# In Pursuit of Patterns in Data Reasoning from Data – The Rough Set Way

Zdzisław Pawlak

Institute of Theoretical and Applied Informatics, Polish Academy of Sciences,
ul. Bałtycka 5, 44 100 Gliwice, Poland
zpw@ii.pw.edu.pl

**Abstract.** This paper concerns some aspects of rough set based data analysis. In particular rough set look on Bayes' formula leads to new methodology of reasoning from data and shows interesting relationship between Bayes' theorem, rough sets and flow graphs. Three methods of flow graphs application in drawing conclusions from data are presented and examined.

*MOTTO:*
"It is a capital mistake to theorise before one has data"
**Sherlock Holmes**
In: A Scandal in Bohemia

## 1 Introduction

No doubt that the most famous contribution to reasoning from data should be attributed to the renowned Mr. Sherlock Holmes, whose mastery of using data in reasoning has been well known world wide for over hundred years.

More seriously, reasoning from data is the domain of inductive reasoning, which uses data about sample of larger reality as a starting point of inference – in contrast to deductive reasoning, where axioms expressing some universal truths are used as a departure point of reasoning.

In the rough set approach granular structure of data imposed by the indiscernibility relation is used do discover patterns in data. In rough set theory patterns in data can be characterized by means of approximations, or equivalently by decision rules induced by the data. With every decision rule in a decision table three coefficients are associated: the *strength*, the *certainty* and the *coverage factors* of the rule. It is shown that these coefficients satisfy Bayes' theorem and the total probability theorem. This enables us to use Bayes' theorem to discover patterns in data in a different way from that offered by standard Bayesian inference technique employed in statistical reasoning, without referring to prior and posterior probabilities, inherently associated with Bayesian inference methodology. Besides, a new form of Bayes' theorem is introduced, based on the strength of decision rules, which simplifies essentially computations.

J.J. Alpigini et al. (Eds.): RSCTC 2002, LNAI 2475, pp. 1–9, 2002.

Furthermore, it is shown that the decision rules define a relation between condition and decision granules, which can be represented by a flow graph. The certainty and coverage factors determine a "flow of information" in the graph, ruled by the total probability theorem and Bayes' theorem, which shows clearly the relationship between condition and decision granules determined by the decision table. This leads to a new class of flow networks, unlike to that introduced by Ford and Fulkerson [1]. The introduced flow graphs may have many applications not necessarily associated with decision tables, but this requires further study.

The decision structure of a decision table can be represented in a "decision space", which is Euclidean space, in which dimensions of the space are determined by decision granules, points in the space are condition granules and coordinates of the points are strengths of the corresponding rules. Distance in the decision space between condition granules allows to determine how "distant" are decision makers in view of their decisions. This idea can be viewed as a generalization of the indiscernibility matrix [7], basic tool to find reducts in information systems. Besides, the decision space gives a clear insight in the decision structure imposed by the decision table.

A simple tutorial example is used to illustrate the basis ideas discussed in the paper.

## 2 Basic Concepts

In this section we recall basic concepts of rough set theory [4,5,6,7].

An *information system* is a pair $S = (U, A)$, where $U$ and $A$, are non-empty finite sets called the *universe*, and the set of *attributes*, respectively such that $a : U \to V_a$, where $V_a$, is the set of all values of $a$ 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 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., a partition determined by $B$, will be denoted by $U/I(B)$, or simply by $U/B$; an equivalence class of $I(B)$, i.e., block of the partition $U/B$, containing $x$ will be denoted by $B(x)$ and called $B$-granule induced by $x$.

If $(x, y)$ belongs to $I(B)$ we will say that $x$ and $y$ are *B-indiscernible* (*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 the information system two disjoint classes of attributes, called *condition* and *decision attributes*, respectively, then the system will be called a *decision table* and will be denoted by $S = (U, C, D)$, where $C$ and $D$ are disjoint sets of condition and decision attributes, respectively and $C \cup D = A$.

$C(x)$ and $D(x)$ will be referred to as the condition granule and the decision granule induced by x, respectively.

An example of a decision table is shown in Table 1.

**Table 1.** An example of decision table

| *Fact no.* | *Driving conditions* | | | *Consequence* | *N* |
|---|---|---|---|---|---|
| | *weather* | *road* | *time* | *accident* | |
| 1 | misty | icy | day | yes | 80 |
| 2 | foggy | icy | night | yes | 140 |
| 3 | misty | not icy | night | yes | 40 |
| 4 | sunny | icy | day | no | 500 |
| 5 | foggy | icy | night | no | 20 |
| 6 | misty | not icy | night | no | 200 |

In the table, 6 facts concerning 980 cases of driving a car in various driving conditions are presented. In the table columns labeled *weather, road* and *time*, called *condition attributes*, represent driving conditions. The column labeled by *accident*, called *decision attribute*, contains information whether an accident has occurred or not. $N$ denotes the number of analogous cases.

## 3 Decision Rules

Each row of the decision table determines a decision rule, e.g., row 1 determines the following decision rule *"if weather is misty and road is icy and time is day then accident occurred"* in 80 cases.

Let $S = (U, C, D)$ be a decision table. Every $x \in U$ determines a sequence $c_1(x), \ldots, c_n(x), d_1(x), \ldots, d_m(x)$ where $\{c_1, \ldots, c_n\} = C$ and $\{d_1, \ldots, d_m\} = D$.

The sequence will be called a *decision rule induced by* $x$ (in $S$) and denoted by $c_1(x), \ldots, c_n(x) \rightarrow d_1(x), \ldots, d_m(x)$ or in short $C \rightarrow_x D$.

The number $supp_x(C, D) = |C(x) \cap D(x)|$ will be called a support of the decision rule $C \rightarrow_x D$ and the number

$$\sigma_x(C, D) = \frac{supp_x(C, D)}{|U|},$$

will be referred to as the *strength* of the decision rule $C \rightarrow_x D$, where $|X|$ denotes the cardinality of $X$.

With every decision rule $C \rightarrow_x D$ we associate a *certainty factor* of the decision rule, denoted $cer_x(C, D)$ and defined as follows:

$$cer_x(C, D) = \frac{|C(x) \cap D(x)|}{|C(x)|} = \frac{\sigma_x(C, D)}{\pi(C(x))},$$

where $C(x) \neq \emptyset$ and $\pi(C(x))$.

The certainty factor may be interpreted as conditional probability that $y$ belongs to $D(x)$ given $y$ belongs to $C(x)$, symbolically $\pi_x(D|C)$, i.e., $cer_x(C, D) = \pi_x(D|C)$.

If $cer_x(C, D) = 1$, then $C \to_x D$ will be called a *certain decision rule*; if $0 < cer_x(C, D) < 1$ the decision rule will be referred to as an *uncertain decision rule.*

Besides, we will also use a *coverage factor* (see [8]) of the decision rule, denoted $cov_x(C, D)$ defined as

$$cov_x\left(C, D\right) = \frac{|C\left(x\right) \cap D\left(x\right)|}{|D\left(x\right)|} = \frac{\sigma_x\left(C, D\right)}{\pi\left(D\left(x\right)\right)},$$

where $D(x) \neq \emptyset$ and $\pi(D(x)) = \frac{|D(x)|}{|U|}$. Similarly

$$cov_x\left(C, D\right) = \pi_x\left(C|D\right).$$

If $C \to_x D$ is a decision rule then $D \to_x C$ will be called an *inverse decision rule.* The inverse decision rules can be used to give *explanations* (*reasons*) for a decision.

In Table 2 the *strength, certainty* and *coverage factors* for Table 1 are given.

**Table 2.** Characterization of decision rules

| *fact no.* | *Strength* | *Certainty* | *Coverage* |
|---|---|---|---|
| 1 | 0.082 | 1.000 | 0.308 |
| 2 | 0.143 | 0.877 | 0.538 |
| 3 | 0.041 | 1.167 | 0.154 |
| 4 | 0.510 | 1.000 | 0.695 |
| 5 | 0.020 | 0.123 | 0.027 |
| 6 | 0.204 | 0.833 | 0.278 |

## 4 Properties of Decision Rules

Decision rules have important probabilistic properties which are discussed next [2,3].

Let $C \to_x D$ be a decision rule. Then the following properties are valid:

$$\sum_{y \in C(x)} cer_y\left(C, D\right) = 1 \tag{1}$$

$$\sum_{y \in D(x)} cov_y\left(C, D\right) = 1 \tag{2}$$

$$\begin{aligned} \pi\left(D\left(x\right)\right) &= \sum_{y \in C(x)} cer_y\left(C, D\right) \cdot \pi\left(C\left(x\right)\right) = \\ &= \sum_{y \in C(x)} \sigma_y\left(C, D\right) \end{aligned} \tag{3}$$

$$\pi\left(C\left(x\right)\right)=\sum_{y\in D(x)}cov_{y}\left(C,D\right)\cdot\pi\left(D\left(y\right)\right)= \tag{4}$$
$$=\sum_{y\in D(x)}\sigma_{y}\left(C,D\right)$$

$$cer_{x}\left(C,D\right)=\frac{cov_{x}\left(C,D\right)\cdot\pi\left(D\left(x\right)\right)}{\pi\left(C\left(x\right)\right)}= \tag{5}$$
$$=\frac{\sigma_{x}\left(C,D\right)}{\pi\left(C\left(x\right)\right)}$$

$$cov_{x}\left(C,D\right)=\frac{cer_{x}\left(C,D\right)\cdot\pi\left(D\left(x\right)\right)}{\pi\left(D\left(x\right)\right)}= \tag{6}$$
$$=\frac{\sigma_{x}\left(C,D\right)}{\pi\left(D\left(x\right)\right)}$$

That is, any decision table, satisfies (1)–(6). Observe that (3) and (4) refer to the well known *total probability theorem*, whereas (5) and (6) refer to *Bayes' theorem.*

Thus in order to compute the certainty and coverage factors of decision rules according to formula (5) and (6) it is enough to know the strength (support) of all decision rules only.

Formulas (5) and (6) can be rewritten as

$$cer_{x}\left(C,D\right)=cov_{x}\left(C,D\right)\cdot\gamma_{x}\left(C,D\right) \tag{7}$$

$$cov_{x}\left(C,D\right)=cer_{x}\left(C,D\right)\cdot\gamma_{x}^{-1}\left(C,D\right) \tag{8}$$

where $\gamma_x(C,D)=\frac{|D(x)|}{|C(x)|}=\frac{cer_x(C,D)}{cov_x(C,D)}$
Let us observe that

$$cov_{x}\left(C,D\right)\cdot\pi\left(D\left(x\right)\right)=\sigma_{x}\left(C,D\right) \tag{9}$$

$$cer_{x}\left(C,D\right)\cdot\pi\left(C\left(x\right)\right)=\sigma_{x}\left(C,D\right) \tag{10}$$

## 5 Granularity of Data and Flow Graphs

With every decision table we associate a *flow graph*, i.e., a directed acyclic graph defined as follows: to every decision rule $C\rightarrow_x D$ we assign a *directed branch* $x$ connecting the *input node* $C(x)$ and the *output node* $D(x)$. Strength of the decision rule represents a *throughflow* of the corresponding branch. The throughflow of the graph is governed by formulas (1),...,(6).

Classification of objects in this representation boils down to finding the maximal output flow in the flow graph, whereas explanation of decisions is connected with the maximal input flow associated with the given decision.

A flow graph for decision table shown in Table 1 is given in Figure 1.

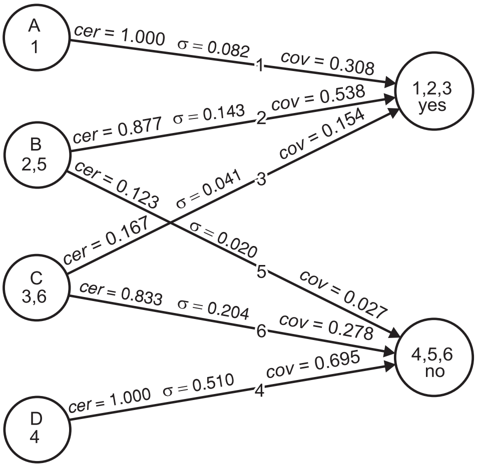

**Fig. 1.** Flow graph

## 6 Decision Space

With every decision table having one $n$-valued decision attribute we can associate $n$-dimensional Euclidean space, where decision granules determine $n$ axis of the space and condition granules determine points of the space. Strengths of decision rules are to be understood as coordinates of corresponding granules.

Distance $\delta(x, y)$ between granules $x$ and $y$ in the $n$-dimensional decision space is defined as

$$\delta(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_j)^2}$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are vectors of strengths of corresponding decision rules.

A decision space for Table 1 is given in Figure 2.

Distances between granules A, B, C and D are shown in Table 3.

**Table 3.** Distance matrix

| | A | B | C | D |
|---|---|---|---|---|
| A | | | | |
| B | 0.064 | | | |
| C | 0.208 | 0.210 | | |
| D | 0.517 | 0.510 | 0.309 | |

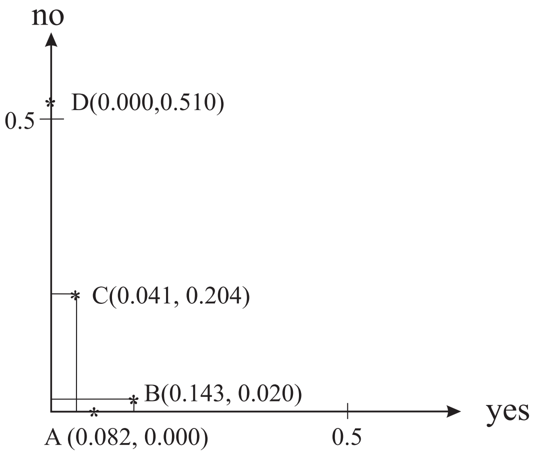

**Fig. 2.** Decision space

## 7 Flow Diagrams, Another Approach

Flow diagrams can be also employed without referring to decision tables, but using other kind of information about the problem we are interested in. We will consider here two cases. In the first case, called in the classical flow network terminology *supply-demand* problem [1], we are given demand of some commodities and we want to find supply of components necessary to produce the commodities. The second case, which will be considered here, is in some sense inverse.

For the sake of simplicity we will explain the problem by means of a simple example for paint demand in a car factory.

Suppose that cars are painted into two colors $Y_1$ and $Y_2$ and that these colors can be obtained by mixing three paints $X_1$, $X_2$ and $X_3$ in the following proportions:

- $Y_1$ contains 20% of $X_1$, 70% of $X_2$ and 10% of $X_3$,
- $Y_2$ contains 30% of $X1$, 50% of $X_2$ and 20% of $X_3$.

We have to find demand of each paint and their distribution among colors $Y_1$ and $Y_2$.

Employing terminology introduced in previous sections we can represent our problem by means of flow graph shown in Figure 3. Thus in order to solve our task first we have to compute strength of each decision rule using formula (9). Next applying formula (4) to each $X_i$ we obtain demand of each paint. Finally, employing formula (5) we get the distribution of each paint among colors of cars.

The final result is presented in Figure 4.

For the sake of simplicity we will use the same numerical data to illustrate the inverse problem. Suppose we want to know distribution of votes of three disjoint group $X_1$, $X_2$ and $X_3$ of voters among two political parties $Y_1$ and $Y_2$ assuming now that we are given data, as shown in Figure 5.

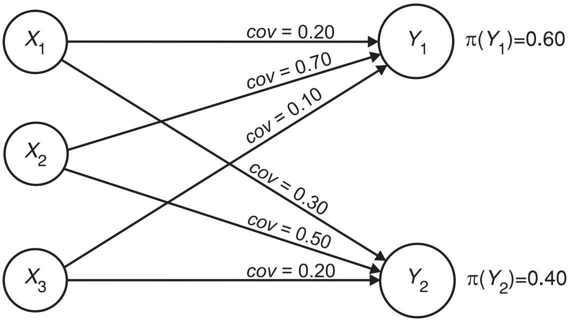

**Fig. 3.** Supply – demand

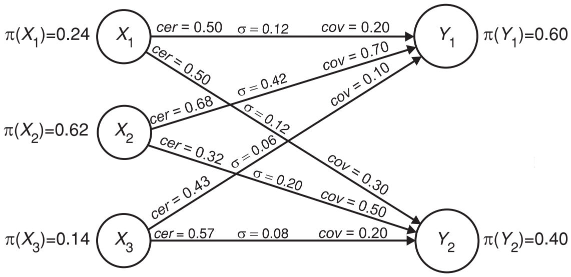

**Fig. 4.** Final results

That is the group $X_1$ consists of 24% of voters, $X_2$ - 62% and $X_3$ - 14%. Votes distribution among parties is as follows:

- group $X_1$ gave 50% of its votes for each party,
- group $X_2$ gave 68% of votes for party $Y_1$ and 32% for party $Y_2$,
- group $X_3$ gave 43% votes for party $Y_1$ and 57% votes for party $Y_2$.

Proceeding in the inverse order as in the previous example we get the final results shown in Figure 4.

That is, first we apply formula (10) and compute strength of each decision rule. Having done this we use formula (6) and compute coverage factors of each decision rule. Next applying formula (3) we obtain the final results, i.e., party $Y_1$ obtained 60% votes, whereas party $Y_2$ obtained 40% votes. Votes distribution for each party is as follows:

- party $Y_1$ obtained 20% votes from group $X_1$, 70% from group $X_2$ and 10% from group $X_3$,
- party $Y_2$ obtained 30% votes from group $X_1$, 50% from group $X_2$ and 20% from group $X_3$.

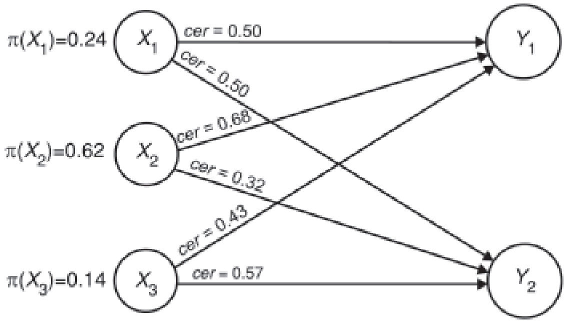

**Fig. 5.** Inverse problem

## 8 Conclusions

Decision tables display interesting probabilistic features, i.e., the obey the total probability theorem and Bayes' theorem. This gives rise to a new perspective on Bayesian inference methodology, leads to new algorithms and new areas of applications.

Furthermore, representation of decision tables by flow graphs and decision spaces gives new insight into the data analysis processes.

## References

1. Ford, L.R., Fulkerson, D.R.: Flows in Networks, Princeton University Press, Princeton. New Jersey
2. Pawlak, Z.: Theorize with Data using Rough Sets (to appear)
3. Pawlak, Z.: Rough Sets, Decision Algorithms and Bayes' Theorem. European Journal of Operational Research 136 (2002) 181–189
4. Polkowski, L., Skowron, A., (eds.).: Rough Set and Current Trends in Computing. Lecture Notes in Artificiale Intelligence 1424, Springer (1998)
5. Polkowski, L., Skowron, A. (eds.).: Rough Sets in Knowledge Discovery. Vol. 1-2, Physica Verlag, Springer (1998)
6. Polkowski, L., Tsumoto, S., Lin, T.Y., (eds.).: Rough Set Methods and applications – New Developments in Knowledge Discovery in Information Systems. Springer (to appear)
7. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems, in: Słowiński (ed.), Intelligent Decision Support, – Handbook of Applications and Advances in Rough Set Theory, Kluwer Academic Publishers, Dordrech (1992) 311–362
8. Tsumoto, S., Tanaka, H.: Discovery of Functional Components of Proteins Based on PRIMEROSE and Domain Knowledge Hierarchy. Proceedings of the Workshop on Rough Sets and Soft Computing (RSSC-94), 1994: Lin, T.Y., and Wildberger, A.M. (eds.) Soft Computing, SCS (1995) 280–285