Decision Rules, Bayes' Rule and Rough Sets

Zdzisław Pawlak

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

Abstract. This paper concerns a relationship between Bayes' inference rule and decision rules from the rough set perspective.
In statistical inference based on the Bayes' rule it is assumed that some prior knowledge (prior probability) about some parameters without knowledge about the data is given first. Next the posterior probability is computed by employing the available data. The posterior probability is then used to verify the prior probability.
In the rough set philosophy with every decision rule two conditional probabilities, called *certainty* and *coverage factors*, are associated. These two factors are closely related with the lower and the upper approximation of a set, basic notions of rough set theory. Besides, it is revealed that these two factors satisfy the Bayes' rule. That means that we can use to data analysis the Bayes' rule of inference without referring to Bayesian philosophy of prior and posterior probabilities.

Keywords: Bayes' rule, rough sets, decision rules, information system

1 Introduction

This paper is an extended version of the author's ideas presented in [5,6,7,8]. It concerns some relationships between probability, logic and rough sets and it refers to some concepts of Łukasiewicz presented in [3].

We will dwell in this paper upon the Bayesian philosophy of data analysis and that proposed by rough set theory.

Statistical inference grounded on the Bayes' rule supposes that some prior knowledge (prior probability) about some parameters without knowledge about the data is given first. Next the posterior probability is computed when the data are available. The posterior probability is then used to verify the prior probability.

In the rough set philosophy with every decision rule two conditional probabilities, called *certainty* and *coverage factors*, are associated. These two factors are closely related with the lower and the upper approximation of a set, basic concepts of rough set theory. Besides, it turned out that these two factors satisfy the Bayes' rule. That means that we can use to data analysis the Bayes' rule of inference without referring to Bayesian philosophy, i.e., to the prior and posterior probabilities. In other words, every data set with distinguished condition

N. Zhong, A. Skowron, S. Ohsuga (Eds.): RSFDGrC'99, LNAI 1711, pp. 1–9, 1999.

and decision attributes satisfies the Bayes' rule. This property gives a new look on reasoning methods about data.

2 Information System and Decision Table

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 labelled by attributes, rows are labelled 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 is 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., partition determined by B, will be denoted by $U/I(B)$, or simple 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* or 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 classes of attributes, called *condition* and *decision attributes*, respectively, then the system will be called a *decision table.*

A simple, tutorial example of an information system (a decision table) is shown in Table 1.

Table 1. An example of a decision table

Car	*F*	*P*	*S*	*M*
1	*med.*	*med.*	*med.*	*poor*
2	*high*	*med.*	*large*	*poor*
3	*med.*	*low*	*large*	*poor*
4	*low*	*med.*	*med.*	*good*
5	*high*	*low*	*small*	*poor*
6	*med.*	*low*	*large*	*good*

The table contains data about six cars, where *F, P, S* and *M* denote *fuel consumption*, *selling price, size* and *marketability*, respectively.

Attributes F, P and S are condition attributes, whereas M is the decision attribute. Each row of the decision table determines a decision obeyed when specified conditions are satisfied.

3 Approximations

Suppose we are given an information system (a datat set) $S = (U, A)$, a subset X of the universe U, and subset of attributes B. 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.

For example, let $C = \{F, P, S\}$ be the set of all condition attributes. Then for the set $X = \{1, 2, 3, 5\}$ of cars with poor marketability we have $C_*(X) = \{1, 2, 5\}$, $C^*(X) = \{1, 2, 3, 5, 6\}$ and $BN_C(X) = \{3, 6\}$.

4 Decision Rules

With every information system $S = (U, A)$ we associate a formal language $L(S)$, written L when S is understood. Expressions of the language L are logical formulas denoted by Φ, Ψ etc. built up from attributes and attribute-value pairs by means of logical connectives $\wedge$ (*and*), $\vee$ (*or*), $\sim$ (*not*) in the standard way. We will denote by $||\Phi||_S$ the set of all objects $x \in U$ satisfying Φ in S and refer to as the *meaning* of Φ in S.

The meaning of Φ in S is defined inductively as follows:

1) $||(a, v)||_S = \{v \in U : a(v) = U\}$ for all $a \in A$ and $v \in V_a$,
2) $||\Phi \vee \Psi||_S = ||\Phi||_S \cup ||\Psi||_S$,
3) $||\Phi \wedge \Psi||_S = ||\Phi||_S \cap ||\Psi||_S$,
4) $|| \sim \Phi||_S = U - ||\Phi||_S$.

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

A *decision rule* in L is an expression $\Phi \rightarrow \Psi$, read *if* Φ *then* Ψ; Φ and Ψ are referred to as *conditions* and *decisions* of the rule, respectively.

An example of a decision rule is given below

$$(F, med.) \wedge (P, low) \wedge (S, large) \rightarrow (M, poor).$$

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

With every decision rule $\Phi \rightarrow \Psi$ we associate a conditional probability $\pi_S(\Psi|\Phi)$ that Ψ is true in S given Φ is true in S with the probability $\pi_S(\Phi)\frac{card(||\Phi||_S)}{card(U)}$, called the *certainty factor* and defined as follows:

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

where $||\Phi||_S \neq 0$.

This coefficient is widly used in data mining and is called "confidence coefficient".

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

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

Besides, we will also need a *coverage factor*

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

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

Certainty and coverage factors for decision rules associated with Table 1 are given in Table 2.

Table 2. Certainty and coverage factors

Car	*F*	*P*	*S*	*M*	*Cert.*	*Cov.*
1	*med.*	*med.*	*med.*	*poor*	1	1/4
2	*high*	*med.*	*large*	*poor*	1	1/4
3	*med.*	*low*	*large*	*poor*	1/2	1/4
4	*low*	*med.*	*med.*	*good*	1	1/2
5	*high*	*low*	*small*	*poor*	1	1/4
6	*med.*	*low*	*large*	*good*	1/2	1/2

More about managing uncertainty in decision rules can be found in [2].

5 Decision Rules and Approximations

Let $\{\Phi_i \to \Psi\}_n$ be a set of decision rules such that:

$$\begin{aligned}&\text{all conditions } \Phi_i \text{ are pairwise mutally exclusive, i.e., } ||\Phi_i \wedge \Phi_j||_S = \emptyset, \text{for any}\\ &1 \leq i, j \leq n,\ i \neq j,\ \text{and} \\ &\sum_{i=1}^{n} \pi_S(\Phi_i|\Psi) = 1.\end{aligned} \tag{1}$$

Let C and D be condition and decision attributes, respectively, and let $\{\Phi_i \to \Psi\}_n$ be a set of decision rules satisfying (1).

Then the following relationships are valid:

$$\text{a) } C_*(||\Psi||_S) = ||\bigvee_{\pi(\Psi|\Phi_i)=1} \Phi_i||_S,$$

$$\text{b) } C^*(||\Psi||_S) = ||\bigvee_{0<\pi(\Psi|\Phi_i)\leq 1} \Phi_i||_S,$$

$$\text{c) } BN_C(||\Psi||_S) = ||\bigvee_{0<\pi(\Psi|\Phi_i)<1} \Phi_i||_S = \bigcup_{i=1}^{n} ||\Phi_i||_S.$$

The above properties enable us to introduce the following definitions:

i) If $||\Phi||_S = C_*(||\Psi||_S)$, then formula Φ will be called the *C-lower approximation* of the formula Ψ and will be denoted by $C_*(\Psi)$;
ii) If $||\Phi||_S = C^*(||\Psi||_S)$, then the formula Φ will be called the *C-upper approximation* of the formula Φ and will be denoted by $C^*(\Psi)$;
iii) If $||\Phi||_S = BN_C(||\Psi||_S)$, then Φ will be called the *C-boundary* of the formula Ψ and will be denoted by $BN_C(\Psi)$.

Let us consider the following example.
The C-lower approximation of (M, *poor*) is the formula

$$\begin{aligned}C_*(M, poor) = &((F, med.) \wedge (P, med.) \wedge (S, med.)) \vee \\ &((F, high) \wedge (P, med.) \wedge (S, large)) \vee \\ &((F, high) \wedge (P, low) \wedge (S, small)).\end{aligned}$$

The C-upper approximation of (M, *poor*) is the formula

$$\begin{aligned}C^*(M, poor) = &((F, med.) \wedge (P, med.) \wedge (S, med.)) \vee \\ &((F, high) \wedge (P, med.) \wedge (S, large)) \vee \\ &((F, med.) \wedge (P, low) \wedge (S, large)) \vee \\ &((F, high) \wedge (P, low) \wedge (S, small)).\end{aligned}$$

The C-boundary of (M, *poor*) is the formula

$$BN_C(M, poor) = ((F, med.) \wedge (P, low) \vee (S, large)).$$

After simplification we get the following approximations

$$C_*(M, poor) = ((F, med.) \wedge (P, med.)) \vee (F, high),$$
$$C^*(M, poor) = (F, med.) \vee (F, high).$$

The concepts of the lower and upper approximation of a decision allow us to define the following decision rules:

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

$$C^*(\Psi) \to \Psi,$$

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

For example, from the approximations given in the example above we get the following decision rules:

$((F,med.) \wedge (P,med.)) \vee (F,high) \to (M,\ poor)$,
$(F,med.) \vee (F,high) \to (M,\ poor)$,
$((F,med.) \wedge (P,low) \wedge (S,large)) \to (M,poor)$.

From these definitions it follows that any decision Ψ can be uniquely discribed by the following two decision rules:

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

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

From the above calculations we can get two decision rules

$((F,med.) \wedge (P,med.)) \vee (F,\ high) \to (M,poor)$,
$((F,med.) \wedge (P,low.) \wedge (S,large)) \to (M,poor)$,

which are associated with the lower approximation and the boudary region of the decision (M, *poor*), respectively and describe decision (M, *poor*).

Obviously we can get similar decision rules for the decision (M, *good*) which are as follows:

$(F,low) \to (M,good)$,
$((F,med.) \wedge (P,low.) \wedge (S,large)) \to (M,good)$.

This coincides with the idea given by Ziarko [15] to represent decision tables by means of three decision rules corresponding to positive region the boundary region, and the negative region of a decision.

6 Decision Rules and Bayes' Rules

If $\{\Phi_i \rightarrow \Psi\}_n$ is a set of decision rules satisfying condition (1), then the well known formula for total probability holds:

$$\pi_S(\Psi) = \sum_{i=1}^{n} \pi_S(\Psi|\Phi_i) \cdot \pi_S(\Phi_i). \tag{2}$$

Moreover for any decision rule $\Phi \rightarrow \Psi$ the following Bayes' rule is valid:

$$\pi_S(\Phi_j|\Psi) = \frac{\pi_S(\Psi|\Phi_j) \cdot \pi_S(\Phi_j)}{\sum_{i=1}^{n} \pi_S(\Psi|\Phi_i) \cdot \pi_S(\Phi_i)}. \tag{3}$$

That is, any decision table or any set of implications satisfying condition (1) satisfies the Bayes' rule, without referring to prior and posterior probablities – fundamental in Baysian data analysis philosophy. Bayes' rule in our case says that: if an implication $\Phi \rightarrow \Psi$ is true to the degree $\pi_S(\Psi|\Phi)$ then the implication $\Psi \rightarrow \Phi$ is true to the degree $\pi_S(\Phi|\Psi)$.

This idea can be seen as a generalization of a *modus tollens* inference rule, which says that if the implication $\Phi \rightarrow \Psi$ is true so is the implication $\sim \Psi \rightarrow \sim \Phi$.

For example, for the set of decision rules

((*F*,*med.*) ∧ (*P*,*med.*)) ∨ (*F*, *high*) → (*M*,*poor*),
((*F*,*med.*) ∧ (*P*,*low*) ∧ (*S*,*large*)) → (*M*,*poor*),
(*F*,*low*) → (*M*,*good*),
((*F*,*med.*) ∧ (*P*,*low*) ∧ (*S*,*large*)) → (*M*,*good*),

we get the values of ceratinty and coverage factors shown in Table 3.

Table 3. Initial decision rules

Rule	*Decision*	*Certainty*	*Coverage*
certain	*poor*	1	3/4
boundary	*poor*	1/2	1/4
certain	*good*	1	1/2
boundary	*good*	1/2	1/2

The above set of decison rules can be "reversed" as

(*M*,*poor*) → ((*F*,*med.*) ∧ (*P*,*med.*)) ∨ (*F*, *high*),
(*M*,*poor*) → ((*F*,*med.*) ∧ (*P*,*low*) ∧ (*S*,*large*)),
(*M*,*good*) → (*F*,*low*),
(*M*,*good*) → ((*F*,*med.*) ∧ (*P*,*low*) ∧ (*S*,*large*)).

Due to Bayes' rule the certainty and coverage factors for inverted decision rules are mutually exchanged as shown in Table 4 below.

Table 4. Reversed decision rules

Rule	*Decision*	*Certainty*	*Coverage*
certain	*poor*	3/4	1
boundary	*poor*	1/4	1/2
certain	*good*	1/2	1
boundary	*good*	1/2	1/2

This property can be used to reason about data in the way similar to that allowed by *modus tollens* inference rule in classical logic.

7 Conclusions

It is shown in this paper that any decision table satisfies Bayes' rule. This enables to apply Bayes' rule of inference without referring to prior and posterior probabilities, inherently associated with "classical" Bayesian inference philosophy. From data tables one can extract decision rules – implications labelled by certainty factors expressing their degree of truth. The factors can be computed from data. Moreover, one can compute from data the coverage degrees expressing the truth degrees of "reverse" implications. This can be treated as generalization of *modus tollens* inference rule.

Acknowledgments

Thanks are due to Prof. Andrzej Skowron for his critical remarks.

References

1. Adams, E. W. The logic of conditionals, an application of probability to deductive logic. D. Reidel Publishing Company, Dordrecht, Boston, 1975.
2. Grzymała-Busse, J. Managing Uncertainty in Expert Systems; Kluwer Academic Publishers, Dordrecht, Boston, 1991,
3. Łukasiewicz, J. Die logishen Grundlagen der Wahrscheinilchkeitsrechnung. Krakow (1913). In: L. Borkowski (ed.): Jan Łukasiewicz – Selected Works, North Holland Publishing Company, Amsterdam, London, Polish Scientific Publishers, Warszawa, 1970.
4. Pawlak, Z. Rough Sets – Theoretical Aspects of Reasoning about Data; Kluwer Academic Publishers: Boston, Dordrecht, 1991.

5. Pawlak, Z. Reasoning about data – a rough set perspective. In: L. Polkowski, A. Skowron (eds.), Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, 1424 Springer, First International Conference, RSCTC'98, Warsaw, Poland, June, Proceedings, (1998) 25–34
6. Pawlak, Z. Rough Modus Ponens. In: Proceedings of Seventh International Conference, International Processing and Management of Uncertainty in Knowledge-Based (IPMU), Paris, France, July 6–10, 1998 1162–1166.
7. Pawlak, Z. Logic, probability and rough sets (to appear).
8. Pawlak, Z. Inference rules, decision rules and rough sets (to appear).
9. Pawlak, Z., Skowron, A. Rough membership functions. In: R.R. Yaeger, M. Fedrizzi, and J. Kacprzyk (eds.), Advances in the Dempster Shafer Theory of Evidence, John Wiley & Sons, Inc., New York, 1994, 251–271.
10. Polkowski, L., Skowron, A. (eds.). Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, 1424 Springer, First International Conference, RSCTC'98, Warsaw, Poland, June, Proceedings, 1998.
11. Polkowski, L., Skowron, A. (eds.). Rough Sets in Knowledge Discovery; Physica-Verlag: Vol. 1, 2, 1998.
12. Skowron, A. Menagement of uncertainty in AI: A rough set approach. In: Proceedings of the conference SOFTEKS, Springer Verlag and British Computer Society. 1994, 69–86.
13. Tsumoto, S., Kobayashi, S., Yokomori, T., Tanaka, H., Nakamura, A. (eds.). Proceedings of the Fourth International Workshop on Rough Sets, Fuzzy Sets, and Machine Discovery (RSFD'96). The University of Tokyo, November 6–8, 1996.
14. Tsumoto, S. Modelling medical diagnostic rules based on rough sets. In: L. Polkowski, A. Skowron (eds.), Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence, 1424 Springer, First International Conference, RSCTC'98, Warsaw, Poland, June, Proceedings, 1998. 475–482.
15. Ziarko, W. Approximation Region -Based Decision Tables. In: L. Polkowski, A. Skowron (eds.): Rough Sets in Knowledge Discovery, Physica-Verlag, 1998, 1, 2. 178 -185.