Journal of Ambient Intelligence and Humanized Computing (2022) 13:5253–5267
https://doi.org/10.1007/s12652-021-03444-x

ORIGINAL RESEARCH

A short-term power load forecasting method based on k-means and SVM

Xia Dong[1] · Song Deng[1] · Dong Wang[2]

Received: 3 December 2020 / Accepted: 5 August 2021 / Published online: 30 August 2021

Abstract

With the continuous development of smart grids, short-term power load forecasting has become increasingly important in the operation of power markets and demand-side management. In order to explore the influence of temperature and holidays on seasonal loads, this paper proposes a short-term SVM power load forecasting method based on K-Means clustering. The method includes the steps of selecting similar days, data preprocessing, SVM prediction model training and parameter adjustment. Among them, the selection of similar days uses K-Means to group seasonal load data into two categories according to temperature characteristics, as the input data to explore the effect of temperature on seasonal load. And divide the data into holidays and working days as the model input data to discover the impact of holidays on seasonal loads by using calendar rules. In order to verify the load forecasting effect of the proposed method, several experiments were carried out on two actual residential load data and two data online, and compared with the LSTM and decision tree load forecasting models in terms of prediction accuracy evaluation index and running time. The results show that the model constructed in this paper has 39.75% improved to the conventional methods for the accuracy and 128.89% improved for the running time.

Keywords Short-term load forecasting · SVM · Seasonal load · K-Means

1 Introduction

In recent years, under the background of power market reform and smart grid construction, the development of smart power grid has promoted the popularization of smart meters on the user side, and the large-scale deployment of various monitoring systems has enabled grid companies to obtain multi-scale and comprehensive users' power consumption information. These large and high-resolution user load data can be applied not only to describe the user's power consumption habits, but also to predict the user's power load (Friedrich and Afshari 2015; Lin et al. 2019; Lu et al. 2020b). With the development of the power industry, the accuracy of power system load prediction becomes particularly important. Mining power load data is of great value for power grid system scheduling optimization, refined management and service to market users (Bozkurt et al. 2017; Lee and Hong 2015; Zhao and Guo 2016). Therefore, load forecasting has become an important research area in power grid operation and management.

Electric load forecasting mainly analyzes its historical data on the basis of considering the influencing factors of electric load, obtains useful information and then establishes a mathematical model, so as to realize the estimation of the future development trend of electric energy and electricity consumption (Hafeez et al. 2020; Haben et al. 2019). Cognitive computing represents a new computing model, which includes a large number of technological innovations in the fields of information analysis, natural language processing and machine learning, which can help decision makers reveal extraordinary insights from large amounts of unstructured data. However, due to the complex randomness of electricity loads, real-time load monitoring and prediction is still a challenging task in the smart grid (Welikala et al. 2017).

Combined with the analysis of electricity consumption behavior, there's a certain relationship between the power load curve and time, and its regular fluctuations provide a research idea for load forecasting. We can try one of the

✉ Song Deng
dengsong@njupt.edu.cn

1 Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

2 State Grid Electric Power Supply Company, Nantong 226000, Jiangsu, China

methods to apply to the field of load forecasting, such as the block diagram shown in Fig. 1 of this paper. This paper proposes a short-term SVM power load forecasting method based on K-Means clustering to establish a load forecasting model (STLF-SK).

The model takes advantage of the SVM algorithm's continuous approximation ability in nonlinear fitting, combined with the K-Means clustering algorithm to consider the characteristics, and constantly looking for the optimal parameters to fit model. This paper also selects two comparison algorithms, decision tree in machine learning and Long and Short Term Memory neural network (LSTM), and compares the prediction accuracy statistical indicators and running time of three algorithms. In summary, the contributions of our research are shown as follows:

- In order to study the influence of temperature and day type on load forecasting, the K-Means is used to cluster the historical load data with the influence characteristics, and the data is divided into several data sets for experiments.
- In consideration of exploring the nonlinear relationship between variables, this paper proposes a cluster-based SVM load forecasting model based on similar days.
- The experimental results of the four stations prove that the accuracy of load prediction based on STLF-SK is more advantageous than some algorithms, in terms of prediction accuracy evaluation index and running time.

The remainder of this paper is organized as follows. Section 2 focuses on introducing related work about short-term load forecasting. Similar day clustering based on K-Means algorithm is presented in Sect. 3. Section 4 proposes load forecasting model based on SVM. Section 5 is simulation and experiment analysis and Sect. 6 is conclusion.

2 Related work

Power load forecasting is one of the important tasks of power system dispatching, power utilization, planning and other management departments. Improving the technical level of load forecasting is conducive to improving the economic and social benefits of the power system, and is of great significance to social development and national stability. According to the characteristics of time series and non-linearity of power load data, short-term load forecasting models can generally be divided into two categories: one is the time series method, which regards historical load data as a time series (Zahid et al. 2019). Commonly used methods include: regression analysis method, pattern recognition method, autoregressive integral moving average model and so on. These traditional time series methods have high requirements on the stability of historical data over time, emphasizing the fitting of historical data (Lu et al. 2019; Herui and Xu 2015). The other is a new type of intelligent method that has emerged with the development of artificial intelligence, including

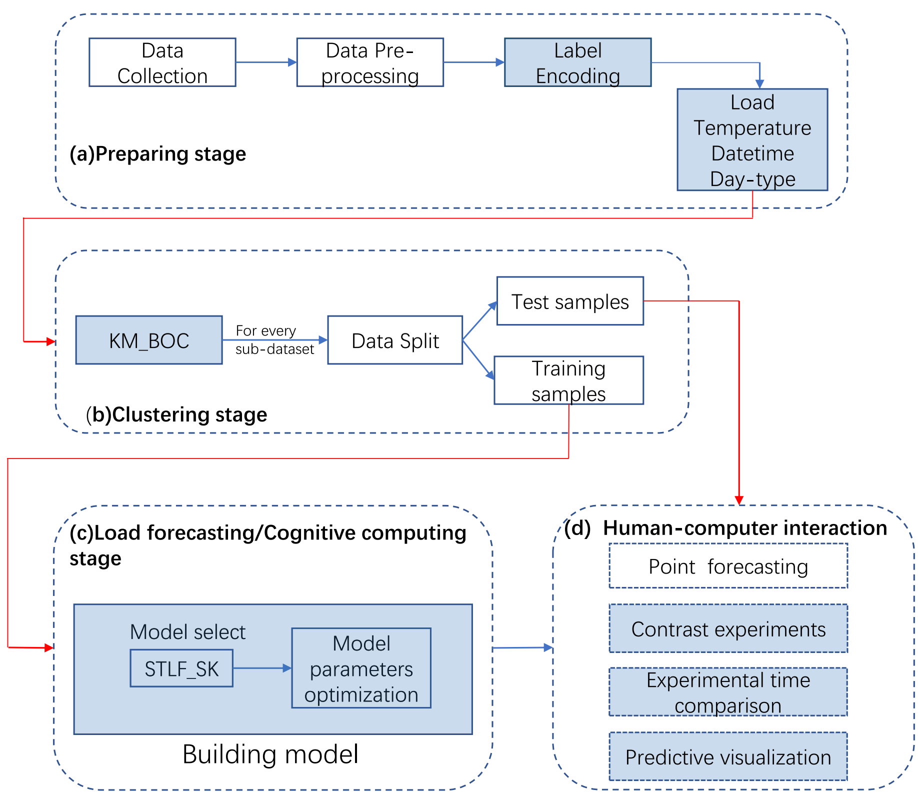


Fig. 1 Block diagram of the proposed forecasting method framework

artificial neural network (Lu et al. 2020a; Wang et al. 2017; Yu and Xu 2014), machine learning such as random forests and support vector machines (Lee and Lin 2017; Vrablecová et al. 2018), time-recurrent neural networks in deep learning methods (Ryu et al. 2017) and other data mining and big data techniques (Zhang et al. 2015) and some other artificial intelligence or brain intelligence technologies (Lu et al. 2018). These methods are widely used in nonlinear regression estimation problems due to their excellent nonlinear fitting ability. Wang et al. (2018) tried regression tree-based models: classification and regression trees, bagging and random forests to identify the variables dominating the marginal price of the commodity as well as for short-term (1 h and day ahead) electricity price forecasting for the Spanish-Iberian market. Ni et al. (2017) combined wavelet changes and extreme learning machines to propose an integrated prediction model. Zhang et al. (2019) proposed a new improved RBF neural network model and used VMD-WT to extract features and removed noise of wind speed data aiming to make accurate short-term wind power forecasting, but there were limitations of model parameter selection depending on previous experience. In 2018, Xia et al. (2018) deployed the combination of wavelet analysis and artificial intelligence machine learning to improve the self learning ability and prediction accuracy, the simulation results showed that the result have better performance. Kong et al. (2017) tried to address the short-term load forecasting problem for individual residential households with a density based clustering technique to evaluate and compare the inconsistency. Due to LSTM's the excellent learning ability of the long-term temporal connections (Muzaffar and Afshari 2019), the result proved to perform better. Although these prediction methods are more effective than time series, regression analysis and other methods in predicting accuracy, they ignore the influence of temperature and holidays on the regional power load within the season range of a single region.

3 Similar day clustering based on K-Means algorithm

Before model training, the accuracy of load forecasting can be effectively improved by selecting historical data that is similar to the temperature conditions on the day to be predicted and the attributes of working days and holidays (Xiao et al. 2015). The essence of selecting historical load data is to select similar characteristics of the load data of the day to be tested. Considering the influence of temperature characteristics on electricity load (Haben et al. 2019), this paper takes the daily weather temperature as the characteristic and adopts the K-Means clustering method to select similar days.

The K-Means algorithm is an unsupervised learning method. The algorithm categorizes the neighboring points through the set center point, and iteratively updates, the value of the cluster center one by one until the best clustering effect is obtained (Huang et al. 2020; Lei et al. 2019). For processing large data sets, K-Means clustering algorithm has high scalability and scalability, so it is widely used. The algorithm based on characteristics (KM_BOC) is described as follows:

Algorithm 1: KM_BOC

Input: Dataset D = $\{x_1, x_2, ..., x_m\}$,k;
Output: $C = \{c_1, c_2, ..., c_k\}$;
Initialize k cluster centroids $\{u_1, u_2, ..., u_k\}$ randomly from D;
repeat
$C_i \neq \emptyset, (1 \leq i \leq k)$
for $j = 1, 2, 3, \ldots, m$ do
  For each x_j, compute the distance to the nearest centroid $u_i (1 \leq i \leq k)$;
  Determine cluster markers base on $\lambda_j = \underset{i \in \{1,2,\ldots,k\}}{\arg\min} ||x_j - u_i||^2$;
  add x_j to the set of centroids:$C_{\lambda_j} = C_{\lambda_j} \cup \{x_j\}$;
end for
for $i = 1, 2, \ldots, k$ do
  Compute the new centroids $u_i{'}$;
  if $u_i{'} \neq u_i$ then
    Update the current u_i to $u_i{'}$;
  else
    Keep the current u_i ;
end for
until the centroids dont change;

4 Load forecasting model based on SVM

SVM is a binary classification model. Its basic model is a linear classifier with the largest interval defined in the feature space. SVM also includes kernel techniques, making it a substantially non-linear classifier. SVM was originally used to solve the problem of pattern recognition, the purpose is to discover decision rules with good generalization performance (Barman et al. 2018).

Solving the regression problem based on SVM is called support vector regression (SVR). Suppose the training data is $D=\left\{\left(x_1,y_1\right),\left(x_2,y_2\right)\ldots\left(x_m,y_m\right)\right\}$, where $y_i \in R$, and the regression model $f(x)=w^T x+b$ based on SVR makes f_x and y as close as possible, where w and b are model parameters.

At the same time, in order to better use SVR for sample data fitting, the choice of kernel function becomes a key issue. In the load forecasting model, the support vector regression RBF function is selected to fit the training sample. The calculation formula of the kernel function is $\kappa(x_i,x_j)=\exp(-\frac{\|x_i-x_j\|^2}{2\sigma^2})$, where $\sigma>0$ is the bandwidth of the kernel, *x_i* and *x_j* are sample data:

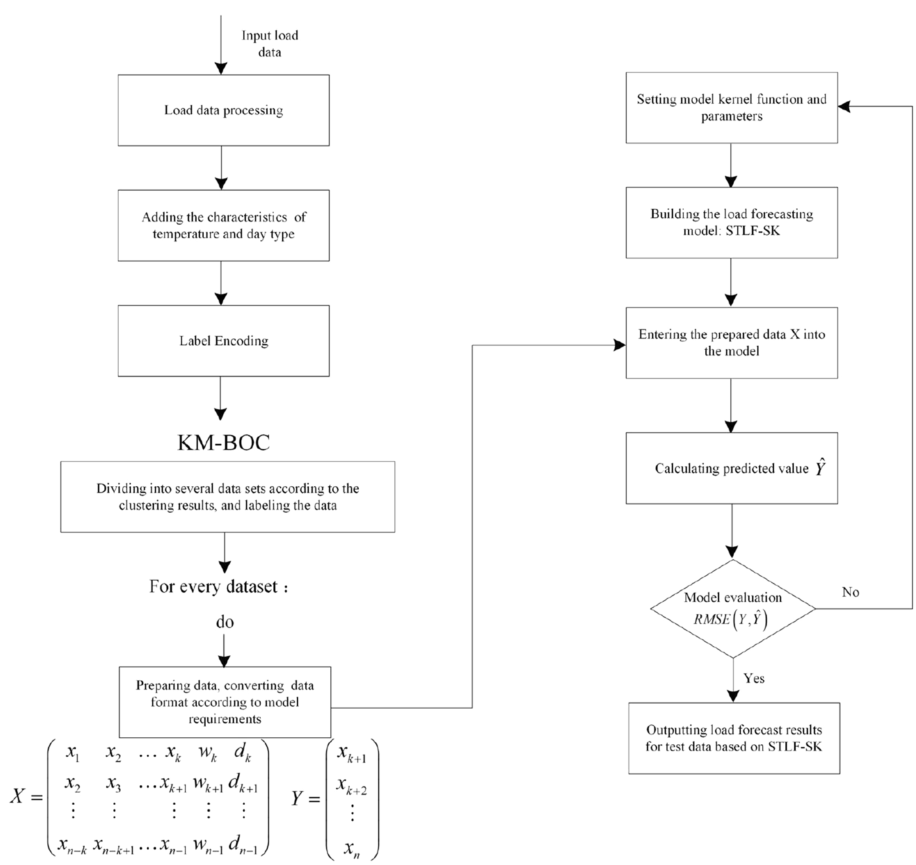
The input data x in the experiment is a one-dimensional vector, $\hat{y}$ represents the actual load data at the time to be predicted. x and $\hat{y}$ constitute the training sample $\{x,\hat{y}\}$ that will be input into the model. Map the sample from the original space to the high-dimensional linear space, so that the sample is linearly separable in the feature space selected by the kernel function. The framework of STLF-SK is shown in Fig. 2.

Algorithm 2: STLF-SK

Input: Dataset , k, C, $\kappa(x_i,x_j)$;
Output: SVM model ;
1. Load data processing;
2. $\{data, temperature, daytype\} \leftarrow$ Adding the characteristics;
3. $\{\mathrm{D}_1,\mathrm{D}_2,\ldots,\mathrm{D}_k\} \leftarrow$ KM-BOC;
/*Getting several small datasets with different labels ;*/
4. Converting data format;
5. For each D_i, Establishing SVM load forecasting model;
6. Determining model parameter selection $C\&\kappa(x_i,x_j)$;
7. Model evaluation;
8. Return SVM model;
9. End for
10. End

In this paper, before building the model, the original data need to be preprocessed, add features then encode them. After this step, the data is input to the KM_BOC algorithm, and the output data is divided into subsets by its label. The

Fig. 2 Framework of STLF-SK



data format of each subset is transformed, and the data is prepared according to the prediction model data format. Enter the test set into the model, evaluate the model, and adjust the model parameters. The load forecasting model $STLF_SK$ is established.

5 Simulation and experiment analysis

5.1 Data description and data preprocessing

In order to verify the validity and applicability of STLF-SK algorithm, load data from four different regions were selected for experiments in this paper. Two district load data from different residential areas of Nantong Power Supply Company of State Grid. Experiments for District 3 are based on EUNITE's 2001 power Load Forecasting competition data from eastern Slovakia Power company's 1998–1999 load data at http://www.eunite.org/. The last collection of dataset is from the project entitled Personalised Retrofit Decision Support Tools for UK Homes using Smart Home Technology (REFIT) at https://pureportal.strath.ac.uk/en/datasets/refit-electrical-load-measurements. Before performing model training and prediction, this experiment first cleans the collected raw load data, including missing value and outlier check. And considering the influence of temperature and holidays, it's necessary to add these features in the original data.

The data after data preprocessing is divided into two parts, 80% is the training sample set, used for the algorithm model training and parameter adjustment proposed in this experiment; 20% is the test sample set, used for the prediction accuracy verification of the algorithm model. The load data set in each season is divided again according to the clustering results and holidays as the data set of the load prediction model. In this article, divide the data set as shown in Table 1.

Data description The historical data are divided into three sub-data sets according to their seasonal attributes, among which summer load data experiment is selected. Each data set is clustered into two types according to temperature through the K-Means clustering algorithm, and labeled with 0 and 1; The data set treats statutory holidays and two-day holidays as holidays according to calendar rules, with a label of 2, and others are treated as working days with a label of 3.

Table 1 Experimental dataset

District	Data description	Datasize/day	Labels	Samples/day
1	Summer load data	51	0	21
			1	30
			2	15
			3	36
2	Winter load data	81	0	27
			1	54
	Spring and Fall load data	119	2	36
			3	83
3	1998/1/1–1999/1/31	396	1	204
			0	192
4	2013/11/1–2015/5	563	2	195
			3	368

5.2 Evaluation indicators

In order to directly and effectively evaluate the prediction accuracy of the model and compare with other methods. In this experiment, $Max\_E = \max \sum_{i=1}^{N} |y_i - \hat{y}_i|$, $Min\_E = \min \sum_{i=1}^{N} |y_i - \hat{y}_i|$, $MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$, $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$ and $MAE = \frac{1}{N} \sum_{i=1}^{N} |(y_i - \hat{y}_i)|$ were selected as evaluation indexes, where N represents the number of samples, y represents the actual load value at the time point, and $\hat{y}$ represents the predicted value at the corresponding time point of the model output.

5.3 Experimental analysis

To predict a specific day of a residential radio zone for a resident of Nantong Power Supply Branch of State Grid Jiangsu Electric Power Co., Ltd., the data was tested using different divided Datasets. Under the premise that the temperature to be predicted is known, every 15 min step size and use the SVM algorithm proposed in this paper to predict the electric load by the k-means algorithm clustered load data, comparing with other two algorithms LSTM and decision tree (DT). The three algorithms compare the predicted results of the actual load curve and the average load curve on the day to be predicted (Figs. 3, 4).

5.3.1 Experiments on District 1

1. *1. Summer overall load forecast* The experiment is based on District 1 for load forecasting. There are 96 load collection time points per day. All models take 80% of all data as a training set for predictive model training and learning. Select September 19, 2019 as the forecast day. The load forecast curve is shown in the following Fig. 3 and Num.E1 in Table 2.

   From the above chart, it can be seen that the three load forecasting models can completely output the forecast load data value and change curve of the whole day on September 19, 2019. The result is consistent with the actual load curve and change trend of the average load

Table 2 Load forecast accuracy index of District 1

Serial number	Model	Max_E	Min_E	MSE	RMSE	MAE
E1	LSTM	42.6919	0.0095	174.8680	13.2237	10.2508
	SVM	**29.2291**	0.2287	**152.0179**	**12.3295**	**9.1093**
	DT	30.4226	0.2304	181.4274	13.4694	11.0755
E2	LSTM	95.5137	0.1434	1193.3058	34.5442	27.6721
	SVM	**55.3093**	**0.0688**	**457.7369**	**21.3947**	**16.5174**
	DT	64.12	1.0240	685.4150	26.1804	20.6391
E3	LSTM	42.9842	0.1354	178.8644	13.3740	10.2505
	SVM	**34.5121**	0.0878	**140.1832**	**11.8399**	**9.2623**
	DT	47.364	0.0388	208.9618	14.3482	11.5669
E4	LSTM	58.6789	1.7999	697.377	26.4078	23.1208
	SVM	**44.8390**	**0.08200**	**366.1865**	**19.1360**	**15.3416**
	DT	85.284	0.2719	972.0590	31.1778	25.5822
E5	LSTM	62.1805	0.0734	388.7386	19.7164	14.3351
	SVM	**35.4379**	**0.1177**	**204.0858**	**14.6157**	**11.5113**
	DT	50.5973	0.5760	256.4050	16.0126	12.5325

Bold values are the best in the experimental results based on the algorithm proposed in this paper

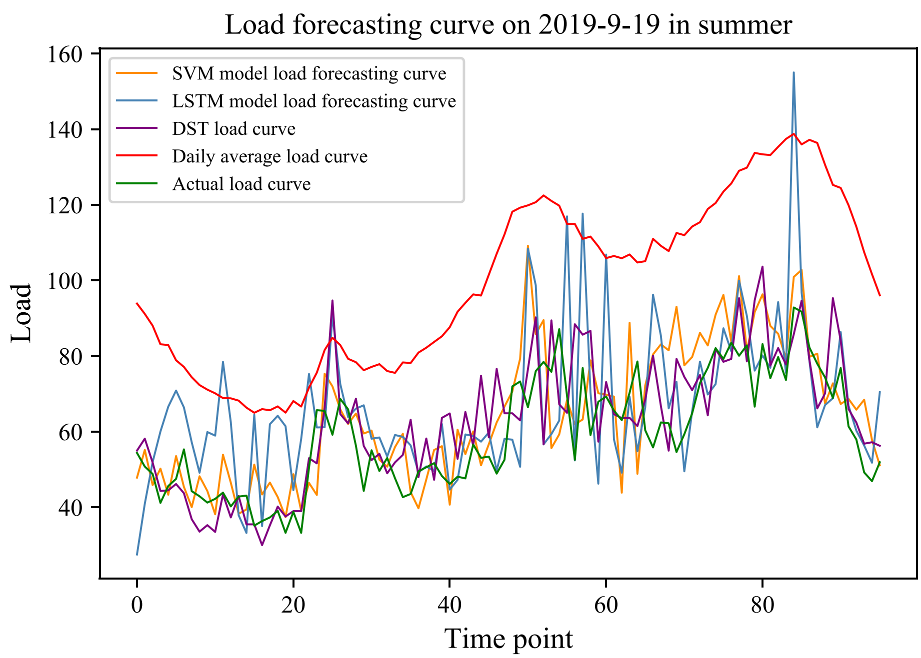


Fig. 3 Load forecasting curve on 2019-9-19 in summer

curve. The SVM prediction model not only has the highest consistency between the prediction result curve and the results of the sister, but also in terms of the statistical indicators of the prediction.

2. *2. Influence of summer temperature on load forecast* This experiment makes load prediction based on the temperature characteristics under Dataset 1. Considering the influence of temperature on the load in the season, the load data value and temperature of 51 days in summer are recorded in a data table. Based on the clustering result, the temperature of the day to be predicted is known, and the same type of data is selected, that is, the similar day is selected to perform the load forecast again. To study the effect of high temperature on the load in summer, select the data with label 0 in the Dataset 1 for the experiment, and select the day of August 8–26 as the day to be predicted. The prediction results are as following Fig. 4 and Num.E2 in Table 2.

The chart can be intuitively reflected: proposed models can also roughly predict the daily load curve of 2019-8-26 under the summer high temperature data set. Both the actual load and the SVM predicted load curve on the day can roughly fit the high-temperature average load curve in the trend, but although the predicted result of the LSTM and

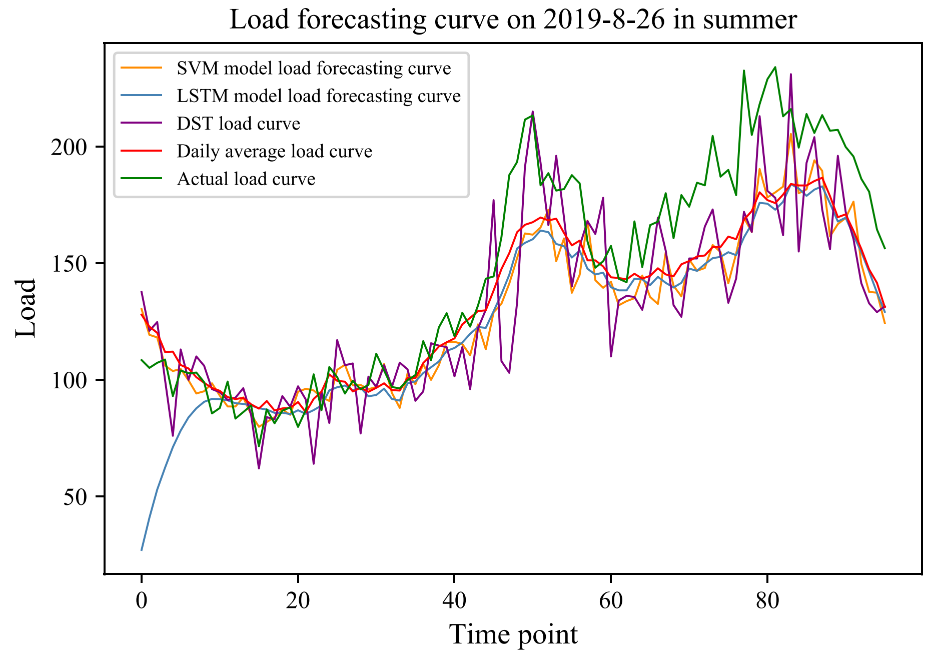


Fig. 4 Load forecasting curve on 2019-8-26 in summer

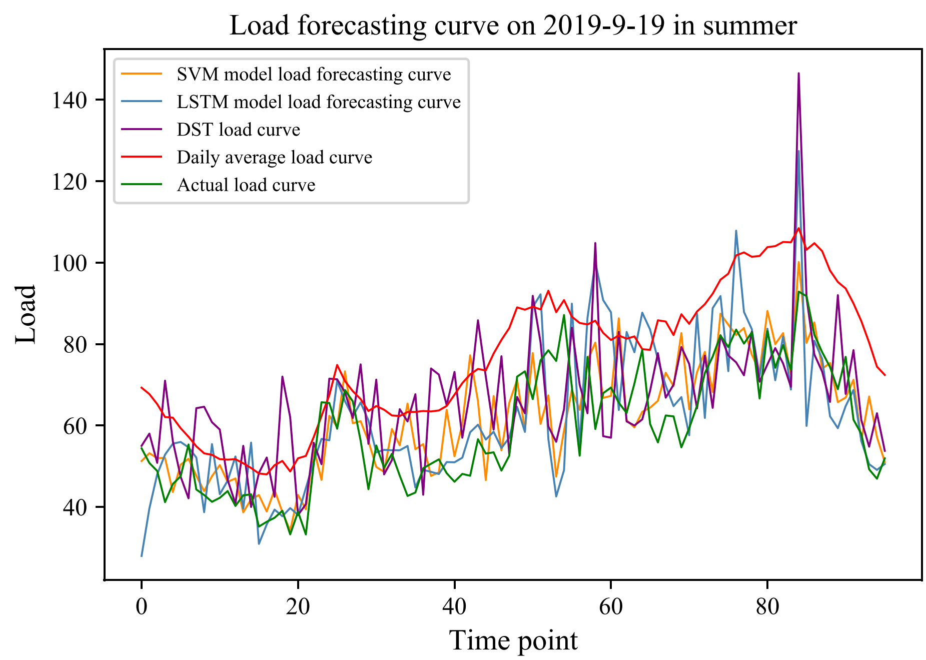


Fig. 5 Load forecasting curve on 2019-9-19 in summer

DT model can reflect the load to a certain extent. So select the data with the label of 1 to re-predict the day, the results are as following Fig. 5 and Num.E3 in Table 2.

Compared with the whole summer data, the accuracy of the SVM prediction model is slightly influenced by the size of dataset when considering the influence of temperature characteristics on the load, but the SVM model still performed best among these models.

3. Effects of summer holiday on load forecasting In order to distinguish the impact of holidays and working days on the daily load curve, all historical sample data in Dataset 1 are labeled as 2 or 3, respectively. Under the premise of more detailed division, the experiment proves that the holiday affects the load. Extract the data with the label 2 in the Dataset 1 for the experiment, select 2019-9-14 as the prediction day. The results are shown in Fig. 6 and Num. E4 in Table 2.

For load forecasting in summer holidays, LSTM , SVM and DT load forecasting models can fully predict the load at 96 time points a day. From the chart, we can see that all the load curves have a high consistency in the load value and LSTM and SVM models are close to the actual conforming

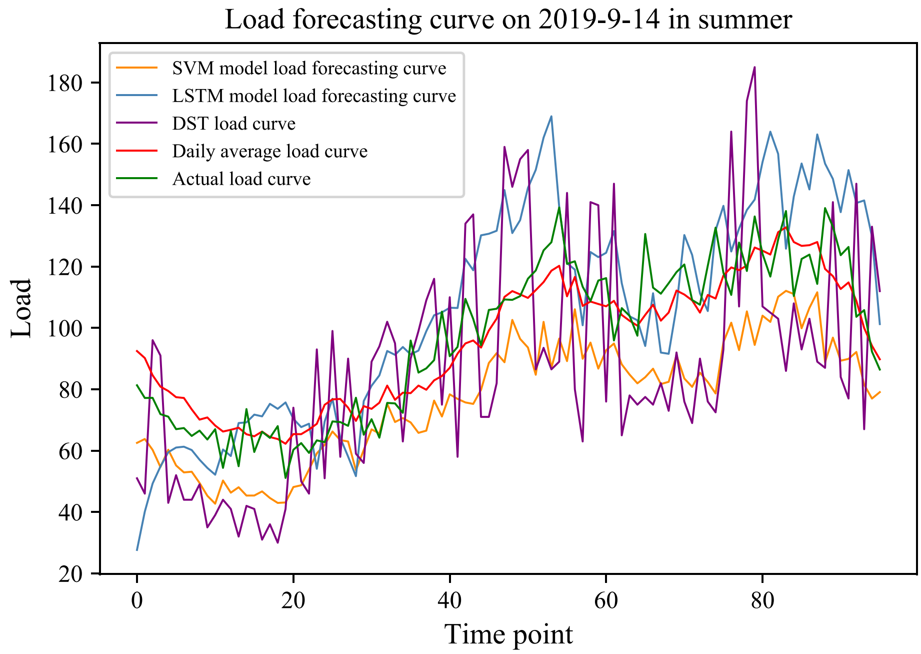


Fig. 6 Load forecasting curve on 2019-9-14 in summer

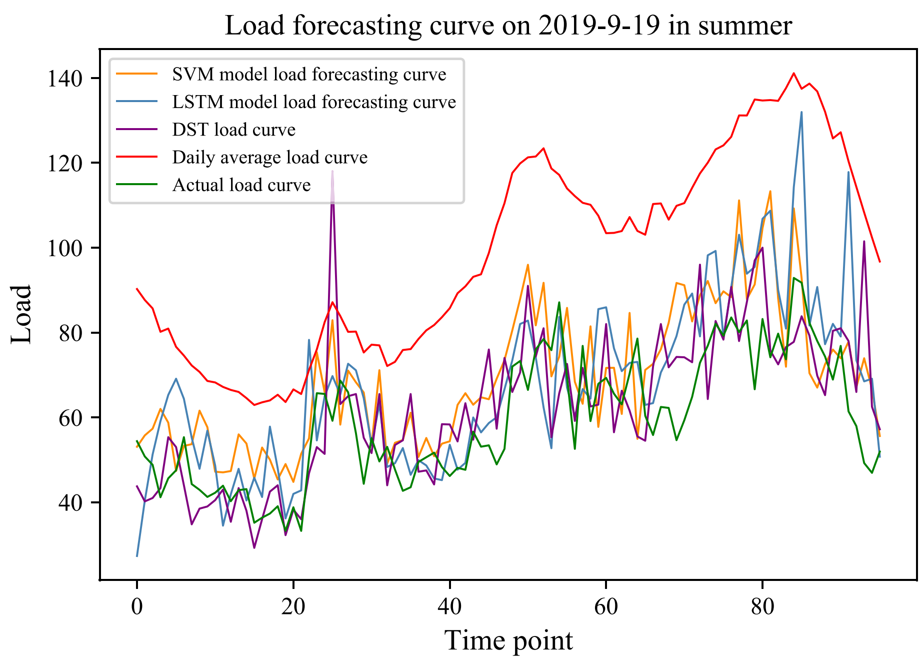


Fig. 7 Load forecasting curve on 2019-9-19 in summer

curve. Compared with the statistical indicators of the prediction results of the other two prediction models, SVM has again achieved better results overall. Based on the working day of September 19, 2019, under the reselection of the data set labeled 3, the load forecast results and accuracy are as presented in Fig. 7 and Num.E5 in Table 2.

It can be seen from the experimental results After re-predicting the working days similar to 2019-9-19 from all the data in summer, the load forecasting effect is less accurate than the forecasting effect in the entire season. The experimental results all prove that the SVM algorithm performs better than the other two comparison algorithms under the experimental indicators.

5.3.2 Experiments on District 2

1. *Influence of winter temperature on load forecast*
   Similar to the above experiments, in order to explore the influence of temperature on the load forecasting level, the winter data with greater temperature influence is selected for experiment. For the first data set, we choose 2020-1-22 as the forecast day. The three pro-

Table 3 Load forecast accuracy index of District 2

Serial number	Model	Max_E	Min_E	MSE	RMSE	MAE
E1	LSTM	72.6524	0.1618	333.8980	20.1836	13.1131
	SVM	**56.559**	**0.1034**	407.3804	**18.2728**	**14.5659**
	DT	67.776	0.456	612.5612	24.7499	18.4699
E2	LSTM	59.967	0.0845	266.039	16.3107	12.1081
	SVM	**38.6156**	0.1827	**218.2383**	**14.7728**	11.7772
	DT	50.8799	0.1111	234.5289	153143	11.3666
E3	LSTM	51.9130	0.0376	212.3354	14.5717	10.3297
	SVM	**39.0448**	**0.0006**	**201.5533**	**14.1969**	10.7980
	DT	82.88	0.088	586.574	24.2193	17.6298
E4	LSTM	33.8232	0.0541	126.799	11.2605	8.4926
	SVM	**30.0282**	0.1389	**92.4481**	**9.6149**	**7.0983**
	DT	61.456	0.136	173.158	13.1589	9.5016

Bold values are the best in the experimental results based on the algorithm proposed in this paper

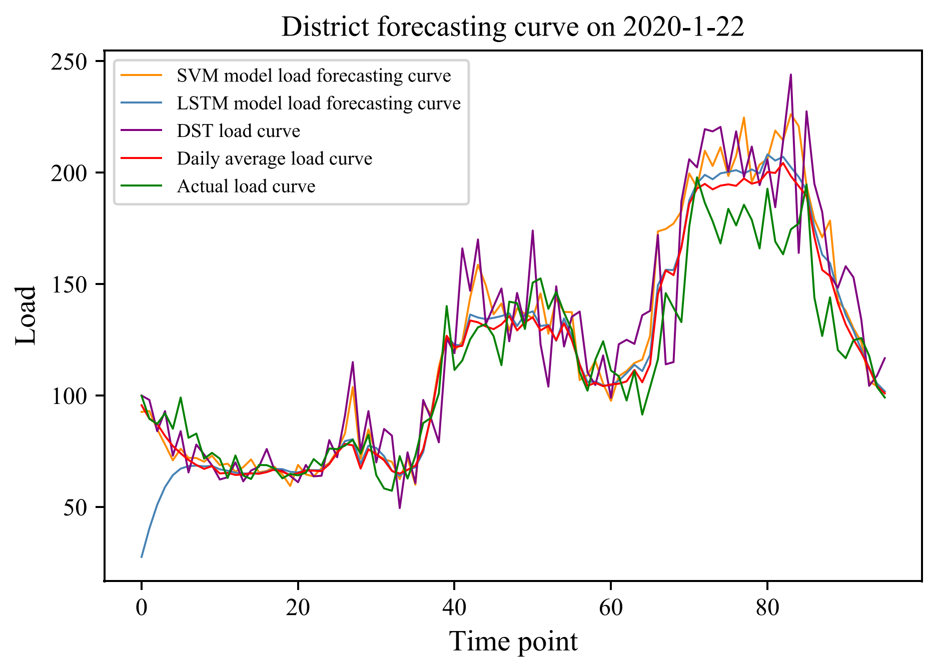


Fig. 8 District 2 forecasting curve on 2020-1-22

posed algorithms are also selected to predict the load curve of the day, and the comparison of evaluation indicators is given in the table below in Fig. 8 and Num.E1 in Table 3.

And for the second dataset, 2020-2-19 is selected as the day to be predicted. The experimental results are shown in the following chart Fig. 9 and Num.E2 in Table 3:

It can be seen from the results of the two experiments that the predicted results of the three algorithms are roughly the same as the actual load curves, and the results are credible. Although the SVM model is slightly inferior to other algorithms on a few indicators, on the whole, its prediction performance is better than the other two models, and it is integrated. This is closely related to residents' electricity consumption behavior.

2. *Influence of holiday on load forecast*

   Experiments on the influence of daily load levels are based on load data in spring and fall. For holidays, choose 2020-4-4 as the day to be forecasted, and for working days, choose 2020-4-6 as the forecast object. The experimental results are still given in Figs. 9, 10 and Num.E3,E4 in Table 3.

After considering the impact of daily types on load forecasting, the experimental results show that SVM is the load forecasting model with the best performance no matter what type of forecasting. The prediction effect of the DT model varies greatly depending on the experimental

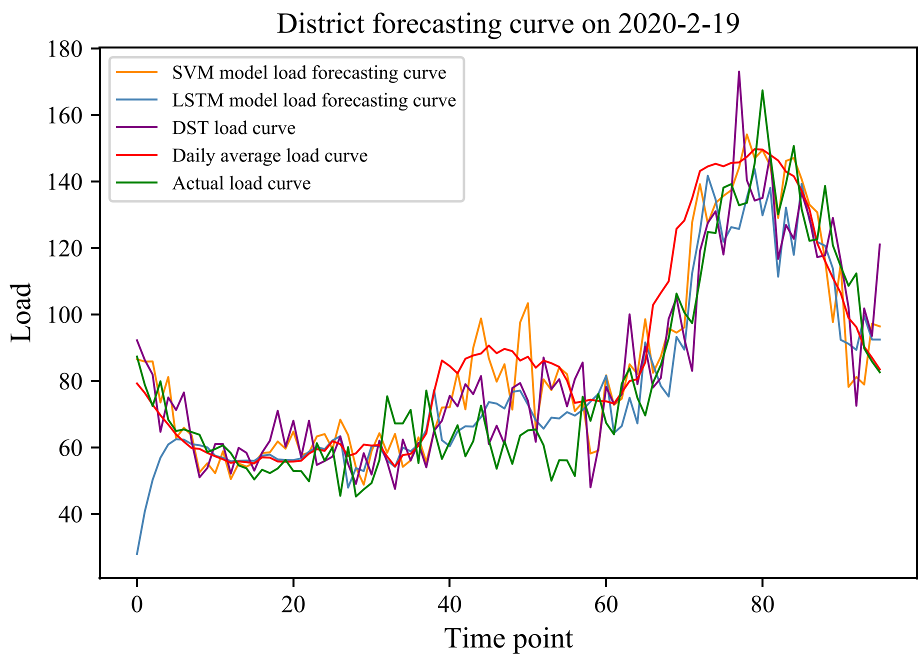


Fig. 9 District 2 forecasting curve on 2020-2-19

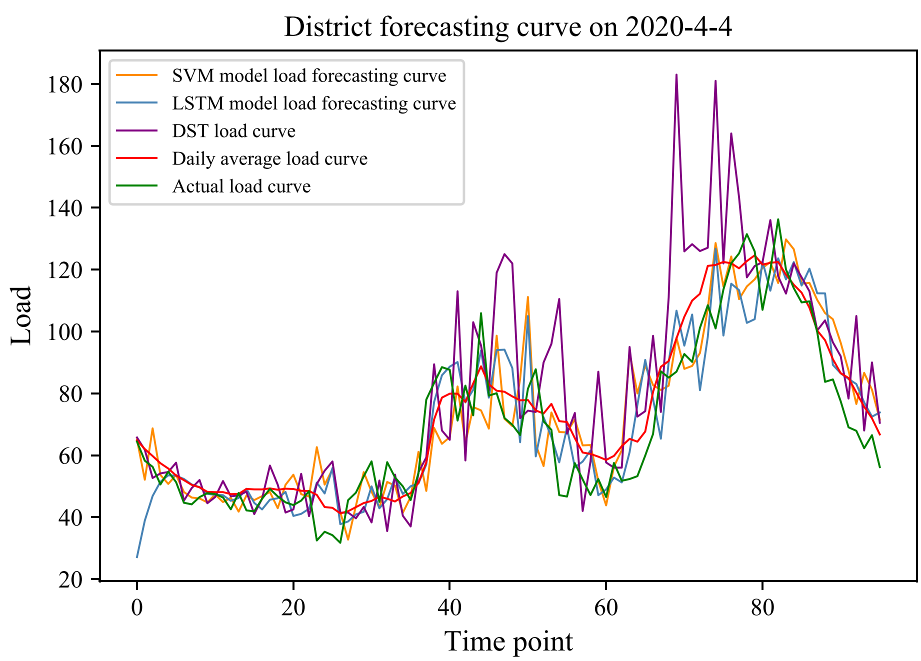


Fig. 10 District 2 forecasting curve on 2020-4-4

data set. Each time the prediction performance of LSTM is consistent with that of svm, it is relatively stable and will not be random due to external conditions such as influencing factors. Although in some experiments, some evaluation indicators are better than svm, svm is the best overall after many experiments (Table 3).

5.3.3 Experiments on District 3

By using k-means clustering algorithm, the 396 days load data in total of district 3 is divided into two data sets. Among them, the temperature clustering center of data set 3–1 is 0.94 when weather is cold, and the temperature clustering center of data set 2 is 15.89 when the weather is warm. And there are 48 time points in total (Figs. 8, 9, 10 and 11).

Similar to the above two regions, in the forecast results, select the load values of 48 time points on any day to draw the forecast curve and visualize it. Two data sets experimental results and evaluation index results are as following Fig. 12 and Table 4.

From the above experimental results, it can be seen that the the SVM has a good performance for load data in

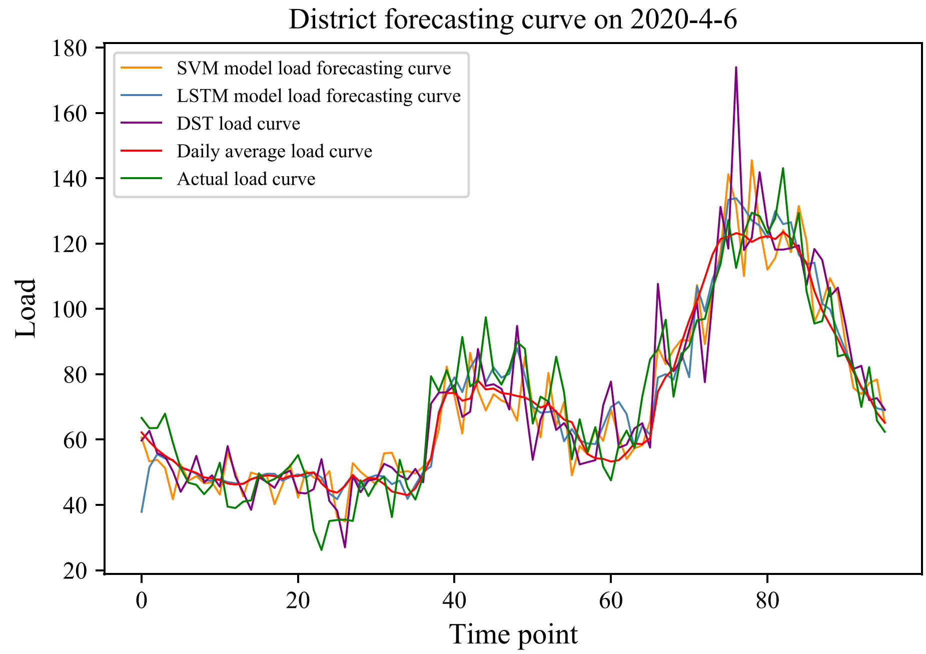


Fig. 11 District 2 forecasting curve on 2020-4-6

Table 4 Load forecast accuracy index of District 3

Serial number	Model	*Max_E*	*Min_E*	MSE	RMSE	MAE
E1	LSTM	115.1185	0.0394	1251.8845	35.3819	25.5323
	SVM	**74.6237**	1.8029	**743.3996**	**27.2653**	**21.8393**
	DT	75.0	0.6666	944.1249	30.7266	24.1334
E2	LSTM	102.8846	0.5783	597.3739	24.4412	17.6102
	SVM	**38.2605**	**0.2971**	**256.6688**	**16.0208**	**11.4949**
	DT	70.4249	0.4186	715.7738	26.7539	20.7083

Bold values are the best in the experimental results based on the algorithm proposed in this paper

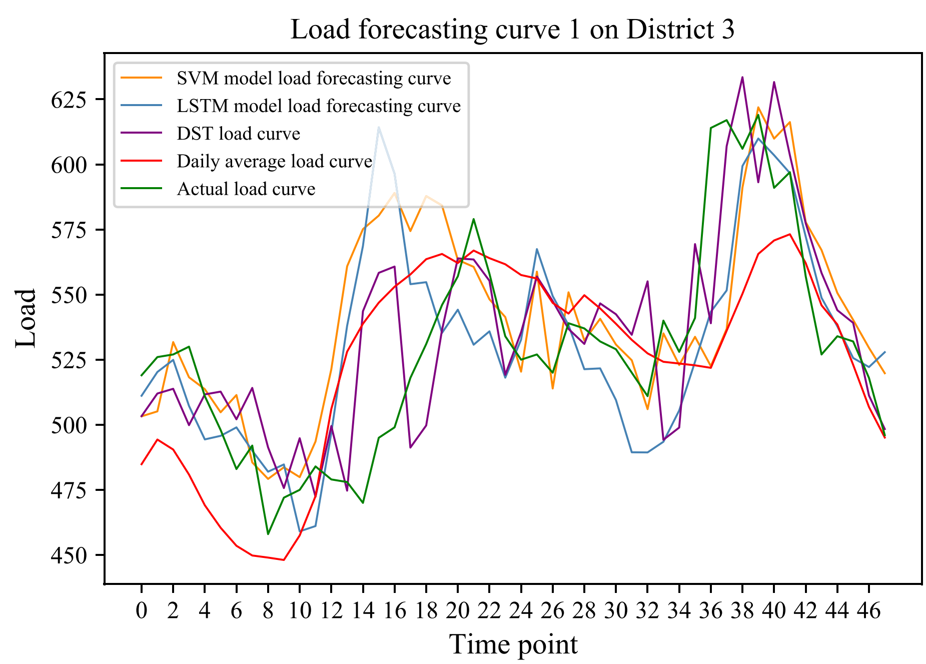


Fig. 12 Forecasting curve 1 of District 3

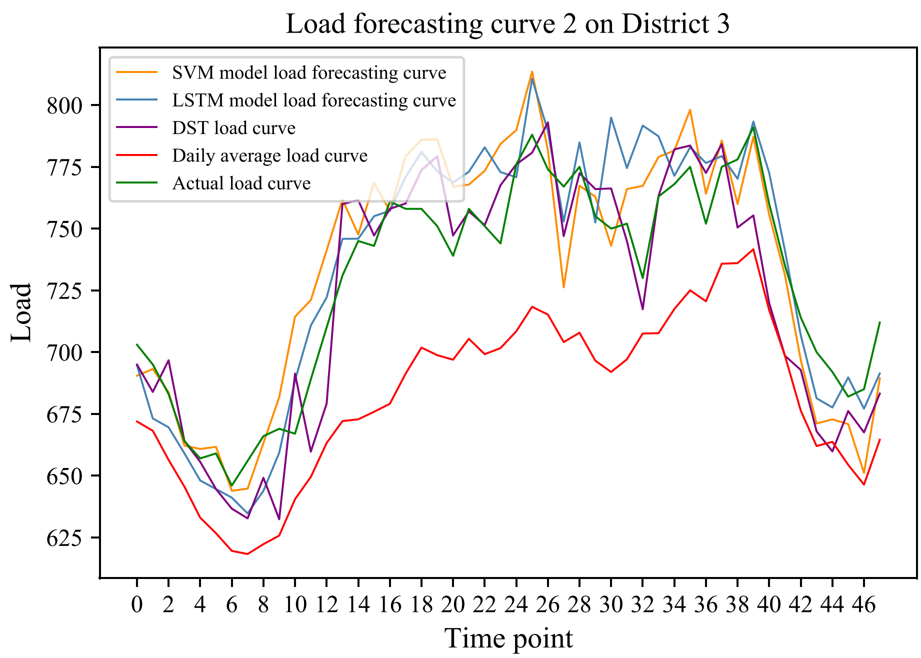


Fig. 13 Forecasting curve 2 of District 3

different regions. Among the selected all-day load forecasting curves, the forecast effect of the SVM model is consistent with the actual curve change of the day, and the difference between the point forecast and the actual value is small. At the same time, four of the five indicators have the best performance (Fig. 13).

For the data set 3–2, the prediction result of the SVM model is very close to the actual value at most time points, and the curve coincidence rate is high. Compared with others, SVM prediction performance is more stable, with the smallest deviation all the time. It is worth mentioning that in this experiment, SVM performed the best in all indicators, which verified the effectiveness of the proposed method.

5.3.4 Experiments on District 4

The load data of the above three regions are all at the regional level. For the purposes of distinction, we selected another country's load data, where we selects a single household electricity load to do experiments. Add holiday features to data according to local statutory holiday standards.

Due to the uncertainty of the load characteristics of a single resident, the user's electricity consumption characteristics are quite random. In order to analyze the impact of characteristic, after adding features, the short-term load forecasting method proposed in this paper are also used for comparative experiments. The load prediction results of experiments were randomly selected as shown in the following figures and Table 5 (Figs.14 and 15).

For holidays, the prediction curve of the model, the average daily load curve and the actual load curve have a similar trend in that day, and there is a daily electricity peak. Combined with the prediction curve and evaluation index, the model this paper proposes is the best. To sum up, the STLF-SK model has the best effect on the short-term prediction of the electricity consumption of the user in this area after considering the daily type characteristics.

Table 5 Load forecast accuracy index of District 4

Serial number	Model	Max_E	Min_E	MSE	RMSE	MAE
E1	LSTM	666.2743	71.424454	134834.1993	367.19776	332.9740
	SVM	**544.1836**	0.55665	**86769.7171**	**294.5669**	**249.7905**
	DT	1208.8481	125.2854	206601.4753	454.5343	391.6774
E2	LSTM	314.6979	2.6541	13048.5935	114.2304	93.8849
	SVM	**230.9553**	**0.01707**	**8337.3050**	**91.3088**	**63.0029**
	DT	394.9248	5.5486	15821.4481	125.7833	95.5443

Bold values are the best in the experimental results based on the algorithm proposed in this paper

Fig. 14 Forecasting curve for holiday of District 4

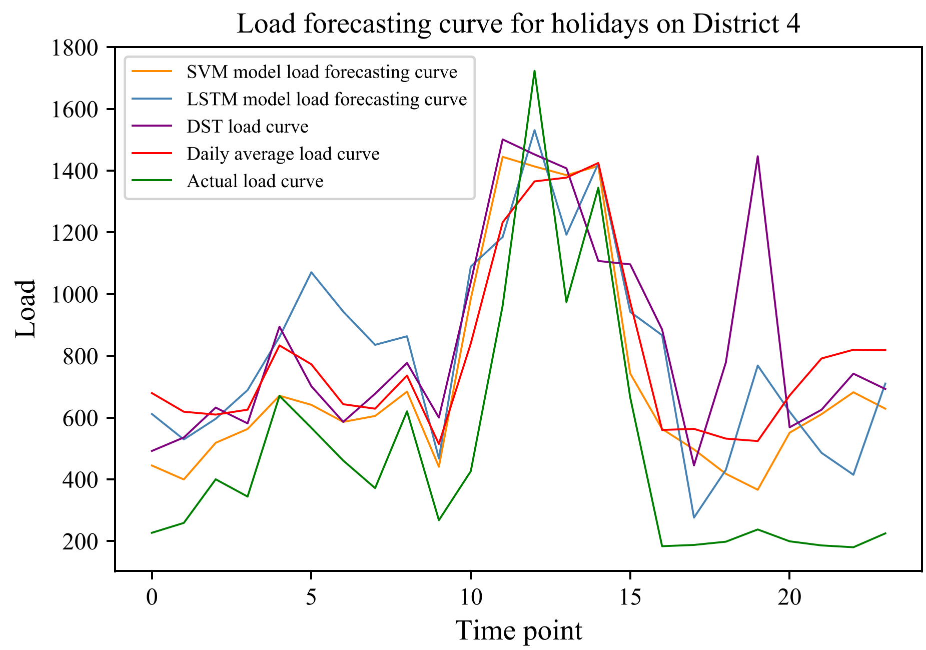


Fig. 15 Forecasting curve for workdays of District 4

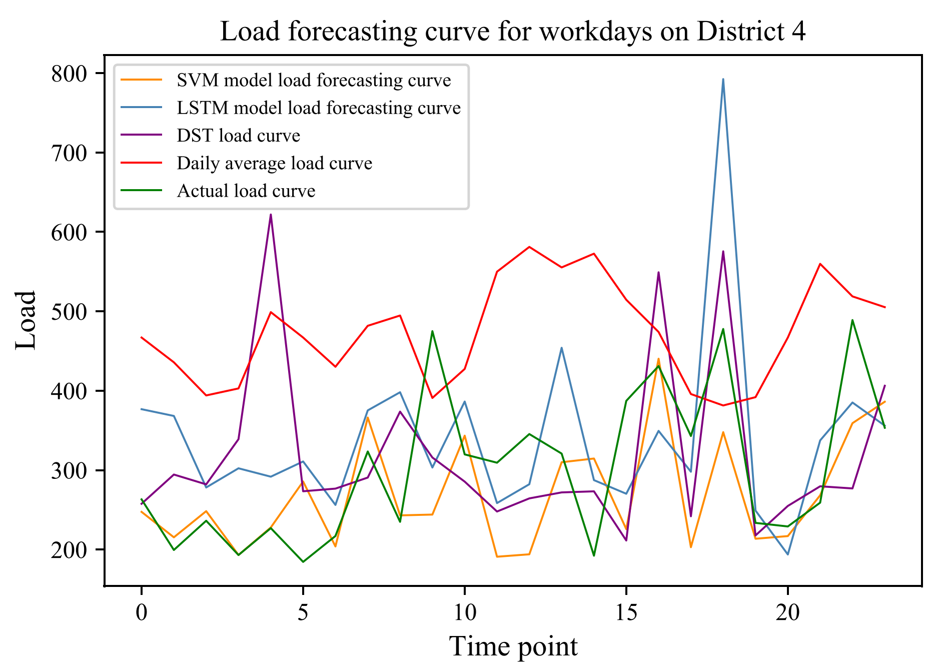


Table 6 The running time of each algorithm on each data set

District	Algorithm	Running time/s				
		1	2	3	4	5
District 1	SVM	**0.6255**	0.568	**0.4744**	**0.5017**	0.5062
	LSTM	293.5582	247.6926	263.0314	258.2809	246.9749
	DT	0.6436	0.5576	0.5227	0.5106	0.4563
District 2	SVM	**0.5496**	0.5317	**0.6758**	0.6014	–
	LSTM	292.6203	266.1459	603.3102	260.2305	–
	DT	0.5709	0.521	0.6954	0.5011	–
District 3	SVM	**98.187**	**90.4043**	–	–	–
	LSTM	629.6685	637.0465	–	–	–
	DT	473.1408	471.2869	–	–	–
District 4	SVM	**101.0842**	98.5964	–	–	–
	LSTM	698.4574	668.2147	–	–	–
	DT	105.2347	97.5628	–	–	–

Bold values are the best in the experimental results based on the algorithm proposed in this paper

5.3.5 Running time of each experiment

In reality, not only the prediction accuracy of the model must be considered, but the running time of the model must also be used as one of the indicators for investigating the prediction model. Taking these into account, we also enumerate the running time spent in all the above experiments, which is showing as following Table 6.

In terms of the running time it takes, although both SVM and DT consume less time, but no obvious difference. As far as the accuracy of all experimental results is concerned, SVM should be better than LSTM and DT. In summary, considering the prediction accuracy and running time of the algorithm, the algorithm STLF-SK proposed in this paper is not only accurate, but also efficient, and has a wide range of application prospects.

6 Conclusion

Aiming at the problem of the influence of temperature and holidays on the load behavior of users in the station area under the seasonal premise, this paper first uses the K-Means clustering algorithm to analyze characteristics. LSTM, SVM and DT are established for historical load data of different temperatures or whether they belong to holidays in the same season. And some experiments and result comparisons have been carried out on four load data sets. The results show that the SVM has a better prediction effect. After considering the influence of temperature and holidays on the load, the prediction effect of the model can be improved to a certain extent by changing the input data. Since the current two models are involved in the selection of functions and the optimization of parameters, more research will be conducted on this issue in the future.

Acknowledgements We would like to thank the anonymous reviewers for their comments and constructive suggestions that have improved the paper. The subject is sponsored by the National Natural Science Foundation of P. R. China (No. 51977113,51507084), BAGUI Scholar Program of Guangxi Zhuang Autonomous Region of China (201979) and NUPTSF (No. NY219095).

References

Barman M, Choudhury ND, Sutradhar S (2018) A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India. Energy 145:710–720

Bozkurt ÖÖ, Biricik G, Tayşi ZC (2017) Artificial neural network and Sarima based models for power load forecasting in Turkish electricity market. PLoS One 12(4):e0175915

Friedrich L, Afshari A (2015) Short-term forecasting of the Abu Dhabi electricity load using multiple weather variables. Energy Procedia 75:3014–3026

Haben S, Giasemidis G, Ziel F, Arora S (2019) Short term load forecasting and the effect of temperature at the low voltage level. Int J Forecast 35(4):1469–1484

Hafeez G, Alimgeer KS, Khan I (2020) Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. Appl Energy 269:114915

Herui C, Xu P (2015) Summer short-term load forecasting based on Arimax model. Power Syst Prot Control 43(4):108–114

Huang N, Wang W, Wang S, Wang J, Cai G, Zhang L (2020) Incorporating load fluctuation in feature importance profile clustering for day-ahead aggregated residential load forecasting. IEEE Access 8:25198–25209

Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y (2017) Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Trans Smart Grid 10(1):841–851

Lee WJ, Hong J (2015) A hybrid dynamic and fuzzy time series model for mid-term power load forecasting. Int J Electr Power Energy Syst 64:1057–1062

Lee CW, Lin BY (2017) Applications of the chaotic quantum genetic algorithm with support vector regression in load forecasting. Energies 10(11):1832

Lei J, Jin T, Hao J, Li F (2019) Short-term load forecasting with clustering-regression model in distributed cluster. Clust Comput 22(4):10163–10173

Lin L, Xin W, Shengyu S (2019) Resident-side intelligent power ubiquitous sensing technology and multi-precision service research. Distrib Util 36(6):10–15

Lu H, Li Y, Chen M, Kim H, Serikawa S (2018) Brain intelligence: go beyond artificial intelligence. Mobile Netw Appl 23(2):368–375

Lu S, Lin G, Liu H, Ye C, Que H, Ding Y (2019) A weekly load data mining approach based on hidden Markov model. IEEE Access 7:34609–34619

Lu H, Zhang M, Xu X, Li Y, Shen HT (2020a) Deep fuzzy hashing network for efficient image retrieval. IEEE Trans Fuzzy Syst. https://doi.org/10.1109/TFUZZ.2020.2984991

Lu H, Zhang Y, Li Y, Jiang C, Abbas H (2020b) User-oriented virtual mobile network resource management for vehicle communications. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2020.2991766

Muzaffar S, Afshari A (2019) Short-term load forecasts using LSTM networks. Energy Procedia 158:2922–2927

Ni Q, Zhuang S, Sheng H, Kang G, Xiao J (2017) An ensemble prediction intervals approach for short-term PV power forecasting. Sol Energy 155:1072–1083

Ryu S, Noh J, Kim H (2017) Deep neural network based demand side short term load forecasting. Energies 10(1):3

Vrablecová P, Ezzeddine AB, Rozinajová V, Šárik S, Sangaiah AK (2018) Smart grid load forecasting using online support vector regression. Comput Electr Eng 65:102–117

Wang D, Luo H, Grunder O, Lin Y, Guo H (2017) Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. Appl Energy 190:390–407

Wang Z, Wang Y, Zeng R, Srinivasan RS, Ahrentzen S (2018) Random forest based hourly building energy prediction. Energy Build 171:11–25

Welikala S, Dinesh C, Ekanayake MPB, Godaliyadda RI, Ekanayake J (2017) Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting. IEEE Trans Smart Grid 10(1):448–461

Xia C, Zhang M, Cao J (2018) A hybrid application of soft computing methods with wavelet SVM and neural network to electric power load forecasting. J Electr Syst Inf Technol 5(3):681–696

Xiao L, Wang J, Hou R, Wu J (2015) A combined model based on data pre-analysis and weight coefficients optimization for electrical load forecasting. Energy 82:524–549

Yu F, Xu X (2014) A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. Appl Energy 134:102–113

Zahid M, Ahmed F, Javaid N, Abbasi RA, Zainab Kazmi HS, Javaid A, Bilal M, Akbar M, Ilahi M (2019) Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids. Electronics 8(2):122

Zhang P, Wu X, Wang X, Bi S (2015) Short-term load forecasting based on big data technologies. CSEE J Power Energy Syst 1(3):59–67

Zhang Y, Chen B, Pan G, Zhao Y (2019) A novel hybrid model based on VMD-WT and PCA-BP-RBF neural network for short-term wind speed forecasting. Energy Convers Manage 195:180–197

Zhao H, Guo S (2016) An optimized grey model for annual power load forecasting. Energy 107:272–286