Prof. Dr. Alfred Toth

Rough semiotic sets

1. The present study aims to introduce the newer mathematical concept of "rough set" into mathematical semiotics.[1] A rough set is a formal approximation of a crisp (conventional) set in terms of a pair of sets, which give the lower and the upper approximation of the original set. The lower and upper approximation sets themselves are crisp sets in the standard version of rough set theory (Pawlak 1991), but in other variations, the approximating sets may be fuzzy sets as well ("rough-fuzzy hybridization").

Let I = (**U**, **A**) be an information system, where **U** is a non-empty set of finite objects (the universe) and **A** is a non-empty finite set of attributes such that a: $\mathbf{U} \rightarrow V_a$ for every $a \in \mathbf{A}$. V_a is the set of values which an attribute may take. With any $P \subseteq \mathbf{A}$ there is an associated equivalence relation IND(P):

$$\mathrm{IND}(P) = \{(x, y) \in \mathbf{U}^2 \mid \forall a \in P, a(x) = a(y)\}$$

The partition of **U** generated by IND(P) is denoted **U**/IND(P) and can be calculated as follows:

$$\mathbf{U}/\mathrm{IND}(P) = \otimes\{\mathbf{U}/\mathrm{IND}(\{a\}) \mid a \in P\},$$

where

$$A \otimes B = \{X \cap Y \mid \forall X \in A, \forall Y \in B, X \cap Y \neq \varnothing\}$$

If $(x, y) \in \mathrm{IND}(P)$, then x and y are **indiscernible** by attributes from P. Thus, for any selected subset of attributes P, there will be sets of objects that are indiscernible based on those attributes. These indistinguishable sets of objects therefore define an equivalence or indiscernibility relation, referred to as the P-indiscernibility relation.

2. The simplest example of a semiotic set of objects together with a semiotic set of attributes is the system of the 10 sign classes and reality thematics (dual systems, DS) and their representation values (Rpv), which we will abbreviate by SS10:

SS10:

1. (3.1 2.1 1.1) × (1.1 1.2 1.3)	Rpv = 9
2. (3.1 2.1 1.2) × (2.1 1.2 1.3)	Rpv = 10
3. (3.1 2.1 1.3) × (3.1 1.2 1.3)	Rpv = 11
4. (3.1 2.2 1.2) × (2.1 2.2 1.3)	Rpv = 11
5. (3.1 2.2 1.3) × (3.1 2.2 1.3)	Rpv = 12

1 The definitions are taken from Pawlak (1991), unless other references are mentioned.

6. (3.1 2.3 1.3) × (3.1 3.2 1.3) Rpv = 13
7. (3.2 2.2 1.2) × (2.1 2.2 2.3) Rpv = 12
8. (3.2 2.2 1.3) × (3.1 2.2 2.3) Rpv = 13
9. (3.2 2.3 1.3) × (3.1 3.2 2.3) Rpv = 14
10. (3.3 2.3 1.3) × (3.1 3.2 3.3) Rpv = 15

Let therefore DS1 ... DS 10 be the set of objects X = $\{O_1, ..., O_{10}\}$ and Rpv = 9 ... Rpv = 15 the set of attributes P = $\{P_1, ..., P_7)$. Then we can build the following seven equivalence classes of SS10:

$$\left\{\begin{array}{l} \{O_1\} \\ \{O_2\} \\ \{O_3, O_4\} \\ \{O_5, O_7\} \\ \{O_6, O_8\} \\ \{O_9\} \\ \{O_{10}\} \end{array}\right.$$

Thus, the two objects in the third, fourth and fifth equivalence classes are indiscernible. The equivalence classes of the P-indiscernibility relation are denoted $[x]_p$.

3. Another simple example of a very elementary semiotic information system containing one set of objects and only one set of attributes together with values is the system of the 27 sign classes (cf. f. ex. Toth 2008b):

SS27:

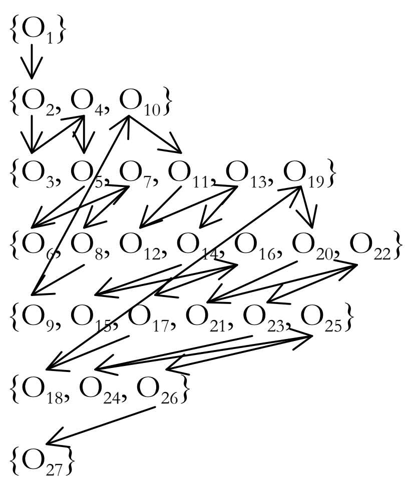
1. (3.1 2.1 1.1) × (1.1 1.2 1.3) Rpv = 9
2. (3.1 2.1 1.2) × (2.1 1.2 1.3) Rpv = 10
3. (3.1 2.1 1.3) × (3.1 1.2 1.3) Rpv = 11
4. *(3.1 2.2 1.1) × *(1.1 2.2 1.3) Rpv = 10
5. (3.1 2.2 1.2) × (2.1 2.2 1.3) Rpv = 11
6. (3.1 2.2 1.3) × (3.1 2.2 1.3) Rpv = 12
7. *(3.1. 2.3 1.1) × *(1.1 3.2 1.3) Rpv = 11
8. *(3.1 2.3 1.2) × *(2.1 3.2 1.3) Rpv = 12
9. (3.1 2.3 1.3) × (3.1 3.2 1.3) Rpv = 13
10. *(3.2 2.1 1.1) × *(1.1 1.2 2.3) Rpv = 10
11. *(3.2 2.1 1.2) × *(2.1 1.2 2.3) Rpv = 11
12. *(3.2 2.1 1.3) × *(3.1 1.2 2.3) Rpv = 12
13. *(3.2 2.2 1.1) × *(1.1 2.2 2.3) Rpv = 11
14. (3.2 2.2 1.2) × (2.1 2.2 2.3) Rpv = 12
15. (3.2 2.2 1.3) × (3.1 2.2 2.3) Rpv = 13
16. *(3.2 2.3 1.1) × *(1.1 3.2 2.3) Rpv = 12
17. *(3.2 2.3 1.2) × *(2.1 3.2 2.3) Rpv = 13

18. (3.2 2.3 1.3) × ($\underline{3.1\ 3.2}$ 2.3) Rpv = 14
19. *(3.3 2.1 1.1) × *($\underline{1.1\ 1.2}$ 3.3) Rpv = 11
20. *(3.3 2.1 1.2) × *($\underline{2.1}$ $\underline{1.2}$ $\underline{3.3}$) Rpv = 12
21. *(3.3 2.1 1.3) × *($\underline{3.1}$ 1.2 $\underline{3.3}$) Rpv = 13
22. *(3.3 2.2 1.1) × *($\underline{1.1}$ $\underline{2.2}$ $\underline{3.3}$) Rpv = 12
23. *(3.3 2.2 1.2) × *($\underline{2.1\ 2.2}$ 3.3) Rpv = 13
24. *(3.3 2.2 1.3) × *($\underline{3.1}$ 2.2 $\underline{3.3}$) Rpv = 14
25. *(3.3 2.3 1.1) × *(1.1 $\underline{3.2\ 3.3}$) Rpv = 13
26. *(3.3 2.3 1.2) × *(2.1 $\underline{3.2\ 3.3}$) Rpv = 14
27. (3.3 2.3 1.3) × (3.1 $\underline{3.2\ 3.3}$) Rpv = 15

In SS27, we have the following seven equivalence classes:

$$\left\{\begin{array}{l} \{O_1\} \\ \{O_2, O_4, O_{10}\} \\ \{O_3, O_5, O_7, O_{11}, O_{13}, O_{19}\} \\ \{O_6, O_8, O_{12}, O_{14}, O_{16}, O_{20}, O_{22}\} \\ \{O_9, O_{15}, O_{17}, O_{21}, O_{23}, O_{25}\} \\ \{O_{18}, O_{24}, O_{26}\} \\ \{O_{27}\} \end{array}\right.$$

Both in SS10 and in SS27 there are thus seven equivalence classes. The structure of the successors of the objects O_i in the equivalence relation is:



Let now X ⊆ **U** be a target set that we wish to represent using an attribute subset P. In general, X cannot be expressed exactly, because the set may include and exclude objects which are indistinguishable based on attributes P. For example, in SS10, consider the target set X = $\{O_2, O_3, O_9\}$, and let the attribute set P = $\{P_1, ..., P_7)$ be the full available set of features. It will be noted that the set X cannot be expressed because in $[x]_p$ objects $\{O_3, O_4\}$ are indiscernible. Thus, there is no way to represent any set X, which includes O_3 but excludes O_4.

4. However, the target set X can be approximated using only the information contained within P by constructing the **P-lower** ($\underline{P}X$) and the **P-upper approximations** ($\overline{P}X$) of X:

$\underline{P}X = \{x \mid [x]_p \subseteq X\}$
$\overline{P}X = \{x \mid [x]_p \cap X \neq \emptyset\}$

The P-lower approximation or **positive region** is the union of all equivalence classes in $[x]_p$ which are contained by the target set. In the above example, $\underline{P}X = \{O_2\} \cup \{O_9\}$. The lower approximation is the complete set of objects in **U**/P that can be positively, i.e. unambiguously, classified as belonging to target set X.

The P-upper approximation is the union of all equivalence classes in $[x]_p$ which have non-empty intersections with the target set. In the example, $\overline{P}X = \{O_2\} \cup \{O_3, O_4\} \cup \{O_9\}$, the union of the three equivalence classes in $[x]_p$ that have non-empty intersections with the target set. The upper approximation is the complete set of objects in **U**/P that cannot be positively (i.e. unambiguously) classified as belonging to the complement of the target set $\overline{X}$. In other words, the upper approximation is the complete set of objects that are **possibly** members of the target set X.

The set **U** – $\overline{P}X$ therefore represents the **negative region**, containing the set of objects that can be definitely ruled out as members of the target set.

If we take as target set SS10, then **U** – $P\overline{X}$ contains all possible triadic relations of the general sign structure <<a.b>, <c.d>, <e.f>> with <a.b>, <c.d>, <e.f> ∈ {<1.1>, <1.2>, <1.3>, ..., <3.3>}, i.e. from the complete relation of prime-signs (Toth 1996). Therefore, with SS10 as target set, the complement SS27\SS10 and thus the 17 dual systems marked above by asterisk, also lie in **U** – $\overline{P}X$. On the other hand, if we take SS27 as target set, then **U** – $\overline{P}X$ contains the high number of 19'683 sign relations, 19'656 of which are not built according to the triadic sign order <<3.b>, <2.d>, <1.f>>.

The **boundary region**, given by the set difference $\overline{P}X - \underline{P}X$, consists of those objects that can neither be ruled in nor ruled out as members of the target set X. In the case we take SS10 as target set, the boundary region thus contains the complement SS27\SS10, i.e. those sign classes (marked by asterisk) which are built according to the triadic sign order <<a.b>, <c.d>, <e.f>> with a = 3., c = 2 and e = 1, but whose trichotomic sign order must not be $b \leq d \leq f$. However, if we take SS27 as target set, the boundary region will be identical with the negative region, unless one defines another semiotic system containing, f. ex., all those sign classes whose general sign structure <<a.b>, <c.d>, <e.f>> fulfills the requirement that a, c, e must be pairwise different but not be ordered in the decreasing (degenerative-semiosic) order (3.a) → (2.b) → (1.c). In other words, this means, that SS27, $\overline{P}X - \underline{P}X$ would contain all possible transformations of the sign classes of SS27 and because of set inclusion also those of SS10. On the other side, SS27 does not contain **any** of the transpositions of the sign classes from SS10, since they must lie, too, in the boundary region of SS27.

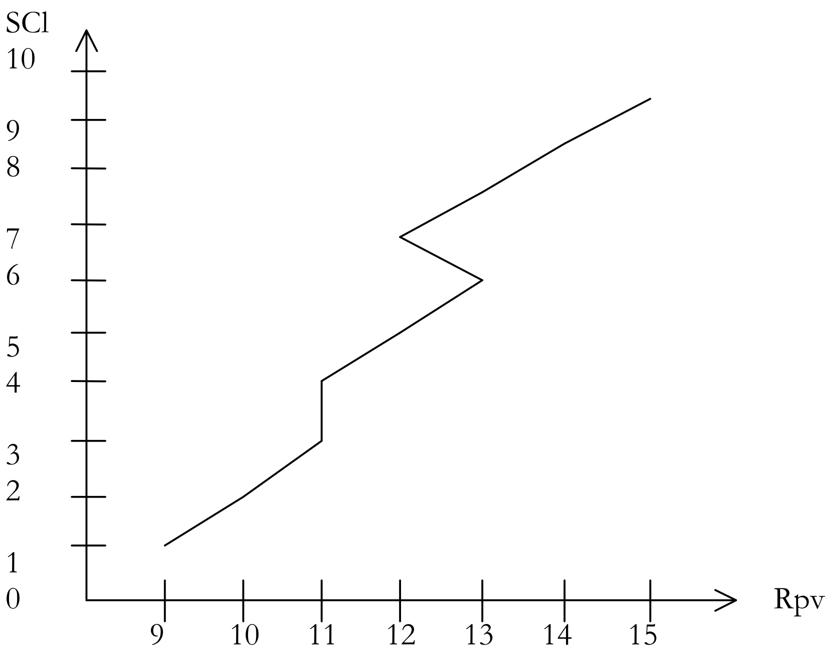
The tuple $<\underline{P}X, \overline{P}X>$, composed of the lower and upper approximation, is called a **rough set**. Thus, a rough set is composed of two crisp sets, one representing a lower boundary of the target set X, and one representing an upper boundary of the target set X.

Therefore, we have

For X = SS10: $<\underline{P}X, \overline{P}X> =$ $<\{O_2\} \cup \{O_9\}, \{O_2\} \cup \{O_3, O_4\} \cup \{O_9\}>$

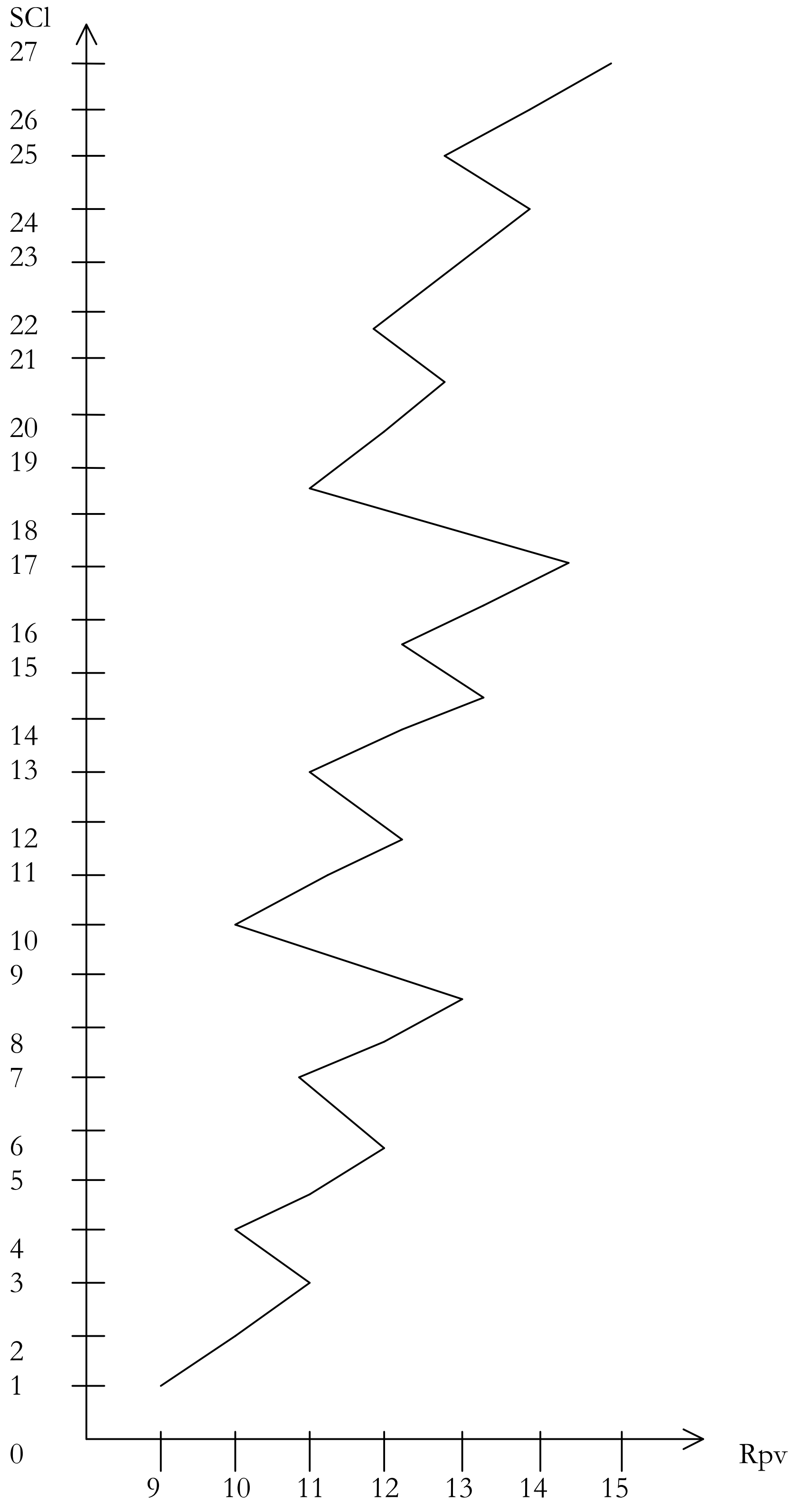
For X = SS27: $<\underline{P}X, \overline{P}X> =$ $<\{O1\} \cup \{O27\}, \{O_1\} \cup \{O_2, O_4, O_{10}\} \cup \{O_3, O_5, O_7, O_{11}, O_{13}, O_{19}\} \cup \{O_6, O_8, O_{12}, O_{14}, O_{16}, O_{20}, O_{22}\} \cup \{O_9, O_{15}, O_{17}, O_{21}, O_{23}, O_{25}\} \cup \{O_{18}, O_{24}, O_{26}\}, \{O_{27}\}>$

We may visualize the rough semiotic set $<\underline{P}X, \overline{P}X>_{SS10}$ in the following graph whose abscissa denotes the interval of representations values (Rpv) |9, 15| and whose ordinate displays the 10 sign classes (or reality thematics) of SS10:



In this graph, the equivalence classes $\{O_3, O_4\}$, $\{O_5, O_7\}$, and $\{O_6, O_8\}$ that contain indiscernible objects are shown very well by the four inflection points.

In order to conclude this first study about rough semiotic sets, we shall also visualize the rough semiotic set $<\underline{P}X, \overline{P}X>_{SS27}$:


SCl
27
26
25
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
9
10
11
12
13
14
15
Rpv

Also in this graph, the equivalence classes $\{O_1\}$, $\{O_2, O_4, O_{10}\}$, $\{O_3, O_5, O_7, O_{11}, O_{13}, O_{19}\}$, $\{O_6, O_8, O_{12}, O_{14}, O_{16}, O_{20}, O_{22}\}$, $\{O_9, O_{15}, O_{17}, O_{21}, O_{23}, O_{25}\}$, $\{O_{18}, O_{24}, O_{26}\}$, and $\{O_{27}\}$ that contain indiscernible objects are shown clearly by the sixteen inflection points.

5. Rough sets thus turn out to be a promising new tool to be applied to mathematical semiotics. For example, we may further construct a semiotic information system on the basis of either SS10 or SS27 and two sets of attributes, f. ex., representation values and structural realities. In this case, both sets of attributes would be fuzzy and rough at the same time. Moreover, we can define a set of objects that does not only contain SS10 or SS27, but also transpositions (Toth 2008a, pp. 177 ss.). Yet anyway, many interesting results are to be awaited.

Bibliography

Pawlak, Zdzislaw, Rough Sets: Theoretical Aspects of Reasoning About Data. Dordrecht 1991

Toth, Alfred, Grundriss einer ordnungstheoretischen Semiotik. In: European Journal for Semiotic Studies 8, 1996, pp. 503-526

Toth, Alfred, Semiotische Strukturen und Prozesse. Klagenfurt 2008 (2008a)

Toth, Alfred, Complete and incomplete fuzzy structural realities. Ch. 29 (2008b)