Reasoning about Data – A Rough Set Perspective

Zdzisław Pawlak

Institute of Theoretical and Applied Informatics
Polish Academy of Sciences
ul. Baltycka 5, 44 000 Gliwice, Poland
e-mail:zpw@ii.pw.edu.pl

[**Abstract.**] **The paper contains some considerations concerning the relationship between decision rules and inference rules from the rough set theory perspective. It is shown that decision rules can be interpreted as a generalization of the** *modus ponens* **inference rule, however there is an essential difference between these two concepts. Decision rules in the rough set approach are used to describe dependencies in data, whereas** *modus ponens* **is used in general to derive conclusions from premises.**

1   Introduction

Data analysis, recently known also as data mining, is, no doubt, a very important and rapidly growing area of research and applications. Historically, data mining methods were first based on statistics, but it is worth mentioning that their origin can be traced back to some ideas of Bertrand Russell and Karl Popper concerning reasoning about data. Recently machine intelligence and machine learning contributed essentially to this domain. Particularly fuzzy sets, rough sets, genetic algorithms, neural networks, cluster analysis and other branches of AI can be considered as a basic tools for knowledge discovery in databases, nowadays.

Main objective of data analysis is finding hidden patterns in data. More specifically, data analysis is about searching for dependencies, or in other words, pursuing "cause-effect" relations, in data.

From logical point of view, data analysis can be perceived as a part of inductive reasoning, and therefore it can be understood as a kind of reasoning about data methods, with specific inference tools.

Reasoning methods are usually classified into three classes: deductive, inductive and common sense reasoning.

Deductive methods are based on axioms and deduction rules, inductive reasoning hinges on data and induction rules, whereas common sense reasoning is based on common knowledge and common sense evident inferences from the knowledge.

Deductive methods are used exclusively in mathematics, inductive methods – in natural sciences, e.g., physics, chemistry etc., while common sense reasoning

is used in human sciences, e.g., politics, medicine, economy, etc. but mainly this method is used almost everywhere in every day life debates, discussions and polemics.

This paper shows that the rough set approach to data analysis bridges somehow the deductive and inductive approach in reasoning about data. The rough set reasoning is also, to some extent, related to common sense reasoning.

Rough set theory gave rise to extensive research in deductive logic, and various logical systems, called rough logics, have been proposed and investigated (see e.g., [3, 6, 7, 10, 11, 18, 20, 23]). However, the basic idea of rough set based reasoning about data is rather of inductive than deductive character. Particularly interesting in this context is the relationship between an implication in deductive logic and a decision rule in the rough set approach.

In deductive logic basic rule of inference, *modus ponens* (*MP*) is based on implication, which can be seen as counterpart of a decision rule in decision rule based methods of data analysis. Although formally decision rules used in the rough set approach are similar to *MP* rule of inference, they play different role to that of *MP* inference rule in logical reasoning. Deduction rules are used to derive true consequences from true premises (axioms), whereas decision rules are description of total or partial dependencies in databases. Besides, in inductive reasoning optimization of decision rules is of essential importance, but in deductive logic we don't need to care about optimization of implications used in reasoning. Hence, implications and decision rules, although formally similar, are totally different concepts and play various roles in both kinds of reasoning methods. Moreover decision rules can be also understood as exact or approximate description of decisions in terms of conditions.

It is also interesting to note a relationship between rough set based reasoning and common sense reasoning methods. Common sense reasoning usually starts from common knowledge shared by domain experts. In the rough set based reasoning the common knowledge is not assumed but derived from data about the domain of interest. Thus the rough set approach can be also seen as a new approach to (common) knowledge acquisition. Also the common rules of inference can be understood in our approach as data explanation methods. Note, that qualitative reasoning, part of common sense reasoning, can be also explained in the rough set philosophy.

Summing up, rough set based reasoning has an overlap with deductive, inductive and common sense reasoning, however it has its own specific features and can be considered in its own right.

2 Data, Information Systems and Decision Tables

Starting point of rough set theory is a set of data (information) about some objects of interest. Data are usually organized in a form of a table called *information system* or *information table*.

A very simple, fictitious example of an information table is shown in Table 1. The table describes six cars in terms of their (attributes) features such as *fuel consumption* (F), *perceived quality* (Q), *selling price* (P) and *marketability* (M).

Table 1. An example of information system

Car	F	Q	P	M
1	high	fair	med.	poor
2	v. high	good	med.	poor
3	high	good	low	poor
4	med.	fair	med.	good
5	v. high	fair	low	poor
6	high	good	low	good

Our main problem can be characterized as determining the nature of the relationship between selected features of the cars and their marketability. In particular, we would like to identify the main factors affecting the market acceptance of the cars.

Information systems with distinguished decision and condition attributes are called *decision tables*.

Each row of a decision table determines a *decision rule*, which specifies *decisions* (*actions*) that should be taken when conditions pointed out by *condition* attributes are satisfied. For example in Table 1 the condition (F,*high*), (Q,*fair*), (P,*med*) determines uniquely the decision (M,*poor*). Decision rules 3) and 6) in Table 1 have the same conditions but different decisions. Such rules are called *inconsistent* (*nondeterministic, conflicting, possible*); otherwise the rules are referred to as *consistent* (*certain, deterministic, nonconflicting, sure*). Decision tables containing inconsistent decision rules are called *inconsistent* (*nondeterministic, etc*); otherwise the table is *consistent* (*deterministic, etc*).

The number of consistent rules to all rules in a decision table can be used as *consistency factor* of the decision table, and will be denoted by $\gamma(C, D)$, where C and D are condition and decision attributes respectively. Thus if $\gamma(C, D) < 1$ the decision table is consistent and if $\gamma(C, D) \neq 1$ the decision table is inconsistent. For example for Table 1 $\gamma(C, D) = 4/6$.

In what follows information systems will be denoted by $S = (U, A)$, where U – is *universe*, A is a set of *attributes*, such that for every $x \in U$ and $a \in A$, $a(x) \in V_a$, and V_a is the domain (set of values of a) of a.

3 Decision Rules and Certainty Factor

Decision rules are often presented as implications and are called "*if... then...*" rules. For example, Table 1 determines the following set of implications:

1) *if (F,high) and (Q,fair) and (P,med) then (M,poor)*,

2) *if (F,v.high) and (Q,good) and (P,med) then (M,poor)*,

3) *if (F,high) and (Q,good) and (P,low) then (M,poor)*,

4) *if (F,med.) and (Q,fair) and (P,med.) then (M,good)*,

5) *if (F,v.high) and (Q,fair) and (P,low.) then (M,poor)*,

6) *if (F,high) and (Q,good) and (P,low) then (M,good)*,

In general decision rules are implications built up from elementary formulas (attribute name, attribute value) and combined together by means of propositional connectives "and", "or" and "implication" in a usual way.

Let Φ and Ψ be logical formulas representing conditions and decisions, respectively and let $\Phi \to \Psi$ be a decision rule, where Φ_S denote the meaning of Φ in the system S, i.e., the set of all objects satisfying Φ in S, defined in a usual way.

With every decision rule $\Phi \to \Psi$ we associate a number, called a *certainty factor* of the rule, and defined as

$$\mu_S(\Phi, \Psi) = \frac{|\Phi_S \cap \Psi_S|}{|\Phi_S|},$$

where $|\Phi|$ denotes the cardinality of Φ. Of course $0 \le \mu_S(\Phi, \Psi) \le 1$; if the rule $\Phi \to \Psi$ is consistent then $\mu_S(\Phi, \Psi) = 1$, and for inconsistent rules $\mu_S(\Phi, \Psi) < 1$. For example, the certainty factor for decision rule 2) is 1, and for decision rule 3) is 0.5.

The certainty factor can be interpreted as a conditional probability of a decision Ψ given the probability of the condition Φ.

It is worth mentioning that association of conditional probability with implication first was proposed by J. Lukasiewicz in the context of multivalued logic and probabilistic logic [4]. This idea has been pursued by other logicians years after [1]. In the rule based knowledge systems many authors also proposed using conditional probability to characterize certainty of the decision rule [2]. In particular in the rough set approach association of condition probabilities with decision rules have been pursued e.g., in [21, 24, 27].

Now the difference between use of implications in classical logic and in data analysis can be clearly seen, particularly in the rough set framework. Implication in deductive logic is used to draw conclusions from premises, by means of *modus ponens* rule of inference. In reasoning about data implications are decision rules used to describe patterns in data. Hence, the role of implications in both cases is completely different. Besides, *modus ponens* is an universal rule of inference valid in any logical system, but decision rules are strictly associated with a specific data and are not valid universally.

However in the rough set approach decision rules can be also exploited in a similar way as *modus ponens* in logic. Let us consider the following formula:

$$\pi_S(\Psi) = \Sigma(\pi_S(\Phi) \cdot \mu_S(\Phi, \Psi)) = \Sigma\pi_S(\Phi \wedge \Psi) \qquad (*)$$

where Σ is taken over all conditions Φ associated with the decision corresponding to Ψ, and $\pi_S(\Phi) = \frac{|\Phi_S|}{|U|}$.

$\pi_S(\Phi)$ is a probability that the condition Φ is satisfied in S. Thus formula (*) shows the relationship between the probability of conditions, certainty factor of a decision rule and the probability of decisions.

Hence the formula (*) allows to compute probability that the decision Ψ is satisfied in S, in terms of the probability of condition Φ and conditional probability of the decision rule $\Phi \Rightarrow \Psi$.

This is a kind of analogous structure to *modus ponens* inference rule and can be treated as its generalization, called *rough modus ponens* (*RMP*) [15]. The certainty factor of a decision rule can be seen as generalization of the rough membership function. It can be also understood as a rough inclusion factor in rough mereology [16, 17] or as a degree of truth of the implication associated with the inclusion.

4   Approximations of Sets

The main problem discussed in the previous section can be also formulated as follows: can we uniquely describe well (poorly) selling cars in terms of their features. Of course, as before, this question cannot be answered uniquely, since cars 3 and 6 have the same features but car 3 sells poorly whereas car 6 sells well, hence we are unable to give unique description of cars selling well or poorly.

But one can observe that in view of the available information we can state that cars 1, 2 and 5 *surely* belong to the set of cars which are selling poorly, whereas cars 1, 2, 3, 5 and 6 *possibly* belong, to the set of cars selling poorly, i.e. cannot be excluded as cars selling poorly. Similarly car 4 surely belongs to well selling cars, whereas cars 3, 4 and 6 possible belong to well selling cars. Hence, because we are unable to give an unique characteristic of cars selling well (poorly), instead we propose to use of two sets, called the *lower* and the *upper approximation* of the set of well (poorly) selling cars.

Now, let us formulate the problem more precisely.

Any subset B of A determines a binary relation I_B on U, which will be called an *indiscernibility relation*, and is defined as follows: xI_By 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., the partition determined by B, will be denoted by U/I_B, or simply U/B; an equivalence class of I_B, i.e., the 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*. Equivalence classes of the relation I_B (or blocks of the partition U/B) are referred to

as *B-elementary concepts* or *B-granules*. As mentioned previously in the rough
set approach the elementary concepts are the basic building blocks (concepts)
of our knowledge about reality.

The indiscernibility relation will be used next to define basic concepts of
rough set theory. Let us define now the following two operations on sets

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

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

assigning to every subset X of the universe U two sets $B_*(X)$ and $B^*(X)$ called
the B-lower and the *B-upper approximation* of X, respectively. 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 the set X is *crisp* (exact) with respect
to B; in the opposite case, i.e., if $BN_B(X) \neq \emptyset$, the set X is referred to as *rough*
(*inexact*) with respect to B.

For example, the lower approximation of the set $\{1,2,3,5\}$, of poorly selling
cars, is the set $\{1,2,5\}$, whereas the upper approximation of poorly selling cars is
the set $\{1,2,3,5,6\}$. The boundary region is the set $\{3,6\}$. That means that cars
1, 2 and 5 can be surely classified, in terms of their features, as poorly selling
cars, while cars 3 and 6 cannot be characterized, by means of available data, as
selling poorly or not. Rough sets can be also defined using a *rough membership
function*, defined as

$$\mu_X^B(x) = \frac{|X \cap B(x)|}{|B(x)|}, \text{ and } \mu_X^B(x) \in [0,1].$$

Value of the membership function $\mu_X(x)$ is kind of conditional probability, and
can be interpreted as a degree of *certainty* to which x belongs to X (or $1 - \mu_X(x)$,
as a degree of *uncertainty*).

For example, car 1 belongs to the set $\{1,2,3,5\}$ of cars selling poorly with
the conditional probability 1, whereas car 3 belongs to the set with conditional
probability 0.5.

5   Dependency of Attributes

Our main problem can be rephrased as whether there is a functional dependency
between the attribute M and attributes F, Q and P. In other words we are asking
whether the value of the decision attribute is determined uniquely by the values
of the condition attributes. It is easily seen that this is not the case for the
example since cars 3 and 6 have the same values of condition attributes but
different value of decision attribute. The consistency factor $\gamma(C,D)$ can be also
interpreted as a *degree of dependency* between C and D. We will say that D
depends on C in a *degree* k $(0 \leq k \leq 1)$, denoted $C \Rightarrow_k D$, if $k = \gamma(C,D)$.

If $k = 1$ we say that D *depends totally* on C, and if $k < 1$, we say that D *depends partially* (in a *degree k*) on C.

For example, for dependency $\{F, P, Q\} \Rightarrow \{M\}$ we get $k = 4/6 = 2/3$.

Dependency of attributes can be also defined using approximations as shown below.

We will say that D depends on C in a degree $k(0 \leq k \leq 1)$, denoted $C \Rightarrow_k D$, if

$$k = \gamma(C, D) = \frac{|POS_C(D)|}{|U|}, \text{ where } POS_C(D) = \bigcup_{X \in U/D} C_*(X),$$

called a *positive region* of the partition U/D with respect to C, is the set of all elements of U that can be uniquely classified to blocks of the partition U/D, by means of C. Obviously

$$\gamma(C, D) = \sum_{X \in U/D} \frac{|C_*(X)|}{|U|}.$$

If $k = 1$ we say that D *depends totally* on C, and if $k < 1$, we say that D *depends partially* (in a *degree k*) on C.

The coefficient k expresses the ratio of all elements of the universe, which can be properly classified to block of the partition U/D, employing attributes C.

6   Reduction of Attributes

We often face a question whether we can remove some data from a data table preserving its basic properties, that is – whether a table contains some superfluous data. This can be formulated as follows.

Let $C, D \subseteq A$, be sets of condition and decision attributes, respectively. We will say that $C' \subseteq C$ is a *D-reduct* (reduct with *respect* to D) of C, if C' is a minimal subset of C such that

$$\gamma(C, D) = \gamma(C', D).$$

Thus reduct enables us to make decisions employing minimal number of conditions.

For example, for Table 1 we have two reducts F, Q and F, P. It means that instead of Table 1 we can use either Table 2 or Table 3, shown below.

These simplifications yield to the following sets of decision rules. For Table 2 we get

1) *if (F,high) and (Q,fair) then (M,poor)*,

2) *if (F,v.high) and (Q,good) then (M,poor)*,

Table 2. Reduced information system   **Table 3.** Another reduced information system

Car	F	Q	M
1	high	fair	poor
2	v. high	good	poor
3	high	good	poor
4	med.	fair	good
5	v. high	fair	poor
6	high	good	good

Car	F	P	M
1	high	med.	poor
2	v. high	med.	poor
3	high	low	poor
4	med.	med.	good
5	v. high	low	poor
6	high	low	good

3) if (F,high) and (Q,good) then (M,poor),

4) if (F,med.) and (Q,fair) then (M,good),

5) if (F,v.high) and (Q,fair) then (M,poor),

6) if (F,high) and (Q,good) then (M,good),

and for Table 3 we have

7) if (F,high) and (P,med) then (M,poor),

8) if (F,v.high) and (P,med.) then (M,poor),

9) if (F,high) and (P,low) then (M,poor),

10) if (F,med.) and (P,med.) then (M,good),

11) if (F,v.high) and (P,low.) then (M,poor),

12) if (F,high) and
(P,low) then (M,good).

Hence, employing the notion of the reduct we can simplify the set of decision rules.

7 Conclusions

Using rough sets to reason about data hinges on three basic concepts of rough set theory: approximations, decision rules and dependencies. All these three notions are strictly connected and are used to express our imprecise knowledge about reality, represented by data obtained from measurements, observations or from knowledgeable expert.

The rough set approach to reasoning about data bridges to some extent the deductive and inductive way of reasoning. Decision rules in this approach can be understood as implications, whose degree of truth is expressed by the certainty factor. Consequently, this leads to generalization of the *modus ponens* inference rule, which in the rough set framework has a probabilistic flavor. It is interesting that the certainty factor of a decision rule is closely related to the rough membership function and to rough inclusion of sets, basic concept of rough mereology.

8   Acknowledgments

References

1. Adams, E. W.: The logic of Conditionals, An Application of Probability to Deductive Logic. D. Reidel Publishing Company, Dordrecht, Boston (1975)
2. Bandler, W. Kohout, L.: Fuzzy power sets and fuzzy implication operators. Fuzzy Sets and Systems **4** (1980) 183–190
3. Banerjee, M., Chakraborty, M.K.: Rough logics: A survay with further directions. In: E. Orłowska (ed.): Incomplete information: Rough set analysis. Physica–Verlag, Heidelberg (1997) 579–600
4. Borkowski, L. (ed.): Jan Łukasiewicz – Selected Works. North Holland Publishing Company, Amsterdam, London, Polish Scientific Publishers, Warszawa (1970)
5. Dempster, A. P.: Upper and lower probabilities induced by induced by the multiplevalued mapping. Ann. Math. Statistics **38** (1967) 325–339
6. Demri, S., Orłowska, E.: Logical analysis of indiscernibility. Institute of Computer Science, Warsaw University of Technology, ICS Research Report **11/96** (1996); see also: E. Orłowska (ed.): Incomplete information: Rough set analysis. Physica–Verlag, Heidelberg (1997) 347–380
7. Gabbay, D., Guenthner, F.: Handbook of Philosophical Logic Vol.1, Elements of Classical Logic, Kluwer Academic Publishers, Dordrecht, Boston, London (1983)
8. Łukasiewicz, J.: Die logischen Grundlagen der Wahrscheinlichkeitsrechnung. Krakow (1913)
9. Magrez, P., Smets, P.: Fuzzy modus ponens: A new model suitable for applications in knowledge-based systems. Information Journal of Intelligent Systems **4** (1975) 181–200
10. Orłowska, E.: Modal logics in the theory of information systems. Zeitschrift für Mathematische Logik und Grundlagen der Mathematik **30** (1984) 213–222
11. Pagliani, P.: Rough set theory and logic–algebraic structures. In: E. Orłowska (ed.): Incomplete information: Rough set analysis. Physica–Verlag, Heidelberg (1997) 109–190
12. Pawlak, Z.: Rough probability. Bull. Polish Acad., Sci. Tech. **33(9-10)** (1985) 499–504
13. Pawlak, Z.: Rough set theory and its application to data analysis. Systems and Cybernetics (to appear)

14. Pawlak, Z.: Granularity of knowledge, indiscernibility and rough sets. IEEE Conference on Evolutionary Computation (1998) 100–103
15. Pawlak, Z.: Rough Modus Ponens. IPMU'98 Conference, Paris (1998)
16. Polkowski, L., Skowron, A.: Rough Mereology. Proc. of the Symphosium on Methodologies for Intelligent Systems **869** (1994) 85–94, Charlotte, N.C., Lecture Notes in Artificial Intelligence, Springer Verlag
17. Polkowski, L., Skowron, A.: Rough Mereology: A New Paradigm for Approximate Reasoning. Journ. of Approximate Reasoning **15(4)** (1996) 333–365
18. Rasiowa, H., Marek, W.: Approximating sets with equivalence relations. Theoret. Comput. Sci. **48** (1986) 145–152
19. Rasiowa, H., Skowron, A.: Rough concepts logic. In: A. Skowron (ed.), Computation Theory, Lecture Notes in Computer Science **208** (1985) 288–297
20. Rauszer, C.: A logic for indiscernibility relations. In: Proceedings of the Conference on Information Sciences and Systems, Princeton University (1986) 834–837
21. Skowron, A.: Management of uncertainty in AI: A rough set approach. In: V. Alagar, S. Belgrer and F.Q. Dong (eds.) Proc. SOFTEKS Workshop on Incompleteness and Uncertainty in Information Systems, Springer Verlag and British Computer Society (1994) 69–86
22. Trillas, E., Valverde, L.: On implication and indistinguishability in the setting of fuzzy logic. Management Decision Support Systems Using Fuzzy Sets and Possibility Theory, Verlag TÜ (1985) 198–212
23. Vakarelov, D.: A modal logic for similarity relations in Pawlak knowledge representation systems. Fundamental Informaticae **15** (1991) 61–79
24. Wong, S.K.M., Ziarko, W.: On learning and evaluation of decision rules in the context of rough sets. Proceedings of the International Symposium on Methodologies for Intelligent Systems (1986) 308–224
25. Zadeh, L.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems **1** (1977) 3–28
26. Zadeh, L.: The role of fuzzy logic in in the management of uncertainty in expert systems. Fuzzy Sets and Systems **11** (1983) 199-277
27. Ziarko, W., Shan, N.: KDD–R: A comprehensive system for knowledge discovery using rough sets. Proceedings of the International Workshop on Rough Sets and Soft Computing (RSSC'94) 164–173, San Jose (1994); see also: T. Y. Lin and A. M. Wildberger (eds.), Soft Computing, Simulation Councils, Inc. (1995) 298–301