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

39 – 42

7/2

Assessment of Reinforcement Phase Shape in MMC Using Decision Trees

K. Gawdzińska [a, *], M. Pelczar [b]

[a] Institute of Basic Technical Sciences, Maritime University of Szczecin, ul. Podgórna 51/53, 70-205 Szczecin, Poland
[b] Faculty of Computer Science and Information Technology, West Pomeranian University of Technology, Szczecin
ul. Żołnierska 49, 71-210 Szczecin, Poland
*Corresponding author. E-mail address: k.gawdzinska@am.szczecin.pl

Received 24.07.2013; accepted in revised form 09.09.2013

Abstract

This paper proposes a description of the reinforcement phase shape in MMCs by means of a decision, or classification tree analysis, recognized as a basic data mining technique. The material under examination was composed of reinforcement particles (SiC) in suspension composites with silumin matrix, made by mechanical stirring method. The use of decision tree method allowed to determine logic rules for the classification of particles to the category *circle* on the basis of its diameter and surface area, taking into account the division into three samples (depending on the location of the analyzed area in the casting space) and a reference sample (representative analysis area – of most desired shape in terms of composite quality - generated by simulation). To assess the accuracy of classification we used a redistribution indicator, that can be a measure used in describing the feature: homogeneity of reinforcement phase particle shape in the space of composite casting.

Keywords: Shape of the reinforcement phase, Data mining, Ddecision trees

1. Introduction

The concept of inhomogeneity is not commonly defined in specialized literature [1-5]. Inhomogeneity is defined in a variety of ways:

– deviation of certain geometric features from a structure conventionally regarded as homogeneous;
– local disturbance of a structure, the intensity of which occurs at a varying probability,
– derivative of geometric features differences of measured elements, where the differences are due to their orientation (anisotropy) or position (gradient) in the examined object.

If we look at composite alloys, in descriptions of quality parameters of these materials the concept of a defect is commonly used, understood as a deviation from desired features. However, it seems more purposeful to employ the concept of material homogeneity. Deviations from this property, that is defects, will be inhomogeneities of, for example, structure density, or amount, distribution of the size and shape of the reinforcement phase. In [1], the author proposed a description of quality by means of so far undefined quality features of these materials, where such features as the homogeneity of shape, size, and distribution of reinforcement phase in the casting space are taken into consideration. The works [6-9] include descriptions of these features. This study presents a method of determining the shape of reinforcement phase and proposes its descrirption, indicating relevant measures and using the decision tree analysis, belonging to basic techniques of data mining. The assumed input variables were those describing particle dimensions: particle area in mm^2 and particle diameter in mm. The output variable, referred to as circularity, defines the percentage of particle area contained in a circle. It has been dichotomized in such a way that it assumes the value 0 when the percentage of particle area encompassed in a circle is smaller than 75%, and the value 1 when the particle percentage in a circle is not smaller than 75%. It has been also assumed that if the variable circularity has a value of 1, it belongs to the category *circle*, while the value 0 means the particle does not belong to the category circle.

Logic conditions generated by means of decision trees enable qualifying the reinforcement phase to the category of circles on the basis of data: surface areas and diameters of particles. We aim at presenting one of the methods for describing one of the features of composites: homogeneity of reinforcement phase shapes in a casting.

2. Description of the method

For the evaluation of reinforcement phase shape we will use a classical classification and regression tree algorithm, developed by Breiman and others. For calculations, the algorithm was implemented in the Statistica PL 10.0 package. The method was used for its hierarchical nature and flexibility, and the ease of interpreting and explaining the obtained results [10-13]. To take account of the dichotomic character of the variable *circularity*, we used the classification tree analysis, which enables generating logic conditions allowing to qualify the output variable to a given category (prediction) based on the values of input (predictive) variables. The process of determining a classification tree comprises four stages [10, 13] :

1. Determination of the prediction reliability criterion.

It is assumed that the most reliable prediction is the one that has the highest quality (according to [10]), while a model that has the highest quality is the one with the least number of wrong classifications – in this case it is the reference sample.

2. The choice of divisions.

We will look for a division that generates groups characteristic of high homogeneity in reference to the output value. We will attempt at improving the degree of homogeneity by maximizing the difference:

$$\Delta Z = Z_0 - \sum_{i=1}^{r} \frac{n_i}{n_0} Z_i \ , \qquad (1)$$

where Z_0 – degree of inhomogeneity of the divided element; n_0 – sample size of the divided element; r – number of elements created by division; Z_i – inhomogeneity of i-th element created by division; n_i – sample size of i-th element created by division.

The degree of inhomogeneity is mostly identified using the following measures:

– Gini coefficient, derived from this formula:

$$1 - \sum_{i=1}^{k} p_i^2 \ , \qquad (2)$$

where k – number of categories assumed by the output variable, p_i – percentage of observations assuming i-th value of output variable; Gini coefficient is preferable due to easy interpretation and standardization to the interval (0, 1);

– entropy coefficient, derived from the formula:

$$-\sum_{i=1}^{k} p_i^2 \log_2 (p_i) , \qquad (3)$$

where: k – number of categories assumed by the output variable, p_i – percentage of observations assuming i-th value of output variable; this coefficient is non-standardized statistic defining a degree of inhomogeneity in a tested group.

3. Determination of the condition for stopping the divisions.

In C&RT two variants for stopping the divisions are in use: minimum sample size (nodes contain a preset minimum number of cases) and the fraction of objects (all terminal nodes are homogeneous).

4. Choice of a tree.

We can apply either of two approaches while choosing a tree in the C&RT algorithm; the tree size is determined by the user. We can also follow the procedure developed by Breiman and others, based on a test sample, v-fold cross test and minimum quality.

The overall assessment of the accuracy of classification obtained by the classification tree method consists of three parts:

1. Assessment by resubstitution. It is a ratio of cases wrongly classified by the model to all cases. This measure is calculated for the same set of data, which constituted a basis for building the classification model.
2. Asssessment based on a test sample. It is a ratio of cases wrongly classified by the classification model to all cases, but this measure is calculated for a set of data different from the one on which the classification model was built.
3. V-fold cross validation. Data are divided into v groups and based on one of them, a classification model is determined, and for the other groups the ratio of wrongly classified cases to all cases is calculated.

3. The research

Reinforcing particles of silicon carbide in silumin matrix were used for the description of the reinforcement phase in composite castings. The tests focused on suspension composites made by mechanical stirring. The use of classification tree method allowed to determine logic rules for classifying a given particle to the category *circle* on the basis of its diameter and surface area, with a division applied to three samples (dpending on the location of the examined area in the space of the casting, Fig.1) and a model sample (representative area of analysis – one with most desired shapes from the quality viewpoint, generated by simulation in the program Statistica PL, Fig. 2).

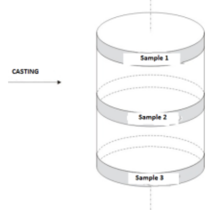

Fig. 1. Areas of sampling for the analysis of the reinforcement structure homogeneity in the casting space

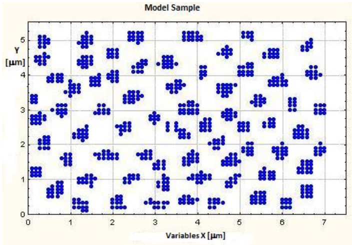

Fig. 2. Visualization of particles distribution in the model sample, made by means of Statistica PL

To classify a particle shape to the category *circle,* the following variables were established:

– input:
  - particle diameter in mm;
  - particle surface area in mm^2 ;
– output:
  - particle percentage contained in a circle; on this basis we determine circularity,
  - circularity, assuming value 0 when the percentage of the particle contained in a circle is smaller than 75%, and value 1 when the percentage of the particle contained in a circle is higher than 75%.

The measures of these variables determined for three samples and for a model (reference) sample (denoted as 4 in Table 1) and basic statistical parameters of these variables are shown in Table 1.

Table 1.
Statistical parameters and measures of the analyzed variables

Measures / Variables	sample	dia-meter	confidence interval for the variable		coefficient of variance
particle surface area	1	49.20	39.90	58.50	94%
particle surface area	2	69.71	57.49	81.93	80%
particle surface area	3	76.72	60.83	92.61	84%
particle surface area	4	58.20	55.43	60.96	20%
particle diameter	1	9.63	8.57	10.70	55%
particle diameter	2	11.03	9.89	12.17	47%
particle diameter	3	10.56	9.35	11.77	47%
particle diameter	4	8.62	8.29	8.95	16%
circularity percentage	1	53.99	49.53	58.45	41%
circularity percentage	2	51.43	46.03	56.83	48%
circularity percentage	3	59.21	52.99	65.43	43%
circularity percentage	4	78.46	74.53	82.39	21%

It follows from Table 1 that sample 1 has the smallest mean surface area of particles and the greatest variability. The variability of particle surface area is the smallest in the model sample. The mean particle diameter is the largest for sample 2, and the shortest for the model sample. The variablity of particle diameter was the greatest in sample 1, and the lowest in the model sample. The mean circularity percentage is the smallest in sample 2, and the greatest for the model sample. The variability of circularity percentage was the greatest in sample 2, and the lowest in the model sample.

The results of classification tree analysis for each sample are presented in Figures 3 to 6.

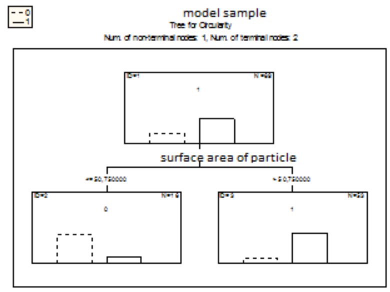

Fig. 3. A diagram of the classification tree for a model sample, denoted as 4 in Table 1

The particle surface area turned out to be a more essential predictor. Particles with an area of more than 50.75mm^2 were qualified to the category *circle*. The assessment of this classification by resubstitution amounts to 16%, that is 84% of cases were qualified correctly.

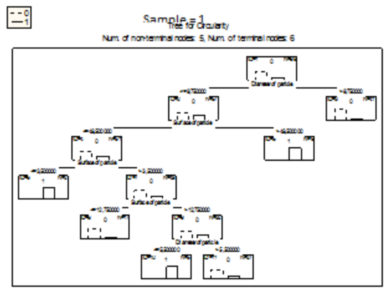
Fig. 4. A diagram of a classification tree for sample 1

For sample 1 the classification algorithm was as follows: if the particle diameter is larger than 9.75 mm, then the particle is not qualified to the category *circle*. When the particle diameter is not larger than 9.75 mm, then the variable particle surface area is taken into consideration. When the particle area is larger than 48.5 mm^2 or not larger than 3.5 mm^2, then the particle is classified as a circle. When a particle area belongs to the interval (3.5 mm^2; 12.75 mm^2], then the particle does not belong to the category *circle*. When a particle area belongs to the interval (12.75mm^2; 48.5mm^2], then the variable *particle diameter* should be considered again. When a particle diameter is not larger than 5.5 mm, then the particle is classed as a circle, otherwise the particle does not belong to the category *circle*.

For sample 1 the assessment of this classification by resubstitution reaches 11%, that is 89% of cases have been qualified correctly.

A simpler algorithm was created for sample 2, and it is as follows: if the particle diameter is not larger than 4.5 mm, then the particle is qualified to the class *circle*; when a particle diameter is larger than 4.5 mm, the variable *particle area* is considered. When a particle surface area is larger than 187.5 mm^2, then the particle is classified to the category *circle*, otherwise the particle does not belong to that category.

For sample 2, the assessment of that classification by resubstitution is 13%, that is 87% of the cases were qualified correctly.

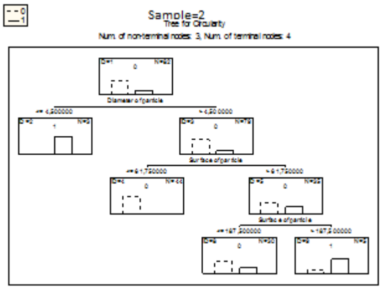

Fig. 5. A diagram of the classification tree for sample 2

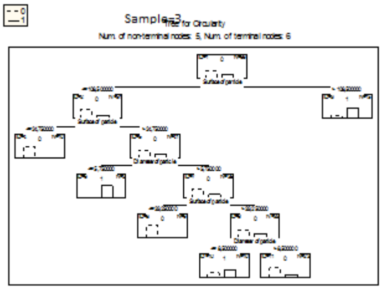

Fig. 6. A diagram of the classification tree for sample 3

For sample 3 a particle is classified to the category *circle* in three cases, i.e. when the particle surface area:

1. belongs to the interval (24.75mm^2; 108.5mm^2] and the particle diameter is not larger than 5.75mm,
2. belongs to the interval (33.25mm^2; 108.5mm^2] and the particle diameter belongs to the interval (5.75mm; 8.5mm],
3. is larger than 108.5mm^2.

In the remaining cases the particle is not classified to the category *circle*.

For sample 3 the assessment of this classification by resubstitution reaches 18%, that is 82% of cases have been qualified correctly.

4. Summary

The article describes an assessment of shape, or geometry, of the reinforcement phase in MMCs, reinforced with SiC particles, by using C&RT analysis. The use of classification trees method allowed to generate logic rules for the classification of individual particles of the reinforcement phase to the category *circle*, based on particle diameter and surface area. There were four particle samples, including a model sample, subjected to the classification procedure. A redistribution coefficient was used in assessing the accuracy of classification. The coefficient did not exceed 18%, which means that over 80% of the particles have been classified correctly.

The assessment of particles shape in MMC is a proposed herein method for the description of the feature referred to as homogeneity of reinforcement phase shape in the space of composite casting.

References

[1] Gawdzińska, K. (2012). Material and technological conditions of quality of metal composite castings. Monograph. *Archives of Foundry Engineering*. 10-50.

[2] Cholewa, M. (2007). Heat flow description during crystallization process of cast dispersive composites. *Archives of Foundry Engineering.* 1(7), 117-122.

[3] Konopka, Z. (2011). Cast metal composites. Częstochowa: University of Technology Publishing House.

[4] Śleziona, J. (1998). *Basics of composite technology.* Glieice: Silesian University of Technology Publishing House.

[5] Orłowicz, W. & Trytek, A. (2007). Shaping of microstructure and service properties of cast iron casting by surface refinement with electric arc plasma. *Archives of Foundry Engineering*. 7(23), 120.

[6] Grabian, J., Gawdzińska, K., Przetakiewicz, W. & Pijanowski, M. (2011). Description of the particle distribution in the space of composite suspension casting by statistical methods. *Archives of Foundry Engineering.* 11(1), 31-34

[7] Gawdzińska, K., Berczyński, S., Chylińska, R., & Grabian, J. (2011). A description of particle shape homogeneity in the space of composite suspension casting. *Archives of Foundry Engineering.* 11(1), 11-14.

[8] Gawdzińska, K. (2010). Homogeneity of particle size in the space of composite suspension casting. *Archives of Foundry Engineering.* 10(spec. 1), 163-168.

[9] Gawdzińska, K. & Maliński, M. (2005). Study of Reinforcement Elements Distribution Exemplified by Composite with AlSi11 Matrix and Carbon Reinforcement. *Metallurgy.* 44, 45-48.

[10] Bronsztejn, I. N., Siemiendiajew, K. A., Musiol, G., Mühlig, H. (2011). *Modern compendium of mathematics.* Warszawa: Wydawnictwo Naukowe PWN.

[11] Breiman, L., Friedman, J., Stone, Ch.J., Olshen, R.A. (1984). *Classification and Regression Trees*. CA: Wadsworth & Brooks/Cole Advanced Books & Software.

[12] Frątczak, E.-red. (2012). *Advanced methods of statistical analysis*. Warszawa: Oficyna Wydawnicza SGH w Warszawie.