Research Article

Ameliorated Austenite Carbon Content Control in Austempered Ductile Irons by Support Vector Regression

Chan-Yun Yang, 1 Liou-Chun Chang, 2 Hooman Samani, 1 and Ryohei Nakatsu 3

1 Department of Electrical Engineering, National Taipei University, San Shia District, New Taipei City 23741, Taiwan
2 Mechanical Engineering and Chemical Technology, British Columbia Institute of Technology, BC, Canada
3 Department of Electrical and Computer Engineering, National University of Singapore, Singapore

Correspondence should be addressed to Hooman Samani; hooman@mail.ntpu.edu.tw

Received 3 February 2013; Accepted 5 April 2013

Academic Editor: Chang-Hua Lien

Copyright © 2013 Chan-Yun Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Austempered ductile iron has emerged as a notable material in several engineering fields, including marine applications. The initial austenite carbon content after austenization transform but before austempering process for generating bainite matrix proved critical in controlling the resulted microstructure and thus mechanical properties. In this paper, support vector regression is employed in order to establish a relationship between the initial carbon concentration in the austenite with austenization temperature and alloy contents, thereby exercising improved control in the mechanical properties of the austempered ductile irons. Particularly, the paper emphasizes a methodology tailored to deal with a limited amount of available data with intrinsically contracted and skewed distribution. The collected information from a variety of data sources presents another challenge of highly uncertain variance. The authors present a hybrid model consisting of a procedure of a histogram equalizer and a procedure of a support-vector-machine (SVM-) based regression to gain a more robust relationship to respond to the challenges. The results show greatly improved accuracy of the proposed model in comparison to two former established methodologies. The sum squared error of the present model is less than one fifth of that of the two previous models.

1. Introduction

Austempered ductile iron (ADI) is a specialty heat-treated material that takes advantage of the near-net shape technology and low-cost manufacturability of ductile iron castings to make a high-strength, low-cost, and excellent abrasion-resistant material. ADI has become an established alternative in many applications that were previously the exclusive domain of steel castings, forgings, weldments, powdered metals, and aluminum forgings and castings [1–6]. This material has been also proven to perform very well under different wear mechanisms such as rolling contact fatigue, adhesion, and abrasion [7, 8]. Considering the low-cost, design flexibility, flexible machinability, high strength-to-weight ratio and good toughness, wear resistance, and fatigue strength of ADI, its usage now is extended into marine application with increasing interest in the study of corrosion and coating of ADI [5, 9–12].

ADI is obtained by heat treating process of ductile irons to have bainite as matrix, which consists of strong bainitic ferrite platelets and tough high-carbon retained austenite, along with spheroidal graphite nodules [1–3]. The typical microstructure of ductile irons, shown in Figure 1(a), includes spheroidal graphite nodules and matrix surrounding them. The bainitic matrix of an austempered ductile iron [8] is illustrated in Figure 1(b). A significant amount of retained austenite is presented as the shape of films and blocks in the matrix.

The heat treatment for developing bainite matrix includes two steps. First, ductile irons are heated to austenitization temperature \( T_r \) around 1550–1750°F to change the original matrix into austenite and then quenched down to the bainite formation temperature range (450–750°F) for one to three hours when bainitic ferrite grows isothermally at the expense of austenite before cooling down to ambient temperature [1–3, 13, 14]. The austenitization reverses the matrix structure to high temperature austenite phase and in the meantime determines the initial carbon concentration in austenite (\( C_0^{\gamma} \)) before austempering process, since the graphite modules
in ductile irons are both a sink and source for carbon atoms. During the isothermal formation of bainite, termed as austempering process, bainitic ferrite forms in a displacive manner at the expense of austenite and partitions excessive carbon into surrounding austenite, which is gradually enriched during the process. The transformation process stops when the carbon content of austenite reaches a certain level which is impossible for the transformation to proceed further thermodynamically, before all austenite is consumed [14–16].

2. Role of Austenization Temperature

The mechanical properties of ADI obtained from a given ductile iron are closely related to the microstructure, which can be controlled by manipulating austenization and austempering temperatures [1–5]. The initial carbon content in austenite ($C_0^\gamma$), which is dictated by austenization temperature ($T_\gamma$), has two significant consequences for the final microstructure of ADI. Firstly, $C_0^\gamma$ affects the choice of austempering temperature since the temperature range for bainite formation is a strong function of carbon and, to a lesser extent, other alloy elements [13, 14, 17]. Furthermore, the austempering temperature is the most important aspect for controlling ADI’s mechanical properties since the nature of the bainitic ferrite formed at different temperatures is variable. Additionally, at a given austempering temperature, higher $C_0^\gamma$ will result in a lower volume fraction of bainite which can be formed during the austempering process [18, 19]. Less bainite formation would lead to more retained austenite in the final microstructure as well as less carbon enrichment in retained austenite, resulting in more blocky shape of retained austenite, which is mechanically unstable and thus detrimental to mechanical properties [14].

There are two established empirical formulas for estimating $C_0^\gamma$. First one involves only austenization temperature and silicon content [3] as follows:

$$C_0^\gamma = \frac{T_\gamma}{420} - 0.17 \ (wt\%Si) - 0.95,$$

where $T_\gamma$ is in °C, and $wt\%Si$ denotes the weight percentage of the silicon content. The other one includes other common alloy contents [20] as follows:

$$C_0^\gamma = 1.61 \times 10^{-6}T_\gamma^2 + 3.35 \times 10^{-4}T_\gamma + 0.006 \ (wt\%Mn) - 0.11 \ (wt\%Si) - 0.07 \ (wt\%Ni) + 0.014 \ (wt\%Cu) - 0.3 \ (wt\%Mo) - 0.435,$$

where $wt\%Mn$, $wt\%Si$, $wt\%Ni$, $wt\%Cu$, and $wt\%Mo$ denote the weight percentages of manganese, silicon, nickel, copper, and molybdenum contents, respectively. It has been indicated that both formulas can only achieve limited accuracy compared to experimental results [5]. The unsatisfactory results of the models are not unexpected due to not only the rarity of available data but also their unspecified variance resulted from the different instruments and measuring methods employed in the respective sources of generating the data. It is clear, nonetheless, that the formulas employing merely linear multivariate regression are incapable of producing an accurate model. Moreover, an examination of the data points shows that the distributions of the corresponding features are seriously contracted and skewed. To gain a more reliable accuracy, an inverted lognormal histogram equalizer and a support-vector-machine-(SVM-) based regression are introduced to manipulate the data points. The histogram equalizer, as a preprocessor, is used to reduce the irregularity and enhance the contrast in the distribution of the data points. After equalizing the attribute contrast, an SVM-based regression is employed to fit the input with highly uncertain variance for a noise-tolerable relationship between the austenitization temperature, $T_\gamma$, and the alloy content with the initial carbon concentration, $C_0^\gamma$. To introduce the model, the procedure of an inverse lognormal histogram equalizer for manipulating input data is outlined first, followed by the procedure of an SVM-based regression model. Finally, the established model and its prediction results are presented and compared with the previous models.
3. Regression Modeling

3.1. Histogram Equalization Preprocessor. In this study, the histogram equalizer is employed in order to derive a balanced contrast of the regression input attributes. Histogram equalizer is often utilized to increase the global contrast of two-dimensional digital images. Through a mapping operation, the corresponding intensities of an image can be distributed expansively on the corresponding histogram, and thus the discrimination of the details could increase. The characteristics of the enhanced expansiveness are therefore conducive to the applications in the areas such as X-ray, thermograph, and face detection images [21, 22]. The present study adopted the two-dimensional mapping to be one-dimensional case to influence the input attributes separately which are chemical compositions and austenite with austenization temperature and alloy contents. Due to a general assertion that skewed data distributions in nature favor a lognormal distribution [23], a lognormal distribution is asserted for fitting the attributes. An ideal histogram equalizer is designated to generate a uniform distributed histogram after the equalization. An inversed lognormal histogram equalizer, mapping the attributes from lognormal to uniform distribution, is, hence, employed to deal with the attributes’ imbalanced tendency.

The lognormal distribution is as follows:

\[
T(\rho) = \frac{1}{\rho \sqrt{2\pi \sigma_{\rho}^2}} \exp \left( -\frac{(\ln \rho - \mu_{\rho})^2}{2\sigma_{\rho}^2} \right) ,
\]

where \( \mu_{\rho} \) and \( \sigma_{\rho}^2 \) are the natural logarithms of average and variance of corresponding \( \theta \), respectively. The inverse lognormal histogram equalizer can be given as

\[
\rho = T^{-1}(\theta).
\]

The mapping is a conversion from lognormal distributed \( \theta \) to uniform distributed \( \rho \) by the inversed function \( T \). The uniform distributed \( \rho \) is seeking wider and more balanced global contrast which is advantageous to the generation of the fitted function.

To compose a set of input features which contribute equivalently in the subsequent regression, the attributes \( \rho \) should furthermore be normalized as follows:

\[
x_i = \frac{\rho_i - \rho_{\text{min}}}{\rho_{\text{max}} - \rho_{\text{min}}}, \quad i = 1, \ldots, 6.
\]

3.2. Support Vector Regression. The SVM-based regression, known to deal with the generalized model of complex uncertain relationships, has been gaining popularity and has shown much improved results [24–26], albeit only few experimental data points, also called observations here, are available. The present study is aimed to employ the support vector regression (SVR) to obtain a more accurate relationship model between the initial carbon concentration in austenite with austenization temperature and alloy contents. Advantageous of the structural risk minimization [27, 28], the SVM-based method for regression [29–32], similar to that for classification [33, 34], simultaneously minimizes both the model complexity and empirical errors, and in turn creates a predictor with a wide margin. In classification, the wide margin represents a high generalization capability for separating unlabeled samples. On the other hand, the wide margin in regression represents a smooth approximation function in which variance from noises will be rejected as much as possible. In contrast to traditional statistical or ANN regression model which derives the approximation function by minimizing the training error between observed and corresponding predicted responses, the SVR attempts to minimize a generalization error which combines the training error and a regularization term to control the model complexity. The generalization error mainly rejects the highly variant noises and achieves a rigid regression.

The SVR is intrinsically a kernel-based method [35]. With a given learning set \( S = \{(x_1, y_1), \ldots, (x_d, y_d)\} \), an approximated function \( \hat{y} = f(x) \) can be established for further prediction. In \( S, x_i \) denotes the \( d \)-dimensional input vector, \( x_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^T, x_i \in \mathbb{R}^d \), and \( y_i \) denotes the corresponding target value of input \( x_i \), \( y_i \in \mathbb{R} \). By using the \( \epsilon \)-insensitive loss function (Figure 2(a)) to regularize the degree of rigidness, the optimized \( f \) can include all input \( x_i \) within the boundary of \( \pm \epsilon \) deviation while keeping the boundary (a tube in space) as straight as possible (Figure 2(b)).

\[y = \langle w, x \rangle + b + \epsilon\]

\[y = \langle w, x \rangle + b - \epsilon\]
The essentially regularized rigidness is beneficial to SVR in finding an optimized generalization for the regression. By introducing the kernel trick [34, 35], the regression function can be described as \( f(x) = \kappa(w, x) + b \), where \( w \) denotes a weight vector, \( w = [w_1, w_2, \ldots, w_d]^T \), \( w \in \mathbb{R}^d \), \( b \) denotes the bias term, and \( \kappa(\cdot, \cdot) \) denotes a kernel function. Here, the kernel function is adopted to deal with the nonlinearity of the regression. Putting the elementary features together, the fitting of SVR can then be formally expressed as a primal convex optimization problem as follows:

\[
\min_{\mathbf{w}, \xi} \frac{1}{2} \| \mathbf{w} \|^2 + \lambda \sum_{i=1}^{l} (\xi_i + \xi_i^*),
\]

subject to

\[
y_i - \kappa(\mathbf{w}, \mathbf{x}_i) - b \leq \varepsilon + \xi_i,
\]

\[
\kappa(\mathbf{w}, \mathbf{x}_i) + b - y_i \leq \varepsilon + \xi_i^*,
\]

\[
\xi_i, \xi_i^* \geq 0, \quad i = 1, \ldots, l,
\]

where \( \lambda \) denotes the regularization factor, and slack variables \( \xi = [\xi_1, \xi_2, \ldots, \xi_l]^T \) and \( \xi^* = [\xi_1^*, \xi_2^*, \ldots, \xi_l^*]^T \) are introduced for allowing errors to cope with some infeasible constraints in the optimization and form a soft margin. Parameter \( \varepsilon \) associated with the \( \varepsilon \)-insensitive loss function, \( \varepsilon(\xi) = \max(0, |\xi| - \varepsilon) \), controls error tolerance of the regression. The loss function defines the \( \varepsilon \)-tube which carries out the rigidness of the approximated function. So, the parameter \( \varepsilon \) affects the smoothness of the induced regression and the number of support vectors as well. On the other hand, the parameter \( \lambda \) controls the tradeoff between keeping the straightness of subsequent \( f \) and limiting the deviations to be less than \( \varepsilon \). Let \( \mathbf{a} = [\alpha_1, \alpha_2, \ldots, \alpha_l]^T \) and \( \mathbf{a}^* = [\alpha_1^*, \alpha_2^*, \ldots, \alpha_l^*]^T \) be the Lagrange multipliers for the first two sets of constraints in (7), and take Lagrange of the primal problem of (6)-(7). The Wolfe dual [36] can be obtained by differentiating the Lagrange with respect to \( \mathbf{w}, b \) as follows:

\[
\max_{\alpha, \mathbf{a}^*} - \frac{1}{2} \sum_{i,j=1}^{l} (\alpha_i - \alpha_j) (\alpha_j - \alpha_j) \kappa(\mathbf{x}_i, \mathbf{x}_j)
\]

\[
- \varepsilon \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i (\alpha_i - \alpha_i^*),
\]

subject to

\[
\sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0, \quad \alpha_i, \alpha_i^* \in [0, \lambda], \quad \forall i.
\]

The functions in (8) and (9) become a quadratic optimization problem. The optimized \( \mathbf{a}^* \) and \( \mathbf{a}^* \) can therefore be obtained after the optimization procedure. To take advantage of the sparseness of support vectors [32, 35], only those \( \mathbf{x}_i \)'s with nonzero \( \alpha_i^* \)'s, called support vectors (SVs), are taken into account to form the consequent \( f \). With the SV, weight vector \( \mathbf{w} \) can be computed by \( \mathbf{w} = \sum_{\mathbf{x}_i \in \mathbf{SV}} (\alpha_i^* - \alpha_i) \mathbf{x}_i \), and therefore

\[
f(x) = \sum_{\mathbf{x}_i \in \mathbf{SV}} (\alpha_i^* - \alpha_i) \kappa(\mathbf{x}_i, \mathbf{x}) + b.
\]

Several kernel functions have been introduced for SVR, including linear, polynomial, and Gaussian kernel [32, 35]. A straightforward way of selecting kernel function is to select the one which can reflect the natural tendency of the distributed data. In the study, a Gaussian kernel

\[
\kappa(\mathbf{x}_i, \mathbf{x}_j) = e^{-|\mathbf{x}_i - \mathbf{x}_j|^2/2\sigma^2},
\]

where \( \sigma \) denotes the width parameter of its corresponding basis function and was employed to adapt the nonlinearity of the present problem.

3.3. Flowchart of Proposed Model. Following the steps mentioned earlier, a flowchart for the whole proposed model is illustrated in Figure 3. Two steps, "data preparation" and "function approximation," are enclosed mainly in the model. With the gathered input raw data, the model can generate an approximated function \( f(x) \) for estimating \( C_y^0 \) automatically.

4. Results and Discussions

Forty-two experimental data points of \( C_y^0 \) were collected from the literature [2, 37–44] for the study. The data points containing six original attributes, including \( \theta_1(T_1^0) \), \( \theta_2(\text{wt}\%\text{Si}) \), \( \theta_3(\text{wt}\%\text{Mn}) \), \( \theta_4(\text{wt}\%\text{Mo}) \), \( \theta_5(\text{wt}\%\text{Ni}) \), and \( \theta_6(\text{wt}\%\text{Cu}) \), were recorded to establish the relationship \( C_y^0 = f(\theta_1, \theta_6) \) for prediction. The corresponding metrics of mean, variance, and skewness of the original attributes were determined, as listed in Table 1. The variance and skewness found are surprisingly undesirable, given the fact that these data points are collected from different sources, in which different instruments and measuring methods were employed.

To enhance the discriminative information contained in the attributes and to ensure their satisfactory global contrast, the inverse lognormal histogram equalizer was applied, mapping original attributes \( \theta_i \) to equalized attributes \( \rho_i \). Figure 4 shows the normalized histograms of the attributes before and after the equalization. Corresponding metrics of the equalized attributes compared to those of the original attributes are also included in Table 1. This table illustrates that the skewness of the attributes, except for \( T_1^0 \), are significantly reduced. With their values close to zero, the remapped attributes are less skewed and more evenly distributed in Figure 4. Moreover, the histogram examination of the attributes' global contrast is followed by reviewing the panels in Figure 4. Compared with the equalized attributes \( \rho_1, \rho_2, \) and \( \rho_3 \), the original attributes \( \theta_1, \theta_2, \) and \( \theta_3 \) are more widely spread in the global contrast. The unexpected spread may be due to the overequalization. To fulfill the objective of the study, a combination of the original and the equalized attributes with wider spread contrast was selected to make up the SVR input. Consequently, three original attributes \( \theta_1, \theta_2, \)
Table 1: Corresponding metrics of the attributes before and after the inversed lognormal histogram equalization.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>𝜃_1</td>
<td>907.00</td>
<td>2285.70</td>
<td>1.38</td>
<td>0.14</td>
<td>0.07</td>
<td>2.79</td>
</tr>
<tr>
<td>𝜃_2</td>
<td></td>
<td>2.56</td>
<td>0.08</td>
<td>0.28</td>
<td>0.17</td>
<td>1.15</td>
</tr>
<tr>
<td>𝜃_3</td>
<td>0.31</td>
<td>0.04</td>
<td>1.16</td>
<td>0.41</td>
<td>0.22</td>
<td>0.42</td>
</tr>
<tr>
<td>𝜃_4</td>
<td>0.16</td>
<td>0.06</td>
<td>1.21</td>
<td>0.35</td>
<td>0.11</td>
<td>0.71</td>
</tr>
<tr>
<td>𝜃_5</td>
<td>0.29</td>
<td>0.28</td>
<td>1.78</td>
<td>0.46</td>
<td>0.03</td>
<td>0.72</td>
</tr>
<tr>
<td>𝜃_6</td>
<td>0.18</td>
<td>0.18</td>
<td>2.87</td>
<td>0.71</td>
<td>0.09</td>
<td>−0.50</td>
</tr>
</tbody>
</table>

and 𝜃_3, together with three equalized attributes 𝜌_4, 𝜌_5, and 𝜌_6, were chosen for further regression.

The attributes, consisting of the equalized 𝜌_4–𝜌_6 and those original sets 𝜌_1 = 𝜃_1, 𝜌_2 = 𝜃_2, and 𝜌_3 = 𝜃_3, were then normalized into the range [0, 1] for consistent contribution to the learning process and form the input vector \( x = [x_1, x_2, \ldots, x_9]^T \).

There are three adjustable parameters \( \sigma, \varepsilon \), and \( \lambda \) during the SVR learning phase. To calibrate the parameters for an optimized model, the method of cross-validation (CV) has been undertaken. Following the cross-validation, the dataset was randomly partitioned into three groups, including 21 data points for training set, 11 data points for validation set, and 10 data points for test set. By the independence of the training and validation datasets, the cross-validation was taken through to pursue the lowest generalization error and obtain the corresponding optimized parameters, namely, \( \sigma^*, \varepsilon^*, \) and \( \lambda^* \). The model parameterized by \( \sigma^*, \varepsilon^*, \) and \( \lambda^* \) is the most generalized model for the prediction of \( C_0^{\gamma} \). As presented in the previous section, the generalized model is resistant to the input noises. To select the most generalized model, the sum squared error and mean squared error

\[
\text{SSE} = \sum_{i=1}^{l_{\text{set}}} (\hat{y}_i - y_i)^2, \quad \text{MSE} = \frac{1}{l_{\text{set}}} \sum_{i=1}^{l_{\text{set}}} (\hat{y}_i - y_i)^2
\]

are adopted to evaluate the errors in the cross-validation, where \( l_{\text{set}} \) denotes the set-length of the chosen set, and \( y_i \) and \( \hat{y}_i \) denote observed and corresponding predicted output responses, respectively.

Since there are three parameters for tuning, adaptively searching the whole parameter-space is inefficient and may converge in a slower rate. In this study, a grid-search method incorporated with priority steps was adopted to speed up the searching for the optimal solution. The width parameter, \( \sigma \), for the basis function determines the nonlinear transformation of the input data. In general, the larger the width is, the more linearity the induced model will be. The parameter \( \lambda \) is a regularization factor for controlling the tradeoff between the training error and complexity of the induced model. A larger \( \lambda \) produces more penalties on the training error and induces a higher complexity for the model [35]. These two parameters, \( \sigma \) and \( \lambda \), are dominant parameters in SV machine learning.
Figure 4: Histograms of attributes before ($\theta_1$–$\theta_6$) and after ($\rho_1$–$\rho_6$) the histogram equalization. For better comparison, all the attributes are normalized for chartings.
and are designated to be determined first in the study. The ranges for $\sigma$ of $[5–25]$ and $\lambda$ of $[10^{-3}–10^{-1}]$ were selected for the grid-searching of $(\lambda, \sigma)$ after some preliminary tests. With the $\sigma^*$ and $\lambda^*$ being determined, the cross-validation then seeks the optimal $\epsilon^*$ to achieve the most generalized model. In this study, a relative small value, ranging between $[10^{-3}–10^{-1}]$, was specified for seeking $\epsilon^*$.

Figure 5 shows the cross-validated SSEs corresponding to the changes of $\sigma$, $\lambda$, and $\epsilon$. From the panels, $\lambda^*$, $\sigma^*$, and $\epsilon^*$ were chosen as $10^{7}$, 20, and $5 \times 10^{-3}$, respectively, with the minimal SSE of $4.55 \times 10^{-2}$. With these optimized parameters, the SVR model was then taken to certify the prediction capability by the standalone test dataset.
As indicated by the unit-slope graphs in Figure 6, which illustrates the discrepancies between predicted and experimental data, the SVR model established with the chosen parameters shows much improved accuracy compared with the two previous models shown in (1) and (2). The MSEs of the three models, not only with the validation dataset but also with the test dataset, are also detailed in Table 2. For example, the error of the test dataset for the present SVR model is only about one third to one fifth of the two previous models.

The present model illustrates a more accurate prediction of the initial carbon concentration in austenite after austenization ($C_0^γ$) with $T^γ$ (austenization temperature) and alloy contents in austempering processes. The results indicate that SVM regression greatly improved the accuracy of $C_0^γ$ prediction in comparison to two established equations using linear regression. Overall accuracy (sum squared error) of the present method is five and eight times of those of the two previous models, respectively. A better control of $C_0^γ$ has been proven to be critical in achieving desired microstructures and mechanical properties for ADI, which has been applied, among numerous fields, for many marine applications.

The present study also demonstrates the possibility of employing a similar procedure to deal with contracted and skewed observations with highly uncertain variance. SVR, characterized by highly noise resistant as well as flexible in compromising between accuracy and complexity, is one of the suitable algorithms to deal with observations collected from multiple sources, in which instruments and measurements vary wildly.

### Acknowledgments

This work was supported by the grants from National Science Council of Taiwan, under Contracts NSC 100-2218-E-149-002 and NSC 99-2221-E-149-009. The authors would like to thank Hsin-Liang Tai for his great contribution to the paper, Professors Jui-Jen Chou and Daw-Kwei Leu for their useful comments for the studies, and Mohammad Arif for his help in formatting this paper.

### References


