Research Article


Jinya Cai, Haiping Zhang, Xinping Yu, and Amir Seraj

1 Zhejiang Guangsha Vocational and Technical University of Construction, Dongyang 322100, China
2 Zhejiang Ruiao Testing Technology Service Co.Ltd., Hangzhou 310027, China
3 Department of Instrumentation and Industrial Automation, Ahwaz Faculty of Petroleum Engineering, Petroleum University of Technology (PUT), Ahwaz, Iran

Correspondence should be addressed to Jinya Cai; caijinya123@126.com and Amir Seraj; amir.seraj@afp.put.ac.ir

Received 15 April 2022; Revised 9 May 2022; Accepted 16 May 2022; Published 29 May 2022

A robust machine learning algorithm known as the least-squares support vector machine (LSSVM) model was used to predict the liquid densities of 48 different refrigerant systems. Hence, a massive dataset was gathered using the reports published previously. The proposed model was evaluated via various analyses. Based on the statistical analysis results, the actual values predicted by this model have high accuracy, and the calculated values of RMSE, MRE, STD, and $R^2$ were 0.0116, 0.158, 0.1070, and 0.999, respectively. Moreover, sensitivity analysis was done on the efficient input parameters, and it was found that CF$_2$H$_2$ has the most positive effect on the output parameter (with a relevancy factor of +50.19). Furthermore, for checking the real data accuracy, the technique of leverage was considered, the results of which revealed that most of the considered data are reliable. The power and accuracy of this simple model in predicting liquid densities of different refrigerant systems are high; therefore, it is an appropriate alternative for laboratory data.

1. Introduction

An isolated subsystem is being cooled lesser than the remainder of the system because of the chilling mechanism [1–3]. Chilling procedures are mainly used in the chemical sectors. Coolant solutions are used in chilling procedures in this regard [4–6]. The thermodynamic characteristics of coolants, such as liquid density, vapor density, enthalpy of evaporation, and vapor pressure, are critical to developing commercially viable low-temperature chilling circuits [7–9]. Research on coolants throughout the publications is extensive, but there is still a lack of empirical evidence available, and also, the data collections are often incongruent [10–12].

Coolants and mixes encounter a plethora of equations about their condition documented in scientific journals and textbooks [13, 14]. Analyze and evaluate four equations about the status and fourteen relationships for calculating the saturated liquid density of coolants. They suggested the chain of rotator group contribution (CORG) equation about the status as the optimal equation [15]. The fourth of eighteen approaches provided for calculating the saturated fluid density of coolants. The Hankinson and Thomson relationship is the strongest among the several correlations [16], followed by the Riedel [17] association and the improved Rackett association recommended by Spencer and Danner [18]. To estimate several of the thermophysical parameters of water-based solutions employed as secondary coolants, an appropriate technique was developed by Lugo and his colleagues [19]. The freezing temperatures, densities, heat capacities, thermal conductions, and dynamic viscosities are determined using the surplus function methodology. As a way to forecast the density of both purified liquids and their combinations, it developed a three-parameter density prototype based on the “corresponding states” method [20]. Hydrofluoroether (HFE) and halogenated
alkane (HA) are two adaptive categories of the investigated novel coolant liquids generation. By employing relationships developed by Boushehri and Mason, Sharafi and Boushehri generalized the ISM equation of condition based on the numerical mechanical perturbation concept for liquefied coolant mixes [21, 22]. On 33 liquid mixtures containing 12 coolants, the equation of status has been validated. Over an extensive temperature range of 170–369 K, the density of liquids may be forecasted with an error percentage of at most 2.8.

It was reported in 2005 that Mafloon-Azad and colleagues could forecast the density of condensed fluid HCFC and HFC coolants using an analytical equation of status. The density of the solution and the warmth of evaporation at the boiling temperature were used as input variables for their status calculation [23]. They predicted the volumetric performance of the six coolants using a modified form of the ISM equation [24]. GMA EOS was used by Goharshadi and Moosavi to anticipate the density of a restricted aqueous coolant combination [25]. A straightforward method for predicting the thermodynamic parameters of liquid coolant mixes, particularly when the same refrigerants are combined, is provided by GMA EOS, according to the researchers [26].

Coolant mixes were studied by Esilami, Mehdipour, and their colleagues, who expanded their prior study on the coolant equation of status [27–30]. Relying on the standard boiling point temperature and fluid density, the temperature-dependent variables of the equation of state were computed using their prior relevant states relationship. They extended their earlier suggested EOS to combinations of coolants by using a quadratic relation established by Nasrifar and Moshfeghian for the standard boiling point constants. With the use of the Goharshadi–Morsali–Abbaspour equation of status (GMA EOS), the density of eleven different hydrochlorofluorocarbon (HCFC) and hydrofluorocarbon (HFC) coolants was estimated. For estimating fluid density, GMA EOS is adequate, according to their findings [13, 31].

Nevertheless, it is necessary to determine the six variables by matching them to empirical evidence to utilize this EOS as a refrigerant. Coolants were further constrained by the usage of GMA EOS. Typically, the crucial constants are required and several other controllable factors to construct the abovementioned relationships or equations of status. In addition to being time-consuming, there is no guarantee that the optimal set of variables can be achieved. While there are still empirical ambiguities, the establishment of statistical methods, including neural networks that can describe and reliably forecast the characteristics of coolants, offers a potential path to completing this project. Several efforts have been undertaken to calculate the thermodynamic characteristics of coolants employing artificial neural networks, for example, Chouai, Laugier, and their colleague employed ANNs to generate PVT models of coolants with temperatures ranging from 240 to 340 K and pressures up to 20 MPa. The experiment by Chouai, Laugier, and their colleague was duplicated by Laugier and Richon, considering six different types of coolants. An artificial neural network (ANN) was created by Sözen, Özalp, and their colleague across the saturated liquid vapor and also the overheated vapor zone to determine the thermodynamic parameters of a substitute coolant (R508b), including specific volume, enthalpy, and entropy [32–34].

In the past few years, the application of attractive methods of modeling and data analysis to facilitate the solution of complex problems has attracted the attention of many researchers and scholars [35–41]. Machine learning algorithms have been extensively employed in a variety of disciplines throughout the last decade [42–47]. Neuronal networks’ capacity to simulate nearly every function steadily and effectively is a significant factor in their fast development and wide range of applications. Neuronal networks are still constructed by a tedious, repetitive trial, and error technique, despite their many possible applications. Consequently, the amount of time and commitment necessary for network development depends solely on the job at the hand and the expertise of the engineer. As a result, a large quantity of time and research is wasted to determine the optimal or nearly optimal topology for a neural network considering the intended mission.

The LSSVM is applied during the building of machine learning algorithms for the purpose of computing the soaked fluid densities of purified and blended coolants throughout this research. Our goal is to achieve an algorithm for more accurate modeling and prediction of output data. Leveraging the empirical evidence, a LSSVM template was developed. In the following, using various statistical analyses, the accuracy of this model in predicting the target data has been investigated. Also, using sensitivity analysis, we determine the impact of each of the input parameters on the target parameter, so that the user can identify effective parameters and use commercial industrial applications with a wider view in a minimum of time.

2. Data Gathering

In this study, 172 data points exist in our database. Notably, the test dataset involves 43 data points (approximately 25%) and the training dataset includes 129 data points (approximately 75%) used for testing and training the considered models’ efficiency, respectively [48]. Also, the normalization of the data points was performed between +1 and −1 to enhance the considered models’ efficiency.

3. LSSVM

For pattern recognition and getting regression, the support vector machine (SVM) is an authoritative method. The function of SVM is defined [49–51] as

\[ f(x) = w^T(x)\phi(x) + b, \]

where \( b \) is the bias function, \( \phi(x) \) is the kernel function, and \( w^T \) is the output layers transposed vector. The number of input variables and data points specifies the dimension of SVM input. Calculating \( w \) and \( b \) parameters is attainable with cost function [52–54]:
Cost function \( \frac{1}{2}w^T w + \frac{1}{2}y \sum_{k=1}^{N} \xi_k^2 \). 

If the cost function is minimized, the result will be the most accurate. For the cost function presented in (2), the restriction is as follows [55–57]:

\[
y_k - \phi(x_k) - b \leq \varepsilon + \xi_k, \quad k = 1, 2, \ldots, N, \\
w^T \phi(x_k) + b - y_k \leq \varepsilon + \xi_k^*, \quad k = 1, 2, \ldots, N, \\
\varepsilon \quad \xi_k, \xi_k^* \geq 0.
\]

\( y_k \) is the output related to \( k \)th data and \( x_k \) assigns to its input. The \( c \) parameter determines the deviation from \( \varepsilon \), \( \xi_k, \xi_k^* \) (slack variables) are the sufficient margins of error, and \( \varepsilon \) is the precision of the function estimations [58, 59]. A quadratic programming problem is desirable to solve the SVM solution. Suykens and Vandewalle simplified the solving process of the SVM solution by presenting the least-squares modification of SVM [60, 61]. The suggested cost function is specified as [62, 63]

\[
\text{Cost function} = \frac{1}{2}w^T w + \frac{1}{2}y \sum_{k=1}^{N} e_k^2.
\]

subjected to

\[
y_k = w^T \phi(x_k) + b + e_k,
\]

where \( e_k \) is the error variable and \( y \) is the tuning parameter of LSSVM. For the LSSVM method, Lagrangian is described as follows (\( a_k \) is the Lagrangian multipliers) [64]:

\[
L(w, b.e.a) = \frac{1}{2}w^T w + \frac{1}{2}y \sum_{k=1}^{N} e_k^2 \\
- \sum_{k=1}^{N} a_k (w^T \phi(x_k) + b + e_k - y_k).
\]

The saddle point of Lagrangian gives the final answer to the optimization problem [65]:

\[
\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{k=1}^{N} a_k \phi(x_k),
\]

\[
\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{k=1}^{N} a_k = 0,
\]

\[
\frac{\partial L}{\partial e_k} = 0 \Rightarrow a_k = \gamma e_k, \quad k = 1, 2, \ldots, N,
\]

\[
\frac{\partial L}{\partial a_k} = 0 \Rightarrow w^T \phi(x_k) + b + e_k - y_k = 0, \quad k = 1, 2, \ldots, N.
\]

Solving the above equations gets us the LSSVM parameters. Furthermore, \( \gamma \) kernel function parameters can be used as the parameters of tuning. In this investigation, the following RBF (radial basis function) is used [66]:

\[
k(x, x_k) = \exp\left(\frac{x_k - x^2}{\sigma^2}\right),
\]

The \( \sigma^2 \) (parameter of RBF) has to be set. By considering regarding minimization of differences between real and estimated data, \( \sigma^2 \) and \( \gamma \) parameters have to be optimized [67]:

\[
MSE = \frac{1}{N} \sum_{i=1}^{N} (H_i^{\text{pred}} - H_i^{\text{act}})^2,
\]

where \( \text{pred.} \) and \( \text{act.} \) are the abbreviations of the forecasted and real data, respectively. \( N \) is the number of data points [68, 69]. It should be noted that all layers in this work are activated using the sigmoid activation function. A schematic of algorithm PSO-LSSVM is shown in Figure 1.

![Figure 1: The general process of the PSO-LSSVM approach [70].](image-url)

4. Results and Discussion

It is worth noting that evaluating the model’s capability was carried out through efficiency analysis. Hence, for examining the considered model’s capability, different statistical analyses were performed among the outputs of the model and the actual values. These statistical analyses involve R-squared \( (R^2) \), mean squared error \( (MSE) \), root mean square error \( (RMSE) \), mean relative errors \( (MRE) \), and standard deviations \( (STD) \) [71–74]. The actual values are illustrated versus the model’s outputs for the study output data at the phases of test and
train, as shown in Figure 2. The purpose is to estimate the considered model with the decent agreement among the model yields and real information and also their capability in the prediction of output becoming prominent. Besides, Figure 3 shows the results of regression analysis at the phases of test and train. According to the related literature, the statistic of $R^2$ is prominent for indicating the relationship between the actual value and model output. The main aim was to carry out a comparative analysis among real values and model yields. The precision of the considered model is enhanced by approaching the fitted line to the bisection line. If $R^2 = 1$, the linear correlation among the actual values and model outputs is remarkable, and it becomes weaker by approaching the value of $R^2$ to zero. The accuracy of models used for prediction is represented through the close-fitting of the data points around the 45-degree line. Based on this figure, this model’s ability to predict the target values in various phases is high [75–78].

Table 1 provides the statistical analysis results of the considered model according to the $R^2$, MRE, STD, and RMSE parameters [73, 79].

Moreover, absolute relative deviation among the model yields of output anticipated and actual values using the examination model is shown in Figure 4 [80–82]. The plot of William was utilized to determine the model’s outliers. The standardized residuals vs. hat values are shown in Figure 5. Based on this figure, down suspected limit, upper limit, and leverage limit are three limited

<table>
<thead>
<tr>
<th>Phase</th>
<th>$R^2$</th>
<th>MRE (%)</th>
<th>RMSE</th>
<th>STD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>1.000</td>
<td>0.182</td>
<td>0.1045</td>
<td>0.1036</td>
</tr>
<tr>
<td>Test</td>
<td>0.999</td>
<td>0.158</td>
<td>0.1077</td>
<td>0.1070</td>
</tr>
<tr>
<td>Total</td>
<td>0.999</td>
<td>0.176</td>
<td>0.1077</td>
<td>0.1044</td>
</tr>
</tbody>
</table>

Figure 2: Estimated density values compared to experimental data using the LSSVM model.

Figure 3: Regression diagram to estimate density using the LSSVM model.

Figure 4: Relative deviation (%) of testing and training data using the LSSVM model.

Figure 5: Detection of the suspicious dataset for the LSSVM model.
boundaries. The standardized residual values of outliers are more than three or less than \(-3\), and their hat is more than \(\hat{h}^*\) (known as the value of warning leverage) and beyond the considered model’s applicability domain [83]. Based on this figure, there are only three suspicious and uncertain points among the entire data points [84, 85].

Eventually, for determining various input parameters’ impact on target parameters, the sensitivity analysis was utilized [86]. More details regarding this analysis are represented elsewhere. Based on Figure 6, the direct effect of \(\text{CF}_2\text{H}_2\) on the considered target parameter is prominent, which relates to the relevancy (\(r\)) factor and is equivalent to \(+50.19\). In contrast, the effect of other input parameters on the considered target parameter is inverse, so temperature with \(r\) equivalent to \(-52.35\) has a prominent negative effect [87,88].

5. Conclusions

The main purpose of the present study was to examine the prediction of the liquid density of the refrigerant systems through a statistical model based on machine learning. For this aim, the implementation of LSSVM was performed in our model. The precision of the estimation versus the actual data points was high, and the calculated values of RMSE, MRE, STD, and \(R^2\) were 0.0116, 0.158, 0.1070, and 0.999, respectively. According to the statistical analysis results, the efficiency and performance of the considered techniques obtained among the actual values and the outputs of the model were verified through an excellent agreement in model assessment during the phases of test and train. Therefore, in contrast to the sophisticated and complex mathematical techniques expanded for predicting output, this model is recognized as a useful and efficient tool for scholars, especially in relevant fields.

Data Availability

The data used to support this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


