Methods of development fuzzy logic driven decision-support models in copper alloys processing

S. Kluska-Nawarecka\textsuperscript{a}, Z. Górn\textsuperscript{a}, B. Mrzygł\textsuperscript{b}, D. Wilk-Kołodziejczyk\textsuperscript{c}, K. Regulski\textsuperscript{b}*  
\textsuperscript{a} Polish Foundry Research Institute, Zakopiańska 73, Krakow, Polska  
\textsuperscript{b} Department of Industrial Computer Science, Faculty of Metals Engineering and Industrial Computer Science, AGH, Mickiewicza 30, Kraków, Polska  
\textsuperscript{c} The Andrzej Frycz Modrzewski Krakow University College, Kraków, Polska  
* Corresponding author. E-mail address: regulski@metal.agh.edu.pl

Received: 26.02.2010; accepted in revised form: 30.03.2010

Abstract

Development of a diagnostic decision support system using different then divalent logical formalism, in particular fuzzy logic, allows the inference from the facts presented not as explicit numbers, but described by linguistic variables such as the "high level", "low temperature", "too much content", etc. Thanks to this, process of inference resembles human manner in actual conditions of decision-making processes. Knowledge of experts allows him to discover the functions describing the relationship between the classification of a set of objects and their characteristics, on the basis of which it is possible to create a decision-making rules for classifying new objects of unknown classification so far. This process can be automated. Experimental studies conducted on copper alloys provide large amounts of data. Processing of these data can be greatly accelerated by the classification trees algorithm which provides classes that can be used in fuzzy inference model. Fuzzy logic also provides the flexibility of allocating to classes on the basis of membership functions (which is similar to events in real-world conditions). Decision-making in foundry operations often requires reliance on knowledge incomplete and ambiguous, hence that the conclusions from the data and facts may be "to some extent" true, and the technologist has to determine what level of confidence is acceptable, although the degree of accuracy for specific criteria is defined by membership function, which takes values from interval \(<0,1>\). This paper describes the methodology and the process of developing fuzzy logic-based models of decision making based on preprocessed data with classification trees, where the needs of the diverse characteristics of copper alloys processing are the scope. Algorithms for automatic classification of the materials research work of copper alloys are clearly the nature of the innovative and promising hope for practical applications in this area.

Keywords: Application of Information Technology to the Foundry Industry, fuzzy logic, copper alloys, classification algorithms, classification and regression trees, C&RT, decision-making rules

1. Introduction

The theory of fuzzy sets and fuzzy logic, author of which was L.A. Zadeh (1965) \cite{1} helped to create models of inference, where we do not have sufficient knowledge of the phenomena which allows the creation of a precise, deterministic mathematical model. Expressed as linguistic variables approximation turns out to be sufficient to control process \cite{2,3,4,5}, where it seems to be the most popular. Much less it works in decision support applications of technology in terms of incomplete and unreliable information on the course of developments in the heat treatment process for elements made of bronze BA1032.
The article describes how to create a fuzzy model based on input from preprocessing algorithm of decision trees (C&RT) to obtain the fuzzy sets, which will be the base for the decision rules. This is the original methodology, which allows to extract knowledge about relationship between the variables of the process from experimental data. Decision trees allow you to apply in situations where a priori knowledge is very small. Previously, to create a fuzzy model, it was necessary to develop rules and a fuzzy sets by an expert field, which was often costly and time consuming. Application of C&RT algorithms, tools from family of data mining, you can get this knowledge automatically.

2. Data mining using classification and regression trees (C&RT).

Analysis is based on the results of tests carried out on four heats of bronze BA1032 (E, F, G, K) under various types of modification (heat E - without modifying, F - modified calcium [Ca], G - modified potassium [K], heat F - modified boron [CuB3]). For total the results for 60 samples were taken under analysis.

Individual samples were subjected to different types of heat treatment:
- H1 - hardening at 950 °C, cooling in water (fast)
- H2 - hardening at 950 °C, cooling in oil (average)
- H3 - hardening at 950 °C, cooling the air (free)

and some of them also:
- S1 - ageing at 500 °C, cooling with the furnace
- S2 - ageing at 500 °C, cooling the air
- S3 - ageing at 700 °C, cooling with the furnace
- S4 - ageing at 700 °C, cooling the air

Changes in mechanical properties were the results of quenching. In search of dependence were taken into account the properties obtained for each sample:
- \( R_m \) [MPa] - strength
- \( R_{0.2} \) [MPa] - conventional yield limit
- A [%] - extension

The purpose of preliminary tests was to find relationships between different types of quenching, and in particular to determine which processes has the most important influence on the desirable mechanical properties.

To acquire this knowledge the algorithm of Classification and Regression Trees (C&RT) was used. Algorithm was popularized by L. Breinman (1984) [6] and others [7,8]. The aim of the model is to set the value of individual classes of dependent variables \( (R_m, R_{0.2}, A) \) formed on the basis of the conduct of the variability of the predictor variables (rate of cooling after hardening, type of modifier, temperature of ageing, the ageing rate of cooling). This will allow for the construction of the rules of inference based on four explanatory variables.

We developed C&RT model using STATISTICA 8.0 (Fig. 1). Cooling rate at hardening or ageing temperature is continuous variable, but the specifics of the process and characteristics of the collected data indicates that the selection was punctual (eg 500°C, 700°C and cooling in air, water, etc.). Therefore we treated these values as categories in qualitative variables in regression trees modeling. As a result, we use a regression tree model in which:
- each arc represents a test,
- each node (leaf) represents a single class.

Tree-building algorithm is based on dividing a set of learning objects (samples) on the partition to the point at which each partition contains the data belonging to one class. Division vertices in the regression trees is based on the criterion of Least Significant Difference (LSD)

\[
R(t) = \frac{1}{N_v(t)} \sum_i w_i (y_i - \bar{y}(t))^2
\]

where
- \( N_v(t) \) - weighted number of cases in node \( t \)
- \( w_i \) - the value of weighing variable for the case \( i \),
- \( f_i \) - frequency variable
- \( y_i \) - the value of the variable of response
- \( \bar{y}(t) \) is a weighted average of the node \( t \).

Application of the algorithm C&RT should lead to the creation of such a model tree, for which the predictive ability is greatest, that is classification accuracy is the highest, that is the variance for each class is the smallest. The cost is used to assess models. Need of minimize costs, not the rate of misclassification, follows that some classification errors can be more disastrous than others [8]. In our case, we use two types of costs: the cost of cross-validation (CV) and the cost of resubstitution. V-fold cross-validation is a viable method for small learning data sets, where it is difficult to separate the test sample. \( V \) sub-samples are selected from the learning sample and the test is carried out. All cases of data are divided among \( v \) groups \( Z_1, Z_2, ..., Z_v \) of the same cardinality, as far as possible. Assessment of the accuracy of classification is the proportion of cases in \( Z \) group among the wrongly classified by the model built on the cases of the \( Z - Z_v \).

\[
R^v(d^{(v)}) = \frac{1}{N_v} \sum_i X(d^{(v)}(x_i) \neq j_v)
\]

where \( d^{(v)}(x) \) is calculated for the sample of \( Z - Z_v \)

The cost of cross-validation is an essential tool for the selection of a tree. One selects the tree with minimal CV cost, or the least complex tree, which CV costs do not differ "significantly" from the minimum [6].
Auxiliary tool in selecting and evaluating appropriate decision tree is the resubstitution cost. It is the proportion of cases wrongly classified by the classifier model built on the basis of all cases. Expected square error is calculated, based on the prediction of continuous dependent variable.

$$R(d) = \frac{1}{N} \sum_{i=1}^{N} (y_i - d(x_i))^2$$

(3)

where the learning sample \( Z \) consists of points \( (x_i, y_i), i=1,2, ..., N \). This measure is calculated for the same data set, which was built based on partition \( d \).

Using a cost index, you can choose the best decision tree for each of the three models \( (R_m, R_{0.2}, A) \) (Fig. 2).

In this way, we created three regression trees for dependent variables \( (R_m, R_{0.2}, A) \) (Fig. 3). All objects (samples) were categorized according to the value of output depending on the stages of heat treatment. These classes will be applied in subsequent stages of construction of the fuzzy model.

Constructed and selected tree allows you to create rules. Interpretation of the tree is straightforward: for each leaf (conclusions) are following all subsequent branches (arcs graph). Encountered each vertex represents a test, and so is the basis for creating the conditions for the rule. And so, on the basis of tree No. 9 for \( R_m \) you can specify rules:

- If the sample not modified with boron (K) was subjected to hardening (H2 or H1), then you will have the strength distribution of the average \( E(X) = 600 \text{ MPa} \) and variance \( D^2(X) = 325 \)

<table>
<thead>
<tr>
<th>ID</th>
<th>Average</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID4</td>
<td>793.25</td>
<td>28.16</td>
</tr>
<tr>
<td>ID5</td>
<td>33.46</td>
<td>5.78</td>
</tr>
<tr>
<td>ID14</td>
<td>42.88</td>
<td>6.55</td>
</tr>
<tr>
<td>ID15</td>
<td>2187.03</td>
<td>46.77</td>
</tr>
<tr>
<td>ID13</td>
<td>324.82</td>
<td>18.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Average</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID4</td>
<td>128.32</td>
<td>11.33</td>
</tr>
<tr>
<td>ID5</td>
<td>426.79</td>
<td>20.66</td>
</tr>
<tr>
<td>ID12</td>
<td>1002.18</td>
<td>31.66</td>
</tr>
<tr>
<td>ID13</td>
<td>137.24</td>
<td>11.71</td>
</tr>
<tr>
<td>ID20</td>
<td>1024.64</td>
<td>32.01</td>
</tr>
<tr>
<td>ID21</td>
<td>40.63</td>
<td>6.37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Average</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID2</td>
<td>29.22</td>
<td>5.41</td>
</tr>
<tr>
<td>ID3</td>
<td>14.132</td>
<td>3.76</td>
</tr>
</tbody>
</table>

Similarly, an analysis of the trees for the remaining dependent variables is leading to achieve a set of further rules. This solution allows you to specify the classes of mechanical properties on the basis of the quenching and tempering processes. Such a model is important in terms of determining the technology and process technology for the production of castings in bronze BA1032.

In the daily operation of the user may be more practical inverse model, namely, corresponding to the question "How should I conduct heat treatment to obtain the mechanical properties of the \( R_m = x, R_{0.2} = y, A = z \)?".

Inverse solution can be made on the basis of the classification tree C&RT. Dependent variable in this case is the qualitative variable, and the predicate is quantitative variable. Such a model will have, however, some significant drawbacks:

- for each stage of processing the samples will be set up a separate tree, which means that even for "dependent" processes as ageing there will be model developed without taking into account other...
inherently related processes. This situation may lead to the creation of contradictory rules. For example: for the ageing will be fixed rule saying that to get some value of extension should be used ageing in 500ºC, or lack thereof, this rule will be true only under the assumption that it has not been carried out hardening H3. If we consider these processes separately, we can not control these dependencies.

- difficulty associated with the expression in a sufficiently expressive form rules in the form of a histogram: for example, a fragment of classification tree for the class number 20 (ID = 20). Its the size is 3 (N = 3). This class is obtained from objects for which \( R_m \leq 606.39 \). Inside the class, frequency histograms of samples from the quenching H1 and H2 indicate that “in order to obtain \( R_m \) less than or equal to 606.4, should be primarily used hardening H1, possibly (though probably less) hardening H2.

To avoid the aforementioned disadvantages associated with the construction of inverse model using C&RT trees, it was decided to use the tool, which is fuzzy logic.

3. Solution of the inverse model using fuzzy logic.

Model was constructed with the environment MATLAB 7.4 (R2007a) Fuzzy Logic Toolbox (Fig. 5). We used model of Mandani inference [9] which is implemented in the module. Fuzzy logic inference model consists of three steps: fuzzyfication, inference and defuzzyfication (Fig. 4).

The final step is defuzzification. It aims to transform the output fuzzy set to a specific value of the actual output of the model. There are many ways of defuzzification Methods most commonly used are: center of maximum, center of gravity and center of sums. The proposed model uses the method of center of gravity (centroid).

Distributions of membership functions adopted for the rate of cooling at hardening stem from the assumption that in the one resort cooling conditions remain unchanged, blurring the borders of the sets (resort boundary) is the result of cooling speed changes depending on the modification of the resort.

As a membership function for the ageing temperature was chosen Gaussian distributions, as the ageing temperature is exactly 500ºC or 700ºC only at one point scale. Surrounding points are only "belonging" to a fuzzy set "to some extent."

Trapezoidal membership functions for the modifier type reflect in best way the character of this variable. Fuzzy sets in this case are similar to those of bivalent logic - a sample is modified or not.
The course of reasoning can be illustrated by diagram showing the rules in schematic form and pointing by color rules that have been activated (Fig. 8).

The ovals indicate inputs and the cumulative fuzzy set for the result of inference, and defuzzified output parameters. Each rule corresponds to one row of matrix, the columns are the fuzzy representation of individual variables. There are defined values of the membership function to fuzzy sets (intersection of red lines to the graph of the membership function) for a given numeric value of variable input (vertical red line). According to the figure of a rule, the value of its left side (condition) is determined by applying the operator MIN so, only those rules for which the value of membership function of all variables are greater than 0 (variables are highlighted in yellow) are active.

According to the received value of the left side of active rules, values of membership functions of output variables (conclusions) are set. Those are represented by the blue columns of the matrix. These results are integrated using the MAX operator. The final result (the lowest position of the last column), subject to defuzzification by applying the Centroid operator, equivalent to coordinate the center of gravity obtained by integration solid (marked red line).

The user can also apply to the inference surfaces created with the possible results of inference on the basis of rules (Fig. 9). Similar surface may be determined for each pair of parameters, and each output variable. Presented figures include the combination of variables which revealed a statistically significant relationship: $R_m - R_{0.2}$; $R_m - A$; $A - R_{0.2}$.
4. Conclusions

The article presents two ways to create models of inference. Based on the algorithms of Classification and Regression Trees C&RT (STATISTICA environment), and based on a Mamdani model of fuzzy logic (environment: MATLAB Fuzzy Logic Toolbox). The first one served as a preliminary analysis of the input material (data from physical measurements), leading to the creation of the classification as a basis for building a fuzzy model.

As shown, when input variables are qualitative in nature (the way of the experiment favored a strong discretization of measurement data), while the output variables are quantitative - regression trees algorithm gives very satisfactory results. For the inverse problem fuzzy logic model was used, so that solved the problem of ambiguity response in the rules of inference, when the output variable is a qualitative.

Should be noted that the input data (in the inverse problem), samples were unevenly distributed in space of the system. However, since the specifics of the experiment indicates that one has some guarantee that the modeled system will be practically in most cases approximately in the states determined by measuring samples, thus this is not a problem for the practice of the model. [4]

Acknowledgments

References