Rough set theory and its applications

Zdzisław Pawlak

Abstract — In this paper rudiments of the theory will be outlined, and basic concepts of the theory will be illustrated by a simple tutorial example, concerning churn modeling in telecommunications. Real life applications require more advanced extensions of the theory but we will not discuss these extensions here. Rough set theory has an overlap with many other theories dealing with imperfect knowledge, e.g., evidence theory, fuzzy sets, Bayesian inference and others. Nevertheless, the theory can be regarded as an independent, complementary, not competing, discipline in its own rights.

Keywords — *rough set, decision rules, churn modeling.*

1. Introduction

Rough set theory can be regarded as a new mathematical tool for imperfect data analysis. The theory has found applications in many domains, such as decision support, engineering, environment, banking, medicine and others.

This paper presents basis of the theory which will be illustrated by a simple example of churn modeling in telecommunications.

Rough set philosophy is founded on the assumption that with every object of the universe of discourse some information (data, knowledge) is associated. Objects characterized by the same information are *indiscernible (similar)* in view of the available information about them. The *indiscernibility relation* generated in this way is the mathematical basis of rough set theory. Any set of all indiscernible (similar) objects is called an *elementary set*, and forms a basic *granule (atom)* of knowledge about the universe. Any union of some elementary sets is referred to as a *crisp (precise)* set – otherwise the set is *rough (imprecise, vague)*. Each rough set has boundary-line cases, i.e., objects which cannot be with certainty classified, by employing the available knowledge, as members of the set or its complement. Obviously rough sets, in contrast to precise sets, cannot be characterized in terms of information about their elements. With any rough set a pair of precise sets, called the *lower* and the *upper approximation* of the rough set, is associated. The lower approximation consists of all objects which *surely* belong to the set and the upper approximation contains all objects which *possibly* belong to the set. The difference between the upper and the lower approximation constitutes the *boundary region* of the rough set. Approximations are fundamental concepts of rough set theory.

Rough set based data analysis starts from a data table called a *decision table*, columns of which are labeled by *attributes*, rows – by *objects* of interest and entries of the table are *attribute values*. Attributes of the decision table are divided into two disjoint groups called *condition* and *decision* attributes, respectively. Each row of a decision table induces a *decision rule*, which specifies decision (action, results, outcome, etc.) if some conditions are satisfied. If a decision rule uniquely determines decision in terms of conditions – the decision rule is *certain*. Otherwise the decision rule is *uncertain*. Decision rules are closely connected with approximations. Roughly speaking, certain decision rules describe lower approximation of decisions in terms of conditions, whereas uncertain decision rules refer to the boundary region of decisions.

With every decision rule two conditional probabilities, called the *certainty* and the *coverage* coefficient, are associated. The certainty coefficient expresses the conditional probability that an object belongs to the decision class specified by the decision rule, given it satisfies conditions of the rule. The coverage coefficient gives the conditional probability of reasons for a given decision.

It turns out that the certainty and coverage coefficients satisfy Bayes' theorem. That gives a new look into the interpretation of Bayes' theorem, and offers a new method data to draw conclusions from data.

In the paper rudiments of the theory will be outlined, and basic concepts of the theory will be illustrated by a simple tutorial example of churn modeling. Real life applications require more advanced extensions of the theory but we will not discuss these extensions in this paper.

Rough set theory has an overlap with many other theories dealing with imperfect knowledge, e.g., evidence theory, fuzzy sets, Bayesian inference and others. Nevertheless, the theory can be regarded as an independent, complementary – not competing discipline, in its own rights.

More information about rough sets and their applications can be found in the references and the Web.

2. Illustrative example

Let us start our considerations from a very simple tutorial example concerning churn modeling in telecommunications, which is a simplified version of an example given in [1]. In Table 1, six facts concerning six client segments are presented.

In the table condition attributes describing client profile are: *In* – incoming calls, *Out* – outgoing calls within the same operator, *Change* – outgoing calls to other mobile operator, the decision attribute describing the consequence is *Churn* and *N* is the number of similar cases.

Each row in the table determine a decision rule. E.g., row 2 determines the following decision rule: "*if the number of incoming calls is high and the number of outgoing calls is high and the number of outgoing calls to the mobile operator is low then these is no churn*".

According to [1]: "*One of the main problem that have to be solved by marketing departments of wireless operators is to find the way of convincing current clients that they continue to use the services. In solving this problems can help churn modeling. Churn model in telecommunications industry predicts customers who are going to leave the current operator*".

Table 1
Client segments

Segment	*In*	*Out*	*Change*	*Churn*	*N*
1	medium	medium	low	no	200
2	high	high	low	no	100
3	low	low	low	no	300
4	low	low	high	yes	150
5	medium	medium	low	yes	220
6	medium	low	low	yes	30

In other words we want to explain churn in terms of clients profile, i.e., to describe market segments $\{4, 5, 6\}$ (or $\{1, 2, 3\}$) in terms of condition attributes *In, Out* and *Change*.

The problem cannot be solved uniquely because the data set is *inconsistent*, i.e., segments 1 and 5 have the same profile but different consequences.

Let us observe that:

- segments 2 and 3 (4 and 6) can be classified as sets of clients who *certainly* do not churn (churn),

- segments 1, 2, 3 and 5 (1, 4, 5 and 6) can be classified as sets of clients who *possibly* do not churn (churn),

- segments 1 and 5 are *undecidable* sets of clients.

This leads us to the following notions:

- the set $\{2,3\}$ ($\{4,6\}$) is the *lower approximation* of the set $\{1,2,3\}(\{4,5,6\})$,

- the set $\{1,2,3,5\}$ ($\{1,4,5,6\}$) is *the lower approximation* of the set $\{1,2,3\}$ ($\{4,5,6\}$),

- the set $\{1,5\}$ is the *boundary region* of the set $\{1,2,3\}(\{4,5,6\})$,

which will be discussed in the next paragraph more exactly.

3. Information systems and approximations

In this section we will examine approximations more exactly. First we define a data set, called an information system.

An *information system* is a pair $S = (U,A)$, where U and A, are finite, nonempty sets called the *universe*, and the set of *attributes*, respectively. With every attribute $a \in A$ we associate a set V_a, of its *values*, called the *domain* of a. Any subset B of A determines a binary relation $I(B)$ on U, which will be called an *indiscernibility relation*, and defined as follows: $(x,y) \in I(B)$ if and only if $a(x) = a(y)$ for every $a \in A$, where $a(x)$ denotes the value of attribute a for element x. Obviously $I(B)$ is an equivalence relation. The family of all equivalence classes of $I(B)$, i.e., a partition determined by B, will be denoted by $U/I(B)$, or simply by U/B; an equivalence class of $I(B)$, i.e., block of the partition U/B, containing x will be denoted by $B(x)$. If (x,y) belongs to $I(B)$ we will say that x and y are *B-indiscernible (indiscernible with respect to B)*. Equivalence classes of the relation $I(B)$ (or blocks of the partition U/B) are referred to as *B-elementary sets* or *B-granules*.

Suppose we are given an information system $S = (U,A)$, $X \subseteq U$, and $B \subseteq A$. Let us define two operations assigning to every $X \subseteq U$ two sets $B_*(X)$ and $B^*(X)$, called the *B-lower* and the *B-upper approximation* of X, respectively, and defined as follows:

$$B_*(X) = \bigcup_{x \in U} \{B(x) : B(x) \subseteq X\},$$

$$B^*(X) = \bigcup_{x \in U} \{B(x) : B(x) \cap X \neq \varnothing\}.$$

Hence, the *B-lower approximation* of a set is the union of all *B-granules* that are included in the set, whereas the *B-upper approximation* of a set is the union of all *B-granules* that have a nonempty intersection with the set. The set

$$BN_B(X) = B^*(X) - B_*(X)$$

will be referred to as the *B-boundary region* of X.

If the boundary region of X is the empty set, i.e., $BN_B(X) = \varnothing$, then X is *crisp (exact)* with respect to B; in the opposite case, i.e., if $BN_B(X) \neq \varnothing$, X is referred to as *rough (inexact)* with respect to B.

Thus, the set of elements is rough (inexact) if it cannot be defined in terms of the data, i.e. it has some elements that can be classified neither as member of the set nor its complement in view of the data.

4. Decision tables and decision rules

If we distinguish in an information system two disjoint classes of attributes, called *condition* and *decision attributes*, respectively, then the system will be called a *decision table* and will be denoted by $S = (U,C,D)$, where C and D are disjoint sets of condition and decision attributes, respectively.

Let $S = (U,C,D)$ be a decision table. Every $x \in U$ determines a sequence $c_1(x), \ldots, c_n(x), d_1(x), \ldots, d_m(x)$, where $\{c_1, \ldots, c_n\} = C$ and $\{d_1, \ldots, d_m\} = D$.

The sequence will be called a *decision rule induced by x* (in *S*) and will be denoted by $c_1(x), \ldots, c_n(x) \rightarrow d_1(x), \ldots, d_m(x)$ or in short $C \rightarrow_x D$.

The number $supp_x(C, D) = |A(x)| = |C(x) \cap D(x)|$ will be called a *support* of the decision rule $C \rightarrow_x D$ and the number

$$\sigma_x(C, D) = \frac{supp_x(C, D)}{|U|},$$

will be referred to as the *strength* of the decision rule $C \rightarrow_x D$, where $|X|$ denotes the cardinality of X.

With every decision rule $C \rightarrow_x D$ we associate the *certainty factor* of the decision rule, denoted $cer_x(C, D)$ and defined as follows:

$$cer_x(C, D) = \frac{|C(x) \cap D(x)|}{|C(x)|} = \frac{supp_x(C, D)}{|C(x)|} = \frac{\sigma_x(C, D)}{\pi(C(x))},$$

where $\pi(C(x)) = \frac{|C(x)|}{|U|}$.

The certainty factor may be interpreted as a conditional probability that *y* belongs to *D(x)* given *y* belongs to *C(x)*, symbolically $\pi_x(D|C)$.

If $cer_x(C, D) = 1$, then $C \rightarrow_x D$ will be called a *certain decision* rule; if $0 < cer_x(C, D) < 1$ the decision rule will be referred to as an *uncertain decision rule*.

Besides, we will also use a *coverage factor* of the decision rule, denoted $cov_x(C, D)$ and defined as

$$cov_x(C, D) = \frac{|C(x) \cap D(x)|}{|D(x)|} = \frac{supp_x(C, D)}{|D(x)|} = \frac{\sigma_x(C, D)}{\pi(D(x))},$$

where $\pi(C(x)) = \frac{|D(x)|}{|U|}$.

Similarly

$$cov_x(C, D) = \pi_x(C|D).$$

If $C \rightarrow_x D$ is a decision rule then $D \rightarrow_x C$ will be called an *inverse decision rule*. The inverse decision rules can be used to give *explanations (reasons)* for a decision.

For Table 1 we have the certainty and coverage factors are as shown in Table 2.

Table 2
Parameters of the decision rules

Decision rule	Strength	Certainty	Coverage
1	0.20	0.48	0.33
2	0.10	1.00	0.17
3	0.30	1.00	0.50
4	0.15	1.00	0.38
5	0.22	0.52	0.55
6	0.03	1.00	0.07

Let us observe that if $C \rightarrow_x D$ is a decision rule then

$$\bigcup_{y \in D(x)} \{C(y) : C(y) \subseteq D(x)\}$$

is the lower approximation of the decision class $D(x)$, by condition classes $C(y)$, whereas the set

$$\bigcup_{y \in D(x)} \{C(y) : C(y) \cap D(x) \neq \emptyset\}$$

is the upper approximation of the decision class by condition classes $C(y)$.

Approximations and decision rules are two different methods to express properties of data. Approximations suit better to express topological properties of data, whereas decision rules describe in a simple way hidden patterns in data.

5. Probabilistic properties of decision tables

Decision tables (and decision algorithms) have important probabilistic properties which are discussed next.

Let $C \rightarrow_x D$ be a decision rule and let $\Gamma = C(x)$ and $\Delta = D(x)$. Then the following properties are valid:

$$\sum_{y \in \Gamma} cer_y(C, D) = 1, \qquad (1)$$

$$\sum_{y \in \Delta} cov_y(C, D) = 1, \qquad (2)$$

$$\pi(D(x)) = \sum_{y \in \Gamma} cer_y(C, D) \cdot \pi(C(y)) =$$
$$= \sum_{y \in \Gamma} \sigma_y(C, D), \qquad (3)$$

$$\pi(C(x)) = \sum_{y \in \Delta} cov_y(C, D) \cdot \pi(D(y)) =$$
$$= \sum_{y \in \Delta} \sigma_y(C, D), \qquad (4)$$

$$cer_x(C, D) = \frac{cov_x(C, D) \cdot \pi(D(x))}{\sum\limits_{y \in \Delta} cov_y(C, D) \cdot \pi(D(y))} = \frac{\sigma_x(C, D)}{\pi(C(x))}, \quad (5)$$

$$cov_x(C, D) = \frac{cer_x(C, D) \cdot \pi(C(x))}{\sum\limits_{y \in \Gamma} cer_y(C, D) \cdot \pi(C(y))} = \frac{\sigma_x(C, D)}{\pi(D(x))}. \quad (6)$$

That is, any decision table satisfies Eqs.(1)–(6). Observe that formulae (3) and (4) refer to the well known *total probability theorem*, whereas (5) and (6) refer to *Bayes' theorem*.

Thus in order to compute the certainty and coverage factors of decision rules according to formula (5) and (6) it is enough to know the strength (support) of all decision rules only. The strength of decision rules can be computed from data or can be a subjective assessment.

6. Decision algorithm

Any decision table induces a set of "*if ... then*" decision rules.

Any set of mutually, exclusive and exhaustive decision rules, that covers all facts in *S* and preserves the indiscernibility relation included by *S* will be called a decision algorithm in *S*.

An example of decision algorithm in the decision Table 1 is given below:

		cer.
1)	*if (In, high) then (Churn, no)*	1.00
2)	*if (In, low) and (Change, low) then (Churn, no)*	1.00
3)	*if (In, med.) and (Out, med.) then (Churn, no)*	0.48
4)	*if (Change, high) then (Churn, yes)*	1.00
5)	*if (In, med.) and (Out, low) then (Churn, yes)*	1.00
6)	*if (In, med.) and (Out, med.) then (Churn, yes)*	0.52

Finding a minimal decision algorithm associated with a given decision table is rather complex. Many methods have been proposed to solve this problem, but we will not consider this problem here.

If we are interested in *explanation* of decisions in terms of conditions we need an *inverse* decision algorithm which is obtained by replacing mutually conditions and decisions in every decision rule in the decision algorithm.

For example, the following inverse decision algorithm can be understood as explanation of churn (no churn) in terms of client profile:

		cer.
1')	*if (Churn, no) then (In, high) and (Out, med.)*	0.33
2')	*if (Churn, no) then (In, high)*	0.17
3')	*if (Churn, no) then (In, low) and (Change, low)*	0.50
4')	*if (Churn, yes) then (Change, yes)*	0.38
5')	*if (Churn, yes) then (In, med.) and (Out, med.)*	0.55
6')	*if (Churn, yes) then (In, med.) and (Out, low)*	0.07

Observe that certainty factor for inverse decision rules are coverage factors for the original decision rules.

7. What the data are telling us

The above properties of decision tables (algorithms) give a simple method of drawing *conclusions* from the data and giving *explanation* of obtained results.

From the decision algorithm and the certainty factors we can draw the following conclusions.

- No churn is implied with *certainty* by:
    - high number of incoming calls,
    - low number of incoming calls and low number of outgoing calls to other mobile operator.

- Churn is implied with *certainty* by:
    - high number of outgoing calls to other mobile operator,
    - medium number of incoming calls and low number of outgoing calls.

- Clients with medium number of incoming calls and low number of outgoing calls within the same operator are *undecided* (no churn, cer. = 0.48; churn, cer. = 0.52).

From the inverse decision algorithm and the coverage factors we get the following explanations:

- the *most probable* reason for no churn is low general activity of a client,

- the *most probable* reason for churn is medium number of incoming calls and medium number of outgoing calls within the same operator.

8. Summary

In this paper the basic concepts of rough set theory and its application to drawing conclusions from data are discussed. For the sake of illustration an example of churn modeling in telecommunications is presented.

References

[1] J. Grant, "Churn modeling by rough set approach", manuscript, 2001.

[2] S. K. Pal and A. Skowron, Eds., *Rough Fuzzy Hybridization*. Springer, 1999.

[3] Z. Pawlak, *Rough Sets – Theoretical Aspects of Reasoning about Data*. Boston, London, Dordrecht: Kluwer, 1991.

[4] Z. Pawlak, "Decision rules, Bayes' rule and rough sets", in *New Direction in Rough Sets, Data Mining, and Granular-Soft Computing*, N. Zhong, A. Skowron, and S. Ohsuga, Eds. Springer, 1999, pp. 1–9.

[5] Z. Pawlak, "New look Bayes' theorem – the rough set outlook", in *Proc. Int. RSTGC-2001*, Matsue Shimane, Japan, May 2001, pp. 1–8; *Bull. Int. Rough Set Soc.*, vol. 5, no. 1/2, 2001.

[6] L. Polkowski and A. Skowron, Eds., *Rough Sets and Current Trends in Computing*. Lecture Notes in Artificial Intelligence 1424, Springer, 1998.

[7] L. Polkowski and A. Skowron, Eds., *Rough Sets in Knowledge Discovery*. Vol. 1–2, Springer, 1998.

[8] L. Polkowski, S. Tsumoto, and T. Y. Lin, Eds., *Rough Set Methods and Applications – New Developments in Knowledge Discovery in Information Systems*. Springer, 2000, to appear.

[9] N. Zhong, A. Skowron, and S. Ohsuga, Eds., *New Direction in Rough Sets, Data Mining, and Granular-Soft Computing*. Springer, 1999.

More info about rough sets can be found at:

http://www.roughsets.org
http://www.cs.uregina.ca/~roughset
http://www.infj.ulst.ac.uk /staff/I.Duentsch
http://www-idss.cs.put.poznan.pl/staff/slowinski/
http://alfa/mimuw.edu.pl
http://www.idi.ntnu.no/~aleks/rosetta/
http://www.infj.ulst.ac.uk/~cccz23/grobian/grobian.html

Zdzisław Pawlak
Institute of Theoretical and Applied Informatics
Polish Academy of Sciences
Bałtycka st 5
44-000 Gliwice, Poland