Predicting Daily Heating Energy Consumption of Rural Residences in Northern China Using Support Vector Machine

P.L. Yuan, L. Duanmu and Z.S. Wang

School of Civil Engineering
Dalian University of Technology, Dalian, Liaoning 116024, China

SUMMARY
It is not feasible to apply the engineering prediction method to predict the daily heating energy consumption of Chinese rural residences due to the limited and uncertain parameters, such as air infiltration. This paper presents a data driven method-support vector machine to predict the daily heating energy consumption of Chinese rural residences. The data of heating energy consumption is collected by testing rather than questionnaire survey. A new input parameter, the cumulative temperature difference between indoor and outdoor air $\Delta T_c$ is proposed instead of average indoor air temperature because of the different temperature variation rules between the rural and urban residences. The results demonstrate that the predicted daily heating energy consumption of rural residences has a good agreement with the measured value. The mean relative error is lower than 10%. This study further forecasts the daily heating energy consumption from 1st January to 30th March.

INTRODUCTION
In China, the energy consumption in northern rural residences accounts for 42% of the national building energy consumption. The average outdoor air temperature in January is below 0 $^\circ$C. Thus, space heating is essential for sustaining the indoor thermal comfort in northern China. The heating energy consumption has reached up to 105 million tce in northern rural residences (BECRC, Tsinghua University 2016). The widely used fuels for heating are bulk coal, wood, cornstalks and so on. These fuels are directly burned in the decentralized heating equipments, including Chinese Kang (Zhuang Z et al. 2009), hot-wall Kang (Duanmu L et al. 2017), and small boiler with radiators and so on. The exhaust gas generated by burning the solid fuels is removed from chimney directly without any treatment measures. Thus, it is imperative to decrease the energy consumption of Chinese rural residences. However, the foundation of energy conservation is the actual energy consumption.

As described above, the heating energy consumption in Chinese rural residences is difficult to measure using meter devices. Generally, there are multi-sources and different channels to obtain the heating energy consumption of Chinese rural residences (Yao C et al 2012). The primary source of rural energy consumption for China is the National Statistics Bureau, which posses the energy consumption of energy sources. However, the end-use energy consumption is not included. The common method for achieving the heating energy consumption of Chinese rural residences is the random stratified sampling with questionnaire survey. Tsinghua University launched several nationwide energy consumption investigations in Chinese rural areas in 2006, 2007 and 2015 (BECRC, Tsinghua University 2012, 2016). Zhou Z (2009) analyzed the household energy consumption in terms of energy sources and energy end-use in villages of Huantal County from 1989 to 2005. The data concerning the energy consumption were all collected using questionnaires. Wang Y (2016) investigated the heating energy consumption through questionnaires in Chinese rural residences. The questionnaire survey is a simple and straight approach to obtain the heating energy consumption of Chinese rural residences. Rural households only need answer the questions in the questionnaire, and then researchers analyze the data through statistical method. Nonetheless, the investigation results have a strong relationship with the number of samples and the quality of the questionnaire. At present, the investigation process is not transparency and the quality of questionnaire survey cannot be assessed.

The heating energy consumption of Chinese rural residences can be achieved by prediction model as well. Developing reliable and accurate model for predicting building energy consumption is a challenge project (William K et al. 2016). Various models for predicting heating energy consumption are based on the dynamic heat transfer (Fischer D et al. 2016), and data driven method (Paudel S et al. 2014). Meanwhile, several building energy consumption simulation software (DOE-2, eQuest, EnergyPlus, Dest) have been developed and applied. The forecasting method for the heating energy consumption of Chinese rural residences concentrates on dynamic heat transfer and simulation software. Some researchers developed the dynamic heat transfer model to predict the indoor air temperature in Chinese rural residences (Yang M et al. 2013; Cao G et al. 2011). Thus, the heating energy consumption can be obtained through setting the demanded indoor air temperature in the dynamic heat transfer model. Dest is widely used to simulate the heating energy consumption of Chinese rural residences. The prediction models based on the dynamic heat transfer require a large number of physical parameters, such as the air infiltration, building size, heat transfer coefficient of building envelope, and so on. Moreover, some physical parameters are uncertain for Chinese rural residences. As we all know, the prediction model of energy consumption should have extensive and convenient feature. Although the models based on the dynamic heat transfer can guide the energy conservation measure effectively, they are complicated for common users to forecast the heating energy consumption of rural residences.

Fortunately, prediction models based on the data driven approaches have been developed gradually with the limited parameters. These models are suitable for the common users. Among these data driven approaches, support vector machine (SVM) possesses higher superiority than artificial neural network in terms of accuracy and generalization ability (Izadyar N et al. 2015; Li Q et al. 2009). Also, SVM can availably tackle the small sample problem. Kavakioglu K
established the Turkey's electricity consumption prediction model by employing the support vector regression. The number of samples is 32. It is very difficult to collect a large number of heating energy consumption data by experiments in Chinese rural areas. Synthesizing the two merits described above, SVM is appropriate for predicting the heating energy consumption of Chinese rural residences. To the best of our knowledge, few studies focus on applying the data driven method to establish the heating energy consumption prediction model of rural residences based on the measured data. SVM models of the published literatures focused on the energy consumption of urban residences. None of previous studies has captured in predicting the heating energy consumption of Chinese rural residences using SVM. Comparing with urban residences, the indoor air temperature of rural residences is unable to hold steady. The reason is that rural households generally go through the process of igniting the coal, filling the coal, damping down the fire and stopping burning every day. Thus, the input parameters such as average outdoor weather data and average indoor air temperature employed in the previous studies is not applicable for rural residences.

The target of this paper is to establish the SVM prediction model of heating energy consumption that can be applied for Chinese rural residences. In this paper, the data of heating energy consumption of rural residences is collected by field test instead of questionnaire survey. Field test has higher than questionnaire survey in terms of the data quality. Two rural residences are selected to validate the feasibility of the SVM prediction model of heating energy consumption in Chinese rural residences.

METHODS

SVM model

SVM guarantees the maximum generalization ability of the model based on the structural risk minimization, which can overcome the dimension disaster and over fitting problem (Vapnik V 1995). Assuming that the input parameters compose the vector \(X_i\), \(Y_i\) is the corresponding output parameters of \(X_i\). The sample set can be defined to be \(\{(X_i, Y_i)\}\) when the number of samples is \(N\). The mathematical structure of SVM is demonstrated by the following equation:

\[
Y = f(X) = W\phi(X) + b
\]  
(1)

\[
\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \phi(X_i) \phi(X_j) - \sum_{i=1}^{N} Y_i (\alpha_i - \alpha_i^*) + \sum_{i \in S} (\alpha_i + \alpha_i^*)
\]

Subject to:

\[
\sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0
\]

(4)

\[
0 < \alpha_i, \alpha_i^* < C
\]

(5)

Thus, function (1) becomes the explicit form:

\[
f(X) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \phi(X_i) \phi(X) + b = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(X_i, X) + b
\]

where \(\alpha_i\) and \(\alpha_i^*\) are the lagrange multiplier vector and should be greater than zero. \(K(X_i, X)\) is the kernel function and equal to \(\phi(X_i) \cdot \phi(X)\).

The conventional kernel function \(K(X_i, X)\) contains the linear function, polynomial function, Gaussian RBF function, exponential base function and sigmoid function. Among these kernel functions, Gaussian RBF function can map the input space into the high dimensional feature space effectively, and makes the complex nonlinear relationship between the input and output parameters present preferably. In this paper, Gaussian RBF function is selected to be the kernel function. The function form can be found in the literature (Izadyar N et al. 2015). In the SVM model, the parameter \(C\), \(\epsilon\) and \(\sigma^2\) should be optimized so that the model can accurately predict the unknown data. This paper first employs particle swarm optimization (PSO) to optimize \(C\) and \(\sigma^2\). Then, \(C\) and \(\sigma^2\) are fixed to be the optimal value, and set \(\epsilon\) as various values in a range and train the SVM model. The optimal value of \(\epsilon\) will be determined based on the prediction performance indices and the number of support vectors.

The performance evaluation indicators of SVM model adopted through this paper are root mean square error \(RMSE\), mean relative error \(MRE\) and the determination coefficient between predicted and measured value \(R^2\). The calculation method can be referenced in the literature (Li Q et al. 2009; Izadyar N et al. 2015)

Input parameters

Many previous researchers adopted the average outdoor air temperature, global solar radiation (GSR) and average indoor air temperature as the input parameters in urban residences (Izadyar N et al. 2015). As we all know, the indoor air temperature of urban residences is persistently steady. The average indoor air temperature can reflect the effect of heating energy consumption. Nonetheless, the indoor air temperature of rural residences has a larger fluctuation because the heating is intermittent. The average indoor air temperature cannot present the function generated by the actual heating energy consumption. Thus, we need propose a new input parameter related to indoor air temperature that can preferably present the function of heating energy consumption in rural residences. Niu S (2010) demonstrated that the difference between the indoor and outdoor air temperature was primarily the result of heating. The cumulative temperature difference between indoor and outdoor air is shown in the following forms:

\[
\Delta T_e = T_{indoor} - T_{outdoor}
\]

(7)
\[
T_{\text{indoor}} = \int_{0}^{n} T_{\text{indoor}}(\tau) d\tau \\
T_{\text{outdoor}} = \int_{0}^{n} T_{\text{outdoor}}(\tau) d\tau
\]

where \(\Delta T_c\) represents the cumulative temperature difference between indoor air and outdoor air, \(\circ\cdot s\). \(T_{\text{indoor}}\) and \(T_{\text{outdoor}}\) are the cumulative indoor air and outdoor air temperature respectively, \(\circ\cdot s\). \(T_{\text{indoor}}\) and \(T_{\text{outdoor}}\) are the indoor air and outdoor air temperature respectively. \(\circ\cdot r\) is time, s; \(n\) is the final time, s. Ref (Niu S et al. 2010) also presented that an increase in the indoor air temperature from \(T_{\text{outdoor}}\) to \(T_{\text{indoor}}\) required the consumption of a certain amount of fuel. In this paper, the input parameters of SVM model are \(\Delta T_c\) and GSR.

Data collection

Through analyzing the input parameters, the hourly indoor and outdoor air temperature and solar radiation should be measured. Two rural residences are selected in a village, located in Chifeng, Inner Mongolia. They are all traditional rural residences shown in Figure 1. The daily heating energy consumption is measured by rural households every day. To our best knowledge, this is first time to measure the daily heating energy consumption in rural areas. The measured data is more objective than questionnaire survey.

![a) Rural residence A b) Rural residence B](Image)

**Figure 1. Two test rural residences**

The field test began in November 21\(^{\text{th}}\) 2017 and lasted 44 days. However, the number of samples is different for the two residences. In rural residence A, the data from 21\(^{\text{th}}\) November to December 25\(^{\text{th}}\) are the training samples, and the residual data are the testing samples. In addition, the data from December 2\(^{\text{th}}\) to December 12\(^{\text{th}}\) are omitted because the data of solar radiation and indoor air temperature are missing during this period. In rural residence B, the data from November 23\(^{\text{th}}\) to December 25\(^{\text{th}}\) are the training samples, and the residual data are the testing samples. The data from December 2\(^{\text{th}}\) to December 8\(^{\text{th}}\) are omitted because the data of solar radiation is missing during this period. The building information and measurement instruments are listed in Table 1 and Table 2.

**Table 1. Building information**

<table>
<thead>
<tr>
<th>Rural residence number</th>
<th>Heating areas (m(^2))</th>
<th>Heating equipment</th>
<th>Number of heating room</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>37.83</td>
<td>Chinese Kang and small boiler with radiators</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>40.37</td>
<td>Chinese Kang and small boiler with radiators</td>
<td>3</td>
</tr>
</tbody>
</table>

**Table 2. Measurement instruments**

<table>
<thead>
<tr>
<th>Measurement parameters</th>
<th>Measurement instruments</th>
<th>Measurement accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuel mass</td>
<td>Electronic scale</td>
<td>5g</td>
</tr>
<tr>
<td>Indoor and outdoor air temperature</td>
<td>Thermo Recorder</td>
<td>±0.5°C, ±5%RH</td>
</tr>
<tr>
<td>Solar radiation</td>
<td>Vantage Pr portable weather station</td>
<td>±5%</td>
</tr>
</tbody>
</table>

During the test, we assumed that the external door and windows were closed a day. Moreover, we didn’t consider the variation of the number of personnel in the room.

**RESULTS**

Heating energy consumption

In the two rural residences, the heating equipment is coupled heating system of Chinese Kang and small boiler with radiators. The fuels for heating contain coal and biomass. The biomass fuels are burned in the stove and the gas heats the Kang surface. Coal is burned in the small boiler. Figure 2 showed that the coal consumption played a dominant role in heating the indoor air. The biomass consumption fluctuated around 5kgce. However, the biomass consumption beyond 10kgce several days, which had larger difference than the other days. The reason is that some visitors come to this rural residence and the rural household has to prepare more food. Thus, the stove is used more frequently than ordinary day.

![Figure 2. Heating energy consumption of the two test rural residences](Image)
sitting or sleeping. The results indicated that the heating energy consumption was approximate to be the coal consumption in the coupled heating system of Chinese Kang and small boiler with radiators. Thus, the coal consumption was regarded as the output variable in SVM models.

Table 3. Partial correlation analysis between coal consumption and $\Delta T_c$, and biomass consumption and $\Delta T_c$

<table>
<thead>
<tr>
<th>Variables</th>
<th>Partial correlation coefficient</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coal consumption</td>
<td>0.579</td>
<td>0.001</td>
</tr>
<tr>
<td>Biomass consumption</td>
<td>-0.112</td>
<td>0.542</td>
</tr>
</tbody>
</table>

Parameters selection

The accuracy of SVM model is influenced by the parameters ($C$, $\varepsilon$ and $\sigma^2$). Firstly, $C$ and $\sigma^2$ were optimized by PSO method, and the value of optimal $C$ and $\sigma^2$ are presented in Table 4.

Table 4. Optimal $C$ and $\sigma^2$ of the two rural residences

<table>
<thead>
<tr>
<th>Rural residence</th>
<th>$C$</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6.70</td>
<td>6.30</td>
</tr>
<tr>
<td>B</td>
<td>0.15</td>
<td>23.90</td>
</tr>
</tbody>
</table>

Then, $C$ and $\sigma^2$ are fixed on the optimum value, and set $\varepsilon$ as various values in a range and train the SVM model. The results of various $\varepsilon$ in the two rural residences are shown in Figure 3.

To evaluate the SVM model performance further, the measured value was compared with the predicted value shown in Figure 4. The relative error between the predicted and measured value of several samples were larger than 15%. However, Figure 4 demonstrated that the predicted heating energy consumption had a good agreement with the measured value. The mean relative errors were all lower than 10%. Thus, the established SVM model can be considered to be reasonable and valid. Also, the results indicated that SVM can be used to predict the heating energy consumption of rural residences. It does not need more prior information and use skill.

Heating energy consumption prediction

The heating energy consumption of rural residences was measured by rural households manually. Furthermore, this established training models should be applied to validate the testing data. Thus, the prediction results on testing data are more significant to evaluate the model performance. The performance evaluation of SVM models are presented in Table 5. Table 5 indicated that the overall model performance of rural residence B was higher than rural residence A.

Table 5. Performance evaluation of SVM models on the test data

<table>
<thead>
<tr>
<th>Rural residence</th>
<th>RMSE</th>
<th>MRE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.122</td>
<td>0.093</td>
<td>0.612</td>
</tr>
<tr>
<td>B</td>
<td>0.089</td>
<td>0.075</td>
<td>0.547</td>
</tr>
</tbody>
</table>

After the optimal value of $C$, $\varepsilon$ and $\sigma^2$ confirmed, the train models of the two rural residences were established. The
paper simply measured the heating energy consumption of 44 days due to the limit of time and manpower. As described above, the established SVM model can be applied to predict the heating energy consumption that is not measured by testing instruments. This study will predict the heating energy consumption of the two rural households from 1st January to 30th March. The weather data derived from the special meteorological data set of China building thermal environment analysis. The hourly outdoor air temperature $t_{\text{out}}$ (°C) and solar radiation $I_{\text{solar}}$ (W/m²) in Chifeng, Inner Mongolia, are shown in Figure 5.

According to the input parameters, the demanded indoor air temperature should be gained before predicting the heating energy consumption. In this study, the demanded indoor air temperature was assumed to be a various value instead of constant value. The reason is that the heating equipment is operated intermittently every day. The indoor air temperature of rural residences cannot sustain steady state. This paper takes the average hourly indoor air temperature during the testing period as the demanded indoor air temperature. The results are shown in Figure 6. Figure 6 showed that there are two peak values. The reason is that rural households generally operate the small boiler in the morning and at sunset.

The heating energy consumption of the two rural residences from 1st January to 31th March is predicted shown in Figure 7. The results showed that the heating energy consumption was fluctuated with the outdoor climatic conditions. The heating energy consumption of rural residence A presented a normal change, which is that, the general tendency of the heating energy consumption decreased with the outdoor air temperature and solar radiation increasing. However, in rural residence B, the general tendency of heating energy consumption presented differently. The reason is possibly that the accuracy of the SVM model is inadequate, which leads to the inaccurate results when the outdoor air temperature and solar radiation is relative high.

Figure 5. Outdoor air temperature and solar radiation in Chifeng

Figure 6. Demanded indoor air temperature of two rural residences

DISCUSSION

This paper applied SVM to predict the heating energy consumption of Chinese rural residences based on the detailed test data. Due to the discrepancy of indoor air temperature between rural and urban residences, the cumulative temperature difference between indoor air and outdoor air $\Delta T_c$ was proposed to be one of input parameters. The heating equipment of the two rural residences is the coupled heating system of Chinese Kang and small boiler with radiators. In general, the heating energy consumption should be the sum of the fuels consumption burned in the stove and small boiler. However, Table 3 showed that the biomass consumption affected the indoor air temperature little. Thus, the heating energy consumption is assumed to be the coal consumption burned in the small boiler.

The SVM models of the two rural residences were established and the model performance was evaluated. The mean relative error of the SVM models was lower than 10%. $R^2$ is around 0.5. The deviation between measured and predicted heating energy consumption is acceptable. However, the prediction performance of the SVM models should be improved according to the source of error. The possible sources of error in the SVM models are presented as follows.

1) The heating energy consumption in Chinese rural residences could not be automatically measured by recorders. Rural households weighed the daily fuels for heating and then recorded the data in the log sheets. The weighting and recording process of the daily fuels relied on manual operation purely, which may result in deviation and unreasonable data.

2) The outside door is opened frequently due to the life style of rural households. Furthermore, rural households visit each other, which results in the variation of the number of indoor occupants. Thus, the measured indoor air temperature may have deviation with the actual indoor air temperature.
Although the SVM models have a certain error, this paper validated the feasibility of SVM to predict the heating energy consumption of Chinese rural residences. Moreover, this paper explored a new avenue to predict the heating energy consumption of Chinese rural residence. SVM model can be applied not only for building designer or researchers, but also for common users.

**CONCLUSIONS**

This study carried out the heating energy consumption test of Chinese rural residences to provide better quality data for developing SVM model. SVM models of the two rural residences were established. The input and output parameters for rural residences have been discussed. The cumulative temperature difference between indoor air and outdoor air $\Delta T_c$ instead of average indoor air temperature was regarded as one of input parameters. Based on partial correlation analysis, the heating energy consumption was assumed to be the coal consumption, rather than the sum of coal and biomass consumption.

The heating energy consumption predicted by SVM models of the two rural residences has a good agreement with the measured heating energy consumption. The mean relative error is lower than 10%, and $R^2$ is around 0.5. This paper also predicts the daily heating energy consumption of the two rural residences from 1st January to 31th March. The prediction results are in consistent with the regular change. Nonetheless, when the outdoor air temperature and solar radiation is relative high, the predicted results present abnormally. The prediction performance of SVM models should be improved in the future.

**ACKNOWLEDGEMENT**

This research is supported by China national key research and development program - Intergovernmental cooperation in science and technology innovation (Project No. 2016YFE0114500).

**REFERENCES**


