Publishing House
AKAPIT

COMPUTER METHODS IN MATERIALS SCIENCE
Informatyka w Technologii Materiałów

Vol. 9, 2009, No. 2

STRUCTURALIZATION OF KNOWLEDGE ABOUT CASTING DEFECTS DIAGNOSIS BASED ON ROUGH SETS THEORY

STANISŁAWA KLUSKA-NAWARECKA[1], DOROTA WILK-KOŁODZIEJCZYK[2], GRZEGORZ DOBROWOLSKI[2], EDWARD NAWARECKI[2]

[1] *Foundry Research Institute, Cracow, Poland*
[2] *AGH University of Science and Technology, Cracow, Poland*
Corresponding Author: nawar@iod.krakow.pl*; (S. Kluska-Nawarecka)*

Abstract

The rough information system is applied in order to structuralize knowledge about casting defects and their reasons, which are faults or lacks of precision in the course of technological processes. The system is fed up with information extracting from Polish, French and Czech standards. The query language and algorithms of the rough sets theory make the knowledge a reach and easy to use source supporting technical assessments and educational activities of various kinds.

Key words: casting defects, knowledge management, rough sets

1. INTRODUCTION

Research at Foundry Research Institute in Cracow dealing with development of Information System for Foundry Industry — INFOCAST [1, 4] embraces, among others, enriching the palette of representation methods of knowledge stored in the system. As a consequence, new ways of application of the knowledge arises that make stronger the system supporting power in the field of assessments and educational activities.

Although rough sets theory was invented over 20 years ago [5], it has not caused such avalanche of application attempts and real applications as e.g. fuzzy sets theory or artificial neural networks. Nevertheless, it can be noticed that the rough sets create a sound base for acquisition, processing and interpretation of incomplete and rough knowledge. So called engineering knowledge, dealing with engineering practice especially, has such characteristic very often.

The presented article reports an idea how to structuralize knowledge about casting defects and their reasons, which are faults or lacks of precision in the course of technological processes. Original sources of the knowledge are observations and analyses of generations of technologists gathered and published in the form of national or European standards. Because the standards in the subject are in fact generalization of experience, it occurs that an attempt towards formalization points at the rough information system as a means.

The article is organized as follows. The next section provides a reader with some definitions and ideas of rough sets theory necessary to show (in section 3) how knowledge included in the standards can be translated into a rough information system. As such system is armed with a powerful query language, the knowledge structuralized this way can be flexible viewed, what can produce semi-products for various kinds of the experts' activities. Section 4 illustrates with examples how the queries are formulated, calculated and evaluated (query results are obviously rough and the "roughness" can be calculated also).

 ISSN 1641-8581

2. ELEMENTS OF ROUGH SETS THEORY

The first articles about rough sets were issued as early as at the beginning of eighties of the last century [5], but wider interest in this theory crowned with the variety of applications can be observed only during the last decade. Notions and definitions of the theory quoted beneath are chosen as necessary to present application of it in formal representation of knowledge about casting defects. Such the idea is not reported yet in the available literature.

As a point of departure the well-known, in various fields of computer science now, structure of information description is taken: $O - A - V$ (object-attribute-value). Interpretation of the above formula is as follows. Having set of objects O and describing them set of attributes A, some value v_j^i is ascribed to each pair (o_i, a_j) of Cartesian product $O \times A$. Possible values $v_j \in V_j$ form a set for attribute a_j and the union for all attributes gives set V.

About thirty years ago, this information description allowed for defining information system S as an aggregate

$$S = \langle\, O,\ A,\ V,\ f : O \times A \longrightarrow V \,\rangle \tag{1}$$

where all sets are finite and f is a complete function called an information one. Due to this, f can be given in the form of a table.

A considered feature of information systems is ability to classify. Rules of the classification are formally introduced by an equivalence relation or, which occurs to be more precise here, indifference relation of objects.

Indifference relation R is defined as follows

$$o_i\ R\ o_k \Longleftrightarrow f(o_i, a_j) = f(o_k, a_j) \quad : \quad o_i, o_k \in O,\ a_j \in A \tag{2}$$

Two objects are indifferent (are in relation R), if values of all their attributes are equal respectively. The indifference relation can be defined also with respect to a restricted subset of attributes or even single one.

Among others, the indifference relation given by (2) is used to define the so-called representation of an information system, which is formulated with respect to the equivalence classes of R, not to objects themselves.

In relation to an information system a little beat more complicated situation is considered also. It is assumed that some objects are represented in the system ambiguously. Although in the reality they are fully characterized by exact values of their attributes, information function f is able to produce just sets of values, which form a kind of approximation. It is certain that each set comprises the appropriate single real value.

In that case an information system becomes rough one and the difference with respect to (1) consist in the definition of an information function, which turns to be

$$(o_i, a_j) \in O \times A \longrightarrow \wp(V) \ni \{v_{j,1}^i, \ldots, v_{j,k}^i\} \quad : \quad v_{j,1}^i, \ldots, v_{j,k}^i \in V_j \tag{3}$$

Summarizing, rough information system $\tilde{S}$ is defined as

$$\tilde{S} = \langle\, O,\ A,\ V,\ \tilde{f} \,\rangle \tag{4}$$

where $\tilde{f}$ is a rough information function (or attribute table).

As the consequence of acceptance of rough descriptions, $\tilde{S}$ loses its ability to classify real objects in the sense of (2). The essence of rough sets theory deals with the above situation. The exact classification, which is carried out with respect to the well-defined set of attribute values for each class, is replaced with special inference that is based on the lower and upper approximations of the equivalence classes. The problem of formulating and calculating queries in rough information systems, which is central for the application reported in the paper, is of this kind.

An appropriate query language is defined for system $\tilde{S}$. Do not going into details, the following terms are introduced syntactically and semantically: descriptors of the form $(a, U), U \subseteq V_a$, classic logical operators, constant terms for *true* and *false* and special operators of lower and upper approximation of a term – $\underline{C}$, $\overline{C}$, respectively. Compound terms are constructed from descriptors joined with operators in a usual way. Processing of terms can be done based on the definitions and axioms of the language.

One can notice that the use of a descriptor with a rough attribute of more then one elementary values relates to objects in the system (rough descriptions, in fact) not in the reality. The reality is not rough.

Two key definitions show what can be derived from information stored in $\tilde{S}$. The lower approximation of term $\underline{C}(a, U)$ has the following interpretation

$$\sigma_S(\underline{C}(a, U)) = \{o \in O \quad : \quad \tilde{f}(o, a) \subseteq U \subseteq V_a\},\quad a \in A \tag{5}$$

It is a set of objects that have the given attribute described in the system exactly, thus they are such "for sure".

The upper approximation of term $\overline{C}(a, U)$ is interpreted as follows

$$\sigma_S(\overline{C}(a,U)) = \{o \in O \\ : \tilde{f}(o,a) \cap U \neq \emptyset; \ U \subseteq V_a \}, \ a \in A \tag{6}$$

It is a set of objects that have the given attribute described in the system roughly, thus it is only possible that they are such.

It can be easily checked that $\sigma_S(\underline{C}(a,U)) \subseteq \sigma_S(\overline{C}(a,U))$, thus a following ratio called precision of the described above approximation is defined soundly as long as $\sigma_S(\overline{C}(a,U)) \neq \emptyset$

$$\mu_S((a,U)) = \frac{\mathrm{card}(\sigma_S(\underline{C}(a,U)))}{\mathrm{card}(\sigma_S(\overline{C}(a,U)))} \tag{7}$$

The sketched above approach shows how to calculate queries in a rough information system together with evaluation of their accuracy ($0 \leq \mu_S((a,U)) \leq 1$).

In the frame of rough sets theory a plenty of interesting notions and features are defined and analyzed also. All that stuff is neglected in this short introduction, because it occurs to be unnecessary for an application that is presented next.

3. ROUGH SETS AND KNOWLEDGE ABOUT CASTING DEFECTS

In order to construct an information system that will process knowledge about casting defects, domain-oriented definitions of the elements of aggregate (1) or its rough counterpart is needed.

Objects in the reported approach are casting defects represented by their names. A part of the set looks like

$$O = \{crack, scratch, crevice, wrinkles, \ldots\} \tag{8}$$

Definition of an attribute set, which in this case represents features of the considered casting defects, is done to some extent arbitrary, but accordingly to knowledge, which engineers could gather in the past and, which can be found in published standardizations and catalogues now. As typical ones the following features can be enumerated here

$$A = \{type, location, severity, shape, \ldots\} \tag{9}$$

Some problems had to be overcome with respect to valuation of these attributes. As reasons the following can be pointed out:

- assignment of numerical values for majority of attributes is impossible (the above examples illustrate that);
- even if it is theoretically possible to give a number (e.g. *number of damages*), engineers prefer to use linguistic characterization (e.g. *single, not numerous*);
- use of too wide spectrum of linguistic descriptions can cause ambiguities and misunderstandings (there is lack of standardization of such descriptions).

In characterized above cases sets of values for attributes have been constructed somewhat arbitrary. Linguistic values are partly proposed by the authors and can be seen also as a kind of experiment leading to common terminology in the field. For attribute *shape* the following values are taken

$$V_5 = \{regular, narrow, zigzagged, oval, \ldots\} \tag{10}$$

Starting from the already presented assumptions, a rough attribute table $\tilde{T}$ (and, in turn, the rough information system $\tilde{S}$), which puts together information about casting defects is constructed. The table consists of several dozens of objects (casting defects) described by 18 attributes. For the sake of editorial reasons, just a part of the table is reproduced in the paper (see table 1).

It is worth to mention that the table, besides information obtained from the Polish standards (PL) [6], comprises description of casting defects based on the French (F) [3] and Czech (CZ) [2] sources. Combination of these three descriptions allows for analyzing their mutual conformity and, in consequence, integrating them with respect to some defects or their groups (e.g. *internal* or *external defects, defects of shape*) in order to create the common ontology. The ontology could not only produce possibility of translation among these sources but point out issues important for more complete and coherent casting knowledge as well.

From the point of view of the rough sets theory the table can be regarded as the representation of a virtual (non-existing but possible) rough information system that gathers all possible cases of casting defects. The representation has been obtained as an effect of analyses carried out by boards of experts preparing the standards, not in the way of the formal procedure defined by the theory.

Table 1. Rough attribute table about casting defects (a part)

Name		a_1 - type	a_2 - spread	a_3 - location	a_4 - severity	a_5 - shape	a_6 - process stage
Cold cracks PL	o_1	v_{11} - crack v_{12} - scratch	v_{21} - local	v_{31} - surface v_{32} - interior	v_{41} - single	v_{51} - regular v_{52} - narrow	v_{61} - transport v_{62} - knocking out v_{63} - fettling
Cold cracks CZ	o_2	v_{11} - crack	v_{21} - local	v_{31} - surface v_{33} - walls	v_{41} - single	v_{51} - regular v_{53} - zigzagged	v_{63} - fettling v_{64} - casting cooling
Cold cracks F	o_3	v_{11} - crack	v_{21} - local v_{22} - extended	v_{31} - surface	v_{41} - single	v_{51} - regular v_{52} - narrow	v_{61} - transport v_{62} - knocking out
Hot cracks PL	o_4	v_{11} - crack v_{13} - crevice	v_{21} - local	v_{31} - surface v_{32} - interior v_{33} - walls	v_{41} - single v_{42} - ramified	v_{52} - narrow v_{53} - zigzagged	v_{62} - knocking out v_{64} - casting cooling v_{65} - designing
Hot cracks CZ	o_5	v_{11} - crack	v_{23} - clustered	v_{32} - interior	v_{41} - single	v_{51} - regular	v_{61} - transport v_{62} - knocking out v_{63} - fettling
Hot cracks F	o_6	v_{13} - crevice	v_{21} - local	v_{31} - surface v_{32} - interior	v_{41} - single	v_{51} - regular	v_{62} - knocking out v_{65} - designing
Folds PL	o_7	v_{13} - crevice	v_{21} - local	v_{31} - surface	v_{41} - single	v_{52} - narrow v_{54} - rounded edges	v_{65} - designing v_{66} - pouring
Folds CZ	o_8	v_{14} - wrinkles v_{15} - washes	v_{21} - local	v_{31} - surface v_{33} - walls	v_{43} - numerous	v_{52} - narrow v_{54} - rounded edges	v_{64} - casting cooling v_{65} - designing v_{66} - pouring
Folds F	o_9	v_{13} - crevice	v_{22} - extended	v_{31} - surface v_{34} - subsurf.	v_{43} - numerous	v_{52} - narrow v_{54} - rounded edges	v_{65} - designing v_{66} - pouring
Surface blowholes PL	o_{10}	v_{16} - recess v_{17} - cavity	v_{21} - local	v_{31} - surface	v_{41} - single v_{43} - numerous	v_{55} - oval v_{56} - spherical	v_{66} - pouring v_{67} - sand preparation v_{68} - metal degassing

Analyzing from the wider perspective possibilities, which rough sets theory application opens, one can conclude that obtained knowledge representation has interesting expression power. In details, it allows for:

- specifying similarities and differences between descriptions of defects that makes easier recognition of them;
- grouping defects according to causes of their arising that helps diagnosis of defects and pointing at critical operations of a technological process;
- integrating domain information coming from different sources;
- constructing a computer tool, which could carry out a dialog with a user in the framework of technical assessments or educational activities.

4. ADVANTAGES OF THE STRUCTURALIZATION

Structuralization of knowledge about casting defects in the shape of the rough information system described in the previous section (see table 1) has interesting advantages. They consist in possibility to formulate even very complicated queries that answer to important practical problems. It ought to be observed that the system under consideration is big enough to make unsupporting inference impractical, while some (mainly logical) calculations imposed by the proposed approach can be done even "by hand".

In order to give some illustration here, two exemplary queries are calculated and discussed below.

Query 1. Which objects (defects) are of type *crevice* and are caused by improper *pouring* or a fault in the process stage of *designing*?

In terms of the system the query is translated into compound term t_1 of the appropriate language

$$\begin{aligned} t_1 = & (a_1, \{v_{13}\}) \cdot ((a_6, \{v_{65}\}) + (a_6, \{v_{66}\})) \\ = & (a_1, \{v_{13}\}) \cdot (a_6, \{v_{65}, v_{66}\}) \end{aligned} \tag{11}$$

Because the system is rough, all what can be calculated are the lower and upper approximations of t_1. Using axioms of the language, the approximations are, respectively

$$\begin{aligned} \sigma_S(\underline{C}(t_1)) &= \{o_6, o_7, o_9\} \cap \{o_7, o_9\} = \{o_7, o_9\} \\ \sigma_S(\overline{C}(t_1)) &= \{o_4, o_6, o_7, o_9\} \\ &\quad \cap \{o_4, o_6, o_7, o_8, o_9, o_{10}\} \\ &= \{o_4, o_6, o_7, o_9\} \\ \mu_S(t_1) &= \tfrac{1}{2} \end{aligned} \tag{12}$$

The results (12) can be annotated in the following way:

- defects *folds PL* and *folds F* have certainly features described by the query, while these features can be possessed at most by defects *hot cracks PL* and *hot cracks F*; the rest of the objects do not share these features surely;

- conformity between the Polish and French standards with respect to the defects under consideration can be observed, while Czech ones comprise a bit different descriptions of them.

Some information about conformity between the standards represented in the system have been obtained "as a byproduct" here. The conformity can be analyzed with a help of especially formulated queries like: Which defects have the same values of attributes in the Polish, French and Czech standards?

Query 2. Which defects located at the *surface* are caused by improper *pouring* or a fault in the process stage of *designing*?

Query t_2 corresponds to the previous one, is a modification obtained by exchanging attribute a_1 with a_3

$$t_2 = (a_3, \{v_{31}\}) \cdot (a_6, \{v_{65}, v_{66}\}) \tag{13}$$

Calculations for this case are as follows

$$\begin{aligned} \sigma_S(\underline{C}(t_2)) &= \{o_7\} \cap \{o_7, o_9\} = \{o_7\} \\ \sigma_S(\overline{C}(t_2)) &= \{o_1, o_2, o_3, o_4, o_6, o_7, o_8, o_9, o_{10}\} \\ &\cap \{o_4, o_6, o_7, o_8, o_9, o_{10}\} \\ &= \{o_4, o_6, o_7, o_8, o_9, o_{10}\} \\ \mu_S(t_2) &= \tfrac{1}{6} \end{aligned} \tag{14}$$

Comparing results (12) and (14), a more general conclusion can be derived. The defects are better characterized by attribute a_1 than a_3. Values of μ_S show it distinctly.

The above exemplary queries do not cover all possibilities of the proposed application of rough sets theory. But, it seems that a reader may feel convinced that:

- knowledge base in the form of a rough attribute table is a reliable and operational representation of knowledge about casting defects included in appropriate standards, here Polish, French and Czech;

- algorithms of the theory make the knowledge a reach and easy to use source supporting technical assessments of various kinds.

5. CONCLUDING REMARKS

The idea of rough sets theory as a means for structuralization of knowledge about casting defects has been presented in rather concise form. The authors have tried to show that such application creates different, but equally interesting possibilities in comparison with rule-based systems, which are widely regarded to be a "natural" way of reusing of the experts' knowledge.

As a key factor in the subject, expression power of the query language, which is available for such structuralization, must be underlined. In consequence, it can supports a reach bunch of assessments with respect to not only single issues of the field — as it has been illustrated with the examples — but subsets of defects, their attributes and reasons as well. It is foreseen that, among others, the second approach can lead to the specific kind of synergy, which can arise as an effect of integration of knowledge covering similar domains but coming from different sources.

Closer identification of the second thread leads to the following indications for the future. It would be valuable to embrace, possible all, sources of information (at least available standards) and to computerized the proposed approach because the system will easily cross the border of manually realized calculation.

REFERENCES

1. Dobrowolski, G., Marcjan, R., Nawarecki, E., Kluska-Nawarecka, S., Dziadus, J., Development of INFOCAST: Information system for foundry industry. TASK Quarterly, 7(2), 2003, 283–289.
2. Elbel, T., Havlicek, F., Rouscek, J., Jelinek, P., Leviand P., Stranski K., Vady odlitkú ze stilin železa (klasifikace, priciny a prevence, MATECS, Brno, 1992 (in Czech).
3. Hèron, G., Masacrè, C., Recherche de la qualitè des pièces de fonderie (in French). Editions Techniques des Industries de la Fonderie, Paris, 1986.
4. Polski Komitet Normalizacji Miar i Jakości, PN-85/H-83105, Odlewy: Podział i terminologia wad, PKNMiJ, Warszawa, 1986 (in Polish).
5. Kluska-Nawarecka, S., Dobrowolski, G., Nawarecki, E., Marcjan R., OntoGRator – an intelligent access to heterogeneous knowledge sources about casting technology, Computer Methods in Materials Science, 7(2), 2007, 324–328.
6. Pawlak, Z., Rough sets, Int. J. of Inf. and Comp. Sci., 11(341), 1982.

STRUKTURALIZACJA WIEDZY O DIAGNOSTYCE WAD ODLEWNICZYCH OPARTA NA TEORII ZBIORÓW PRZYBLI˙ZONYCH

Streszczenie

Koncepcja przybliżonego systemu informacyjnego została zastosowana do strukturalizacji wiedzy o wadach odlewów i ich przyczynach, którymi są najczęściej błędy lub niedokładności popełniane w trakcie procesu technologicznego. System załadowany jest informacjami pochodzącymi z polskich, francuskich i czeskich norm. Język zapytań i algorytmy teorii zbiorów przybliżonych udostępniają tę wiedzę stanowiąc bogate i łatwe w użyciu źródło wspomagające ekspertyzy techniczne i działalność dydaktyczną różnych rodzajów.

Submitted: November 3, 2008
Submitted in a revised form: December 3, 2008
Accepted: December 4, 2008