

PROCEEDINGS / ACTES

**SEVENTH INTERNATIONAL CONFERENCE
SEPTIEME CONFERENCE INTERNATIONALE**

IPMU

**INFORMATION PROCESSING AND MANAGEMENT
OF UNCERTAINTY IN KNOWLEDGE-BASED
SYSTEMS**

**TRAITEMENT D'INFORMATION ET GESTION
D'INCERTITUDES DANS LES SYSTEMES A BASE
DE CONNAISSANCES**

**July 6-10, 1998
6-10 Juillet 1998**

Volume II

**La Sorbonne
PARIS**

Rough Modus Ponens

Zdzisław Pawlak

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

Abstract

The paper proposes a new approach to certainty factor of decision rules in knowledge based systems. The approach is based on rough set theory, and can be viewed as a generalization of Lukasiewicz's ideas connected with multivalued logic and probability. In particular *Rough Modus Ponens* inference rule is defined and briefly discussed. Connection of the proposed concepts with rough mereology is pointed out.

1 Introduction

Classical deductive reasoning is based on *Modus Ponens* inference rule, which states that if a formula Φ is true and the implication $\Phi \rightarrow \Psi$ is true then the formula Ψ is also true. Lukasiewicz first proposed to extend *Modus Ponens* to the case when instead of true values probabilities are associated with Φ , $\Phi \rightarrow \Psi$ and Ψ [3], [5]. Later, independently, various probabilistic logics have been proposed and investigated by many logicians and philosophers [1], [6].

Recently the generalization of *Modus Ponens* become a very important issue in connection with knowledge based systems. Particularly interesting in this context is the *Generalized Modus Ponens*, introduced by Zadeh in the setting of fuzzy sets [16], [17], which next has been investigated by various authors [2], [7], [14].

Skowron has proposed generalization of *Modus Ponens* in the framework of rough set theory [13]. In this paper we also propose a generalization of *Modus Ponens* within rough set theory, called a *Rough Modus Ponens (RMP)*, however different to that given in [13], and referring to Lukasiewicz's ideas. The essence of our approach consists in association with the implication $\Phi \rightarrow \Psi$ a conditional probability, whereas with Φ and Ψ unconditional probabilities are associated. The as-

sumption that the probability of implication $\Phi \rightarrow \Psi$ is a conditional probability is due to Ramsey [1] but similar ideas can be also found in Lukasiewicz, however, not expressed *explicitly* [3]. Association of conditional probability with decision rules in the context of rough sets has been proposed also by other authors (cf. [15], [18]) but our aim is entirely different. We try to set this issue rather in the frame work of logical research, establish sound logical foundations for this kind of research and show that decision rules used in the rough set approach play different role as *MP* inference rule in logical reasoning, and thus they cannot be in fact treated as a simple generalization of *MP*. Although association of conditional probabilities to implications is quite obvious it leads to logical and philosophical difficulties. Extensive discussion of this problem can be found in [1].

Implication is strongly related to inclusion, i.e., if $\Phi \rightarrow \Psi$ is true then every x satisfying Φ satisfies also Ψ , or in other words $|\Phi| \subseteq |\Psi|$, where $|\Phi|$ denotes the set of all x satisfying Φ i.e., the meaning of Φ . To define *RMP* we will need partial (rough) inclusion of sets and to this aim we will adopt the idea of rough mereology proposed by Polkowski and Skowron [11], [12]. Thus the proposed *RMP* has also connection with rough mereology, which can be understood as a natural theory of rough inclusion, and consequently - rough implication.

This paper contains extended version of some ideas presented in [8], [9].

2 Multivalued logics as probability logics – a Łukasiewicz's approach

In this section we present briefly basic ideas of Lukasiewicz's approach to multivalued logics as probability logics.

Lukasiewicz associates with every so called *indefinite proposition* of one variable x , $\Phi(x)$ a true value

Table 1: Exemplary data table

| Patient | (H) | (M) | (T) | (F) |
|---------|-----|-----|-----------|-----|
| p1 | no | yes | high | yes |
| p2 | yes | no | high | yes |
| p3 | yes | yes | very high | yes |
| p4 | no | yes | normal | no |
| p5 | yes | no | high | no |
| p6 | no | yes | very high | yes |

$\pi(\Phi(x))$, which is the ratio of the number of all values of x which satisfy $\Phi(x)$, to the number of all possible values of x . For example, the true value of the proposition "x is greater than 3" for $x = 1, 2, \dots, 5$ is $2/5$. It turns out that assuming the following three axioms

- 1) Φ is false if and only if $\pi(\Phi) = 0$;
- 2) Φ is true if and only if $\pi(\Phi) = 1$;
- 3) if $\pi(\Phi \rightarrow \Psi) = 1$ then $\pi(\Phi) + \pi(\sim \Phi \wedge \Psi) = \pi(\Psi)$;

one can show that

- 4) if $\pi(\Phi \equiv \Psi) = 1$ then $\pi(\Phi) = \pi(\Psi)$;
- 5) $\pi(\Phi) + \pi(\sim \Phi) = 1$;
- 6) $\pi(\Phi \vee \Psi) = \pi(\Phi) + \pi(\Psi) - \pi(\Phi \wedge \Psi)$;
- 7) $\pi(\Phi \wedge \Psi) = 0$ iff $\pi(\Phi \vee \Psi) = \pi(\Phi) + \pi(\Psi)$.

Obviously, the above properties have probabilistic flavour. With every implication $\Phi \rightarrow \Psi$ one can associate conditional probability $\pi(\Psi|\Phi) = \frac{\pi(\Phi \wedge \Psi)}{\pi(\Phi)}$. In what follows the above ideas will be used to define the *Rough Modus Ponens*. Let us mention that in applications we are often interested in properties more specific than (1)-(7), related to properties of π defined by data tables.

3 Decision tables and decision rules

Usually we start considerations on rough sets from the concept of a data table. An example of a simple data table is shown in Table 1. In the table H , M , T and F are abbreviations of *Headache*, *Muscle-pain*, *Temperature* and *Flu* respectively.

Columns of the table are labelled by *attributes* (symptoms) and rows – by *objects* (patients), whereas entries of the table are *attribute values*. Such tables are known as *information systems*, *attribute-value tables* or *information tables*. We will use here the term *information table*. Often we distinguish in an information table

two classes of attributes, called *condition* and *decision* (*action*) attributes. For example in Table 1 attributes *Headache*, *Muscle-pain* and *Temperature* can be considered as condition attributes, whereas the attribute *Flu* – as a decision attribute.

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 $(H, no), (M, yes), (T, high)$ determines uniquely the decision (F, yes) . Objects in a decision table are used as labels of decision rules. Decision rules 2) and 5) in Table 1 have the same conditions by different decisions. Such rules are called *inconsistent* (*nondeterministic, conflicting*); otherwise the rules are referred to as *consistent* (*certain, deterministic, nonconflicting*). Sometimes consistent decision rules are called *sure* rules, and inconsistent rules are called *possible* rules. Decision tables containing inconsistent decision rules are called *inconsistent* (*nondeterministic, conflicting*); otherwise the table is *consistent* (*deterministic, non conflicting*). 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 $\gamma(C, D) = 4/6$. 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 (H, no) and (M, yes) and $(T, high)$ then (F, yes) ,
- 2) if (H, yes) and (M, no) and $(T, high)$ then (F, yes) ,
- 3) if (H, yes) and (M, yes) and $(T, very high)$ then (F, yes) ,
- 4) if (H, no) and (M, yes) and $(T, normal)$ then (F, no) ,
- 5) if (H, yes) and (M, no) and $(T, high)$ then (F, no) ,
- 6) if (H, no) and (M, yes) and $(T, very high)$ then (F, yes) ,

From logical point of view decision rules are implications built up from elementary formulas of the form (attribute name, attribute value) and combined together by means of propositional connectives "and", "or" and "implication" in a usual way.

4 Dependency of attributes and decision rules

Intuitively, a set of attributes D depends totally on a set of attributes C , denoted $C \Rightarrow D$, if all values of attributes from D are uniquely determined by values of attributes from C . In other words, D depends totally on C , if there exists a functional dependency between values of D and C . In Table 1 there are no total dependencies whatsoever. If in Table 1, the value of the attribute *Temperature* for patient p_5 were *normal* instead of *high*, there would be a total dependency $\{T\} \Rightarrow \{F\}$, because to each value of the attribute *Temperature* there would correspond an unique value of the attribute *Flu*.

We would need also a more general concept of dependency of attributes, called a *partial dependency* of attributes. Let us depict the idea by example, referring to Table 1. In this table, for example, the attribute *Temperature* determines uniquely only some values of the attribute *Flu*. That is, $(T, \textit{very high})$ implies (F, \textit{yes}) , similarly (T, \textit{normal}) implies (F, \textit{no}) , but (T, \textit{high}) does not imply always (F, \textit{yes}) . Thus the partial dependency means that only some values of D are determined by values of C .

Formally dependency can be defined in the following way. Let D and C be subsets of A .

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 dependency $\{H, M, T\} \Rightarrow \{F\}$ we get $k = 4/6 = 2/3$, because four out of six patients can be uniquely classified as having flu or not, employing attributes *Headache*, *Muscle-pain* and *Temperature*.

The set of decision rules associated with a decision table $S = (U, C, D)$ can be viewed as a description of the dependency $C \Rightarrow D$.

For example the set of decision rules 1), ..., 6) associated with Table 1 can be understood as a description of the dependency $\{H, M, T\} \Rightarrow \{F\}$.

5 Certainty factor of a decision rule

In order to express certainty of decision specified by a decision rule we would need numerical characterization of the rule, showing to what extend the decision can be trusted. To this end we define a *certainty factor* of the rule.

Let Φ and Ψ be logical formulas representing condi-

tions and decisions, respectively and let $\Phi \rightarrow \Psi$ be a decision rule, where $|\Phi|_S$ denote the meaning of Φ in S , i.e., the set of all objects satisfying Φ in S , defined in a usual way, where $S = (U, C, D)$ is a decision table and U, C, D are objects, condition and decision attributes, respectively.

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

$$\mu_S(\Phi, \Psi) = \frac{\text{card}(|\Phi|_S \cap |\Psi|_S)}{\text{card}(|\Phi|_S)},$$

assuming $\text{card}(|\Phi|_S) \neq 0$. Of course $0 \leq \mu_S(\Phi, \Psi) \leq 1$; if the rule $\Phi \rightarrow \Psi$ is deterministic then $\mu_S(\Phi, \Psi) = 1$, and for nondeterministic rules $\mu_S(\Phi, \Psi) < 1$. We will write $\mu(\Phi, \Psi)$ instead of $\mu_S(\Phi, \Psi)$ if S is understood. For example, the certainty factor for decision rules considered in section 3 are as follows:

$$\begin{aligned} \mu(\Phi_1, \Psi_1) &= 1, & \mu(\Phi_4, \Psi_4) &= 1, \\ \mu(\Phi_2, \Psi_2) &= 1/2, & \mu(\Phi_5, \Psi_5) &= 1/2, \\ \mu(\Phi_3, \Psi_3) &= 1, & \mu(\Phi_6, \Psi_6) &= 1, \end{aligned}$$

where Φ_i, Ψ_i denote conditions and decisions of the rule i . Let us notice that the certainty factor can be viewed as a conditional probability that an object x satisfies decision, provided it satisfies condition of the rule, i.e.,

$$\mu(\Phi, \Psi) = \pi(\Psi|\Phi) = \frac{\pi(\Phi \wedge \Psi)}{\pi(\Phi)}, \text{ where } \pi(\Phi) = \frac{\text{card}(|\Phi|_S)}{\text{card}(U)}$$

is an unconditional probability that object of the universe U satisfies Φ .

It is also interesting to observe that the certainty factor of a decision rule is a generalization of a rough membership function

$$\mu_X^B(x) = \frac{\text{card}(X \cap B(x))}{\text{card}(B(x))},$$

where $B(x)$ is an equivalence class, of the equivalence relation generated by the set of attributes B , containing x .

The certainty factor can be also viewed as a degree to what a set X is included in a set Y , i.e., it defines rough inclusion of sets – the basic concept in rough mereology, introduced by Polkowski and Skowron [11], [12]. Hence rough mereology seems to be a very good candidate for a natural theory of rough (approximate) implications, or in other words – as basic theory for reasoning about uncertain inference (decision) rules.

6 Rough Modus Ponens

With every decision rule $\Phi \rightarrow \Psi$ one can associate a rough deduction rule called *Rough Modus Ponens*

$$\frac{\pi(\Phi); \mu(\Phi, \Psi)}{\pi(\Psi)},$$

where

$$\pi(\Psi) = \pi(\sim \Phi \wedge \Psi) + \pi(\Phi) \cdot \mu(\Phi, \Psi). \quad (*)$$

This formula allows us to compute the probability of decisions in terms of probability of conditions and conditional probability of the decision rule, and leads to a generalization of Lukasiewicz's axiom 3), which now has the form:

if $\pi(\Phi \rightarrow \Psi) \neq 0$, then $\pi(\Psi) = \pi(\sim \Phi \wedge \Psi) + \pi(\Phi) \cdot \mu(\Phi, \Psi)$.

Probabilities involved in the *Rough Modus Ponens* inference rule can be computed from data set or can be obtained from a knowledgeable expert.

For example, if $\Phi = (H, \text{yes})$ and (M, no) and (T, high) , $\Psi = (F, \text{yes})$ then $\pi(\Phi) = 1/3$, $\pi(\sim \Phi \wedge \Psi) = 1/2$, $\mu(\Phi, \Psi) = 1/2$ and consequently we get $\pi(\Psi) = 2/3$. For $\Phi = (H, \text{yes})$ and $\Psi(F, \text{no})$ we obtain $\pi(\Phi) = 1/2$, $\pi(\sim \Phi \wedge \Psi) = 1/6$, $\mu(\Phi, \Psi) = 1/3$ and finally $\pi(\Psi) = 1/3$.

The problem considered above is a part of a wider question pursued for many years in AI and is related to common-sense reasoning methods. In classical logic basic rule of inference is grounded on the assumption that if a premise Φ and the implication $\Phi \rightarrow \Psi$ are true then the conclusion Ψ must be also true. This deduction rule is known as *Modus Ponens*. However in the common-sense reasoning methods we must admit that a premise and an inference rule are often not known with certainty, but with some probability and therefore the conclusion must be also equipped with a proper probability measure. Classical logic does not offer methods to solve this dilemma and paradigm of classical logic is no more valid in this case. Consequently *Modus Ponens* cannot be postulated as a fundamental reasoning rule for common-sense reasoning. *Rough Modus Ponens* appears to be one of the possible answers to this dilemma.

Although, formally *RMP* can be considered as a generalization of *MP*, yet there are essential practical and philosophical differences between the to rules of inference.

Classical *MP* is an universal logical rule of inference valid in any system, whereas *RMP* is restricted to a specific data table and can be used to reason about specific experimental knowledge hidden in the data table. Moreover, in the rough set approach we are rather interested not with using single decision rules but we have to use a set of decision rules determined by a dependency between condition and decision attributes. Thus we have to deal not with a single *RMP* rule of inference but with a family of *RMP*'s induced by the dependency. Consequently, the *RMP* needs fur-

ther modification since in the rough set approach we deal not with sets but with families of sets (partitions). Therefore the term $\sim \Phi \wedge \Psi$ may be omitted in every decision rule and the formula (*) can be replaced by the formula

$$\pi(\Psi) = \Sigma(\pi(\Phi) \cdot \mu(\Phi, \Psi)) = \Sigma\pi(\Phi \wedge \Psi) \quad (**)$$

where Σ is taken over all conditions Φ associated with the decision corresponding to Ψ .

This process should be repeated for each decision class for a given table.

7 Conclusions

Very many mathematical models have been proposed to deal with uncertainty of decision rules in knowledge based systems and master uncertainty in reasoning. The presented rough set approach seems to be a very natural answer to this problem and it has a very inherent interpretation in data sets.

Particularly interesting seems the connection of the proposed approach with rough mereology, which appear to be a very well suited basis for managing uncertainty of inconsistent decision rules in knowledge based systems, but this question requires further research.

It seems also interesting to employ to the *Rough Modus Ponens* interval probability, in sense proposed in [4], [8]. This would allow to replace in the *Rough Modus Ponens* crisp probabilities by rough probabilities [8] and thus relax strict restriction on probabilities involved in the inference rules.

8 Acknowledgments

Thanks are due to Prof. Andrzej Skowron, Dr. Marzena Kryszkiewicz and Prof. W. Ziarko for critical remarks.

References

- [1] E. W. Adams. The logic of Conditionals, An Application of Probability to Deductive Logic. D. Reidel Publishing Company, Dordrecht, Boston, 1975.
- [2] W. Bandler, L. Kohout. Fuzzy power sets and fuzzy implication operators. *Fuzzy Sets and Systems*, 4.
- [3] L. Borkowski (ed.). Jan Lukasiewicz - Selected Works, North Holland Publishing Company, Amsterdam, London, Polish Scientific Publishers, Warszawa, 1970.

- [4] A. P. Dempster. Upper and lower probabilities induced by induced by the multiplevalued mapping. *Ann. Math. Statistics*, 38.
- [5] J. Lukasiewicz. Die logischen Grundlagen der Wahrscheinlichkeitsrechnung. Krakow, 1913.
- [6] Handbook of Philosophical Logic, Vol.1, Elements of Classical Logic, D. Gabbay and F. Guenther, Kluwer Academic Publishers, Dordrecht, Boston, London, 1983.
- [7] P. Magrez, P. Smets. Fuzzy modus ponens: A new model suitable for applications in knowledge-based systems. *Information Journal of Intelligent Systems*, 4.
- [8] Z. Pawlak. Rough probability. *Bull. Polish Acad. Sci. Tech.*, 33(9-10).
- [9] Z. Pawlak. Rough set theory and its application to data analysis. *Systems and Cybernetics*, (to appear).
- [10] Z. Pawlak. Granularity of knowledge, indiscernibility and rough sets. In *IEEE Conference on Evolutionary Computation*, pages 100–103, 1998 (to appear).
- [11] L. Polkowski, A. Skowron. Rough Mereology. In *Proc. of the Symposium on Methodologies for Intelligent Systems*, volume 869, pages 85–94, Charlotte, N.C., Lecture Notes in Artificial Intelligence, Springer Verlag, Berlin, 1994.
- [12] L. Polkowski, A. Skowron. Rough mereology: a new paradigm for approximate reasoning. *Journ. of Approximate Reasoning*, 15(4).
- [13] A. Skowron. Management of uncertainty in AI: A rough set approach. In *Proceedings of the conference SOFTEKS*, pages 69–86, Springer Verlag and British Computer Society, 1994.
- [14] E. Trillas, L. Valverde. On implication and indistinguishability in the setting of fuzzy logic. In *Management Decision Support Systems Using Fuzzy Sets and Possibility Theory*, pages 198–212, Verlag TÜ, Rheinland, Köln, 1985.
- [15] S.K.M. Wong, W. Ziarko. On learning and evaluation of decision rules in the context of rough sets. In *Proceedings of the International Symposium on Methodologies for Intelligent Systems*, pages 308–224, Knoxville, 1986.
- [16] L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1.
- [17] L. Zadeh. The role of fuzzy logic in in the management of uncertainty in expert systems. *Fuzzy Sets and Systems*, 11.
- [18] W. Ziarko, N. Shan. KDD–R: A comprehensive system for knowledge discovery using rough sets. In *Proceedings of the International Workshop on Rough Sets and Soft Computing, RSSC'94*, pages 164–173, San Jose, 1994; see also: T. Y. Lin and A. M. Wildberger (eds.), *Soft Computing, Simulation Councils, Inc.*, pages 298–301, San Diego 1995.