VH, H, M, VL, H, VH \rangle$, $\langle L, VH, H, M, VL \rangle$ and $\langle VL, VH, M, H, VH \rangle$ are assumed.

1st Order	VL	L	M	H	VH
$S_1 = \{VL\}$	0	0	0	1	2
$S_2 = \{M\}$	4	0	0	1	0
$S_3 = \{L\}$	0	0	1	1	1
$S_4 = \{VH\}$	0	0	1	2	0
$S_5 = \{H\}$	0	0	3	0	2

2st Order	VL	L	M	H	VH	2st Order	VL	M	L	VH	H
$S_1 = \{VL, H\}$	0	0	0	0	1	$S_7 = \{L, H\}$	0	1	0	0	0
$S_2 = \{VL, VH\}$	0	0	1	0	0	$S_8 = \{L, VH\}$	0	0	0	0	1
$S_3 = \{M, VL\}$	0	0	0	1	1	$S_9 = \{VH, M\}$	0	0	0	0	1
$S_4 = \{M, H\}$	0	0	0	0	1	$S_{10} = \{VH, H\}$	0	2	0	0	0
$S_5 = \{VL, VH\}$	0	0	1	0	0	$S_{11} = \{H, M\}$	3	0	0	0	0
$S_6 = \{L, M\}$	1	0	0	0	0	$S_{12} = \{H, VH\}$	0	0	0	0	0

Fig.3. An example of a transfer probability matrix for the Markov model

After creating the transfer matrix, a new transaction is predicted for the subscribers. For each order of the Markov model corresponding to the transfer matrix, a transfer probability matrix is created. Finally, the probability is added based on the consumption of the same subscribers for all sequences and the amount of consumption is considered with the highest probability. Therefore, in this paper, the largest order of Markov model that can cover all sequences is used. In this regard, first the matrix is created with orders of 1 to k. Then, with the entry of a subscriber consumption, all sequences from 1 to k are calculated based on the last k of consumption. In the next step, the probability of transfer losses of each state is extracted from the transfer matrix, and finally the total probability for similar states for all sequences is considered. Eq. (9) shows how the probability is summed up for similar sequences for all sequences.

$$p(a_{n+1}|a_n a_{n-1} \dots a_{n-k}) = p(a_{n+1}|a_n) + p(a_{n+1}|a_n a_{n-1}) + \dots + p(a_{n+1}|a_n a_{n-1} \dots a_{n-k}) \tag{9}$$

Where, n is the number of levels and k is the number of modes examined in a subscriber consumption.

4. Results and Discussion

In this section, we conduct an experimental study to evaluate the performance of the proposed method. All experiments are carried out on an Intel Core i7 CPU at 3.2GHz and 16GB of RAM and Windows 10 64-bit operating system. In addition, MATLAB R2019a is used for implementation [18]. In order to achieve more reliable results, all experiments have been reported by 10-fold cross validation [19]. In this paper, the real data of a number of subscribers

of Bushehr-Iran Electricity Distribution Company have been used for simulation work. Available data includes 263860 samples and 11 features. Examples include electricity subscribers and features include region, computer code, year, period, course start date, end date, number of days, medium load consumption, high load consumption, low load consumption and total consumption. In addition, the last column of this dataset refers to the subscriber label. The proposed method consists of several parameters whose values are adjusted based on the literature reference or trial and error [20], where, $N_p = 75$, $th = 0.15$, $\alpha = 0.4$, $\rho = 0.25$, $\gamma = 0.2$, $Max_{Iter} = 100$, and $K = 2\ to\ 10$.

In this paper, the silhouette index is used to evaluate the results of the K-means clustering algorithm. This index is a combination of intra-cluster and inter-cluster similarity [2], as shown in Eq. (10).

$$S_i = \frac{b_i - a_i}{\max{(b_i, a_i)}} \tag{10}$$

Where, a_i represents the mean distance between observation i and other observations in a similar cluster, and b_i represents the mean distance of observation i to all observations in other clusters.

The results of the silhouette index with the number of different clusters from 2 to 10 are examined in Fig. 4. The results show that the best performance of the proposed method is obtained with 7 clusters and a silhouette index of 0.758.
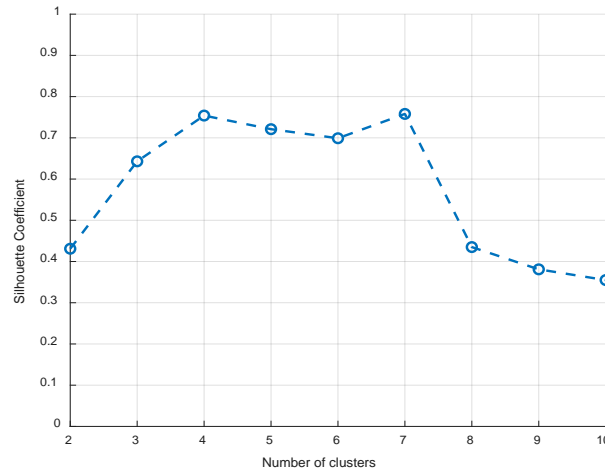

Fig.4. Comparison of the silhouette index of K-means with the number of different clusters

In this paper, effective features were selected/extracted using a hybrid algorithm. The proposed algorithm, in addition to selecting the features, also determines the optimal number of them. Fig. 5 shows the prediction accuracy for each number of selected feature subsets. The results reported the best accuracy with 7 features and 98.71% accuracy.
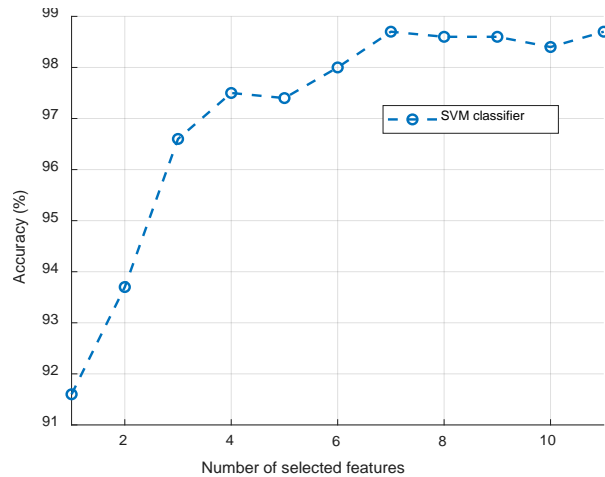

Fig.5. Prediction accuracy of the proposed method with a number of different features
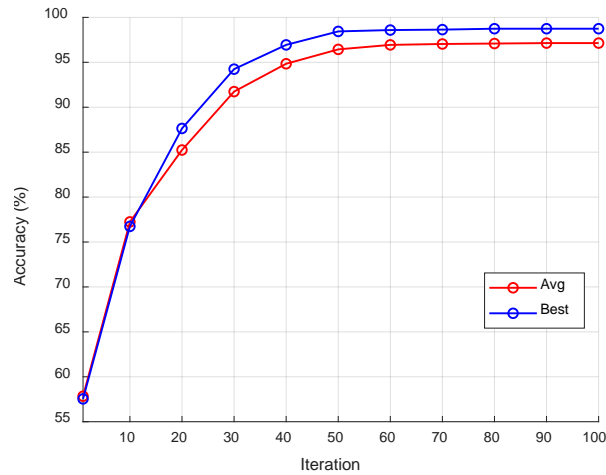
Fig.6. Convergence of the proposed method for selecting the optimal feature subset

In another experiment, the convergence of the Firefly algorithm was investigated to find the optimal feature subset. The results of the accuracy for the proposed method in iterations 1 to 100 are reported in Fig. 6. These results are calculated and shown in two modes of mean and best accuracy during iteration. The results show that the proposed method converges in 50 iterations with an accuracy of about 98% for the best case and an accuracy of about 97% for the average case, after which no significant change in accuracy is obtained.

Regarding the use of PCA method for extracting optimal features and Firefly algorithm for selecting features, in Table 2, the features used for modeling are reported. Here, the first 5 features are selected from the original dataset and the other 2 features are extracted by PCA. Therefore, a total of 7 features have been used by the proposed method.

Table 2. Features of selection/extraction by the proposed method

Row	Feature number	Feature name	Selected
1	7	number of days	Original dataset
2	8	Intermediate consumption	Original dataset
3	9	Fertile consumption	Original dataset
4	10	Low load consumption	Original dataset
5	11	Total consumption	Original dataset
6	-	-	By PCA
7	-	-	By PCA

In the following, the simulation results of the proposed method are compared with LASSO-QRNN and HyFIS methods. The results of this comparison are reported based on different evaluation criteria in Table 3. The results of these comparisons showed that the proposed method is relatively superior to the compared methods. Here, the proposed method has the best performance with an average accuracy of 98.71% compared to the other two methods. After the proposed method, LASSO-QRNN and HyFIS methods with 98.03% and 98.31% accuracy, respectively. The rate of improvement of the proposed method compared to LASSO-QRNN is about 1% and compared to HyFIS method is about 0.5%.

Table 3. Comparison of the proposed method with other algorithms

Methods	Number of features used	Accuracy	F-measure	Run-time
LASSO-QRNN	9	98.03	97.95	1024
HyFIS	9	98.31	98.17	881
Proposed method	7	98.71	98.36	852

Due to the lack of hardware facilities and limited data available, the proposed method is a simulation for quantitative data. However, according to the simulation results, it can be predicted that the results will be similar for big data and real system.

5. Conclusion

Electricity is a very important but scarce resource that is available to humans. To avoid wasting and unnecessary use of energy, intelligent solutions are needed to ensure energy efficiency is optimized. Given the problem of electrical

energy storage, it is critical to predict the amount of load required in order to have a reliable and stable power distribution network. Therefore, forecasting the electricity consumption of subscribers and analyzing their consumption behavior under the influence of various factors and time variables is important. In this paper, feature selection techniques, clustering and Markov process are used to model and predict the power consumption data of subscribers. The results clearly prove the efficiency and effectiveness of the proposed method for predicting power consumption. Due to the complexity of the issue of high-consumption subscriber forecasting, this study uses a clustering approach to identify subscriber associations based on past consumption Features. This makes the search space smaller and modeling can be applied to smaller segments of the subscribers. To evaluate the superiority of the proposed method, extensive experiments were performed based on the data of electricity subscribers in Bushehr-Iran Power Distribution Company. The proposed method has achieved 98.71% prediction accuracy with only 7 features. Here, k-means is used for clustering. Since this algorithm depends on the number of clusters, it is recommended to use other clustering methods such as Louvain for future work.

References

[1]   Munkhammar, J., van der Meer, D., & Widén, J. (2021). Very short-term load forecasting of residential electricity consumption using the Markov-chain mixture distribution (MCM) model. *Applied Energy*, 282, 116180.

[2]   Kaytez, F., Taplamacioglu, M. C., Cam, E., & Hardalac, F. (2015). Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. *International Journal of Electrical Power & Energy Systems*,67, 431-438.

[3]   Rezaeipanah, A., Mojarad, M., & Fakhari, A. (2020). Providing a new approach to increase fault tolerance in cloud computing using fuzzy logic. *International Journal of Computers and Applications*, 1-9.

[4]   Bianco, V., Manca, O., & Nardini, S. (2009). Electricity consumption forecasting in Italy using linear regression models. *Energy*, 34(9), 1413-1421.

[5]   Rezaeipanah, A., Matoori, S. S., & Ahmadi, G. (2021). A hybrid algorithm for the university course timetabling problem using the improved parallel genetic algorithm and local search. *Applied Intelligence*, 51(1), 467-492.

[6]   Mohamed, Z., & Bodger, P. (2005). Forecasting electricity consumption in New Zealand using economic and demographic variables. *Energy*, 30(10), 1833-1843.

[7]   Devaine, M., Gaillard, P., Goude, Y., & Stoltz, G. (2013). Forecasting electricity consumption by aggregating specialized experts. *Machine Learning*, 90(2), 231-260.

[8]   Eseye, A. T., Lehtonen, M., Tukia, T., Uimonen, S., & Millar, R. J. (2019, July). Short-term forecasting of electricity consumption in buildings for efficient and optimal distributed energy management. In *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)* (Vol. 1, pp. 1103-1110). IEEE.

[9]   Rakhmonov, I., Berdishev, A., Niyozov, N., Muratov, A., & Khaliknazarov, U. (2020, July). Development of a scheme for generating the predicted value of specific electricity consumption. *In IOP Conference Series: Materials Science and Engineering* (Vol. 883, No. 1, p. 012103). IOP Publishing.

[10] Ye, Z., & Kim, M. K. (2018). Predicting electricity consumption in a building using an optimized back-propagation and Levenberg–Marquardt back-propagation neural network: Case study of a shopping mall in China. *Sustainable Cities and Society*, 42, 176-183.

[11] Rahman, A., Srikumar, V., & Smith, A. D. (2018). Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Applied energy*, 212, 372-385.

[12] Ding, S., Hipel, K. W., & Dang, Y. G. (2018). Forecasting China's electricity consumption using a new grey prediction model. *Energy*, 149, 314-328.

[13] Wu, L., Gao, X., Xiao, Y., Yang, Y., & Chen, X. (2018). Using a novel multi-variable grey model to forecast the electricity consumption of Shandong Province in China. *Energy*, 157, 327-335.

[14] Fan, G. F., Wei, X., Li, Y. T., & Hong, W. C. (2020). Forecasting electricity consumption using a novel hybrid model. *Sustainable Cities and Soci*ety, 61, 102320.

[15] He, Y., Qin, Y., Wang, S., Wang, X., & Wang, C. (2019). Electricity consumption probability density forecasting method based on LASSO-Quantile Regression Neural Network. *Applied energy*, 233, 565-575.

[16] Tang, L., Wang, X., Wang, X., Shao, C., Liu, S., & Tian, S. (2019). Long-term electricity consumption forecasting based on expert prediction and fuzzy Bayesian theory. *Energy*, 167, 1144-1154.

[17] Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4), 433-459.

[18] Amit Pandey, Achin Jain, "Comparative Analysis of KNN Algorithm using Various Normalization Techniques", International Journal of Computer Network and Information Security, Vol.9, No.11, pp.36-42, 2017.

[19] Tran Son Hai, Le Hoang Thai, Nguyen Thanh Thuy, "Facial Expression Classification Using Artificial Neural Network and K-Nearest Neighbor", International Journal of Information Technology and Computer Science, vol.7, no.3, pp.27-32, 2015.

[20] Saptarsi Goswami, Amlan Chakrabarti, "Feature Selection: A Practitioner View", International Journal of Information Technology and Computer Science, vol.6, no.11, pp.66-77, 2014.

Authors' Profiles

Hadis Dalkani received his B.E. degree in computer software engineering from Kherad Institute of Higher Education Bushehr, Iran, and Department of Computer Engineering in 2018, and has received his M.Sc. degree in computer software engineering from of Lian Bushehr Institute, Iran, in 2021. His hobbies are Data Visualization, Data Science, Data Analysis, Web Development and Object-Oriented Programming.

Mousa Mojarad received his PhD in Computer-Software Engineering in 2020. He is currently a lecturer and faculty member of the Islamic Azad University, Firoozabad Branch. His hobbies are Big Data, Face Recognition, Machine Learning, Pattern Recognition and Feature Extraction.

Hassan Arfaeinia obtained her B.E in Computer Science from Rafsanjan University, Kerman, Iran in 2009. HE received her M.S in Computer Engineering from the Amirkabir University of Technology, Tehran, Iran in 2011 and PhD in Computer Engineering from Islamic Azad University, North Tehran Branch, Iran in 2016. His main paper interests consist of Software Development, Machine Learning, Statistical Modeling and Computer Programming.