ELSEVIER

Information Sciences 147 (2002) 1–12

INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL

www.elsevier.com/locate/ins

# Rough sets and intelligent data analysis ☆

Zdzisław Pawlak

*University of Information Technology and Management ul., Newelska 6, 01-447 Warsaw, Poland*

Received 27 November 2000; received in revised form 8 May 2001; accepted 20 August 2001

**Abstract**

Rough set based data analysis starts from a data table, called an *information system*. The information system contains data about objects of interest characterized in terms of some attributes. Often we distinguish in the information system condition and decision attributes. Such information system is called a *decision table*. The decision table describes decisions in terms of conditions that must be satisfied in order to carry out the decision specified in the decision table. With every decision table a set of decision rules, called a *decision algorithm* can be associated. It is shown that every decision algorithm reveals some well-known probabilistic properties, in particular it satisfies the total probability theorem and the Bayes' theorem. These properties give a new method of drawing conclusions from data, without referring to prior and posterior probabilities, inherently associated with Bayesian reasoning.


## 1. Introduction

Rough set theory is a new mathematical approach to intelligent data analysis and data mining. After almost 20 years of pursuing rough set theory and its applications the approach reached a certain degree of maturity. In recent years we witness a rapid grow of interest in rough set theory and its applications, worldwide. Many international workshops, conferences and

☆ This is an extended version of the paper presented at the Konan-IIASA Joint Workshop on Natural Environment Management and Applied Systems Analysis held September 6–9, 2000 in Laxemburg, Austria.

0020-0255/02/$ - see front matter 
PII: S0020-0255(02)00197-4

seminars included rough sets in their programs. About 2000 papers and several books have been published until now on various aspects of rough sets.

Rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (data, knowledge). 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 two basic operations in the rough set theory.

The rough set approach seems to be of fundamental importance to AI and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems, inductive reasoning and pattern recognition. Rough set theory has been successfully applied in many real-life problems in medicine, pharmacology, engineering, banking, finances, market analysis, environment management and others.

The rough set approach to data analysis has many important advantages. Some of them are listed below:

- provides efficient algorithms for finding hidden patterns in data;
- finds minimal sets of data (data reduction);
- evaluates significance of data;
- generates sets of decision rules from data;
- offers straightforward interpretation of obtained results;
- most algorithms based on the rough set theory are particularly suited for parallel processing;
- it is easy to understand.

In the paper basic concept of rough set based data analysis will be outlined and will be illustrated by a simple tutorial example, concerning global warming.

Basics of rough sets can be found in [1,4].

More about rough sets can by found in [3,5–9].

## 2. Information systems and decision tables

Starting point of rough set based data analysis is a data set, called an information system.

An information system is a data table, whose columns are labeled by attributes, rows are labeled by objects of interest and entries of the table are attribute values.

Formally, by an *information system* we will understand 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 $a \in B$, 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 $U/(IB)$, 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*.

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.

An example of a decision table is shown in Table 1.

In the table *Solar Energy*, *Volcanic Activity* and *Residual* $CO_2$ are condition attributes, and *Temperature* is a decision attribute.

The example concerns global warming and is taken after some simplifications from [2].

We want to explain what causes the high (low) temperatures, i.e., to describe the set of facts {1, 2, 3, 5} ({4, 6}) in terms of condition attributes: *Solar Energy*, *Volcanic Activity* and *Residual* $CO_2$.

Table 1
An example of a decision table

| Fact | Solar energy | Volcanic activity | Residual $CO_2$ | Temperature | Days count |
|---|---|---|---|---|---|
| 1 | Medium | High | Low | High | 20 |
| 2 | High | High | High | High | 30 |
| 3 | Medium | Low | High | High | 90 |
| 4 | Low | Low | Low | Low | 120 |
| 5 | High | High | Medium | High | 70 |
| 6 | Medium | Low | High | Low | 34 |

The data set is *inconsistent* because facts 3 and 6 are contradictory, therefore the problem cannot be solved exactly but only approximately. Let us observe what the data are telling us:

- facts 1, 2, 5 can be *certainly* classified as causing high temperature;
- fact 4 can be *certainly* classified as causing low temperature;
- facts 3, 6 can be *possibly* classified as causing high or low temperature.

## 3. Approximation of sets

Suppose we are given an information system $S = (U, A)$, $X \subseteq U$, and $B \subseteq A$. Our task is to describe the set $X$ in terms of attribute values from $B$. To this end we 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:

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

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

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., $BN_B(X) = \emptyset$, then $X$ is *crisp*(*exact*) with respect to $B$; in the opposite case, i.e., if $BN_B(X) \neq \emptyset$, $X$ is referred to as *rough* (*inexact*) with respect to $B$.

In our illustrative example we have, with respect to the condition attributes:

- the set $\{1, 2, 5\}$ is the *lower approximation* of the set $\{1, 2, 3, 5\}$;
- the set $\{1, 2, 3, 5, 6\}$ is the *upper approximation* of the set $\{1, 2, 3, 5\}$;
- the set $\{3, 6\}$ is the *boundary region* of the set $\{1, 2, 3, 5\}$.

## 4. Decision rules

Decision rules constitute a formal language to describe approximations in logical terms.

Decision rules are expressions in the form "*if... then...*", in symbols $\Phi \rightarrow \Psi$.

Examples of decision rules are shown below:
(*Solar Energy*, *medium*) $\wedge$ (*Volcanic Activity*, *high*) $\rightarrow$ (*Temperature*, *high*);
(*Solar Energy*, *high*) $\rightarrow$ (*Temperature*, *high*).

Formally the language of decision rules, called a *decision language*, is defined as shown below.

Let $S = (U, A)$ be an information system. With every $B \subseteq A$ we associate a formal language, i.e., a set of formulas $For(B)$. Formulas of $For(B)$ are built up from attribute-value pairs $(a, v)$ where $a \in B$ and $v \in V_a$ by means of logical connectives $\wedge$ (*and*), $\vee$ (*or*), $\sim$ (*not*) in the standard way.

For any $\Phi \in For(B)$ by $\|\Phi\|_S$ we denote the set of all objects $x \in U$ satisfying $\Phi$ in $S$ defined inductively as follows:

$$\|(a, v)\|_S = \{x \in U : a(x) = v\} \quad \text{for all a} \in B \text{ and } v \in V_a,$$

$$\|\Phi \vee \Psi\|_S = \|\Phi\|_S \cup \|\Psi\|_S, \quad \|\Phi \wedge \Psi\|_S = \|\Phi\|_S \cap \|\Psi\|_S,$$
$$\| \sim \Phi\|_S = U - \|\Phi\|_S.$$

A formula $\Phi$ is *true* in $S$ if $\|\Phi\|_S = U$.

A *decision rule* in $S$ is an expression $\Phi \rightarrow \Psi$, read *if* $\Phi$ *then* $\Psi$, where $\Phi \in For(C)$, $\Psi \in For(D)$ and $C$, $D$ are condition and decision attributes, respectively; $\Phi$ and $\Psi$ are referred to as *conditions* and *decisions* of the rule, respectively.

A decision rule $\Phi \rightarrow \Psi$ is *true* in $S$ if $\|\Phi\|_S \subseteq \|\Psi\|_S$.

The number $supp_S(\Phi, \Psi) = card(\|\Phi \wedge \Psi\|_S)$ will be called the *support* of the rule $\Phi \rightarrow \Psi$ in $S$. We consider a probability distribution $p_U(x) = 1/card(U)$ for $x \in U$ where $U$ is the (non-empty) universe of objects of $S$; we have $p_U(X) = card(X)/card(U)$ for $X \subseteq U$. For any formula $\Phi$ we associate its probability in $S$ defined by

$$\pi_S(\Phi) = p_U(\|\Phi\|_S).$$

With every decision rule $\Phi \rightarrow \Psi$ we associate a conditional probability called the *certainty factor*, and denoted $cer_S(\Phi, \Psi)$ defined as

$$cer_S(\Phi, \Psi) = \pi_S(\Psi|\Phi) = p_U\big(\|\Psi\|_S|\|\Phi\|_S\big).$$

We have

$$cer_S(\Phi, \Psi) = \pi_S(\Psi|\Phi) = \frac{card\big(\|\Phi \wedge \Psi\|_S\big)}{card\big(\|\Phi\|_S\big)},$$

where $\|\Phi\|S \neq \emptyset$.

This coefficient is now widely used in data mining and is called *confidence coefficient*.

Obviously, $\pi_S(\Psi|\Phi) = 1$ if and only if $\Phi \to \Psi$ is true in $S$.

If $\pi_S(\Psi|\Phi) = 1$, then $\Phi \to \Psi$ will be called a *certain decision* rule; if $0 < \pi_S(\Psi|\Phi) < 1$ the decision rule will be referred to as a *uncertain decision* rule.

Besides, we will also use a *coverage factor* denoted $cov_S(\Phi, \Psi)$ and defined as

$$cov_S(\Phi, \Psi) = \pi_S(\Phi|\Psi) = p_U\left(\|\Phi\|_S | \|\Psi\|_S\right).$$

Obviously we have

$$cov_S(\Phi, \Psi) = \pi_S(\Phi|\Psi) = \frac{card\left(\|\Phi \wedge \Psi\|_S\right)}{card\left(\|\Psi\|_S\right)}.$$

The certainty factors in $S$ can be interpreted as the frequency of objects having the property $\Psi$ in the set of objects having the property $\Phi$ and the coverage factor – as the frequency of objects having the property $\Phi$ in the set of objects having the property $\Psi$.

For example, for the decision rule

(*Solar Energy*, *medium*) ∧ (*Volcanic Activity*, *low*) → (*Temperature*, *high*)

we have: *support* = 90, *strength* = 0.25, *certainty* = 0.74, *coverage* = 0.43.

The number

$$\sigma_S(\Phi, \Psi) = \frac{\operatorname{sup} p_S(\Phi, \Psi)}{card(U)} = cer_S(\Phi, \Psi) \cdot \pi_S(\Phi)$$

will be called the *strength* of the decision rule $\Phi \to \Psi$ in $S$.

Summing up, decision rules, which are in fact logical implications, constitute a logical counterpart of approximations: certain rules correspond to the lower approximation, whereas the uncertain rules correspond to the boundary region. Thus we have two formal tools to deal with vagueness: approximations and implications. Mathematically approximations are basic operations (interior and closure) in a topology generated by a data set. Thus if we want to prove properties of the data (find patterns in the data) the topological language of approximations is the right tool. However, in order to describe the patterns in the data for practical use the logical language of implications is the proper one.

The certainty and the coverage factors of decision rules are conditional probabilities which express how exact is our knowledge (data) about the considered reality. Let us remain that the factors are not assumed arbitrarily but computed from the data, thus they are in a certain sense objective.

From logical point of view the certainty factor can be interpreted as a degree of truth of the decision rule, i.e., how strongly the decision can be trusted in view of the data. On the contrary, the coverage factor can be viewed as a degree of truth of the "inverted" decision rule, i.e., to what degree the reasons for a decision can be trusted in view of the data.

Statistically, the certainty factor reveals simply the frequency of facts satisfying conditions, among the facts satisfying decision of the decision rule, whereas the interpretation of the coverage factor is converse.

Finally let us briefly comment the concept of the strength of a decision rule. This number simply expresses the ratio of all facts which can be classified by the decision rule to all facts in the data table. It will be shown in the next sections that this coefficient plays essential role in further considerations, and will be used to new formulation of Bayes' theorem.

## 5. Decision algorithms

In this section we define the notion of a decision algorithm, which is a logical counterpart of a decision table.

Informally, a decision algorithm is a set of mutually exclusive and exhaustive decision rules associated with a given decision table.

An example of a decision algorithm associated with Table 1 is given below.

1. (*Solar Energy*, *medium*) $\wedge$ (*Volcanic Activity*, *high*) $\rightarrow$ (*Temperature*, *high*).
2. (*Solar Energy*, *high*) $\rightarrow$ (*Temperature*, *high*).
3. (*Solar Energy*, *medium*) $\wedge$ (*Volcanic Activity*, *low*) $\rightarrow$ (*Temperature*, *high*).
4. (*Solar Energy*, *low*) $\rightarrow$ (*Temperature*, *low*).
5. (*Solar Energy*, *medium*) $\wedge$ (*Volcanic Activity*, *low*) $\rightarrow$ (*Temperature*, *low*).

Formally a decision algorithm is defined as follows.

Let $Dec(S) = \{\Phi_i \rightarrow \Psi_i\}_{i=1}^m$, $m \geqslant 2$, be a set of decision rules in a decision table $S = (U, C, D)$.

(1) If for every $\Phi \rightarrow \Psi$, $\Phi' \rightarrow \Psi' \in Dec(S)$ we have $\Phi = \Phi'$ or $\|\Phi \wedge \Psi'\|_S = \emptyset$, and $\Psi = \Psi'$ or $\|\Psi \wedge \Psi'\|_S = \emptyset$, then we will say that $Dec(S)$ is the set of pairwise *mutually exclusive* (*independent*) decision rules in $S$.

(2) If $\| \bigvee_{i=1}^m \Phi_i\|_S = U$ and $\| \bigvee_{i=1}^m \Psi_i\|_S = U$ we will say that the set of decision rules $Dec(S)$ *covers* $U$.

(3) If $\Phi \rightarrow \Psi \in Dec(S)$ and $supp_S(\Phi, \Psi) \neq 0$ we will say that the decision rule $\Phi \rightarrow \Psi$ is *admissible* in $S$.

(4) If $\bigcup_{X \in U/D} C_*(X) = \| \bigvee_{\Phi \rightarrow \Psi \in Dec^+(S)} \Phi\|_S$ where $Dec^+(S)$ is the set of all certain decision rules from $Dec(S)$, we will say that the set of decision rules $Dec(S)$ preserves the *consistency* of the decision table $S = (U, C, D)$.

The set of decision rules $Dec(S)$ that satisfies (1)–(4), i.e., is independent, covers $U$, preserves the consistency of $S$ and all decision rules $\Phi \rightarrow \Psi \in Dec(S)$ are admissible in $S$ – will be called a *decision algorithm* in $S$.

If $\Phi \rightarrow \Psi$ is a decision rule then the decision rule $\Psi \rightarrow \Phi$ will be called an *inverse* decision rule of $\Phi \rightarrow \Psi$.

Let $Dec^*(S)$ denote the set of all inverse decision rules of $Dec(S)$.

It can be shown that $Dec^*(S)$ satisfies (1)–(4), i.e., it is an decision algorithm in $S$.

If $Dec(S)$ is a decision algorithm then $Dec^*(S)$ will be called an *inverse* decision algorithm of $Dec(S)$.

The inverse decision algorithm for the decision algorithm (1)–(5) is as follows:
(1′) (*Temperature*, *high*) → (*Solar Energy*, *medium*) ∧ (*Volcanic Activity*, *high*).
(2′) (*Temperature*, *high*) → (*Solar Energy*, *high*).
(3′) (*Temperature*, *high*) → (*Solar Energy*, *medium*) ∧ (*Volcanic Activity*, *low*).
(4′) (*Temperature*, *low*) → (*Solar Energy*, *low*).
(5′) (*Temperature*, *low*) → (*Solar Energy*, *medium*) ∧ (*Volcanic Activity*, *low*).

The inverse decision algorithm can be used as an *explanation* of decision in terms of conditions, i.e., it gives *reasons* for decisions.

As mentioned at the beginning of this section decision algorithm is a counterpart of a decision table. Properties (1)–(4) have been chosen in such a way that the decision algorithm preserves basic properties of the data in the decision table, in particular approximations and boundary regions of decisions.

Crucial issue in the rough set based data analysis is the generation of "optimal" decision algorithms from the data. This is a complex task, particularly when large databases are concerned. Many methods and algorithms have been proposed to deal with this problem but we will not dwell upon this issue here, for we intend to restrict this paper to rudiments of rough set theory only. The interested reader is advised to consult the references and the web.

## 6. Decision algorithms and approximations

Decision algorithms can be used as a formal language for describing approximations.

For example, *certain* decision rules
1. (*Solar Energy*, *medium*) ∧ (*Volcanic Activity*, *high*) → (*Temperature*, *high*).
2. (*Solar Energy*, *high*) → (*Temperature*, *high*) describe the lower approximation of the decision (*Temperature*, *high*) and uncertain decision rule.
3. (*Solar Energy*, *medium*) ∧ (*Volcanic Activity*, *low*) → (*Temperature*, *high*) describes the boundary region of the decision (*Temperature*, *high*).

The above relationships can be defined more precisely as follows:

Let $Dec(S)$ be a decision algorithm in $S$ and let $\Phi \to \Psi \in Dec(S)$. By $C(\Psi)$ we denote the set of all conditions of $\Psi$ in $Dec(S)$ and by $D(\Phi)$ – the set of all decisions of $\Phi$ in $Dec(S)$.

Then we have the following relationships:
(a) $C_*(\|\Psi\|_S) = \| \bigvee_{\Phi' \in C(\Psi), \pi(\Psi|\Phi')=1} \Phi' \|_S$,
(b) $C^*(\|\Psi\|_S) = \| \bigvee_{\Phi' \in C(\Psi), 0<\pi(\Psi|\Phi') \leqslant 1} \Phi' \|_S$,
(c) $BN_C(\|\Psi\|_S) = \| \bigvee_{\Phi' \in C(\Psi), 0<\pi(\Psi|\Phi')<1} \Phi' \|_S$.

From the above properties we can get the following definitions:
(i) If $\|\Phi\|_S = C_*(\|\Psi\|_S)$, then formula $\Phi$ will be called the *C-lower approximation* of the formula $\Psi$ and will be denoted by $C_*(\Psi)$.
(ii) If $\|\Phi\|_S = C^*(\|\Psi\|_S)$, then the formula $\Phi$ will be called the *C-upper approximation* of the formula $\Phi$ and will be denoted by $C^*(\Psi)$.

(iii) If $\|\Phi\|_S = BN_C = (\|\Psi\|_S)$, then $\Phi$ will be called the *C-boundary* of the formula $\Psi$ and will be denoted by $BN_C(\Psi)$.

The above properties say that any decision $\Psi \in Dec(S)$ can be uniquely described by the following certain and uncertain decision rules respectively:

$$C_*(\Psi) \to \Psi,$$

$$BN_C(\Psi) \to \Psi.$$

Thus decision algorithms can be viewed as a logical counterpart of approximations, or more exactly as a formal language to describe approximations. The language of decision rules is more convenient to describe decisions in terms of conditions than the topological language of approximations. However, approximations give better insight into vagueness and uncertainty of data.

## 7. Some properties of decision algorithms

Decision algorithms have interesting probabilistic properties which are discussed next.

Let $Dec(S)$ be a decision algorithm and let $\Phi \to \Psi \in Dec(S)$. Then the following properties are valid:

$$\sum_{\Phi' \in C(\Psi)} cer_S\left(\Phi', \Psi\right) = 1, \tag{1}$$

$$\sum_{\Psi' \in D(\Phi)} cov_S\left(\Phi, \Psi'\right) = 1, \tag{2}$$

$$\pi_S(\Psi) = \sum_{\Phi' \in C(\Psi)} cer_S\left(\Phi', \Psi\right) \cdot \pi_S\left(\Phi'\right) = \sum_{\Phi' \in C(\Psi)} \sigma_S\left(\Phi', \Psi\right), \tag{3}$$

$$cer_S(\Phi, \Psi) = \frac{cov_S(\Phi, \Psi) \cdot \pi_S(\Psi)}{\sum_{\Psi' \in D(\Phi)} cov_S\left(\Phi, \Psi'\right) \cdot \pi_S\left(\Psi'\right)} = \frac{\sigma_S(\Psi, \Phi)}{\pi_S(\Phi)}. \tag{4}$$

That is, any decision algorithm, and consequently any decision table, satisfies (1)–(4). Observe that (3) is the well-known *total probability theorem* and (4) is the *Bayes' theorem.*

Note that we are not referring to prior and posterior probabilities – fundamental in Bayesian data analysis philosophy.

In other words the Bayes' theorem in our case reveals some relationships between decisions and their reasons, or more exactly it discovers some relationships in every set of data.

Thus in order to compute the certainty and coverage factors of decision rules it is enough to know the strength (support) of all decision rules in the decision algorithm only.

Table 2
Certainty and coverage factors

| Decision rule | Support | Strength | Certainty | Coverage |
|---|---|---|---|---|
| 1 | 20 | 0.06 | 1.00 | 0.10 |
| 2 | 100 | 0.27 | 1.00 | 0.47 |
| 3 | 90 | 0.25 | 0.74 | 0.43 |
| 4 | 120 | 0.33 | 1.00 | 0.79 |
| 5 | 34 | 0.09 | 0.26 | 0.21 |

The certainty and coverage factors for the decision algorithm (1)–(5) are given in Table 2.

The strength of decision rules can be computed from the data or can be a subjective assessment.

From the certainty factors of the decision algorithm we can conclude the following:

1. If the solar energy is medium and the volcanic activity is high then the temperature is *certainly* high.
2. If the solar energy is high then the temperature is *certainly* high.
3. If the solar energy is medium and the volcanic activity is low then the *probability* that the temperature is high equals to 0.74.
4. If the solar energy is low then the temperature is *certainly* low.
5. If the solar energy is medium and the volcanic activity is low then the *probability* that the temperature is low equals to 0.26.

The coverage factors of the decision algorithm leads us to the following explanation of global warming:

1′. If the temperature is high then the *probability* that the solar energy is medium and the volcanic activity is high amounts to 0.10.

2′. If the temperature is high then the *probability* that the solar energy is high to equals 0.47.

3′. If the temperature is high then the *probability* that the solar energy is medium and the volcanic activity is low equals to 0.43.

4′. If the temperature is low then the *probability* that the solar energy is low equals to 0.79.

5′. If the temperature is low then the *probability* that the solar energy is medium and the volcanic activity is low equals to 0.21.

Summing up, from the data we can conclude that:

- medium solar energy and high volcanic activity or high solar energy *certainly* cause high temperature;
- low solar energy *certainly* causes low temperature;
- medium solar energy and low volcanic activity cause:
  - high temperature with (probability = 0.74);
  - low temperature with (probability = 0.26).

Whereas the data lead to the following explanation of global warming.

The reasons for high temperature are:

- medium solar energy and high volcanic activity (probability = 0.10);
- high solar energy (probability = 0.47);
- medium solar energy and low volcanic activity (probability = 0.43).

The reasons for low temperature are:

- low solar energy (probability = 0.79);
- medium solar energy and low volcanic activity (probability = 0.21).

In short, we can derive from the data the following conclusions:

- medium solar energy and high volcanic activity or high solar energy *certainly* cause high temperature;
- low solar energy *certainly* causes low temperature;
- medium solar energy and low volcanic activity *most probably* cause high temperature

and the following explanations:

- the *most probable* reason for high temperature is high solar energy;
- the *most probable* reason for low temperature is low solar energy.

Summing up, from the rough set view Bayes' theorem reveals probabilistic structure of a data set (i.e., any decision table or decision algorithm) without referring to either prior or posterior probabilities, inherently associated with Bayesian statistical inference methodology. In other words, it identifies probabilistic relationships between conditions and decisions in decision algorithms, in contrast to classical Bayesian reasoning, where data are employed to verify prior probabilities. This is not the case in rough set based data analysis.

Let us also stress that Bayes' theorem in the rough set approach has a new mathematical form based on strength of decision rules, which simplifies essentially computations and gives a new look on the theorem.

## 8. Conclusions

Approximations, basic concepts of rough set theory have been defined and discussed. Some probabilistic properties of approximation have been revealed, in particular the relationship with the total probability theorem and the Bayes' theorem. These relationships give a new efficient method of drawing conclusions from data, without referring to prior and posterior probabilities intrinsically associated with Bayesian reasoning. The application of the proposed method, by means of a simple tutorial example, concerning global warming has been outlined.

## Acknowledgements

Thanks are due to Professor Hirotaka Nakayama and the anonymous referee for critical remarks.

## References

[1] I. Düntsch, G. Gediga, Rough Set Data Analysis – A Road to Non-invasive Knowledge Discovery, Metoδos Publisher, Bangor, Bissendorf, 2000.
[2] J.D. Gunn, J.W. Grzymała-Busse, Global temperature stability by rule induction: an interdisciplinary bridge, Human Ecology 22 (1994) 59–81.
[3] S.K. Pal, A. Skowron (Eds.), Rough Fuzzy Hybridization, Springer, Berlin, 1999.
[4] Z. Pawlak, Rough Sets – Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Boston, London, Dordrecht, 1991.
[5] Z. Pawlak, Decision rules, Bayes' rule and rough sets, in: N. Zhong, A. Skowron, S. Ohsuga (Eds.), New Direction in Rough Sets, Data Mining, and Granular-Soft Computing, Springer, Berlin, 1999, pp. 1–9.
[6] L. Polkowski, A. Skowron (Eds.), Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, 1424, Springer, Berlin, 1998.
[7] L. Polkowski, A. Skowron (Eds.), Rough Sets in Knowledge Discovery, vols. 1–2, Physica Verlag, A Springer Company, Berlin, 1998.
[8] L. Polkowski, S. Tsumoto, T.Y. Lin (Eds.), Rough Set Methods and Applications – New Developments in Knowledge Discovery in Information Systems, Springer, Berlin, 2000, to appear.
[9] N. Zhong, A. Skowron, S. Ohsuga (Eds.), New Direction in Rough Sets Data Mining and Granular-Soft Computing, Springer, Berlin, 1999.