Rough Set Theory and Formal Concept Analysis - a brief comparison

Piotr Wasilewski
Department of Logic
Jagiellonian University
email: wasilew@grodzki.phils.uj.edu.pl

We shall investigate into connections between Rough Set Theory (R.S.T.) and Formal Concept Analysis (F.C.A.). These two important tools in Artificial Intelligence (A.I.) are founded on the same basis - lattice theory, and are applicable in similar domains such as data analysis, classification of objects, approximate reasoning. However, they quite different approaches to these issues

1. Preliminaries

By term *space* we understand any nonvoid set. If U is any set by $\mathcal{P}(U)$ we denote the family of all subsets of U. By $Eq(U)$ we denote the family of equivalence relations on U. We will use terms *set* and *property* as they are used in classical set theory and A.I. *Sets of sets* and *sets of properties* will be called *families*. Let us note that despite some similarities sets and properties are quite different notions: properties violate The Extension Principle which for sets says that any two sets consisting of the same elements are equal (identical).

Let X be any set and $x, y \in X$. For any set $A \subseteq X$ by $\approx_A$ we denote a relation of indiscernibility with respect to A and define it as follows:

$$x \approx_A y \ \Leftrightarrow_{def} \ x \in A \Leftrightarrow y \in A.$$

For any family $\mathcal{A} = \{A_i\}_{i \in I} \subseteq \mathcal{P}(X)$ by $\approx_{\mathcal{A}}$ we denote a relation of indiscernibility with respect to this family, and define it in the following way:

$$(x, y) \in \approx_{\mathcal{A}} \Leftrightarrow_{def} \ (x, y) \in \bigcap_{i \in I} \approx_{A_i} .$$

2. Rough set theory

Rough set theory was invented by Zdzisław Pawlak in the beginning of eighties. He introduced the notions of approximation space and rough sets, and with E. Orłowska and others, developed R.S.T. theory and applied it in a wide range of problems within A.I. and philosophical logic e.g. in semantics of vague concepts.

Let U be a space and $R \in Eq(U)$, then (U, R) is called *an approximation space*. Equivalence classes of R are called *R-atoms*. We assume that $\emptyset$ is also an atom. A set $X \subseteq U$ is called *R-composed* iff is a sum of some atoms. By $\mathrm{Com}_R(U)$ we denote the set of all R-composed subsets of U. For each $X \subseteq U$ *lower* and *upper approximation* of X in

(U,R) are defined respectively:

$$\underline{R}(X) = \bigcup\{Y \in U_{/R} : Y \subseteq X\},$$

$$\overline{R}(X) = \bigcup\{Y \in U_{/R} : Y \cap X \neq \emptyset\}$$

The following theorem holds:

Theorem 1. (Pawlak) *Let (U,R) be an approximation space, then $(\mathrm{Com}_R(U); \cup, \cap, ^*, \emptyset, U)$ is a complete, atomic Boolean algebra (where * is a complement of a set).*

Let (U,R) be an approximation space, for $X, Y \subseteq U$ we say that X is *roughly included in* Y ($X \subseteq_R Y$ iff $\underline{R}(X) \subseteq \underline{R}(Y)$ and $\overline{R}(X) \subseteq \overline{R}(Y)$. We say that a set X is *roughly equal* to a set Y ($X =_R Y$) iff $\underline{R}(X) = \underline{R}(Y)$ and $\overline{R}(X) = \overline{R}(Y)$. The relation $=_R$ is called a rough equality of the approximation space (U,R).
One can easily show that $=_R$ is an equivalence relation on $\mathcal{P}(U)$, whence the pair $(\mathcal{P}(U), =_R)$ is also an approximation space. Equivalence classes of $=_R$) are called *rough sets*. For any set $X \subseteq U$ the equivalence class of the relation $=_{(U,R)}$ containing X is denoted by $[X]_{=_R}$, thus

$$[X]_{=_R} = \{A \subseteq R : \underline{R}(A) = \underline{R}(X) \text{ and } \overline{R}(A) = \overline{R}(X)\}.$$

The family of rough sets of an approximation space (U,R) we denote by $\mathcal{R}_R(U)$. On family $\mathcal{R}_R(U)$ we define the following relation:

$$[X]_{=_R} \leqslant_R [Y]_{=_R} \Leftrightarrow X \subseteq_R Y.$$

One can show, that $\leqslant_R$ is a partial order on $\mathcal{R}_R(U)$. It also has been proved that a partially ordered set $(\mathcal{R}_R(U), \leqslant_R)$ is a lattice. Moreover, the following theorem holds:

Theorem 2. (J.A. Pomykała, J. Pomykała) *The lattice of rough sets of any approximation space is a complete, atomic Stone algebra.*

3. Formal Concept Analysis

Since 1979 Rudolf Wille and the members of "Forschungsgruppe Begriffsanalyse" at the Technische Hochschule Darmstadt have been developing Formal Concept Analysis. F.C.A. proved to be fruitful in a great variety of domains such as: conceptual data analysis, conceptual knowledge processing or classification and identification of objects over a given family of properties.

A formal context is a triple $(U, \Phi, \lhd)$ where U and Φ are sets, while $\lhd$ is a binary relation between U and Φ ($\lhd \subseteq U \times \Phi$). The elements of U are called objects, the elements of Φ are called properties. Let $a \in U$ and $\phi \in \Phi$, the expressions $a \lhd \phi$ is read as follows: the object a has the property ϕ or equivalently the property ϕ is possessed by the object a . We introduce between power sets of U and Φ two different operators ($X \subseteq U$, $Y \subseteq \Phi$):

$$X \mapsto X' = \{\phi \in \Phi : a \lhd \phi \ for\ all\ a \in X\},$$

$$Y \mapsto Y' = \{a \in U : a \lhd \phi \ for\ all\ \phi \in Y\}.$$

Although the operators defined above are different, adopting common practice (see Wille [6]), we denote them using the same symbol.

A formal concept of a formal context $(U, \Phi, \lhd)$ is a pair (A, B) with $A \subseteq U$, $B \subseteq \Phi$, where $A = B'$ and $B = A'$. A and B are called *the extent* and *the intent* of the formal concept (A, B). The family of all formal concepts of a formal context $(U, \Phi, \lhd)$ we denote by $W(U, \Phi)$. On family $W(U, \Phi)$ we define relation $\precsim$ in the following way:

$$(A_1, B_1) \precsim (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2.$$

Let us note that the relation $\precsim$ is a partial order on $W(U, \Phi)$. The following theorem holds:

Theorem 3. R.Wille (The Basic Theorem on Concept Lattices)
Let $(U, \Phi, \lhd)$ be a formal context. Then $W(U, \Phi)$ is a complete lattice called the concept lattice of a formal context $(U, \Phi, \lhd)$ for which suprema and infima have the following forms:

$$\bigwedge_{i \in I} (A_i, B_i) = (\bigcap_{i \in I} A_i, (\bigcup_{i \in I} B_i)''),$$

$$\bigvee_{i \in I} (A_i, B_i) = ((\bigcup_{i \in I} A_i)'', \bigcap_{i \in I} B_i).$$

4. Comparison

For comparison between R.S.T. and F.C.A we will present an equivalent formulation of the notion of *approximation space* using indiscernibility relations. This formulation enables us to produce an approximation space for any family of properties. On this base the differences between R.S.T. and F.C.A. will be investigated. The first difference is that each approximation space produces two special complete distributive lattices: a Boolean algebra of composed sets and a Stone algebra of rough sets, both of distributive while some formal contexts produce complete lattices, which are nondistributive. We will also try to find sufficient and (or) necessary conditions for a concept lattice to be a Boolean algebra and discuss their meaning for conceptual data analysis.

References

[1] Z. Bonikowski. A certain conception of the calculus of rough sets. *Notre Dame Journal of Formal Logic*, 3(3):412–421, Summer 1992.

[2] E. Orłowska. *Semantics of vague concepts. In G. Dorn and P. Weingartner, editors, Foundations of Logic and Linguistics. Problems and Solutions. Selected Contributions to the 7th International Congress of Logic, Methodology.* Plenum Press, 1983.

[3] Z. Pawlak. Rough sets. *International Journal of Computing and Information Sciences*, 11:341–356, 1982.

[4] Z. Pawlak. Rough sets: An algebraic and topological approach. *ICS PAS Report, Polish Academy of Sciences*, (482), 1982.

[5] B. Ganter R. Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer, Heidelberg, 1999.

[6] R. Wille. *Introduction to Formal Concept Analysis*. Tehnische Hochschule Darmstadt, preprint, 1996.