Generalized Decision Algorithms, Rough Inference Rules, and Flow Graphs

Salvatore Greco[1], Zdzisław Pawlak[2], and Roman Słowiński[3]

[1] Faculty of Economics, University of Catania,
Corso Italia, 55, 95129 Catania, Italy
salgreco@mbox.unict.it
[2] Institute of Theoretical and Applied Informatics, Polish Academy of Sciences,
Baltycka 5, 44-100 Gliwice, Poland
zpw@ii.pw.edu.pl
[3] Institute of Computing Science, Poznan University of Technology,
Piotrowo 3a, 60-965 Poznan, Poland
slowinsk@sol.put.poznan.pl

Abstract. Some probabilistic properties of decision algorithms composed of "*if*..., *then*..." decision rules are considered. With every decision rule three probabilities are associated: the *strength*, the *certainty* and the *coverage* factors of the rule. It has been shown previously that the certainty and the coverage factors are linked by Bayes' theorem. Bayes' theorem has also been presented in a simple form employing the strength of decision rules. In this paper, we relax some conditions on the decision algorithm, in particular, a condition on mutual exclusion of decision rules, and show that the former properties still hold. We also show how the total probability theorem is related with *modus ponens* and *modus tollens* inference rules when decision rules are true in some degree of the certainty factor. Moreover, we show that under the relaxed condition, with every decision algorithm a flow graph can be associated, giving a useful interpretation of decision algorithms.

1 Introduction

We are considering some probabilistic properties of decision algorithms being finite sets of "*if*..., *then*..." decision rules. The rules are induced from a data table where a finite set of objects is described by a finite set of condition and decision attributes. With every decision rule three probabilities are associated: the *strength*, the *certainty* and the *coverage* factors of the rule. Pawlak (2002a) has shown that the certainty and the coverage factors are linked by Bayes' theorem. Moreover, Bayes' theorem in the proposed setting can be presented in a simple form employing the strength of decision rules. These properties have been derived under specific conditions imposed on decision rules, in particular, a mutual exclusion (or independence) condition. In this paper we relax these conditions and show that the former properties still hold. We also show how the total probability theorem is related with *modus ponens* and *modus tollens*

J.J. Alpigini et al. (Eds.): RSCTC 2002, LNAI 2475, pp. 93–104, 2002.

inference rules when decision rules are true in degree of the certainty factor and the decision algorithm satisfies the relaxed conditions.

Moreover, we show that under the relaxed conditions, with every decision algorithm a flow graph can be associated. The through-flow in the graph is related to above-mentioned probabilities and is ruled by the total probability theorem and Bayes' theorem. The flow graph satisfies the usual properties of network flows, i.e. conservation of flow in each node and in the whole network. Simple tutorial examples illustrate the interest of the flow graph for practical interpretation of decision algorithms.

2 Decision Rules and Decision Algorithms

Let S = (U, A) be an information system, where U and A are finite, non-empty sets called the *universe* and the set of *attributes*, respectively. If in the set A two disjoint classes of attributes, called *condition* and *decision attributes*, are distinguished, then the system is called a *decision table* and is denoted by S = (U, C, D), where C and D are sets of condition and decision attributes, respectively. With every subset of attributes, one can associate a formal language of formulas **L** defined in a standard way and called the *decision language*. Formulas for a subset B⊆A are build up from attribute-value pairs (a, v), where a∈B and v∈V_a (set V_a is domain of a), by means of logical connectives $\wedge$ (*and*), $\vee$ (*or*), $\neg$ (*not*). We assume that the set of all formula sets in **L** is partitioned into two classes, called *condition* and *decision formulas*, involving condition and decision attributes, respectively.

A *decision rule* induced from S and expressed in **L** is an implication $\Phi \rightarrow \Psi$, read "*if* Φ, *then* Ψ", where Φ and Ψ are condition and decision formulas in **L**, respectively.

Let $||\Phi||$ denote the set of all objects from universe U, having the property Φ in S.

If $\Phi \rightarrow \Psi$ is a decision rule, then $supp_S(\Phi,\Psi) = \text{card}(||\Phi \wedge \Psi||)$ will be called the *support* of the decision rule and $\sigma_S(\Phi,\Psi) = \frac{supp_S(\Phi,\Psi)}{\text{card}(U)}$ will be referred to as the *strength* of the decision rule.

With every decision rule $\Phi \rightarrow \Psi$ we associate a *certainty factor* $cer_S(\Phi,\Psi) = \frac{supp_S(\Phi,\Psi)}{\text{card}(||\Phi||)}$ and a *coverage factor* $cov_S(\Phi,\Psi) = \frac{supp_S(\Phi,\Psi)}{\text{card}(||\Psi||)}$.

If $cer_S(\Phi,\Psi)=1$, then the decision rule $\Phi \rightarrow \Psi$ will be called *certain*, otherwise the decision rule will be referred to as *uncertain*.

A set of decision rules covering all objects of the universe U creates a *decision algorithm* in S. Pawlak (2002a) points out that every decision algorithm associated with S displays well-known probabilistic properties, in particular it satisfies the total probability theorem and Bayes' theorem. As a decision algorithm can also be interpreted in terms of the rough set concept, these properties give a new look on Bayes' theorem from the rough set perspective. In consequence, one can draw conclusions from data without referring to prior and posterior probabilities, inherently associated with Bayesian reasoning. The revealed relationship can be

used to invert decision rules, i.e., giving reasons (explanations) for decisions, which is useful in decision analysis.

The relationship revealed by Pawlak (2002a) uses, however, some restrictive assumptions that we want to relax in the present study.

3 Some Properties of Decision Algorithms

Pawlak (2002a) defines the decision algorithm as a set of decision rules $Dec_S(\Phi,\Psi) = \{\Phi_i \to \Psi_i\}_{i=1,\ldots,m}$, m≥2, associated with a decision table S = (U, C, D), satisfying the following conditions:

1. *Mutual exclusion* (*independence*):
   for every $\Phi_i \to \Psi_i$ and $\Phi_j \to \Psi_j \in Dec_S(\Phi,\Psi)$, $\Phi_i = \Phi_j$ or $||\Phi_i \wedge \Phi_j|| = \emptyset$, and $\Psi_i = \Psi_j$ or $||\Psi_i \wedge \Psi_j|| = \emptyset$,
2. *Admissibility*: $supp_S(\Phi,\Psi) \neq \emptyset$ for any $\Phi \to \Psi \in Dec_S(\Phi,\Psi)$,
3. *Covering*: $\bigcup_{i=1}^{m} \|\Phi_i\|$=U and $\bigcup_{i=1}^{m} \|\Psi_i\|$=U.

Under these conditions, the following properties of decision algorithms hold:

$$\sum_{\Psi' \in D(\Phi)} cer_S(\Phi,\Psi') = \sum_{\Psi' \in D(\Phi)} \frac{\text{card}(\|\Phi \wedge \Psi'\|)}{\text{card}(\|\Phi\|)} = 1, \tag{1}$$

$$\sum_{\Phi' \in C(\Psi)} cov_S(\Phi',\Psi) = \sum_{\Phi' \in C(\Psi)} \frac{\text{card}(\|\Phi' \wedge \Psi\|)}{\text{card}(\|\Psi\|)} = 1, \tag{2}$$

$$\pi_S(\Psi) = \sum_{\Phi' \in C(\Psi)} cer_S(\Phi',\Psi)\pi_S(\Phi') = \sum_{\Phi' \in C(\Psi)} \sigma_S(\Phi',\Psi), \tag{3}$$

$$\pi_S(\Phi) = \sum_{\Psi' \in D(\Phi)} cov_S(\Phi,\Psi')\,\pi_S(\Psi') = \sum_{\Psi' \in D(\Phi)} \sigma_S(\Phi,\Psi'), \tag{4}$$

$$cer_S(\Phi,\Psi) = \sigma_S(\Phi,\Psi)/\pi_S(\Phi), \tag{5}$$

$$cov_S(\Phi,\Psi) = \sigma_S(\Phi,\Psi)/\pi_S(\Psi) \quad , \tag{6}$$

where $\pi_S(\Phi) = \frac{\text{card}(\|\Phi\|)}{\text{card}(U)}$, $\pi_S(\Psi) = \frac{\text{card}(\|\Psi\|)}{\text{card}(U)}$, while
$D(\Phi) = \{\Psi\colon \Phi \to \Psi \in Dec_S(\Phi,\Psi)\}$ and $C(\Psi) = \{\Phi\colon \Phi \to \Psi \in Dec_S(\Phi,\Psi)\}$ denote the set of all decisions of Φ and the set of all conditions of Ψ in $Dec_S(\Phi,\Psi)$, respectively.

It can be observed that (3) and (4) refer to the total probability theorem, whereas (5) and (6) refer to Bayes' theorem, without using prior and posterior probabilities. In other words, if we know the ratio of Φ_S in Ψ, thanks to Bayes' theorem we can compute the ratio of Ψ_S in Φ.

We want to generalize formulae (1)–(6) to the case where condition 1) on mutual exclusion (independence) of the decision rules in the decision algorithm is not satisfied. This relaxation means that there may exist at least two decision rules Φ'$\rightarrow \Psi$' and Φ"$\rightarrow \Psi$"$\in Dec(\Phi,\Psi)$ such that $||\Phi' \wedge \Phi''|| \neq \emptyset$ or $||\Psi' \wedge \Psi''|| \neq \emptyset$.

We claim that if the independence condition does not hold with respect to decisions (i.e. if there exist at least two decisions ψ' and ψ" such that $||\psi' \wedge \psi''|| \neq \emptyset$), formula (1) becomes:

$$\sum_{\Psi' \in D(\Phi)} cer_S(\Phi, \Psi') - \sum_{\Psi', \Psi'' \in D(\Phi)} cer_S(\Phi, \Psi' \wedge \Psi'') +$$
$$+ \sum_{\Psi', \Psi'', \Psi''' \in D(\Phi)} cer_S(\Phi, \Psi' \wedge \Psi'' \wedge \Psi''') + ... =$$
$$= \sum_{\Psi' \in D(\Phi)} \frac{\mathrm{card}(\|\Phi \wedge \Psi'\|)}{\mathrm{card}(\|\Phi\|)} - \sum_{\Psi'', \Psi' \in D(\Phi)} \frac{\mathrm{card}(\|\Phi \wedge \Psi' \wedge \Psi''\|)}{\mathrm{card}(\|\Phi\|)} +$$
$$+ \sum_{\Psi''', \Psi'', \Psi' \in D(\Phi)} \frac{\mathrm{card}(\|\Phi \wedge \Psi' \wedge \Psi'' \wedge \Psi'''\|)}{\mathrm{card}(\|\Phi\|)} \ldots = 1. \quad (1')$$

Analogously, if independence condition does not hold with respect to conditions (i.e. if there exist at least two conditions Φ' and Φ" such that $||\Phi' \wedge \Phi''|| \neq \emptyset$), formula (2) becomes:

$$\sum_{\Phi' \in C(\Psi)} cov_S(\Phi', \Psi) - \sum_{\Phi', \Phi'' \in C(\Psi)} cov_S(\Phi' \wedge \Phi'', \Psi) +$$
$$+ \sum_{\Phi', \Phi'', \Phi''' \in C(\Psi)} cov_S(\Phi' \wedge \Phi'' \wedge \Phi''', \Psi) \ ... =$$
$$= \sum_{\Phi' \in C(\Psi)} \frac{\mathrm{card}(\|\Phi' \wedge \Psi\|)}{\mathrm{card}(\|\Psi\|)} - \sum_{\Phi', \Phi'' \in C(\Psi)} \frac{\mathrm{card}(\|\Phi' \wedge \Phi'' \wedge \Psi\|)}{\mathrm{card}(\|\Psi\|)} +$$
$$+ \sum_{\Phi', \Phi'', \Phi''' \in C(\Psi)} \frac{\mathrm{card}(\|\Phi' \wedge \Phi'' \wedge \Phi''' \wedge \Psi\|)}{\mathrm{card}(\|\Psi\|)} ... = 1 \quad (2')$$

Similar transformation can be performed on formula (3):

$$\pi_S(\Psi) = \sum_{\Phi' \in C(\Psi)} cer_S(\Phi', \Psi)\pi_S(\Phi') - \sum_{\Phi', \Phi'' \in C(\Psi)} cer_S(\Phi' \wedge \Phi'', \Psi)\pi_S(\Phi' \wedge \Phi'') +$$
$$+ \sum_{\Phi', \Phi'', \Phi''' \in C(\Psi)} \mathrm{cer}_S(\Phi' \wedge \Phi'' \wedge \Phi''', \Psi)\pi_S(\Phi' \wedge \Phi'' \wedge \Phi''') ... =$$
$$= \sum_{\Phi' \in C(\Psi)} \sigma_S(\Phi', \Psi) - \sum_{\Phi', \Phi'' \in C(\Psi)} \sigma_S(\Phi' \wedge \Phi'', \Psi) +$$
$$+ \sum_{\Phi', \Phi'', \Phi''' \in C(\Psi)} \sigma_S(\Phi' \wedge \Phi'' \wedge \Phi''', \Psi) ... \quad (3')$$

and on formula (4)

$$\pi_S(\Phi) = \sum_{\Psi' \in D(\Phi)} cov_S(\Phi, \Psi')\pi_S(\Psi') - \sum_{\Psi', \Psi'' \in D(\Phi)} cov_S(\Phi, \Psi' \wedge \Psi'')\pi_S(\Psi' \wedge \Psi'') +$$
$$+ \sum_{\Psi', \Psi'' \in D(\Phi)} cov_S(\Phi, \Psi' \wedge \Psi'' \wedge \Psi''')\pi_S(\Psi' \wedge \Psi'' \wedge \Psi''') \ ... =$$
$$= \sum_{\Psi' \in D(\Phi)} \sigma_S(\Phi, \Psi') - \sum_{\Psi', \Psi'' \in D(\Phi)} \sigma_S(\Phi, \Psi' \wedge \Psi'') +$$
$$+ \sum_{\Psi', \Psi'', \Psi''' \in D(\Phi)} \sigma_S(\Phi, \Psi' \wedge \Psi'' \wedge \Psi''') \ + ... \quad (4')$$

4 Total Probability Theorems and Rough Inference Rules

Remark that formulae (3') and (4') referring to total probability theorems are closely related with *modus ponens* (MP) and *modus tollens* (MT) inference rules in some specific way.

Classically, MP has the following form:

if	$\Phi \rightarrow \psi$	is true
and	Φ	is true
then	ψ	is true

If we replace truth values by corresponding probabilities, we can generalize the inference rule as *rough modus ponens* (RMP):

if	$\Phi \rightarrow \psi$	is true with probability $cer_S(\Phi,\psi)$
and	Φ	is true with probability $\pi_S(\Phi)$
then	ψ	is true with probability $\pi_S(\Psi)$ given by (3').

RMP enables us to calculate the probability of conclusion ψ of a decision rule $\Phi \rightarrow \psi$ in terms of strengths of all decision rules in the form $\Phi' \rightarrow \psi$, $\Phi' \wedge \Phi'' \rightarrow \psi$, $\Phi' \wedge \Phi'' \wedge \Phi''' \rightarrow \psi$ and so on. In comparison with the *rough modus ponens* of Pawlak (2002a), the above RMP handles a set of rules suggesting the same decision and such that the intersection of supports of their condition parts can be non-empty.

Classically, MT has the following form:

if	$\Phi \rightarrow \psi$	is true
and	$\neg\psi$	is true
then	$\neg\Phi$	is true

If we replace truth values by corresponding probabilities, we can generalize the inference rule as *rough modus tollens* (RMT):

if	$\Phi \rightarrow \psi$	is true with probability $cer_S(\Phi,\psi)$
and	ψ	is true with probability $\pi_S(\psi)$
then	Φ	is true with probability $\pi_S(\Phi)$ given by (4').

RMT enables us to calculate the probability of condition Φ of a decision rule $\Phi \rightarrow \psi$ in terms of strengths of all decision rules in the form $\Phi \rightarrow \psi'$, $\Phi \rightarrow \psi' \wedge \psi''$, $\Phi \rightarrow \psi' \wedge \psi'' \wedge \psi'''$ and so on. Again, in comparison with the *rough modus tollens* of Pawlak (2002a), the above RMT handles a set of rules having the same condition and such that the intersection of supports of their decision parts can be non-empty.

5 Decision Algorithms and Flow Graphs

Pawlak (2002b, 2002c) has shown recently that a decision algorithm (decision table) can be represented by a flow graph in which the flow is ruled by the total

probability theorem and by the Bayes' theorem. The graph is acyclic, directed and connected; there are two layers of nodes – input nodes, corresponding to particular conditions of decision rules, and output nodes, corresponding to decisions of particular decision rules. To every decision rule $\Phi \rightarrow \psi$ there is assigned an arc connecting the input node Φ and the output node ψ. Strength of the decision rule represents the through-flow of the corresponding arc. The through-flow of the graph is governed by formulas (1)-(6) that can be considered as flow conservation equations. In particular, formula (3) states that the outflow of the output node amounts to the sum of its inflows, whereas formula (4) says that the sum of outflows of the input node equals to its inflow. Moreover, formulas (5) and (6) reveal how through-flow in the flow graph is distributed between its inputs and outputs.

In this section we propose an interpretation of the decision table in terms of the flow graph when the independence condition does not hold.

The generalized flow graph will be explained using two simple examples inspired by Berthold and Hand (1999) and Pawlak (2002b).

Table 1. Statistical summary of a sample of cases

Fact	T_1	T_2	D	Number of cases
1	-	-	-	9320
2	-	-	+	200
3	+	-	-	150
4	+	-	+	20
5	-	+	-	5
6	-	+	+	140
7	+	+	-	5
8	+	+	+	300

Example 1. Consider two physician's diagnostic tests T_1 and T_2, for presence of disease D. Table 1 presents the results of the tests and the presence or absence of the disease on a sample of 10140 cases (660 with and 9480 without disease D). Table 1 represents decision table S.

One can induce from Table 1 a set of decision rules relating results of the tests with the presence of disease D. The rules are presented in Table 2 using the following notation: 1 means positive and -1 means negative result of the corresponding test, and 0 means that the corresponding test is not considered. For example, rule #2 can be read as: "in 97.8% of cases in which test T_2 is positive, disease D is present". Analogously, rule #5 can be read as: "in 11.8% of cases in which test T_1 is positive and test T_2 is negative, disease D is present".

As the ratio of the number of cases with disease D to the total number of cases in the sample is 0.065, there is also a "default" rule #0 having the following interpretation: "without considering any test, in 6.5% of cases disease D is present". Another decision rule is also interesting for interpretation in terms of flow graph:

Rule #6: "in 74.2% of cases in which test T_1 or test T_2 is positive, disease D is present" (coverage=0.697, support=460, strength=.045).

Table 2. Decision rules concluding the presence of disease D, induced from Table 1

Rule	T_1	T_2	D	Certainty factor	Coverage	Support	Strength
Rule #1	1	0	+	.674	.485	320	.032
Rule #2	0	1	+	.978	.667	440	.043
Rule #3	1	1	+	.983	.454	300	.030
Rule #4	1	-1	+	.118	.061	20	.002
Rule #5	-1	1	+	.966	.212	140	.014

The flow graph corresponding to decision algorithm composed of the six decision rules is presented in Figure 1. The graph is composed of three input nodes, corresponding to performed tests (T_1 alone, T_2 alone, T_1 and T_2 together) and of one output node corresponding to the presence of disease D. The input and the output nodes have circular shapes in the flow graph, while the rectangular boxes on the arcs include information on the through-flow of the arcs. Let us remark that the through-flows of the arcs in Figure 1 represent the strength of the corresponding decision rules. In other words, the flow graph can be seen as a decomposition of the output flow, equal to probability $\pi_S(\mathrm{D})=\frac{\mathrm{card}(\|D\|)}{\mathrm{card}(U)}$ of disease D in S, into subsets of patients supporting particular decision rules. This decomposition is as follows.

The flow leaving node T_1 and entering node D is equal to the strength of Rule #1:

$$\sigma(\#1) = \sigma(T_1, D) = \frac{\mathrm{card}(\|T_1 \wedge D\|)}{\mathrm{card}(U)}.$$

It represents the contribution to $\pi_S(\mathrm{D})$ of a subset of patients having positive result of test T_1 (with no regard to test T_2 that may give positive or negative result). Analogously, the flow leaving node T_2 and entering node D is equal to the strength of Rule #2:

$$\sigma(\#2) = \sigma(\mathrm{T}_2,\mathrm{D}) = \frac{\mathrm{card}(\|T_2 \wedge D\|)}{\mathrm{card}(U)}.$$

It represents the contribution to $\pi_S(\mathrm{D})$ of a subset of patients having positive result of test T_2 (with no regard to test T_1 that may give positive or negative result).

Since the subset of patients having simultaneously positive result of T_1 and T_2 is included in both $\sigma(\#1)$ and $\sigma(\#2)$, to obtain $\pi_S(\mathrm{D})$ its contribution must be subtracted from the sum of $\sigma(\#1)$ and $\sigma(\#2)$. The subtraction is represented by the flow leaving node D and entering node (T_1,T_2), having strength of Rule #3:

$$\sigma(\#3) = \sigma(\mathrm{T}_1 \wedge \mathrm{T}_2,\mathrm{D}) = \frac{\mathrm{card}(\|T_1 \wedge T_2 \wedge D\|)}{\mathrm{card}(U)},$$

It represents the contribution to $\pi_S(\mathrm{D})$ of a subset of patients having simultaneously positive result of T_1 and T_2.

Thus, the algebraic sum $\sigma(\#1)+\sigma(\#2)-\sigma(\#3)$ is equal to the probability $\pi_S(\mathrm{D})$ corresponding to the output flow. Remark that the output flow is equal to the strength of the most general decision rule in S, i.e. Rule #6:

$\sigma(\#6) = \sigma(\mathrm{T}_1 \vee \mathrm{T}_2, \mathrm{D}) = \frac{\mathrm{card}(\|T_1 \vee T_2 \wedge D\|)}{\mathrm{card}(U)}$.

The output flow is returned to the input nodes of the graph, giving another decomposition of the probability $\pi_S(\mathrm{D})$. In fact, the output flow is split among:

- the input node T_1, in amount equal to the strength of Rule #4:
  $\sigma(\#4) = \sigma(\mathrm{T}_1 \wedge \neg \mathrm{T}_2, \mathrm{D}) = \frac{\mathrm{card}(\|T_1 \wedge \neg T_2 \wedge D\|)}{\mathrm{card}(U)}$;
  it represents the contribution to $\pi_S(\mathrm{D})$ of a subset of patients having positive result of T_1 and negative result of T_2;
- the input node T_2, in amount equal to the strength of Rule #5:
  $\sigma(\#5) = \sigma(\neg \mathrm{T}_1 \wedge \mathrm{T}_2, \mathrm{D}) = \frac{\mathrm{card}(\|\neg T_1 \wedge T_2 \wedge D\|)}{\mathrm{card}(U)}$;
  it represents the contribution to $\pi_S(\mathrm{D})$ of a subset of patients having negative result of T_1 and positive result of T_2;
- the input node $(\mathrm{T}_1,\mathrm{T}_2)$, in amount equal to the strength of Rule #3, representing the contribution to $\pi_S(\mathrm{D})$ of a subset of patients having simultaneously positive result of T_1 and T_2.

The balance of flows in the input node T_1 can be interpreted as follows. The contribution to $\pi_S(\mathrm{D})$ of a subset of patients with positive result of T_1 and negative result of T_2 is equal to the difference between flow $\sigma(\#1)$ from node T_1 to node D, representing the contribution to $\pi_S(\mathrm{D})$ of a subset of patients with positive result of T_1, and flow $\sigma(\#3)$ from node $(\mathrm{T}_1,\mathrm{T}_2)$ to node D, representing the contribution to $\pi_S(\mathrm{D})$ of a subset of patients having simultaneously positive result of T_1 and T_2. This follows from the observation that $||\mathrm{T}_1|| - ||\mathrm{T}_1 \wedge \mathrm{T}_2|| = ||\mathrm{T}_1 \wedge \neg \mathrm{T}_2||$.

Analogously, the contribution to $\pi_S(\mathrm{D})$ of a subset of patients with positive result of T_2 and negative result of T_1 is equal to the difference between flow $\sigma(\#2)$ from node T_2 to node D, representing the contribution to $\pi_S(\mathrm{D})$ of a subset of patients with positive result of T_2, and flow $\sigma(\#3)$ from node $(\mathrm{T}_1,\mathrm{T}_2)$ to node D, representing the contribution to $\pi_S(\mathrm{D})$ of a subset of patients having simultaneously positive result of T_1 and T_2. This follows from the observation that $||\mathrm{T}_2|| - ||\mathrm{T}_1 \wedge \mathrm{T}_2|| = ||\neg \mathrm{T}_1 \wedge \mathrm{T}_2||$.

The flow graphs representing the coverage and the support of the rules have the same structure and the through-flows of the arcs are proportional to those in Figure 1. Indeed, given rule $\Phi \rightarrow \Psi$, coverage $cov(\Phi,\Psi)$ and support $supp(\Phi,\Psi)$ can be calculated from the strength by a linear transformation:

$cov(\Phi,\Psi) = \sigma(\Phi,\Psi)\pi_S(\Psi)$ and $supp(\Phi,\Psi) = \sigma(\Phi,\Psi)\mathrm{card}(\mathrm{U})$.

Let us remark that the graph presented in Figure 1 satisfies two important properties of the flow graphs: (i) at each node of the flow graph there is a zero algebraic sum of the inflow and the outflow; (ii) the sum of flows entering the input nodes is equal to the sum of flows leaving the output node.

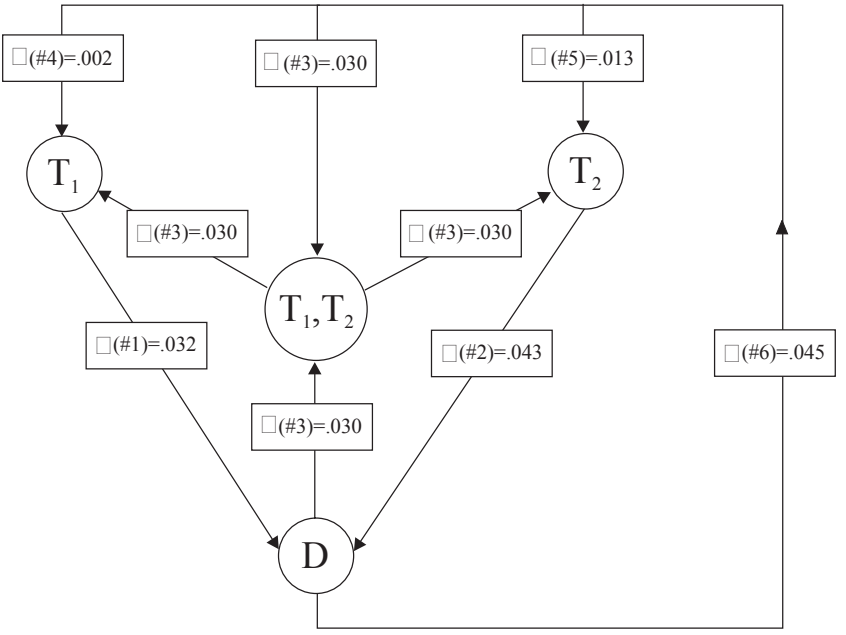

Fig. 1. The flow graph of the decision algorithm representing a decomposition of the output flow, equal to probability $\pi_S(\mathrm{D})$, in terms of the strength of particular decision rules.

Example 2. While in Example 1 we considered strength, coverage and support of decision rules, the present example underlines the interest of the flow graph in representation of the certainty factor.

Consider one physician's diagnostic test T, for presence of two diseases D_1 and D_2. Table 3 presents the results of the test and the presence or absence of the diseases on a sample of 3850 cases (all with positive result of test T). Table 3 represents decision table S.

Table 3. Statistical summary of a sample of cases

Fact	T	D_1	D_2	Number of cases
1	+	-	-	200
2	+	+	-	400
3	+	-	+	250
4	+	+	+	3000

One can induce from Table 3 a set of decision rules relating results of the test with the presence or absence of disease D_1 and/or D_2. The rules are presented in Table 4 using the following notation: 1 means presence and -1 means absence of the corresponding disease, and 0 means that the corresponding disease is not considered. For example, rule #1 can be read as: "in 88.3% of cases in which the

result of test T is positive, disease D_1 is present". Analogously, rule #5 can be read as: "in 6.5% of cases in which the result of test T is positive, disease D_1 is present while disease D_2 is absent".

Table 4. Decision rules

Rule	T	D_1	D_2	Certainty factor
Rule #1	+	1	0	.883
Rule #2	+	0	1	.844
Rule #3	+	1	1	.779
Rule #4	+	-1	1	.104
Rule #5	+	1	-1	.065

Another decision rule is also interesting for interpretation in terms of flow graph:

Rule #6: "in 94.8% of cases in which test T is positive, one of the diseases is present".

The flow graph corresponding to decision algorithm composed of the six decision rules is presented in Figure 2. The graph is composed of one input node, corresponding to test T, and of three output nodes corresponding to the presence of diseases (D_1 alone, D_2 alone, and D_1 and D_2 together). The input and the output nodes have again circular shapes in the flow graph, while the rectangular boxes on the arcs include information on the through-flow of the arcs. Let us remark that the through-flows of the arcs in Figure 2 represent the certainty factors of the corresponding decision rules. In other words, the flow graph can be seen as a decomposition of the input flow, equal to certainty factor of the most general decision rule in S, i.e. Rule #6. This decomposition is as follows.

The input flow equal to certainty of Rule #6,

$$cer_S(\#6) = cer_S(\mathrm{T},\mathrm{D}_1\vee\mathrm{D}_2) = \frac{\mathrm{card}(\|T\wedge(D_1\vee D_2)\|)}{\mathrm{card}(\|T\|)}$$

is a sum of output flows corresponding to certainties of Rule #3, Rule #4 and Rule #5, respectively, i.e.

$$cer_S(\mathrm{T},\mathrm{D}_1\vee\mathrm{D}_2) = cer_S(\mathrm{T},\mathrm{D}_1\wedge\mathrm{D}_2) + cer_S(\mathrm{T},\mathrm{D}_1 \wedge \neg\mathrm{D}_2) + cer_S(\mathrm{T},\neg\mathrm{D}_1\wedge\mathrm{D}_2).$$

This is based on the observation that $||\mathrm{T}\wedge(\mathrm{D}_1\vee\mathrm{D}_2)|| = ||\mathrm{T}\wedge\mathrm{D}_1\wedge\mathrm{D}_2|| \cup ||\mathrm{T}\wedge\mathrm{D}_1 \wedge \neg\mathrm{D}_2|| \cup ||\mathrm{T}\wedge\neg\mathrm{D}_1\wedge\mathrm{D}_2||$.

The graph shows also another decomposition of the input flow and, therefore, of the certainty of Rule #6. This new decomposition is as follows.

The input flow, equal to certainty of Rule #6, is a sum of flows leaving node D, i.e. certainties of Rule #1 and Rule #2, minus flows entering node D, i.e. certainty of Rule #3:

$$cer_S(\mathrm{T},\mathrm{D}_1\vee\mathrm{D}_2) = cer_S(\mathrm{T},\mathrm{D}_1) + cer_S(\mathrm{T}, \mathrm{D}_2) - cer_S(\mathrm{T},\mathrm{D}_1\wedge\mathrm{D}_2).$$

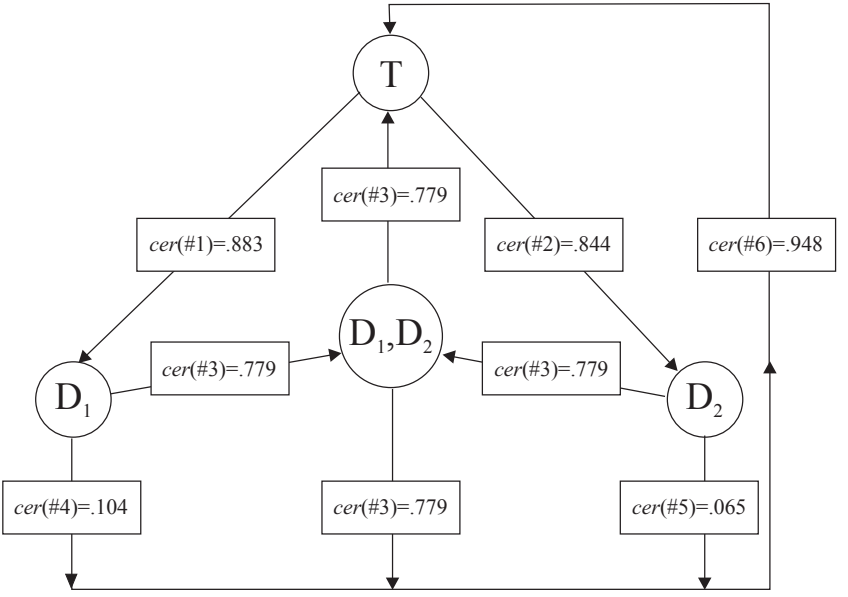

Fig. 2. The flow graph of the decision algorithm representing a decomposition of the input flow in terms of the certainty factor of particular decision rules.

This is based on the observation that $||\mathrm{T}\wedge(\mathrm{D}_1\vee\mathrm{D}_2)||=(||\mathrm{T}\wedge\mathrm{D}_1||\cup||\mathrm{T}\wedge\mathrm{D}_2||)-||\mathrm{T}\wedge\neg\mathrm{D}_1\wedge\mathrm{D}_2||$.

Moreover, in the output node D_1, the certainty factor of Rule #4:

$$cer_S(\#4) = cer_S(\mathrm{T},\mathrm{D}_1 \wedge \neg\mathrm{D}_2) = \frac{\mathrm{card}(\|T\wedge D_1\wedge\neg D_2\|)}{\mathrm{card}(\|T\|)}$$

is equal to the difference between flow

$$cer_S(\#1) = cer_S(\mathrm{T},\mathrm{D}_1) = \frac{\mathrm{card}(\|T\wedge D_1\|)}{\mathrm{card}(\|T\|)},$$

leaving node T and entering node D_1 (Rule #1), and flow

$$cer_S(\#3) = cer_S(\mathrm{T},\mathrm{D}_1\wedge\mathrm{D}_2) = \frac{\mathrm{card}(\|T\wedge D_1\wedge D_2\|)}{\mathrm{card}(\|T\|)},$$

leaving node D_1 and entering node $(\mathrm{D}_1,\mathrm{D}_2)$ (Rule #3). This is based on the observation that $||\mathrm{T}\wedge\mathrm{D}_1|| - ||\mathrm{T}\wedge\mathrm{D}_1\wedge\mathrm{D}_2|| = ||\mathrm{T}\wedge\mathrm{D}_1 \wedge \neg\mathrm{D}_2||$.

Analogously, in the output node D_2, the certainty factor of Rule #5:

$$cer_S(\#5) = cer_S(\mathrm{T},\neg\mathrm{D}_1\wedge\mathrm{D}_2) = \frac{\mathrm{card}(\|T\wedge\neg D_1\wedge D_2\|)}{\mathrm{card}(\|T\|)}$$

is equal to the difference between flow

$$cer_S(\#2) = cer_S(\mathrm{T},\mathrm{D}_2) = \frac{\mathrm{card}(\|T\wedge D_2\|)}{\mathrm{card}(\|T\|)},$$

leaving node T and entering node D_2 (Rule #2), and flow

$$cer_S(\#3) = \frac{\text{card}(\|T \wedge D_1 \wedge D_2\|)}{\text{card}(\|T\|)},$$

leaving node D_1 and entering in node (D_1,D_2) (Rule #3). This is based on the observation that $||T \wedge D_2|| - ||T \wedge D_1 \wedge D_2|| = ||T \wedge \neg D_1 \wedge D_2||$.

Finally, let us remark that the flow graph presented in Figure 2 also satisfies the properties 1) and 2) of the flow graph from Figure 1.

6 Conclusions

This paper shows some interesting probabilistic features of decision rules inferred from data tables. It extends some previous results in this field by relaxing the assumption of mutual exclusion (independence) of decision rules. Due to some interesting theoretical developments, this relaxation enables interpretation of decision rules encountered in real-life applications where the independence property of a decision algorithm is often violated.

The interpretation of the probabilistic features in terms of flow graphs gives an interesting representation of the relations between the strength, support, coverage and certainty of decision rules induced from one data table. This permits the user to have a deeper comprehension of the fundamental relations in the data.

Acknowledgement. The first author wishes to acknowledge financial support from Italian Ministry of Education, University and Scientific Research (MIUR) The research of the third author has been supported by the State Committee for Scientific Research (KBN), research grant no. 8T11F 006 19, and by the Foundation for Polish Science, subsidy no. 11/2001.

References

Berthold, M., Hand, D. J.: *Intelligent data analysis, an introduction*. Springer-Verlag, Berlin, Heidelberg, New York (1999)

Pawlak, Z.: *Rough Sets - Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Boston Dordrecht, London (1991)

Pawlak, Z.: Rough sets, decision algorithm and Bayes' theorem. *European Journal of Operational Research* 136 (2002a) pp. 181–189

Pawlak, Z.: *Bayes' Theorem – the Rough Sets Perspective*. Working paper, Warsaw (2002b)

Pawlak, Z.: *Decision Algorithms, Bayes' Theorem and Flow Graph*. Working paper, Warsaw (2002c)