A Semi-Boolean Generalization of Rough Sets

Marcin Wolski
Department of Logic and Methodology of Science,
Maria Curie-Skłodowska University, Poland
marcin.wolski@umcs.lublin.pl

1. Introduction

The present paper applies Rauszer's investigations on semi-Boolean algebras (SBAs) [6, 7, 8] to the mathematical foundations of Rough Set Theory (RST) [9]. The foundations of RST can be approached from several directions (e.g., [10]). Here, we use the theory of Galois connections [3, 12] to prove that the operators of RST are "only one half" of their respective SBAs. The "second half" is given by operators fundamental for another important theory of data analysis and knowledge discovery, namely Formal Concept Analysis (FCA) [11]. However, the paper is confined only to "metamathematics" of RST and FCA, leaving aside a more general presentation of these theories. Moreover, the theory of SBAs allows us to introduce the concept of *$\mathcal{R}$-rough set* generalizing the Pawlak's concept of rough set. We prove that operators involved in the concept of $\mathcal{R}$-rough set provide a semantics for tense logic S4.t in the same way, as the operators of RST provide a sematics for modal logic S5.

2. Semi-Boolean Algebras

In the present section we introduce the basic concepts of the theory of semi-Boolean algebras [6, 7, 8].

Definition 2.1. (Semi-Boolean algebra)
A *semi-Boolean algebra* (SBA) $(\mathcal{U}, \leq, \vee, \wedge, \overset{\wedge}{\to}, \overset{\vee}{\to}, \top, \bot)$ is a Heyting algebra $(\mathcal{U}, \leq, \vee, \wedge, \overset{\wedge}{\to}, \top, \bot)$ equipped with the operation of *pseudo-difference* $\overset{\vee}{\to}$, that is:

$$a \vee b \geq c \text{ if and only if } b \geq a \overset{\vee}{\to} c, \text{ for all } a, b, c \in \mathcal{U}$$

A detailed exposition of the properties of semi-Boolean algebras may be found in [7, 8]. The standard method of producing SBAs is based on preorders (i.e. reflexive and transitive relations). There is a bijective correspondence between preorders and Alexandroff topologies (i.e. topologies closed under arbitrary intersections and arbitrary unions). Given an Alexandroff topological space $(\mathcal{U}, \tau)$ we define a preorder $\leq$, called the *specialization order* of $(\mathcal{U}, \tau)$, as follows:

$$a \leq b \text{ iff } \nabla b \subseteq \nabla a, \text{ where } \nabla a = \{b \in \mathcal{U} : a \leq b\}$$

A preordered set $\mathcal{S} = (\mathcal{U}, \leq)$ induces a topological space $(\mathcal{U}, \tau_{\mathcal{S}})$, where sets ∇a, for all $a \in \mathcal{U}$, form a basis of $\tau_{\mathcal{S}}$. Rauszer defines two operations on members of $\tau_{\mathcal{S}}$:

$$A \overset{\wedge}{\to} B = \{a \in \mathcal{U} : (\forall b \geq a)(b \in A \Rightarrow b \in B)\}$$

$$A \overset{\vee}{\rightarrow} B = \{a \in \mathcal{U} : (\exists b \leq a)(b \in A \ \& \ b \notin B)\}$$

Proposition 2.1. (C. Rauszer)
Let $\mathcal{S} = (\mathcal{U}, \leq)$ be a preordered set and $(\mathcal{U}, \tau_{\mathcal{S}})$ its Alexandroff topological space, then the algebra $(\tau_{\mathcal{S}}, \subseteq, \cap, \cup, \overset{\wedge}{\rightarrow}, \overset{\vee}{\rightarrow}, \emptyset, G)$ is a SBA.

3. Galois connections

This section is concerned with mathematical foundations of Rough Set Theory (RST) [9] and Formal Concept Analysis (FCA) [11]. The presentation of Galois connections is based on [3].

Definition 3.1. (Galois Connection)
Let $(\mathcal{U}, \leq)$ and $(\mathcal{V}, \preceq)$ be partially ordered sets (posets). If $\pi_* : \mathcal{U} \to \mathcal{V}$ and $\pi^* : \mathcal{V} \to \mathcal{U}$ are functions such that for all $a \in \mathcal{U}$ and $b \in \mathcal{V}$, $a \leq \pi^* b$ iff $\pi_* a \preceq b$, then the quadruple $\pi = \langle (\mathcal{U}, \leq), \pi_*, \pi^*, (\mathcal{V}, \preceq) \rangle$ is called a *Galois connection*, where π_* and π^* are called the coadjoint and adjoint part of π, respectively.

Given two posets $(\mathbf{P}\mathcal{U}, \subseteq)$ and $(\mathbf{P}\mathcal{V}, \subseteq)$, where $\mathbf{P}\mathcal{U}$ is the power set of $\mathcal{U}$, we may introduce two types of Galois connections: a covariant Galois connection π (called axiality) between $(\mathbf{P}\mathcal{U}, \subseteq)$ and $(\mathbf{P}\mathcal{V}, \subseteq)$, and a contravariant Galois connection π (called polarity) between $(\mathbf{P}\mathcal{U}, \subseteq)$ and $(\mathbf{P}\mathcal{V}, \subseteq)^{\mathrm{op}} = (\mathbf{P}\mathcal{V}, \supseteq)$.

Proposition 3.1. Any relation $R \subseteq \mathcal{U} \times \mathcal{V}$ induces a covariant Galois connection (axiality) ${R_\exists}^\forall = \langle (\mathbf{P}\mathcal{U}, \subseteq), R_\exists, R^\forall, (\mathbf{P}\mathcal{V}, \subseteq) \rangle$, where $R_\exists$ and $R^\forall$ are defined as follows: for any $A \subseteq \mathcal{U}$ and $B \subseteq \mathcal{V}$,

$$R_\exists(A) = \{b \in \mathcal{V} : (\exists a \in \mathcal{U}) \langle a, b \rangle \in R \ \& \ a \in A\}$$

$$R^\forall(B) = \{a \in \mathcal{U} : (\forall b \in \mathcal{V}) \langle a, b \rangle \in R \Rightarrow b \in B\}$$

The theoretical dual of ${R_\exists}^\forall$, defined as ${R^\exists}_\forall = \langle R^\exists, R_\forall \rangle = {(R^{-1})_\exists}^\forall$, is also an axiality but from $(\mathbf{P}\mathcal{V}, \subseteq)$ to $(\mathbf{P}\mathcal{U}, \subseteq)$. R^{-1} means the converse relation of R, that is, $bR^{-1}a$ iff aRb. Now, we recall basic concepts of RST.

Definition 3.2. (Approximation Operators, Rough Sets)
Let $\mathcal{U}$ be a set, E an equivalence relation on $\mathcal{U}$, and $[a]_E$ – the equivalence class containing $a \in \mathcal{U}$. With each $A \subseteq \mathcal{U}$, we can associate its E-lower and E-upper *approximations*, $\underline{A}$ and $\overline{A}$, respectively, defined as follows:

$$\underline{A} = \{a \in \mathcal{U} : [a]_E \subseteq A\} \text{ and } \overline{A} = \{a \in \mathcal{U} : [a]_E \cap A \neq \emptyset\}$$

A *rough set* is an equivalence class of sets which are indistinguishable by their upper and lower approximations. A pair $(\mathcal{U}, E)$ is called *approximation space*. A subset $A \subseteq \mathcal{U}$ is called *definable* if $A = \bigcup B$ for some $B \subseteq \mathcal{U}/E$, where $\mathcal{U}/E$ is the family of equivalence classes of E.

Each approximation space $(\mathcal{U}, E)$ may be converted into topological space $(\mathcal{U}, \tau_E)$, where the family $\mathcal{U}/E$ forms a basis of τ_E. In the literature topological spaces induced by equivalence relations are called *topological approximation spaces* [5]. On this view, the lower approximation $\underline{A}$ is the interior of $A \subseteq \mathcal{U}$ and the upper approximation $\overline{A}$ is the closure of A. A set $A \subseteq \mathcal{U}$ is definable only if $A \in \tau_E$.

Proposition 3.2. (I. Düntsch and G. Gediga)
Let $(\mathcal{U}, \tau_E)$ be a topological approximation space and $A \subseteq \mathcal{U}$ then:

$$E^{\forall}E_{\exists}(A) = \overline{A} \text{ and } E^{\exists}E_{\forall}(A) = \underline{A}$$

This result – dressed differently – has been proved in [1, 2]. The direct relationship between Galois connections and RST has been observed in [12]. Since SBAs are related to Alexandroff topological spaces, we shall generalize this proposition.

Proposition 3.3. Let $(\mathcal{U}, \tau)$ be an Alexandroff topological space, R its specialization order, $\mathcal{I}$ and $\mathcal{C}$ the interior and closure operators induced by τ, respectively. Then

$$R^{\forall}R_{\exists}(A) = \mathcal{C}(A) \text{ and } R^{\exists}R_{\forall}(A) = \mathcal{I}(A) \text{ for all } A \subseteq \mathcal{U}$$

This proposition shows that the approximation operators of RST, when interpreted on the basis of Alexandroff topological spaces, preserve their original topological meanings in terms of interior and closure operators. However, the upper approximation may be indefinable. The solution to this problem is brought by the concept of bitopological space.

Definition 3.3. (Bitopological Space)
A *bitopological space* $(\mathcal{U}, \mathcal{I}, \mathcal{C})$ is a non-empty set equipped with an interior operation $\mathcal{I}$ and a closure operation $\mathcal{C}$ satisfying:

$$\mathcal{I}(A) = \mathcal{C}\mathcal{I}(A) \text{ and } \mathcal{C}(A) = \mathcal{I}\mathcal{C}(A), \text{ for all } A \subseteq \mathcal{U}$$

Please note that any toplogical approximation space $(\mathcal{U}, \tau)$ gives rise to a bitopological space $(\mathcal{U}, \mathcal{I}, \mathcal{C})$, where both operators, $\mathcal{I}$ and $\mathcal{C}$, are induced by τ. But in case of an Alexandroff topological space $\mathcal{S} = (\mathcal{U}, \tau)$, the lower approximation operator $R^{\exists}R_{\forall}$ is "only one half" of its bitopological space $(\mathcal{U}, R^{\exists}R_{\forall}, \mathcal{C})$, since $\mathcal{C}$ – in generall – may be different from $R^{\forall}R_{\exists}$. It is clear that $\mathcal{C}$ returns τ-open sets, i.e. definable sets, as it is required by RST. The same argument applies to the upper approximation operator $R^{\forall}R_{\exists}$ and $(\mathcal{U}, \mathcal{I}, R^{\forall}R_{\exists})$. Now we find the "second half" of RST.

Proposition 3.4. Any relation $R \subseteq \mathcal{U} \times \mathcal{V}$ induces a contravariant Galois connection (polarity) ${R_+}^+ = \langle(\mathbf{P}\mathcal{U}, \subseteq), R_+, R^+, (\mathbf{P}\mathcal{V}, \subseteq)\rangle$, where R_+ and R^+ are defined as follows: for any $A \subseteq \mathcal{U}$ and $B \subseteq \mathcal{V}$,

$$R_+A = \{b \in \mathcal{V} : (\forall a \in A)\langle a, b\rangle \in R\}$$

$$R^+B = \{a \in \mathcal{U} : (\forall b \in B)\langle a, b\rangle \in R\}$$

Below, we shall present FCA in a very concise way to give the reader at least a "taste" of this theory. For a detailed exposition of FCA see [11].

A triple $\langle\mathcal{U}, \mathcal{V}, R\rangle$, where $R \subseteq \mathcal{U} \times \mathcal{V}$, is called a *context*. Each context is associated with two operators R_+ and R_+ called *derivation operators*. These operators allows one to build concepts, i.e. meaningful entities which constitute our knowledge about the context.

Definition 3.4. (Concept)
A concept of a given context $\langle\mathcal{U}, \mathcal{V}, R\rangle$ is a pair (A, B), where $A \subseteq \mathcal{U}$ and $B \subseteq \mathcal{V}$ such that $A = R^+B$ and $B = R_+A$.

The collection of all concepts of a given context is ordered by a subconcept-superconcept relation defined as follows: $(A_1, B_1) \leq (A_2, B_2)$ iff $A_1 \subseteq A_2$ (equivalently, $B_1 \supseteq B_2$). The set of all concepts of a given context $\langle \mathcal{U}, \mathcal{V}, R \rangle$ together with the defined order $\leq$ is denoted by $\mathbb{C}\langle \mathcal{U}, \mathcal{V}, R \rangle = \{(A, B) : A = R^{+}B \; \& \; B = R_{+}A\}$. The fundamental theorem of FCA states that:

Proposition 3.5. (Wille)
For any formal context $\langle \mathcal{U}, \mathcal{V}, R \rangle$, $\mathbb{C}\langle \mathcal{U}, \mathcal{V}, R \rangle$ is a complete lattice, called the concept lattice of $\langle \mathcal{U}, \mathcal{V}, R \rangle$, for which infima (meet) and suprema (join) are respectively:

$$\bigwedge_{t \in T} (A_t, B_t) = (\bigcap_{t \in T} A_t, R_{+}R^{+} \bigcup_{t \in T} B_t) \text{ and } \bigvee_{t \in T} (A_t, B_t) = (R^{+}R_{+} \bigcup_{t \in T} A_t, \bigcap_{t \in T} B_t).$$

Basically, FCA deals with hierarichies (i.e. lattices) of concepts induced by contexts. Here, we are more interested in derivation operators rather than concepts.

Proposition 3.6. Let $(\mathcal{U}, \tau)$ be an Alexandroff topological space and R its specialization order, then $(\mathcal{U}, R^{\exists}R_{\forall}, R^{+}R_{+})$ is a bitopological space.

Proposition 3.7. Let $(\mathcal{U}, \tau)$ be an Alexandroff topological space and R its specialization order, then the algebra $(\tau, \subseteq, \cup, \cap, \overset{\wedge}{\to}, \overset{\vee}{\to}, \mathcal{U}, \emptyset)$ is a SBA:

$$A \overset{\wedge}{\to} B = R^{\exists}R_{\forall}(-A \cup B)$$

$$A \overset{\vee}{\to} B = R^{+}R_{+}(A \cap -B)$$

for all $A, B \in \tau$, where $-$ is the set complement.

The same result may be obtained for the upper approximation operator $R^{\forall}R_{\exists}$. Its semi-Boolean counterpart is given by the dual of $R^{+}R_{+}$, denoted by $dualR^{+}R_{+}$. It means that for $A \subseteq \mathcal{U}$ we have $dualR^{+}R_{+}(A) = -(R^{+}R_{+}(-A))$.

Proposition 3.8. Let $(\mathcal{U}, \tau)$ be an Alexandroff topological space and R its specialization order, then $(\mathcal{U}, dualR^{+}R_{+}, R^{\forall}R_{\exists})$ is a bitopological space.

Proposition 3.9. Let $(\mathcal{U}, \tau)$ be an Alexandroff topological space and R its specialization order, then the algebra $(-\tau, \subseteq, \cup, \cap, \overset{\wedge}{\to}, \overset{\vee}{\to}, \mathcal{U}, \emptyset)$ is a SBA:

$$A \overset{\wedge}{\to} B = dualR^{+}R_{+}(-A \cup B)$$

$$A \overset{\vee}{\to} B = R^{\forall}R_{\exists}(A \cap -B)$$

for all $A, B \in -\tau$, where $-\tau = \{A \subseteq \mathcal{U} : -A \in \tau\}$.

Let $(\mathcal{U}, \tau_E)$ be an approximation topological space and let $(\mathcal{U}, \mathcal{I}, \mathcal{C})$ be the induced bitopological space. Then a rough set may be represented as $(\mathcal{I}(A), \mathcal{C}(A))$, for some $A \subseteq \mathcal{U}$. In this context the proposition 3.6 suggests an easy generalization of rough sets by means of the induced bitopological space: a $\mathcal{R}$-*rough set* of an Alexandroff topological space $(\mathcal{U}, \tau)$ is as a pair os sets $(R^{\exists}R_{\forall}(A), R^{+}R_{+}(A))$, for some $A \subseteq \mathcal{U}$. Please note that both $\mathcal{R}$-approximations of A, namely $R^{\exists}R_{\forall}(A)$ and $R^{+}R_{+}(A)$, are definable in $(\mathcal{U}, \tau)$.

Proposition 3.10. Let $(\mathcal{U}, \tau_E)$ be a topological approximation space induced by an equivalence relation E. Then, the set of rough sets induced by $(\mathcal{U}, \tau_E)$ is equal to the set of $\mathcal{R}$-rough sets induced by this space.

This proposition demonstrates that $\mathcal{R}$-rough sets are in fact generalized rough sets in the sense of RST.

4. $\mathcal{R}$-Rough Sets and Rough Approximations

In this section we discuss the issues arising around RST and its logics. As has been observed by Marek and Truszczyński [4], in most applications sets we want to reason about are not completely specified. In consequence, their approximations may be not rough sets in the sense of Pawlak's definition. In order to cope with such situations, we need a concept much more flexible, than that of rough set, e.g., a *rough aproximation* [4].

Definition 4.1. (Rough Approximations)
Let $(\mathcal{U}, E)$ be an approximation space and $\mathcal{U}/E$ be the family of equivalence classes of E, a pair (A, B) of sets $A \neq B$ is a *rough approximation* if $A \subseteq B$ and both sets A and B are definable.

Proposition 4.1. (W. Marek, M. Truszczyński)
Let $(\mathcal{U}, E)$ be an approximation space and $\leq$ be the following order of its rough approximations: $(A_1, B_1) \leq (A_2, B_2)$ iff $A_1 \subseteq A_2$ and $B_2 \subseteq B_1$, then rough sets are $\leq$-maximal rough approximations.

Now we generalize the concept of rough approximation for Alexandroff topological spaces. It is easy to see that a rough approximation of an Alexandroff space (X, τ) is a pair (U, V) of sets $U \neq V$ such that $U \subseteq V$ and both sets are open.

Proposition 4.2. Let $(\mathcal{U}, \tau)$ be an Alexandroff topological space and let $\leq$ be the order of its rough approximations defined as in the proposition 4.1. Then, $\mathcal{R}$-rough sets are $\leq$-maximal rough approximations.

Hence, $\mathcal{R}$-rough sets are not merely one of the possible generalizations of rough sets. They are also compatible with the "surroundings" of RST. It is well known that Pawlak's approximation operators provide a semantics for modal logic S5. Below, we prove that $\mathcal{R}$-approximation operators have their own logical counterparts as well.

The Heyting-Brouwer sentential calculus (HBL) is formulated in the propositional language $\mathcal{L}_{HB}$ with connectives $\wedge, \vee, \xrightarrow{\wedge}, \xrightarrow{\vee}, \top, \bot$. Its axiomatization was delivered by C. Rauszer in [7].

Proposition 4.3. (C. Rauszer)
A formula of $\mathcal{L}_{HB}$ is provable in HBL iff it is valid in all semi-Boolean algebras.

By the proposition 3.7 R-operators provide a semantics for HBL. We can "copy" these operators from the model of HBL to the language of tense logic S4.t by the extended Gödel translation. The key part is defined as follows:

$$(\alpha \xrightarrow{\wedge} \beta)^t = \Box_F(\alpha^t \xrightarrow{\wedge} \beta^t)$$

$$(\alpha \xrightarrow{\vee} \beta)^t = \Diamond_P(\alpha^t \wedge \neg\beta^t)$$

Definition 4.2. (**R-Topological Model**)
A *R-topological model* is a touple $(\mathcal{U}, \tau, R, V)$ where $(\mathcal{U}, \tau)$ is an Alexandroff topological space, R its specialization order and V the valuation function which assigns propositional letters subsets of $\mathcal{U}$. V is extended on Boolean connectives in the standard way, for modal operators the extension is as follows:

$$V(\Box_F \alpha) = R^{\exists} R_{\forall}(V(\alpha)), V(\Diamond_F \alpha) = R^{\forall} R_{\exists}(V(\alpha))$$

$$V(\Box_P \alpha) = dual R^{+} R_{+}(V(\alpha)), V(\Diamond_P \alpha) = R^{+} R_{+}(V(\alpha))$$

The definition of truth is as usual: $a \vDash \alpha$ iff $a \in V(\alpha)$.

Proposition 4.4. A formula is provable in S4.t iff it is valid in all R-topological models.

As it has been demonstrated [13], a topological space $\mathcal{S}$ is indistinguishable from its finite approximation by means of S4.t. Basically, the finite approximation of $\mathcal{S}$ represents our finite, or better still, incomplete knowledge about the possibly infinite space $\mathcal{S}$. Then, S4.t allows us to reason correctly about $\mathcal{S}$, despite the fact that our knowledge is incomplete. It is a very strong – yet approximate – logic, which fits topology very well. It makes $\mathcal{R}$-rough sets deserve the further scientific attention.

References

[1] Düntch, I., Gediga, G. (2002), Modal-style operators in qualitative data analysis, in *Proc. of the 2002 IEEE Conference on Data Mining* (ICDM'2002), 155-162.

[2] Düntch, I., Gediga, G. (2003), Approximation operators in qualitative data analysis, in de Swart, H. et al. (Eds.), *Theory and Applications of Relational Structures as Knowledge Instruments*, LNCS, 216-233, Springer, Heidelberg.

[3] Erne, M., Klossowski, E., Melton, A., Strecker, G. E. (1991), *A primer on Galois connections*, www.math.ksu.edu/ ˜ strecker/primer.ps.

[4] Marek, W, Truszczyński, M. (1999), Contributions to the theory of rough sets, *Fundamenta Informaticae* 34, 1-21.

[5] Rasiowa, H. (2001), *Algebraic models of logics*, University of Warsaw.

[6] Rauszer, C. (1971), Representation theorem for semi-Boolean algebras I, *Bulletin de l'Académie Polonaise des Sciences*, Vol. XIX. No. 10, 881-887.

[7] Rauszer, C. (1974), Semi-boolean algebras and their applications to intuitionistic logic with dual operations, *Fundamenta Mathematicae* 83, 81-87.

[8] Rauszer, C. (1980), An algebraic and Kripke-style approach to a certain extensions of intuintionistic logic, *Dissertationes Mathematicae*, vol CLXVII.

[9] Pawlak, Z. (1982), Rough sets, *Int. J. Computer and Information Sci.*, 11, 341-356.

[10] Skowron, A., Stepaniuk, J. (1994), Generalized approximation spaces. In *Proceedings of 3rd International Workshop on Rough Sets and Soft Computing*, San Jose, 156-163.

[11] Wille, R. (1982), Reconstructing lattice theory: An approach based on hierachies of concepts, in Rival, I. (Ed.), *Ordered Sets*, NATO Advanced Studies Institute, 445-470, Dordrecht, Reidel.

[12] Wolski, M. (2004), Galois connections and data analysis, *Fundamenta Informaticae* 60, 401-415.

[13] Wolski, M. (2005), Formal Concept Analysis and Rough Set Theory from the perspective of finite topological approximations, *Transactions on Rough Sets* III, LNCS, Springer-Verlag, Berlin Heidelberg, 230-244.