Decision and flow networks

Zdzisław Pawlak

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

For late Professor Henryk Greniewski, my mentor

Abstract: This paper, which is a continuation of series of authors papers on the relationship between decision algorithms and Bayes' theorem, is related to Łukasiewicz's ideas concerning the relationship between multivalued logic, probability and Bayes' theorem. We proposed in this paper a new mathematical model of a flow network different to that introduced by Ford and Fulkerson. Basically, the model is intended to be used rather as a mathematical model of decision processes than as flow analysis and it concerns rather flow of information than material media. Branches of the network are interpreted as decision rules, whereas the whole network represents a decision algorithm. It is shown that flow in such networks is governed by Bayes' formula. In this case the formula describes deterministic information flow distribution among branches of the network, without referring to its probabilistic character. This leads to a new look on Bayes' formula and many new applications.

Keywords: Decision analysis, Flow networks, Bayes theorem

1. Introduction

This paper is an extension of the article [7] presented at the RSTCT 2002 and is a continuation of ideas presented in author's previous papers on rough sets, Bayes' theorem and decision tables [8].

In [4] a mathematical model of flow in a network has been introduced and studied. The model was intended to capture the nature of flow in transportation or communication network.

In this paper we present another kind of mathematical model for flow networks, which may be interpreted rather as a model of a deterministic, steady state flow in a plumbing network − than a transportation network. Although, essentially the model is intended to be used as a description of decision processes. Branches of the network are interpreted as decision rules, whereas the network is supposed to describe a decision algorithm. It is shown that flow in such a network is governed by Bayes' rule. Furthermore, this interpretation bring to light another understanding of Bayes' rule: the rule may be interpreted entirely in a deterministic way, without referring to its probabilistic nature, inherently associated with classical Bayesian philosophy. This leads to new philosophical and practical consequences. Some of them will be discussed in this paper.

2. Flow graphs

A flow graph is a *directed, acyclic*, *finite* graph $G = (N, \mathcal{B}, \varphi)$, where N is a set of *nodes*, $\mathcal{B} \subseteq N \times N$ is a set of *directed branches, $\varphi : \mathcal{B} \rightarrow R^{+}$* is a *flow function* and R^{+} is the set of non-negative reals.

Input of $x \in N$ *is the set* $I(x)=\{y \in N: (x, y) \in \mathscr{B}\}$; *output* of $x \in N$ is defined as $O(x) = \{y \in N: (x, y) \in \mathscr{B}\}$.

We will also need the concept of *input* and *output* of a graph *G*, defined respectively as follows: $I(G) = \{x \in N : I(x) = \emptyset\}$, $O(G) = \{x \in N : O(x) = \emptyset\}$.

Inputs and outputs of *G* are *external nodes* of *G*; other nodes are *internal nodes* of *G*.

If $(x, y) \in \mathscr{B}$ then $\varphi(x, y)$ is a *troughflow* from *x* to *y*.

With every node of a flow graph we associate its *inflow* and *outflow* defined as $\varphi_+(y) = \sum_{x \in I(y)} \varphi(x, y)$, $\varphi_-(x) = \sum_{y \in O(x)} \varphi(x, y)$ respectively.

Similarly we define an inflow and an outflow for the whole flow graph, which are defined as follows $\varphi_+(G) = \sum_{x \in I(G)} \varphi_-(x)$, $\varphi_-(G) = \sum_{x \in O(G)} \varphi_+(x)$, respectively.

We assume that for any internal node *x*, $\varphi_+(x) = \varphi_-(x) = \varphi(x)$, where $\varphi(x)$ is a *troughflow* of *x*.

Obviously $\varphi_+(G) = \varphi_-(G) = \varphi(G)$, where $\varphi(G)$ is a *troughflow* of *G*.

The above formulas can be considered as *flow conservation equations* [4].

3. Strength, certainty and coverage of flow

With every branch (x, y) we associate its strength defined as $\sigma(x, y) = \frac{\varphi(x, y)}{\varphi(G)}$.

Certainly, $0 \leq \sigma(x, y) \leq 1$ and it can be considered as a normalized flow of the branch (x, y). The strength of a branch expresses simply the percentage of a total flow, flowing trough the branch.

A flow graph can be also interpreted as a decision graph [7]; branches of the flow graph can be then interpreted as decision rules, and strength of a decision rule represents normalized support of the decision rule.

We define now two important coefficients assigned to every branch of a flow graph − the *certainty* and the *coverage* factors.

The *certainty* and the *coverage* of (x, y) are defined as follows $cer(x, y) = \frac{\sigma(x, y)}{\sigma(x)}$, and the *coverage* of (x, y) $cov(x, y) = \frac{\sigma(x, y)}{\sigma(y)}$, respectively, where $\sigma(x)$ is the *normalized troughflow* of *x*, defined as $\sigma(x) = \sum_{y \in O(x)} \sigma(x, y) = \sum_{y \in I(x)} \sigma(y, x)$.

The certainty and the coverage factors have been for a long time used in machine learning and data bases, see e.g. [9,10], but in fact these coefficients have been first used by J. Lukasiewicz in connection with his study of multivalued logic, probability and Bayes' theorem [5].

Both coefficients have a probabilistic flavor, and can be interpreted as a kind of conditional probability, however in the flow graph setting they can be understood entirely in a deterministic way, and denote simply a deterministic distribution of flow between branches of the flow graph. However to stress the relationship of these coefficients to decision analysis we will use notation employed in the above said disciplines.

This paper, which is a continuation of series of authors papers on the relationship between decision algorithms and Bayes' theorem, is related to Lukasiewicz's ideas and not to those considered in machine learning and databases.

4. Properties of flow

The below properties are immediate consequences of definitions given in the preceding sections.

$$\sum_{y \in O(x)} cer(x, y) = 1 \qquad (1)$$

$$\sum_{x \in I(y)} cov(x, y) = 1 \qquad (2)$$

$$cer(x, y) = \frac{cov(x, y)\sigma(y)}{\sigma(x)} \qquad (3)$$

$$cov(x, y) = \frac{cer(x, y)\sigma(x)}{\sigma(y)} \qquad (4)$$

It is easily seen that the above properties have a probabilistic flavor. In particular, equations (3) and (4) are well known Bayes' formulas. However, in our case the properties can be interpreted in entirely deterministic way. They simply describe some features of steady flow in a flow network, i.e., flow distribution among branches in the network.

The properties can be also interpreted as features of decision rules in a decision graph.

Let us also observe that Bayes' formula is, in our setting, expressed by means of a strength coefficient. This leads to very simple computations and give also new insight into the meaning of Bayesian methodology.

5. An example

Suppose that cars are painted in 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 x_1, 50% of x_2 and 20% of x_3.

We have to find demand of each paint and its distribution among colors y_1 and y_2.

Employing terminology introduced in previous section we can represent our problem by means of flow graph shown in Figure 1.

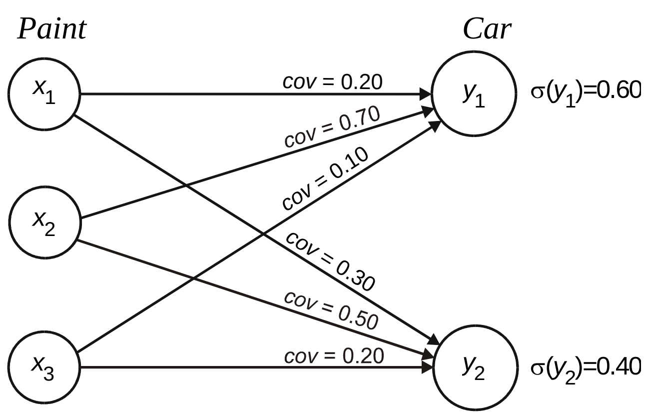

Fig. 1

Thus in order to solve our task first we have to compute strength of each decision rule. Next we compute demand of each paint. Finally, we compute the distribution of each paint among colors of cars.

The final result is presented in Figure 2.

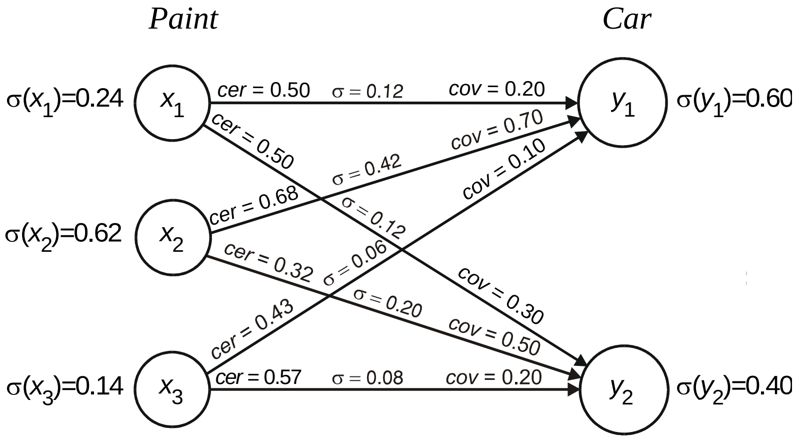

Fig. 2.

Suppose now that the cars are produced by three manufacturers z_1, z_2 and z_3 in proportions as shown in Fig. 3,

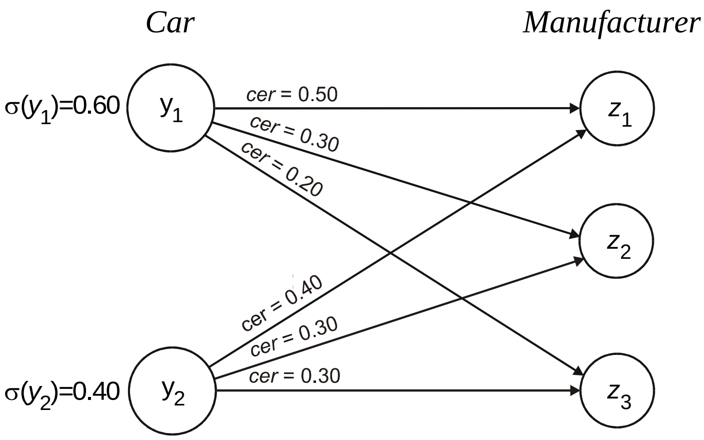

Fig. 3.

i.e.,

– 50% of cars y_1 are produced by manufacturer z_1
– 30 % of cars y_1 are produced by manufacturer z_2
– 20% of cars y_1 are produced by manufacturer z_3

and

– 40% of cars y_2 are produced by manufacturer z_1
– 30% of cars y_2 are produced by manufacturer z_2
– 30% of cars y_2 are produced by manufacturer z_3

Employing technique used previously we can compute car production distribution among manufacturers as shown in Fig. 4, e.g., manufacturer z_1 produces 65% of cars y_1 and 35% of cars y_2, etc. Finally, the manufacturer z_1 produces 46% cars, manufacturers z_2 – 30% cars and manufacturer z_3 – 24% of cars.

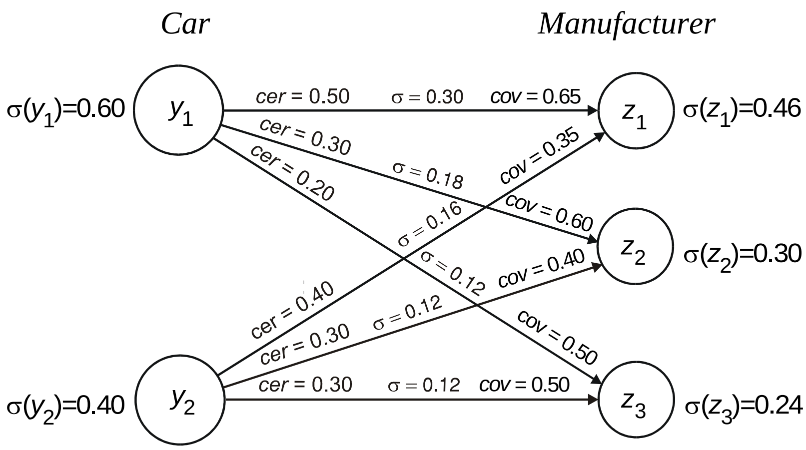

Fig. 4.

We can combine graphs shown in Fig 2 and Fig.4 and we obtain the flow graph shown in Fig.5.

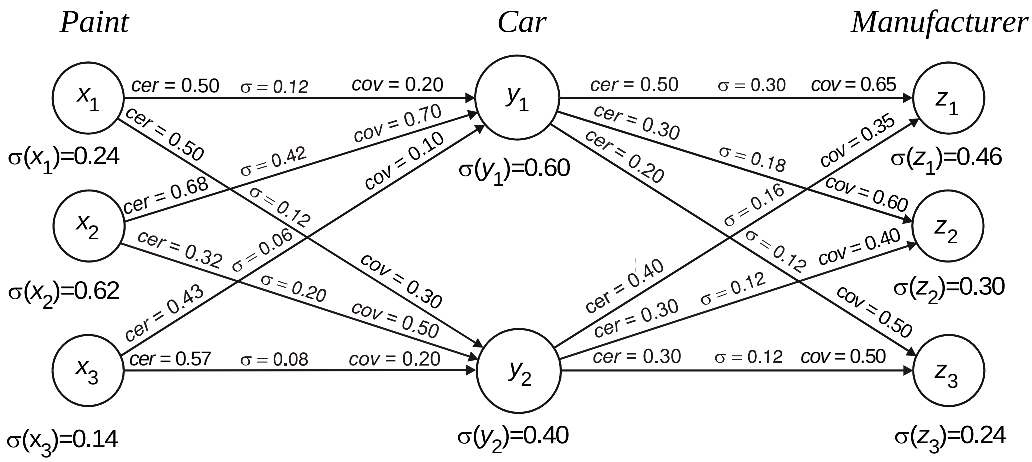

Fig. 5.

This flow graph can be understood as a composition of two flow graphs shown in Fig. 2 and Fig. 3.

The graph shows clearly the flow structure of the whole production process. From this graph it is easily seen how the flow of decisions is structured.

6. Simplification of flow graphs

We can ask what is the paint demand by each manufacturer. In order to answer this question we have to eliminate in the flow graph nodes y_1 and y_2 from the flow graph. To this end we have to replace each sub-graph determined by each pair x, y of nodes such that x and y are input and output of the graph respectively − by a single branch (x, y) which preserves the same flow between point x and y, as the whole sub-graph determined by these nodes.

In order to solve this problem we need some auxiliary notions, which are defined next.

A (*directed*) *path* from x to y, $x \neq y$ denoted $[x, y]$, is a sequence of nodes $x_1,\ldots,x_n$ such that $x_1 = x$, $x_n = y$ and $(x_i, x_{i+1}) \in \mathcal{B}$ for every i, $1 \leq i \leq n\text{-}1$.

Now we extend the concept of certainty, coverage and strength from single branch to a path, as shown below.

The *certainty* of a path $[x_1, x_n]$ is defined as

$$cer[x_1, x_n] = \prod_{i=1}^{n-1} cer(x_i, x_{i+1}),$$

the *coverage* of a path $[x_1, x_n]$ is the following

$$cov[x_1, x_n] = \prod_{i=1}^{n-1} cov(x_i, x_{i+1}),$$

the *strength* of a path $[x, y]$ is $\sigma[x, y] = \sigma(x)\, cer[x, y] = \sigma(y)\, cov[x, y]$.

The set of all paths from x to y ($x \neq y$) denoted $< x, y >$, will be called a *connection* from x to y. In other words, connection $< x, y >$ is a sub-graph determined by nodes x and y.

We will also need extension of the above coefficients for connections (i.e., sub-graphs determined by nodes x and y) as shown in what follows:

The *certainty* of connections $< x, y >$ is

$$cer < x, y > = \sum_{[x,y] \in < x,y >} cer[x, y],$$

the *coverage* of connections is $< x, y >$

$$cov < x, y > = \sum_{[x,y] \in < x,y >} cov[x, y],$$

the *strength* of connections is $< x, y >$

$$\sigma < x, y > = \sum_{[x,y] \in < x,y >} \sigma[x, y].$$

Let x, y ($x \neq y$) be nodes of G. If we substitute the sub-graph $< x, y >$ by a single branch (x, y) such that $\sigma(x, y) = \sigma < x, y >$ then $cer(x, y) = cer < x, y >$, $cov(x, y) = cov < x, y >$ and $\varphi(G) = \varphi(G')$, where G' is the graph obtained from G by substituting $< x, y >$ by (x, y).

The example (cont.)

The final result for the considered example is shown in Fig. 6.

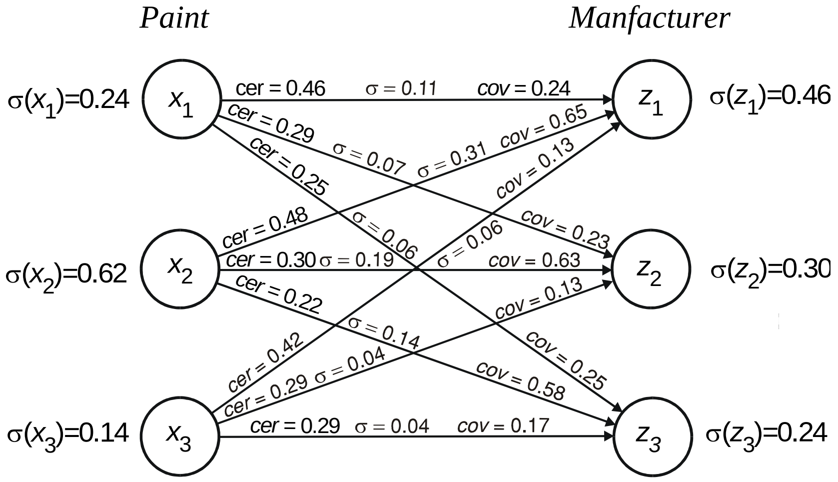

Fig. 6

It is easily seen from the flow graph how paint supply is distributed among manufacturers and what the demand for each paint is by every manufacturer.

For example, supply of paint x_1 is distributed among manufacturers z_1, z_2 and z_3 in the proportions 46%, 29% and 25% respectively, whereas demand for paints x_1, x_2 and x_3 by manufacturer z_1 is 24%, 65% and 13% respectively.

7. Flow graphs and decision tables

With every flow graph we can associate a decision table (and conversely). In particular, if the flow graph has only external nodes (i.e., input and output nodes) we can interpret the input nodes as conditions and output nodes as decision of the decision table. In other words the set of inputs determines a condition attribute, whereas the output nodes determine the decision attribute.

The example (cont.).

For example the flow graph presented in Fig. 5 can be depicted by two decision tables given in Table 1 and Table 2, respectively.

	Paint	*Car*	*Strength*
1	x_1	y_1	0.12
2	x_1	y_2	0.12
3	x_2	y_1	0.42
4	x_2	y_2	0.20
5	x_3	y_1	0.06
6	x_3	y_2	0.08

Table 1

	Car	*Manu.*	*Strength*
1	y_1	z_1	0.30
2	y_1	z_2	0.18
3	y_1	z_3	0.12
4	y_2	z_1	0.16
5	y_2	z_2	0.12
6	y_2	z_3	0.12

Table 2

The decision table corresponding to the flow graph shown in Fig. 6 is given in Table 3.

	Paint	*Manu.*	*Strength*
1	x_1	z_1	0.11
2	x_1	z_2	0.07
3	x_1	z_3	0.06
4	x_2	z_1	0.31
5	x_2	z_2	0.19
6	x_2	z_3	0.14
7	x_3	z_1	0.06
8	x_3	z_2	0.04
9	x_3	z_3	0.04

Table 3

This table can be understood as a result of operation performed on the constituent decision tables Table 1 and Table 2. It is somehow similar to the join operation widely used in data bases, however the operation is in our case of course, totally different.

8. Conclusions

We presented in this paper a new approach to flow networks. This approach is basically meant as a new tool to decision analysis. It is also shown that the flow in the flow graph is governed by Bayes' formula, however the meaning of the Bayes' formula has entirely deterministic character and does not refer to any probabilistic interpretation. Thus our approach is entirely free from the mystic flavor of Bayesian reasoning raised by many authors, e.g., [1,2]. Besides, it gives clear interpretation of obtained results and simple computational algorithms.

References

1. Bernardo, J. M., Smith, A. F. M.: Baysian theory, Wiley series in probability and mathematical statistics. John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore (1994)
2. Box, G.E.P., Tiao, G.C.: Bayesiaon inference in statistical analysis. John Wiley and Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore (1992)
3. Berthold, M., Hand, D.J.: Intelligent data analysis, an introduction. Springer-Verlag, Berlin , Heidelberg, New York (1999)
4. Ford, L.R., Fulkerson, D.R.: Flows in Networks, Princeton University Press, Princeton. New Jersey
5. Łukasiewicz, J.: Die logishen Grundlagen der Wahrscheinilchkeitsrechnung. Kraków (1913), in: L. Borkowski (ed.), Jan łukasiewicz - Selected Works, North Holland Publishing Company, Amsterdam, London, Polish Scientific Publishers, Warsaw (1970)
6. Greco, S., Pawlak, Z., Słowiński, R.: Generalized Decision Algoritms, Rough Inference Rules, and Flow Graphs, in: J.J. Alpigini et al. (eds.), Lecture Notes in Artificial Intelligence 2475 (2002) 93-104
7. Pawlak, Z.: In Pursuit of Patterns in Data Reasoning from Data – The Rough Set Way, in: J.J. Alpigini et al. (eds.), Lecture Notes in Artificial Intelligence 2475 (2002) 1-9
8. Pawlak, Z.: Rough Sets, Decision Algorithms and Bayes' Theorem. European Journal of Operational Research 136 (2002) 181-189
9. 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
10. Wong, S.K.M., Ziarko, W.: Algorithm for Inductive Learning. Bull. Polish Academy of Sciences 34, 5-6 (1986) 271-276