ARCHIVES of FOUNDRY ENGINEERING

Published quarterly as the organ of the Foundry Commission of the Polish Academy of Sciences

ISSN (1897-3310)
Volume 13
Special Issue
2/2013

70 – 73

13/2

Applying Rough Set Theory for the Modeling of Austempered Ductile Iron Properties

A. Kochański *, A. Soroczyński, J. Kozłowski
Department of Plastic Forming and Foundry Engineering, Warsaw University of Technology,
Narbutta 85, 02-524Warszawa, Poland
*Corresponding author. E-mail address: akochans@wip.pw.edu.pl

Received 30.07.2013; accepted in revised form 09.09.2013

Abstract

The article discusses the possibilities of employing an algorithm based on the Rough Set Theory for generating engineering knowledge in the form of logic rules. The logic rules were generated from the data set characterizing the influence of process parameters on the ultimate tensile strength of austempered ductile iron. The paper assesses the obtained logic rules with the help of the rule quality evaluation measures, that is, with the help of the measures of confidence, support, and coverage, as well as the proposed rule quality coefficient.

Keywords: Data mining, Knowledge rules, Rough Set Theory, Casting

1. Introduction

Austempered ductile iron is a material which competes with materials so diverse as aluminum [1] or cast steel [2]. Very good mechanical properties in combination with relatively good elongation make it possible to employ this cast iron as a replacement material in many practical uses, in the car industry [3], in mining [4] or in food processing [5]. Very good mechanical properties are obtained thanks to appropriately selected heat treatment parameters. The two-stage heat treatment consists of the following processes: the austenitizing and the austempering. Austenitizing temperatures fall within the 850^0C to 1050^0C range. Two kinds of austenitizing may be differentiated: high temperature austenitizing within the temperature range of 950^0C÷1050^0C and low temperature austenitizing within the temperature range of 850^0C÷950^0C. Austempering temperatures typically fall within the range from 250^0C to 400^0C, although non-typical solutions may also be found [6]. Also in this case the two temperature ranges may be differentiated: the high temperature range of 325^0C÷400^0C and the low temperature range of 250^0C÷325^0C. Despite the fact that all process parameters are continuous values, in industrial conditions discrete values are, nevertheless, used: for instance, low carbon content; low, medium, and high silicone content; or low austempering temperature. This kind of approach suggests the employment of modeling tools which use discrete values for constructing decision rules. The Rough Set Theory put forward in the 80s by Pawlak [7] makes it possible to generate logic rules from large data sets. Rules derived by induction with the use of the Rough Set Theory have been employed in solving important problems within the area of fault diagnosis [8], as well as in the manufacturing process control [9]. The possibilities of using the RST for the modeling of complex processes, as well as a comparison between the models under consideration and the decision tree models with respect to the modeling quality are discussed in [10].

2. The state of the art

Logical (decision) rules take the following form:

IF attribute 1 = … AND attribute 2 = …, THEN output = …

The rule formulated in this way has its conditional constituent, containing the conjunction of conditional attributes which describe observations and which are located on the left side of the word „THAN", and its decision constituent, which characterizes the decision class is satisfied by the conditional constituent of the rule and which is located on the right side of the word „THAN". Generating rules with the use of the Rough Set Theory requires that not only the outputs, but also the inputs assume discrete values, that is, nominal values or ordinal values. Each differentiable example (observation record) generally constitutes a rule. The rule set obtained in this way may usually be reduced, or the rules within it may be simplified (that is, the conditional constituent may be shortened). This is done via crossing out those attributes which do not contribute in any way to the classification, that is, after whose omission the rule still always points to the same decision class (for all records within the learning set). Rules are evaluated primarily with respect to the unequivocality of classification, which is expressed as the so-called rule confidence. This parameter is defined as the ratio of the number of observations in which a particular combination of inputs and, at the same time, a particular output class is found to the number of all observations with the same combination of inputs (that is, observations including also those in which the output class is different). Another parameter of rule evaluation is the number (the proportion) of observations corresponding to a particular rule within the learning set – this parameter is called the rule support. If it is not possible to obtain from a data set rules with 100% confidence, then less unequivocal rules are employed, which are usually evaluated on the basis of the confidence and support combination or some other, more complex criteria. The RST also makes possible an easy evaluation of input significance, which is based on the assessment of how the classification unequivocality decreases with the omission of a particular input in all rules. The rules induction was performed with the help of the logic rule generating program which is described and made available in [10], and which is a modification of the EXPLORE algorithm characterized in [11].

3. The modeling set

The database covers a wide range of data features pertaining to the production of ductile cast iron which was subjected to the austempering heat treatment. The database contains information covering more than 1300 melts and conducted heat treatments, including nearly 250 melts with different chemical compositions. Each melt was characterized in terms of 26 input parameters, such as: chemical composition (14 elements), properties of ductile cast iron (as cast – 8 variables), heat treatment parameters (4 variables), as well as 8 output parameters, such as mechanical and structural properties. Selected variables together with their variability ranges and mean values are represented in Table 1.

4. The modeling set

The generation of rules requires a discretization of the data set. The number of the generated rules depends on the number of parameters and the number of discretization levels. Each observation may become a rule and thus the number of the generated rules may equal n^k, where: n – the number of inputs , k – the number of input levels. Generating the rules for the complete data set only for five levels is beyond the capabilities of the currently developed computer program. For this reason, in the preliminary research the number of inputs characterizing the melt and the heat treatment was reduced. On the basis of the experience from other modelings, 8 element contents and the austenitizing temperature were selected as the inputs and the ultimate tensile strength was selected as the single output parameter. A reduced learning set containing 924 complete observations was used for rule generation. The discretization was performed through dividing the parameter variability range into easily interpretable sets, for instance, silicon content: low, medium, high; austenitization temperature: low, high. The number of levels for all parameters is represented in Table 2.

Table 1.
Selected parameters together with their minimal, average, and maximal values

Parameters	C [%]	Si [%]	Mg [%]	Mn [%]	Ni [%]	Cu [%]	Mo [%]	Cr [%]	Al [%]	Graphite volume fraction [%]	Nodularity [%]	Perlit volume fraction [%]	Ferrite volume fraction [%]	UTS [MPa] (as cast)	Elongation [%] (as cast)
Min	3.00	1.06	0.020	0.00	0.00	0.00	0.00	0.00	0.00	7.60	60	12.0	0.0	370	2.8
Mean	3.56	2.58	0.049	0.26	0.65	0.44	0.11	0.01	0.01	10.22	90	91.3	8,7	620	7.8
Max	3.97	3.58	0.100	1.23	3.52	1.60	0.63	0.33	4.88	14.53	100	100.0	88.0	890	27.7

Parameters	Austenitizing		Austenitizing		UTS [MPa]	Elongation A_5 [%]	Brinell Hardness [HB]	Charpy impact strength [J]	Martensite volume fraction Xa'	Retained austenite volume XyR
	Temperature [^{o}C]	Time [min]	Temperature [^{o}C]	Time [min]						
Min	816	20	200	0	450	0.0	144	3	0.0	0.0
Mean	899	97	340	109	1102	6.5	344	70	12.7	23.4
Max	1050	300	454	480	2000	33.0	610	157	64.0	66.0

Table 2.
The number of parameter discretization levels in the data set

C [%]	Si [%]	Mn [%]	Cu [%]	Ni [%]	Cr[%]	Mg [%]	S [%]	T_{aust} [^{0}C]	UTS[MPa]
2	3	2	4	4	2	3	2	2	7

The output value (the ultimate tensile strength) was divided into 7 levels, which corresponds to a natural and easily comprehensible division [12] into the following values: very low, low, low medium, medium, high medium, high, very high.
For the set prepared in this way 171 rules were generated with the confidence of above 50%. The set of generated rules contained rules with the confidence of nearly 85%. The length of the generated rules (that is, the number of conditions) ranged from 1 to 6. The numbers of rules with particular numbers of conditions are represented in Fig. 1. For each of the rules the support and the coverage were calculated. The rule support falls within the range from 1 to 10%. The rule coverage falls within the range from 2 to 73%. Three rules with the highest confidence equaling 84,6% contained 4 or 5 inputs. However, these rules had relatively low support and coverage of, respectively, 1,2% and 2,9%.

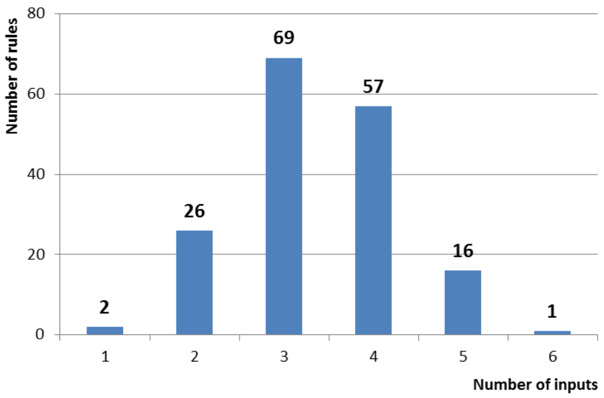

Fig. 1. Rule distribution with respect to the rule length

With the help of the rule quality coefficient Q, which is represented with the following correlation:

$$Q = A \cdot S \cdot C \tag{1}$$

where: A – confidence (accuracy), S – support, C – coverage the rules were decreasingly ordered.

The highest value of the Q coefficient was obtained for seven rules with the confidence within the 50 – 60% range, the support within the 8 – 10% range, and the coverage within the 25 – 30% range. All the rules had the length (the number of inputs) of 2 or 3 and indicated only two levels of outputs. They predicted the ultimate tensile strength on the level: low medium or medium (respectively, level 3 and 4). In all the rules one of the inputs was the magnesium content assuming the low or medium level. A juxtaposition of selected rules is offered in Table 3. The rules differ with respect to the number of inputs (rule length). The same prediction effect may be obtained via the use of a primary rule or a rule with redundant conditions. Primary rules are the rules with the smallest number of inputs.

5. Summary

The obtained rules do not cover completely all the possible output levels. No rule with the confidence of above 50% was obtained for the output levels 2, 6, and 7, which corresponds to the ultimate tensile strength on the levels: low, high, and very high. The basic reason is the fact that the number of observations involving such a strength is too limited in the employed set using only complete records. In accordance with the methodology discussed in [13] the empty data should be replaced. A solution of the problem at hand requires coming up with a special method for replacing missing or empty data which is to be used in the RST modeling, since so far no such method has been offered in any publications. Reducing the number of output levels would surely decrease prediction quality, without at the same time guaranteeing an increase in confidence.
As suggested in [10], low rule confidence may reflect some inconsistency or incoherence in the collected data. This seems to be confirmed by an analysis of the data set quality discussed in [14].

Table 3.
A juxtaposition of selected rules with the highest rule quality coefficient Q

Rule No	Mg [%]	Cu [%]	S [%]	Si [%]	Cr [%]	T_{aust} [^{o}C]	UTS [MPa]	Confidence	Quality coefficient Q
1	**L**	**HM**					**M**	**53,8**	**1,43**
2	L	HM	L				M	54,0	1,33
3	L	HM		L			M	54,2	1,21
4	**M**		**L**				**LM**	**52,5**	**1,38**
5	M		L			L	LM	53,2	1,29
6	M				L		LM	51,7	1,23
7	M		L		L		LM	57,2	1,36

where: L – low, M – medium, LM – low medium, HM – high medium

The obtained results suggest that the RST has a considerable potential as far as the control of advanced industrial processes is concerned and that further research in this area is necessary. Logic rules obtained with the help of association rules, decision trees and the RST are easy to interpret and offer a deeper insight into the process under analysis. As demonstrated in [15], the RST methods perform considerably better than DTs. This applies particularly to cases of sets from which rules with redundant conditions are generated.

References

[1] Fraś, E., Górny, M. & Stachurski, W. (2006). The problem supercienkościennych of ductile iron castings. *Archives of Foundry*, 6(21), 43-57.

[2] Reimer, D. (2006). Applications for austempered ductile iron castings, *Ductile Magazine*. Retrieved from http://www.ductile.org/magazine/2005_3/reimer.pdf

[3] Hayrynen, K. L., Brandenberg, K. R. & Keough J. R., Applications of Austempered Cast Irons. *AFS Transactions.* 02-084, 1-10.

[4] Raghavendra, H., Bhat, K., Rajendra, L., Udupa K., Rajath Hegde, M. M. (2010). Grinding Wear Behavior of Stepped Austempered Ductile Iron as Media Material During Comminution of Iron Ore in Ball Mills, International Conference on Advances in Materials and Processing Technologies, AMPT, Conference Proceedings eds.: Chinesta F., Chastel Y., and El Mausori M.

[5] Laino, S., Sikora, J., & Dommarco, R. C., (2011). Advances in the Development of Carbidic ADI. *Key Engineering Materials.* 457, 187-192.

[6] Idirs, U. D., Aigbodion, V. S. & Shehu M. A. (2011). Potential of using black palm kernel oil in austempering of ductile cast iron use in the production of agricultural implements. *Journal of Materials Design and Application, Proceedings of the Institution of Mechanical Engineers.* 225(L), 340-346.

[7] Pawlak, Z. (1982). Rough Set. *International Journal of Parallel Programming.* 11, 341-356.

[8] Shen, L., Francis, E., Tay, H., Liangsheng, Qu., & Shen, Y. (2000). Fault diagnosis using rough sets theory. *Computers in Industry.* 43(1), 61-72.

[9] Sadoyan, H., Zakarian, A., & Mohanty P. (2006). Data mining algorithm for manufacturing process control. *The International Journal of Advanced Manufacturing Technology.* 28, 342-50.

[10] Soroczynski, A. (2011). *Generate engineering knowledge rules based on recorded cases foundry*. Unpublished doctoral dissertation. Wydział Inżynierii Produkcji Politechnika Warszawska.

[11] Mienko, R., Stefanowski, J., Toumi, K. & Vanderpooten, D. (1996). Discovery-oriented induction of decision rules. *Cahier du LAMSADE.* 141.

[12] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*. 63(2), 81-97.

[13] Kochański, A. (2010). Data preparation. *Computer Methods in Materials Science.* 10(1), 25-29.

[14] Kochański, A., Perzyk, M., Kłębczyk, M. (2012). Knowledge in imperfect data, in: *Advances in Knowledge Representation*, ed.: Carlos Ramirez Gutiérrez, Publisher: InTech, Wien, ISBN 978-953-51-0597-8.

[15] Perzyk, M., Biernacki, R., Kochański, A., Kozłowski, J., Soroczyński, A. (2011). *Applications of data mining to diagnosis and control of manufacturing processes*, in: Knowledge - oriented applications in Data mining, eds.: Kimito Funatsu and Kiyoshi Hasegawa, Publisher: InTech, Wien, ISBN 978-953-307-154-1.