On Soft Sets and Formal Concept Analysis (FCA) as mathematical categorization systems - An Engineering Application

Anca C. Pascu, Laurent Nana, Jean Vareille, François Monin
University of Brest, Lab-STICC (UMR CNRS 6285)
Computer Science Department
20, Avenue Le Gorgeu C.S. 93837 - 29238 Brest Cedex 3
{pascu, nana, vareille, monin}@univ-brest.fr

Abstract

This paper deals with an analysis of the reasoning for the cognitive problem of categorization in systems. Such reasoning requires the construction of a formal model of the system that is performed in 3 steps: cognitive description of the system, extraction of system's meaningful concepts, and design of the mathematical model based on the concepts. Our goal is to explore the cognitive features and their predominance in some old models of categorization well known in literature. We highlight the relation between Formal Concept Analysis (FCA) model and soft sets and we analyze the notion of "parameter" from the cognitive point of view. Based on this analysis, we propose the notion of *double soft sets* as a mathematical notion more adequate in the engineering applications. All our study is conducted in the framework of material selection in mechanical engineering.

Keywords: Mathematical Theory of Categorization; Soft set; Double soft set; Formal Concept Analysis; Systems parameters; Durability of materials

1 Introduction

For more than 30 years, one of the directions developed by applied mathematics regarding categorization is the following: starting from the classical notion of *set* in mathematics, already axiomatized a long time ago, one defines several types of *sets* in order to capture in their definition more cognitive features that were erased in the standard definition of set. In this context, the basic type of set is called sometime *crisp set* to be distinguished from other types of sets. This is how one has defined the *fuzzy sets* (Zadeh 1965), the *rough sets* (Pawlak 1982; Irwinski 1987; Pomykala 1987), the *locologies* (De Glas 2010), the *quasi-topologies* (Desclés, Pascu, and Biskri 2017), *formal concepts* in Formal Concept Analysis (FCA) (Ganter, Stumme, and Wille 2005). Soft sets (Molodtsov 1999; Alcantud et al. 2024) are the last type in this list.

FCA and soft sets provide several features such as concept lattices, related dependencies and association rules, that can be used in the context of artificial intelligence for knowledge discovery, learning, etc. In (Vinogradov 2020), the authors proposed a supervised FCA-based machine learning using formal contexts as datasets. In this paper, we investigate FCA and soft sets for knowledge discovery with application to a specific domain, namely the selection of material in engineering. The main ideas are the following:

- An analysis of FCA model versus soft sets model. This analysis is performed on a real problem of categorization arisen in selection of material in engineering.
- Introduction of a new notion called *double soft set*, based on the results of this analysis. We propose this notion and investigate its application in engineering.

The paper is organized as follows:

In Section 2, we recall the definition of soft sets and we analyze this notion from both mathematical point of view and application needs. In Section 3, we compare the FCA model with the soft sets model from a mathematical point of view. In Section 4, we present and discuss an example of categorization issued from a problem of technology, namely materials selection in engineering. In Section 5, we discuss the example of Section 4 from the point of view of its cognitive features and propose the definition of *double soft set*. In Section 6, we discuss possibilities of application of double soft sets in categorization in engineering. In Section 7, we present our conclusions and future work.

2 Soft sets from cognitive and mathematical point of view

In this section, we recall the basic definition of a soft set.

Definition 1 *(Molodtsov 1999) A pair (F, E) is called soft set (over X) if and only if F is a mapping of E into the set of all subsets of the set X.*
In other words, the soft set is a parameterized family of subsets of the set X. Every set F(ε), $\varepsilon \in E$, from this family may be considered as the set of ε - elements of the soft set (F, E), or as the set of ε-approximate elements of the soft set.

This definition is the genuine definition of Molodtsov that we can rewrite as follows:

Definition 2 *A soft set is a 3-uple (X, E, F) if and only if X is an initial space, E another space, and F a mapping of E into the set of all subsets of the set X that is: $F : E \longrightarrow \mathcal{P}(X)$.*

Remark 1 *In the two definitions above:*

- *The nature of mapping F is not specified;*

Copyright © 2024 by the authors.
This open access article is published under the Creative Commons Attribution-NonCommercial 4.0 International License.

- *The nature of parameters and their status with respect to the space X is not specified.*

We can see that in different fields of sciences, technologies or economics, for a real decision making, the notion of *parameter* receive different meanings. This fact is important in real applications. That's why we need to sharpen the notion of soft set without taking it off some degree of generality. Therefore, we introduce the notion of *double soft set* .

3 Relation between soft sets and Formal Concept Analysis

The Formal Concept Analysis (FCA) model is presented in (Ganter, Stumme, and Wille 2005).

Definition 3 *A formal context is a triple (O, A, I) where O is a set of objects, A is a set of attributes and I is a binary relation from O to A defined by:*
$\forall o \in O, a \in A,$
$I(o, a) = 1$ *when the object o has the attribute a*
$I(o, a) = 0$ *otherwise.*

Starting from a formal context, by defining two derivation operators, a *formal concept* is defined as being a pair of sets: its *extension* and its *intension.* The extension is a set of objects, the intension a set of attributes such that each object from extension has all attributes of its intension and conversely. In (Ganter, Stumme, and Wille 2005), it is shown that all formal concepts in this way defined form a Galois lattice.

We establish the relation between the two models of categorization namely, Formal Concept Analysis (FCA) and soft sets. This relation is given in theorem 1.

Theorem 1 *The concept lattice generated by a FCA model is a particular soft set.*

Proof 1 *Let us consider a FCA model given by a set of objects* $O = \{o_1, o_2, ...o_n\}$*, a set of attributes* $A = \{a_1, a_2, ...a_m\}$ *and the formal context matrix, Conx. Conx is the binary relation I from O to A defined by:*
$\forall o \in O, a \in A,$
$I(o, a) = 1$ *when the object o has the attribute* a*,*
$I(o, a) = 0$ *otherwise.*

The soft set (X, E, F) *associated to this model can be defined as follows:*

- $X = O$*,* $E = A$
- $F : A \to O$ *defined by:* $\forall a_i \in A,$
  $F(a_i) = 1$ *if* $I(o, a_i) = 1$*, and* $F(a_i) = 0$ *otherwise.*

Remark 2 *The following aspects are important for a proper use of soft sets and FCA model, and for relating both:*

- *Attributes are considered as parameters.*
- *The particularity of FCA is that* F *defined by FCA conditions is not a mapping but a relation in mathematical sense. Otherwise, for each attribute* a_i*,* $F(a_i)$ *is a subset of O, but* F *is a relation not a mapping.*
- *From an applicative point of view, it will be interesting to see which cognitive element is given in addition by the Galois lattice interpreted in the frame of soft sets.*
- *We can see that the Molodtsov condition for* F *to be a mapping is more restrictive. It will be interesting to analyse in which kind of applications we keep the condition for* F *to be a mapping and in which kind of applications we relax this condition.*

4 An example of application to materials selection in engineering

In this section, we look at the choice of materials for ordinary products. We show that with little information, a rational choice can be made from two points of view: products' resistance to the environment and their environmental qualities. We use elements of FCA.

We look at the resistance (durability) of technical products in the following environments: ultraviolet radiation (photodegradation), aerated water or salt water (oxidation), acid or basic/alkali aqueous solutions (chemical attacks), and by organic solvents (alteration). The idea is to select materials with which to manufacture products that will withstand their intended use for a sufficiently long time.

In (Ashby 1999) an analysis of material selection in mechanical design is given. Chapter 4 is dedicated to analysis of existing charts of material selection. He explains:
In mechanical engineering, the selection of materials can be done by a method which consists in looking at the material first as a set of properties, from which the engineer makes his design.

We worked on the durability diagram from (Ashby 1999) presented in Figure 1. It's a representation of the comparative ranking of resistance of materials to attacks by six environments. Ashby considers this diagram as an introduction to a problem which "requires a detailed and complex expertise". Figure 1 presents a categorization of four classes of materials, Alloys, Polymers, Ceramics and Composites depending on their resistance against attacks by six environments: Aerated Water, Salt Water, Strong Acids, Strong Alkalis, Organic Solvents and U-V Radiation. Their resistance in terms of lifespan (durability) is ranked in four levels: A (Excellent), B (Good), C (Poor) and D (Bad).

We analyzed the diagram in Figure 1 using FCA model. We applied the tool ConExp-ng (Kiss 2013) to different formal contexts corresponding to various situations. We present one of them denoted "context 1" and represented in Figure 2. It corresponds to the case where, for each material group, the maximum strength level of the strength level spectrum of the elements in the group is selected. In this approach, there is no guarantee that this maximum level will be satisfied at the end of the selection process, since the actual level depends on the strength level spectrum of the elements in the group.

We adopted the following notations:
For FCA objects:
Allo <=> Alloys; Poly <=> Polymers;
Compo <=> Composites; Cera <=> Ceramics.

For FCA attributes:
AeWA-{A,B,C,D} <=> Aerated Water-{A,B,C,D};
SaWa-{A,B,C,D} <=> Salt Water-{A,B,C,D};

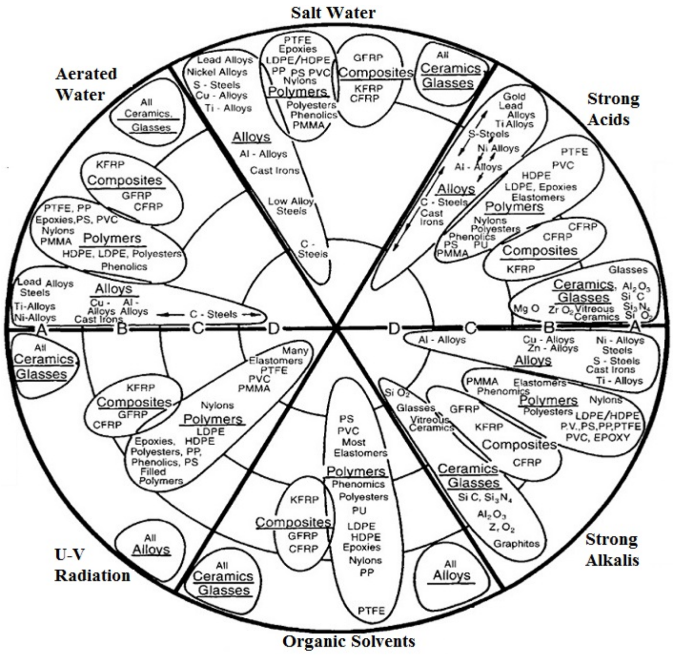

Figure 1: The resistance of materials in environments. Ashby durability diagram (see Ashby 1999).

StAc-{A,B,C,D} <=> Strong Acids-{A,B,C,D};
StAl-{A,B,C,D} <=> Strong Alkalis-{A,B,C,D};
OrSo -{A,B,C,D} <=> Organic Solvents-{A,B,C,D};
UVRA-{A,B,C,D} <=> U-V-{A,B,C,D}.

	AeWa-{A, B, C, D}	SaWa-{A, B, C, D}	StAc-A	StAc-{B, C, D}	StAl-A	StAl-{B, C, D}	OrSo-A	Orso-{B, C, D}	UVRa-A	UVRa-{B, C, D}
Allo	x	x	x	x	x	x	x	x	x	x
Poly	x	x	x	x	x	x	x	x		x
Compo	x	x		x		x		x		x
Cera	x	x	x	x	x	x	x	x	x	x

Figure 2: Ashby diagram - context 1.

In Figure 3, the lattice corresponding to the context 1 is presented.

Analyzing this case, we can make the following conclusions. From the mathematical point of view, the Galois lattice is reduced at a single path. The Alloys and Ceramics are excellent in U-V radiation. Polymers, Alloys and Ceramics have an excellent resistance to Strong Acids, Strong Alkalis and Organic Solvents. The other materials are rather good in all other environments. The single path indicates a complete hierarchy on the concepts and the presence of alloys and ceramics at the entry point of the lattice gives an indication of their higher resistance to environment attacks in comparison to the other groups of materials. Apparently, no additional information is deducible from this Galois lattice.

From the cognitive point of view and mathematical point of view as well, some conclusions resulting from the application of FCA model to above example can be drawn. These conclusions are presented in remark 3.

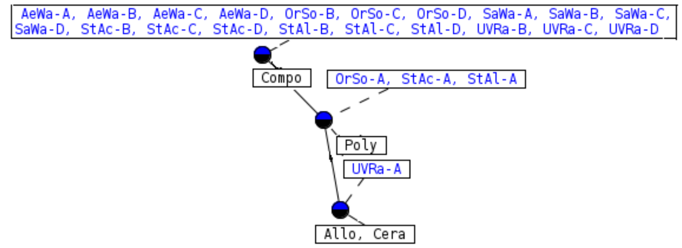

Figure 3: Ashby diagram - lattice 1.

Remark 3 *From the semantic type of parameters and from mathematical modeling, we remark that:*

- *In the above example, there are two types of parameters in the categorization: those related to the environments and those related to the evaluation. These two sets of parameters are completely independent. Environment parameters are more related to materials as objects to be classified. In this example, each material can be related to each environment. There are other cases where it is not the case, meaning that there can be objects not related to some of the parameters. However, evaluation parameters are completely independent from the environments themselves.*
- *From the mathematical point of view, the categorization of materials according to environment parameters and evaluation parameters is a matter of Cartesian product of type* $(X \times Y) \times Z$ *which can not be directly treated by a standard FCA.*
- *This example proves the limit of the FCA model, and therefore the need to conceive a 3-dimensional FCA in order to extend the standard notion of soft set into that of "double soft set".*

5 Double soft sets

5.1 A short analysis of "parameters"

In Molodtsov's definition (Molodtsov 1999), the space of parameters is viewed as a basic set in the sense of axiomatic set theory, without other properties. In practice, depending on applications the problem of parameters is more complex. There are several types of parameters classification following each science or technology.

A first cognitive splitting of parameters is in: quantitative parameters and qualitative parameters. Then, the classification can continue on several levels.

In example from section 4, we have 2 levels of parameters each one of qualitative type (type of environment and level of resistance to the environment).

5.2 Double soft sets definition

The "double soft set" conceptually models an objects space categorization using two sets of parameters. The relation between objects and parameters is more complex than a simple function F. In Figure 4, we represent three possible alternatives.

These three possible alternatives concern the relationship between objects and parameters, and between the parameter

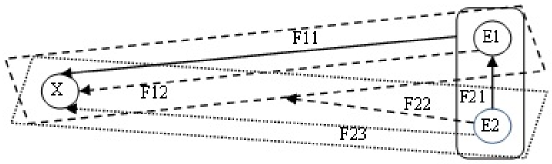

Figure 4: Three cases of double soft set.

levels themselves. Let us suppose a space X of objects and two spaces of parameters E_1 and E_2. They are:

1. The space E_2 creates a categorization on the space E_1 (F21) and this last one creates the categorization on X (F11). In Figure 4, this alternative is represented with solid lines. It is the case of *a priori* classification of parameters in derived parameters often encountered in physics and technology.
2. The space E_1 creates a categorization on the space X (F12) and at this categorization one applies the categorization imposed by the space E_2 (F22). In Figure 4, this alternative is represented with dashed lines. It is the case of the categorization in the example of section 4.
3. The space E_1 creates a categorization on the space X (F12). Independently, in parallel, the space E_2 creates a categorization on the space X (F23). Finally, one compares the two obtained categorizations. In Figure 4, this alternative is represented dashed lines and dotted lines.

Before defining the notion of double soft set, we first introduce the following preliminary definition.

Definition 4 *A categorization of a space X depending on a space E, denoted by $\mathcal{C}$(X, E), is a subset of $\mathcal{P}$(X) depending on E.*

Definition 5 *A* **double soft set** *is a 5-uple (X, E_1, E_2, F_1, F_2) where F_2 and F_1 are defined using one of the following options:*

1. *F_2 is a mapping of E_2 into the set of all subsets of the set E_1, let it be $\mathcal{C}(E_1,E_2)$ and F_1 is a mapping of $\mathcal{C}(E_1,E_2)$ into the set of all subsets of the set X.*
2. *F_1 is a mapping of E_1 into the set of all subsets of the set X, let it be $\mathcal{C}$(X, E_1) and F_2 is a mapping of E_2 into $\mathcal{C}$(X, E_1).*
3. *F_1 and F_2 are respectively defined as the mapping of E_1 into X and the mapping of E_2 into X. The mapping results are then combined to obtain the final categorization of X.*

Remark 4 *Definition 5 is a double "*softification*". Functions F_1, F_2 can be viewed each as a 3-uple relation establishing a link between parameters and objects, in one of the following three cases:*

1. *p_2 to p_1 and then, to x;*
2. *p_1 to x and then, to p_2;*
3. *p_1 to x, p_2 to x and then combining.*

Remark 5 *The example in section 4 is an example of a double soft set. We can link materials with environments firstly (F_1) and, then, obtained classes with ranking levels (F_2).*

6 Application of the double soft sets method in engineering

In several engineering applications there are a lot of parameters to take into account in the design of a product. Materials satisfying some strong constraints on the parameters have to be selected. Categorization is necessary in this purpose. The FCA model is not sufficient in most cases. Modeling a problem of categorization by soft sets allows to take better account of parameters properties, by applying an "a priori" parameters' categorization corresponding to the needs. Often the parameters are ordered. This means that the notion of order relation is important to take into account here. As presented in the previous sections, a bridge can be built between the notion of Galois lattice and that of soft set. The following theorem can be exploited in the case of modeling under ordered parameters.

Theorem 2 *Any lattice can be the Galois lattice of a binary relation. Conversely, two binary relations can have the same Galois lattice (or, more rigorously, two isomorphic Galois lattices) (Ganter, Stumme, and Wille 2005).*

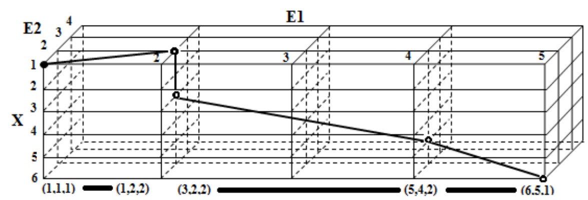

Figure 5: Relation on (X, E_1, E_2).

Another classical order is "lexicographic order". It also leads to a Galois lattice, but a special one. In Figure 5, a 3-D Galois lattice is represented as soft set.

7 Conclusions and future work

The soft set theory extends the possibilities of analysis because of its categorization "guided by parameters". The complexity of parameters justify the introduction of a new notion. In this paper, we extend the notion of soft set to the derived notion of "double soft set". We stay in the general context where a soft set is viewed as a categorization tool. The application of this notion to material selection in engineering shows its benefit for explainability, contrarily to other existing approaches. It could be complementary to artificial intelligence approaches such as deep learning.
For future work, we plan to investigate further the quantitative - qualitative aspects in order to better model the notion of parameter (e.g. physical variables and parameters in engineering) in the context of the double soft sets approach for categorization problems in engineering.

References

Alcantud, J.C.R. et al. (2024). "A systematic literature review of soft set theory". In: *Neural Comput & Applic*.
Ashby, M. (1999). *Material Selection in Mechanical Design*. Oxford, Butterworth-Heinemann.

De Glas, M. (2010). “Sortir de l’enfer cantorien”. In: *Intellectica*.
Desclés, J-P., A. C. Pascu, and I. Biskri (2017). “A topological approach for the notion of quasi topology structure”. In: *South American Journal of Logic* 10.
Ganter, B., G. Stumme, and R. Wille (2005). *Formal Concept Analysis. Foundations and Application*. Heidelberg, Springer.
Irwinski, T.B. (1987). “Algebraic approach to rough sets”. In: *Bulletin of the Polish Academy of Science and Mathematics* 35.
Kiss, E. (2013). “Concept Explorer Next Generation”. In: URL: `https://github.com/fcatools/conexp-ng/wiki/About`.
Molodtsov, D. (1999). “Soft Set Theory – First Results”. In: *Computers and Mathematics with Applications* 37.
Pawlak, Z. (1982). “Rough Sets”. In: *International Journal of Computer and Information Sciences* 11(5).
Pomykala, P. (1987). “Approximation operations in approximation space”. In: *Bulletin of the Polish Academy of Science and Mathematics* 35.
Vinogradov, D. (2020). “Continuous Attributes for FCA-based Machine Learning”. In: *Proceedings of Eighth International Workshop “What can FCA do for Artificial Intelligence?” FCA4AI, 24th European Conference on Artificial Intelligence ECAI 2020*, pp. 103–111.
Zadeh, L. (1965). “Fuzzy Sets”. In: *Information and Control 8*.