Logic, Probability and Rough Sets

Zdzislaw Pawlak

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

1 Introduction

This paper is concerned with some considerations concerning inference rules and decision rules, from the rough set perspective.

Decision rules play an important role in various branches of AI, e.g., data mining, machine learning, decision support and others. Interference rules play a fundamental role in logic, and attracted attention of logicians and philosophers for many years. From logical point of view both, decision rules and interference rules are implications.

However, there are essential differences between using decision rules in AI and inference rules in logic. Inference rules (modus ponens) must be true, in order to guarantee to draw true conclusions from true premises. In contrast, in AI decision rules are meant as prescription of decisions that must be taken when some conditions are satisfied. In this case, in order to express to what degree the decision can be trusted, instead of truth, a credibility factor of the decision rule is associated. The relationship between truth and probability first was investigated by Łukasiewicz [3, 16], who showed that probabilistic interpretation of implication leads to Bayes’ theorem.

In rough set theory decision rules are of special interest, since they are inherently connected with the basic concepts of the theory – approximations and partial dependencies. The rough set approach bridges some extend the logical and AI views on decision rules, and can be seen as generalization of Łukasiewicz’s probabilistic logic associated with multivalued logic [3, 16]. Some considerations on this subject can be also found in [22, 23, 24, 25, 30].

For more information about rough set theory and its generalizations the reader is advised to consult the enclosed references. In particular an overview of current state of the theory and its applications can be found in [29].

2 The Łukasiewicz’s Approach

In this section we present briefly basic ideas of Łukasiewicz’s approach to multivalued logics as probabilistic logic.

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 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 \( x = 1, 2, \ldots, 5 \) is 2/5. It turns out that assuming the following three axioms

1) \( \Phi \) is false if and only if \( \pi(\Phi) = 0 \);
2) \( \Phi \) is true if and only if \( \pi(\Phi) = 1 \);
3) if \( \pi(\Phi \rightarrow \Psi) = 1 \) then \( \pi(\Phi) + \pi(\neg \Phi \land \Psi) = \pi(\Psi) \);

one can show that

4) if \( \pi(\Phi \equiv \Psi) = 1 \) then \( \pi(\Phi) = \pi(\Psi) \);
6) \( \pi(\Phi \lor \Psi) = \pi(\Phi) + \pi(\Psi) - \pi(\Phi \land \Psi) \);
7) \( \pi(\Phi \land \Psi) = 0 \) iff \( \pi(\Phi \lor \Psi) = \pi(\Phi) + \pi(\Psi) \).

Obviously, the above properties have probabilistic flavour.

The idea that implication should be associated with conditional probability is attributed to Ramsey (cf. [1]), but as mentioned in the introduction, is can be traced back to Łukasiewicz [3, 16], who first formulated this idea in connection with his multivalued logic and probability. More extensive study of connection of implication and conditional probability can be found in [1].

3 Rough Sets – the Intuitive Background

Rough set theory is based on the indiscernibility relation. The indiscernibility relation identifies objects displaying the same properties, i.e., groups together elements of interest into granules of indiscernible (similar) objects. These granules, called elementary sets (concepts), are basic building blocks (concepts) of knowledge about the universe. For example, if our universe of discourse were patients suffering from a certain disease, then patients displaying the same symptoms would be indiscernible in view of the available information and form clusters of similar patients.

Union of elementary concepts is referred to as a crisp or precise concept (set); otherwise a concept (set) is called rough, vague or imprecise. Thus rough concepts cannot be expressed in terms of elementary concepts. However, they can be expressed approximately by means of elementary concepts using the idea of the lower and the upper approximation of a concept. The lower approximation of the concept is the union of all elementary concepts which are included in the concept, whereas the upper approximation is the union of all elementary concepts which have nonempty intersection with the concept, i.e., the lower and the upper approximations of a concept are the union of all elementary concepts which are surely and possibly included in the concept, respectively. The difference between the lower and the upper approximation of the concept is its boundary region. Hence a concept is rough if it has nonempty boundary region.

Approximations are basic operations in rough set theory. They are used to deal with rough (vague) concepts, since in the rough set approach we replace
rough concepts by pairs of precise concepts – the lower and the upper approximations of the rough concept. Thus approximations are used to express precisely our knowledge about imprecise concepts.

The problem of expressing vague concepts in terms of precise concepts in rough set theory can be also formulated differently, by employing the idea of dependency (partial) between concepts. We say that a concept (set) depends totally on a set of concepts if it is the union of all those concepts; if it is the union of some concepts it depends partially on these concepts. Thus partial dependency can also be used to express vague concepts in terms of precise concepts.

Both, approximations and dependencies are defined using decision rules, which are implications in the form "if...then..."

Approximations, dependencies and decision rules are basic tools of rough set theory and will be discussed in details in the next sections.

4 Database

Rough set theory is mainly meant to be used to data analysis, therefore in what follows the theory will be formulated not in general terms but with reference to data. Hence we will start our consideration from a database. Intuitively by the database we will understand a data table whose columns are labelled by attributes (e.g., color, temperature, etc.), rows are labelled by objects of interest (e.g., patients, states, processes etc.) and entries of the table are attribute values (e.g., red, high, etc.). A very simple example of a database is shown below: The table contains data about six cars where $F,$

<table>
<thead>
<tr>
<th>Car</th>
<th>F</th>
<th>C</th>
<th>P</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>med.</td>
<td>black</td>
<td>med.</td>
<td>poor</td>
</tr>
<tr>
<td>2</td>
<td>high</td>
<td>white</td>
<td>med.</td>
<td>poor</td>
</tr>
<tr>
<td>3</td>
<td>med.</td>
<td>white</td>
<td>low</td>
<td>poor</td>
</tr>
<tr>
<td>4</td>
<td>low</td>
<td>black</td>
<td>med.</td>
<td>good</td>
</tr>
<tr>
<td>5</td>
<td>high</td>
<td>red</td>
<td>low</td>
<td>poor</td>
</tr>
<tr>
<td>6</td>
<td>med.</td>
<td>white</td>
<td>low</td>
<td>good</td>
</tr>
</tbody>
</table>

$C,$ $P$ and $M$ denote fuel consumption, color, selling price and marketability, respectively.

Formally the database is defined as follows.

By a database we will understand a pair $S = (U, A),$ where $U$ and $A,$ are finite, nonempty sets called the universe, and a set of attributes respectively.
With every attribute $a \in A$ we associate a set $V_a$, of its \textit{values}, called the \textit{domain} of $a$. Any subset $B$ of $A$ determines a binary relation $I(B)$ on $U$, which will be called an \textit{indiscernibility relation}, and is defined as follows:

$$(x,y) \in I(B) \text{ if and only if } a(x) = a(y) \text{ for every } a \in A,$$

where $a(x)$ denotes the value of attribute $a$ for element $x$.

It can easily be seen that $I(B)$ 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 \textit{$B$-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 \textit{$B$-elementary sets} or \textit{$B$-granules}.

For example, cars 1, 3 and 6 are pairwise indiscernible with respect to the attribute $F$. If $B = \{C,P\}$ and $x = 3$, then $B(x) = \{3,6\}$.

Instead of an equivalence relation as a basis for rough set theory many authors proposed another relations, e.g., a tolerance relation, an ordering relations and others. However in this paper we will stay by the equivalence relation.

5 Approximations of Sets

Having defined the notion of a database we are now in the position to put forth our basic notions of approximation of a set by other sets, which is defined next.

Let us define two following operations on sets:

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

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

assigning to every $X \subseteq U$ two sets $B_*(X)$ and $B^*(X)$ called the $B$-\textit{lower} and the $B$-\textit{upper approximation} of $X$, respectively.

Hence, the $B$-lower approximation of a concept is the union of all $B$-granules that are included in the concept, whereas the $B$-upper approximation of a concept is the union of all $B$-granules that have a nonempty intersection with the concept. The set

$$BN_B(X) = B^*(X) - B_*(X)$$

will be referred to as the $B$-\textit{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, for the set of cars $X = \{4, 6\}$ selling well the $B$-lower and the $B$-upper approximations of $X$ are $\{4\}$ and $\{3, 4, 6\}$, respectively, where $B = \{F, C, P\}$.

6 Dependency of Attributes

Another important issue in data analysis is discovering dependencies between attributes. 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 $C$.

We would need also a more general concept of dependency, called a partial dependency of attributes. Intuitively, 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

$$k = \gamma(C, D) = \frac{\sum_{X \in U/D} \text{card}(C, (X))}{\text{card} U},$$

If $C \Rightarrow_k D$, we will call $C$-condition and $D$-decision attributes, respectively. Any database with distinguished condition and decision attributes is usually called a decision table.

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 $C$, and if $k = 0$, does not depend on $C$.

The coefficient $k$ expresses the ratio of all elements of the universe, which can be properly classified to blocks of the partition $U/D$, employing attributes $C$ and will be called the degree of the dependency.

For example, the degree of dependency between the set of condition attributes $\{F, C, P\}$ and the decision attribute $\{M\}$ in Table 1 is $2/3$.

The notion of dependency of attributes is used to express relationships hidden in the database. It expresses the global properties of the database in contrast to approximations which express local properties of the database.

7 Decision Rules

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

By $\Phi, \Psi$ etc. we will denote logical 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 objects \(x \in U\) satisfying \(\Phi\) in \(S\) and refer to as the meaning of \(\Phi\) in \(S\).

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

A decision rule is an expression in the form "if...then...", written \(\Phi \rightarrow \Psi\); \(\Phi\) and \(\Psi\) are referred to as conditions and decisions 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 \(\Psi\) is true in \(S\) given \(\Phi\) is true in \(S\) with the probability \(\pi_S(\Phi)\), called the certainty factor and defined as follows:

\[
\pi_S(\Psi|\Phi) = \frac{\text{card}([\Phi \wedge \Psi]_S)}{\text{card}([\Phi]_S)},
\]

where \([\Phi]_S \neq \emptyset\) denotes the set of all objects satisfying \(\Phi\) in \(S\).

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

\[
\pi_S(\Phi|\Psi) = \frac{\text{card}([\Phi \wedge \Psi]_S)}{\text{card}([\Phi]_S)},
\]

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

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

Let \(\{\Phi_i \rightarrow \Psi\}_n\) be a set of decision rules such that all conditions \(\Phi_i\) are pairwise mutually exclusive, i.e., \([\Phi_i \wedge \Phi_j]_S = \emptyset\), for any \(1 \leq i, j \leq n, i \neq j\), and

\[
\sum_{i=1}^{n} \pi_S(\Phi_i|\Psi) = 1. \tag{1}
\]

Then the following property holds:

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

For any decision rule \(\Phi \rightarrow \Psi\) the following formula is valid:

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

Formula (2) is well known as the formula for total probability and it can be seen as generalization of axiom 3) in Lukasiewicz’s probabilistic logic whereas formula (3) is the Bayes’ theorem.
This means that any database, with distinguished condition and decision attributes (a decision table) or any set of implications satisfying condition (1) satisfies the Bayes' theorem. Thus databases or 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.

8 Conclusions

The papers shows that any set of data satisfying some simple conditions satisfies the Bayes' theorem — without referring to prior and posterior probabilities inherently associated with Baysian statistical inference philosophy. The result bridges somehow rough set theory and some ideas provided by Lukasiewicz in context of his multivalued logic and probability.

References

16. Łukasiewicz J. (1913) Die logischen Grundlagen der Wahrscheinlichkeitsrechnung, Kraków
42. Tsumoto S. (Ed.) (1996) Bulletin of International Rough Set Society 1/1
43. Tsumoto S. (Ed.) (1997) Bulletin of International Rough Set Society 1/2
44. Tsumoto S. (Ed.) (1998) Bulletin of International Rough Set Society 2/1