Inference Rules, Decision Rules and Rough Sets

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

Inference rules, like e.g. modus ponens, play an essential role in logical reasoning and are fundamental in deductive logic, whereas decision rules are basic tools of reasoning in many branches of AI, particularly in data mining, machine learning decision support and others.

Both inference rules and decision rules are implication, but there are essential differences between these two concepts described by premisses and by conclusions of implication. Inference rules are used to draw true conclusions from true premises, whereas decision rules are prescription of decisions (actions) that must be made when some condition are satisfied. Therefore some probabilistic, fuzzy or rough measures must be associated

with decision rules, to measure the closenes of concepts – in contrast to inference rules where truth values are propagated from from premisses to conclusions. The rough set approach bridges somehow both concepts – inference and decision rules.

With every implication two conditional probabilities are associated, called *credibility* and *coverage factors* respectively. The credibility factor may be considered as partial truth value of the implication and was first introduced by Łukasiewicz in 1913 – whereas the covering factor, introduced recently by Tsumoto, shows how strongly a decision rule covers decision of the decision rule.

It can be shown that the relationship between this two factors is disclosed by the Bayes' Theorem. However, the meaning of Bayes' Theorem in this case differs from that postulated in statistical inference. In statistical data analysis based on Bayes' Theorem, we assume that prior probability about some parameters without knowledge about the data is given. The posterior probability is computed next, which tells us what can be said about prior probability in view of the data. In the rough set approach the meaning of Bayes' Theorem is different. It reveals some relationships in the database, without referring to prior and posterior probabilities, and it can be used to reason about data in terms of approximate (rough) implications. Thus, the proposed approach can be seen as a new model for Bayes' Theorem.

Thus the rough set approach combines together both logical and probabilistic aspects of implications. This idea is due to Lukasiewicz who first pointed out the relationship between implications and Bayes' Theorem. In the lecture the above ideas will be formulated precisely and discussed from the rough set perspective.

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. Łukasiewicz 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 refering to Łukasiewcz's ideas. The essence of Łukasiewcz's idea consists in association with the implication $\Phi \rightarrow \Psi$ a conditional probability, whereas with Φ and Ψ unconditional probabilities are associated. The assumption 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 Łukasiewicz, 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 leeds 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 Łukasiewicz's approach to multivalued logics as probability logics.

Łukasiewicz associates with every so called *indefinite proposition* of one variable x, $\Phi(x)$ a true value $\pi(\Phi(x))$, which is the ratio of the number of all values of x which satify $\Phi(x)$, to the number of all possible values of x. For example, the true value of the proposition "x is greater than 3" for $x = 1, 2, \ldots, 5$ 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 anly if $\pi(\Phi) = 1$;

3) if $\pi(\Phi \to \Psi) = 1$ 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 $\Phi \to \Psi$ 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 aften interested in properties more specific than (1)-(7)s related to properties of π defined by data tables.

3 Decision tables and decision rules

Usually we start considerartions 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.

Table 1: Exemplary data table

Patient	*Headache* (*H*)	*Muscle-pain* (*M*)	*Temperature* (*T*)	*Flu* (*F*)
*p*1	*no*	*yes*	*high*	*yes*
*p*2	*yes*	*no*	*high*	*yes*
*p*3	*yes*	*yes*	*very high*	*yes*
*p*4	*no*	*yes*	*normal*	*no*
*p*5	*yes*	*no*	*high*	*no*
*p*6	*no*	*yes*	*very high*	*yes*

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 form elementary formulas of the from (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 $p5$ 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,very high*) implies (*F,yes*), similarly (*T,normal*) implies (*F,no*), but (*T,high*) does not imply always (*F,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 $1), \ldots, 6)$ associated with Table 1 can be understood as a description of the dependency $\{H, M, T\} \Rightarrow \{F\}$.

5 Decision rules

Let S be a database and let C and D be condition and decision attributes, respectively.

By Φ, Ψ etc. we will denote logicals formulas built up from attributes, attribute-values and logical connectives (*and, or, not*) in a standard way. We will denote by $|\Phi|_S$ the set of all object $x \in U$ satisfying Φ in S and refer to as the *meaning* of Φ in S.

The expression $\pi_S(\Phi) = \frac{card(|\Phi|_S)}{card(U)}$ can be interpreted the probability that the formula Φ is true in S.

A *decision rule* is an expression in the form *"if...then..."*, written $\Phi \rightarrow \Psi$; Φ and Ψ are refered to as *condition* and *decision* of the rule respectively.

A decision rule $\Phi \rightarrow \Psi$ is *admissible* in S if $|\Phi|_S$ is the union of some C-elementary sets, $|\Psi|_S$ is the union of some D-elementary sets and $|\Phi \wedge \Psi|_S \neq \emptyset$. In what follows we will consider admissible decision rules only.

With every decision rule $\Phi \rightarrow \Psi$ we associate the conditional probability that Ψ is true in S given Φ is true in S with the probability $\pi_S(\Phi)$, a called *certainty factor*

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

where $|\Phi|_S$ denotes the set of all objects satisfying Φ in S.

Besides, we will also need a *coverage factor* [46]

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

which is the conditional probability that Φ is true in S given Ψ is true in S with the probability $\pi_S(\Psi)$.

For example, (*F,low*) *and* (*C,black*) *and* (*P,med.*) $\rightarrow$ (*M,good*) is an admissible rule in Table 1 and the certainty and coverage factors for this rule are $1/2$ and $1/4$ respectively.

Let $\{\Phi_i \to \Psi\}_n$ be a set of decision rules such that all conditions Φ_i are pairwise mutally exclusive, i.e., $|\Phi_i \wedge \Phi_j|_S$, for any $1 \leq i, j \leq n$, $i \neq j$, and

$$\sum_{i=1}^{n} \pi_S(\Phi_i|\Psi) = 1. \qquad (*)$$

Then the following property holds:

$$\pi_S(\Psi) = \sum_{i=1}^{n} \pi_S(\Psi|\Phi_i) \cdot \pi_S(\Phi_i). \qquad (**)$$

For any decision rule $\Phi \to \Psi$ the following formula is valid:

$$\pi_S(\Phi|\Psi) = \frac{\pi_S(\Psi|\Phi) \cdot \pi_S(\Phi)}{\sum_{i=1}^{n} \pi_S(\Psi|\Phi_i) \cdot \pi_S(\Phi_i)}. \qquad (***)$$

Formula (***) can be seen as generalization of .... 3) in Łukasiewicz's probabilistic logic.

This means that any database, with distinguished condition and decision attributes (a decision table) or any set of implications satisfying condition (*) satisfies the Bayes' Theorem. Thus databases and set of decision rules can be perceived as a new model for the Bayes' Theorem. Let us note that in both cases we do not refer to prior or posterior probabilities and the Bayes' Theorem simple reveals some patterns in data. This property can be used to reason about data, by inverting implications valid in the database.

6 Conclusions

7 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 (1980) 183-190.

[3] L. Borkowski (ed.). Jan Łukasiewicz – Selected Works, North Holland Publishing Company, Amsterdam, London, Polish Scientific Publushers, Warszawa, 1970.

[4] A. P. Dempster. Upper and lower probabilities induced by induced by the multipl-evalued mapping. *Ann. Math. Statistics*, 38, (1967), 325-339.

[5] J. Łukasiewicz. Die logishen Grundlagen der Wahrscheinilchkeitsrechnung. Krakow, 1913.

[6] Handbook of Philosophical Logic, Vol.1, Elements of Classical Logic, D.Gabbay and F. Guenthner, 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 (1975) 181-200.

[8] Z. Pawlak. Rough probability. *Bull. Polish Acad, Sci. Tech.* 33(9-10), (1985) 499-504.

[9] Z. Pawlak. Rough set theory and its application to data analysis. *Systems and Cybernetics*, (1997), (to appear)

[10] Z. Pawlak. Granurality 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 Symphosium 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*, 1996, 15(4), 333–365.

[13] A. Skowron. Menagement of uncertainty in AI: A rough set approach. In V.Alagar, S. Belgrer and F.Q. Dong (eds.) Proc. SOFTEKS Workshop on Incompletness and Uncertainty in Information Systems, Springer Verlag and British Computer Society (1994), 69-86.

[14] E. Trillas, L. Valverde. On implication and indistiguishability in the setting of fuzzy logic. In *Management Decision Support Systems Using Fuzzy Sets and Possibility Theory*, Verlag TÜ, Rheinland, Köln (1985), 198–212.

[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 possibiliy. *Fuzzy Sets and Systems*, 1 (1977) 3-28.

[17] L. Zadeh. The role of fuzzy logic in in the management of uncertainty in expert systems. *Fuzzy Sets and Systems*, 11 (1983) 199-277

[18] W. Ziarko, N. Shan. KDD–R: A comprehensive system for knowledge discovery using rough sets. *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.