

# Inference Rules and Decision Rules

Zdzisław Pawlak

Institute for Theoretical and Applied Informatics  
Polish Academy of Sciences  
ul. Bałtycka 5, 44-100 Gliwice, Poland  
and

University of Information Technology and Management  
ul. Newelska 6, 01-447 Warsaw, Poland  
zpw@ii.pw.edu.pl

**Abstract.** Basic rules of inference used in classical logic are *Modus Ponens* (MP) and *Modus Tollens* (MT). These two reasoning patterns start from some general knowledge about reality, expressed by true implication, "if  $\Phi$  then  $\Psi$ ". Then basing on true *premise*  $\Phi$  we arrive at true *conclusion*  $\Psi$  (MP), or from negation of true conclusion  $\Psi$  we get negation of true premise  $\Phi$  (MT).

In reasoning from data (data mining) we also use rules "if  $\Phi$  then  $\Psi$ ", called *decision rules*, to express our knowledge about reality, but in this case the meaning of the expression is different. It does not express general knowledge but refers to partial facts. Therefore decision rules are not true or false but probable (possible) only.

In this paper we compare inference rules and decision rules in the context of decision networks, proposed by the author as a new approach to analyze reasoning patterns in data.

**Keywords:** *Modus Ponens*, *Modus Tollens*, decision rules

## 1 Introduction

Basic rules of inference used in classical logic are *Modus Ponens* (MP) and *Modus Tollens* (MT). These two reasoning patterns start from some general knowledge about reality, expressed by true implication, "if  $\Phi$  then  $\Psi$ ". Then basing on true *premise*  $\Phi$  we arrive at true *conclusion*  $\Psi$  (MP), or if negation of conclusion  $\Psi$  is true we infer that negation of premise  $\Phi$  is true (MT).

In reasoning from data (data mining) we also use rules "if  $\Phi$  then  $\Psi$ ", called *decision rules*, to express our knowledge about reality, but the meaning of decision rules is different. It does not express general knowledge but refers to partial facts. Therefore decision rules are not true or false but probable (possible) only.

In this paper we compare inference rules and decision rules in the context of decision networks, proposed by the author as a new approach to analyze reasoning patterns in data.

Decision network is a set of logical formulas  $\mathcal{F}$  together with a binary relation over the set  $\mathcal{R} \subseteq \mathcal{F} \times \mathcal{F}$  of formulas, called a *consequence relation*. Elements of

the relation are called *decision rules*. The decision network can be perceived as a directed graph, nodes of which are formulas and branches – are decision rules. Thus the decision network can be seen as a knowledge representation system, revealing data structure of a data base.

Discovering patterns in the database represented by a decision network boils down to discovering some patterns in the network. Analogy to the *modus ponens* and *modus tollens* inference rules will be shown and discussed.

## 2 Decision Networks

In this section we give after [3] basic notations of decision networks.

Let  $U$  be a non empty finite set, called the *universe* and let  $\Phi, \Psi$  be logical formulas. The meaning of  $\Phi$  in  $U$ , denoted by  $|\Phi|$ , is the set of all elements of  $U$ , that satisfies  $\Phi$  in  $U$ . The truth value of  $\Phi$  denoted  $val(\Phi)$  is defined as  $card|\Phi|/card(U)$ , where  $cardX$  denotes cardinality of  $X$  and can be interpreted as probability that  $\Phi$  is true [1].

By *decision network* over  $S = (U, \mathcal{F})$  we mean a pair  $N = (\mathcal{F}, \mathcal{R})$ , where  $\mathcal{R} \subseteq \mathcal{F} \times \mathcal{F}$  is a binary relation, called a *consequence relation*.

Any pair  $(\Phi, \Psi) \in \mathcal{R}, \Phi \neq \Psi$  is referred to as a *decision rule* (in  $N$ ).

We assume that  $S$  is known and we will not refer to it in what follows.

A decision rule  $(\Phi, \Psi)$  will be also presented as an expression  $\Phi \rightarrow \Psi$ , read if  $\Phi$  then  $\Psi$ , where  $\Phi$  and  $\Psi$  are referred to as *premise (conditions)* and *conclusion (decisions)* of the rule, respectively.

If  $\Phi \rightarrow \Psi$  is a decision rule, then  $\Psi \rightarrow \Phi$  will be called an *inversed* decision rule. If we invert all decision rules in a decision network, than the resulting decision network will be called *inverted*.

The number  $supp(\Phi, \Psi) = card(|\Phi \wedge \Psi|)$  will be called a *support* of the rule  $\Phi \rightarrow \Psi$ . We will consider nonvoid decision rules only, i.e., rules such that  $supp(\Phi, \Psi) \neq 0$ .

With every decision rule  $\Phi \rightarrow \Psi$  we associate its *strength* defined as

$$str(\Phi, \Psi) = \frac{supp(\Phi, \Psi)}{card(U)}. \tag{1}$$

Moreover, with every decision rule  $\Phi \rightarrow \Psi$  we associate the *certainty factor* defined as

$$cer(\Phi, \Psi) = \frac{str(\Phi, \Psi)}{val(\Phi)} \tag{2}$$

and the *coverage factor* of  $\Phi \rightarrow \Psi$

$$cov(\Phi, \Psi) = \frac{str(\Phi, \Psi)}{val(\Psi)}, \tag{3}$$

where  $val(\Phi) \neq 0$  and  $val(\Psi) \neq 0$ .

We assume that

$$val(\Phi) = \sum_{\Psi \in Con(\Phi)} str(\Phi, \Psi) \tag{4}$$

and

$$val(\Psi) = \sum_{\Phi \in Pre(\Psi)} str(\Phi, \Psi), \quad (5)$$

where  $Con(\Phi)$  and  $Pre(\Psi)$  are sets of all conclusions and premises of the corresponding formulas respectively.

Consequently we have

$$\sum_{Con(\Phi)} car(\phi, \Psi) = \sum_{Pre(\Psi)} cov(\Phi, \Psi) = 1 \quad (6)$$

If a decision rule  $\Phi \rightarrow \Psi$  uniquely determines decisions in terms of conditions, i.e., if  $cer(\Phi, \Psi) = 1$ , then the rule is *certain*, otherwise the rule is *uncertain*.

If a decision rule  $\Phi \rightarrow \Psi$  covers all decisions, i.e., if  $cov(\Phi, \Psi) = 1$  then the decision rule is *total*, otherwise the decision rule is *partial*.

Immediate consequences of (2) and (3) are:

$$cer(\Phi, \Psi) = \frac{cov(\Phi, \Psi)val(\Psi)}{val(\Psi)}, \quad (7)$$

$$cov(\Phi, \Psi) = \frac{cer(\Phi, \Psi)val(\Phi)}{val(\Psi)}. \quad (8)$$

Note, that (7) and (8) are Bayes' formulas. This relationship first was observed by Łukasiewicz [1].

Any sequence of formulas  $\Phi_1 \dots \Phi_n$ ,  $\Phi_i \in \mathcal{F}$  and for every  $i$ ,  $1 \leq i \leq n-1$ ,  $(\Phi_i, \Phi_{i+1}) \in \mathcal{R}$  will be called a *path* from  $\Phi_1$  to  $\Phi_n$  and will be denoted by  $[\Phi_1 \dots \Phi_n]$ .

We define

$$cer[\Phi_1 \dots \Phi_n] = \prod_{i=1}^{n-1} cer[\Phi_i, \Phi_{i+1}], \quad (9)$$

$$cov[\Phi_1 \dots \Phi_n] = \prod_{i=1}^{n-1} cov[\Phi_i, \Phi_{i+1}], \quad (10)$$

and

$$str[\Phi_1 \dots \Phi_n] = val(\Phi_1)cer[\Phi_1 \dots \Phi_n] = val(\Phi_n)cov[\Phi_1 \dots \Phi_n]. \quad (11)$$

The set of all paths from  $\Phi$  to  $\Psi$ , denoted  $\langle \Phi, \Psi \rangle$ , will be called a *connection* from  $\Phi$  to  $\Psi$ .

For connection we have

$$cer \langle \Phi, \Psi \rangle = \sum_{[\Phi \dots \Psi] \in \langle \Phi, \Psi \rangle} cer[\Phi \dots \Psi], \quad (12)$$

$$cov \langle \Phi, \Psi \rangle = \sum_{[\Phi \dots \Psi] \in \langle \Phi, \Psi \rangle} cov[\Phi \dots \Psi], \quad (13)$$

$$\begin{aligned}
 str \langle \Phi, \Psi \rangle &= \sum_{[\Phi \dots \Psi] \in \langle \Phi, \Psi \rangle} str[\Phi \dots \Psi] = \\
 &= val(\Phi)cer \langle \Phi, \Psi \rangle = val(\Psi)cov \langle \Phi, \Psi \rangle . \tag{14}
 \end{aligned}$$

With every decision network we can associate a flow graph [2, 3]. Formulas of the network are interpreted as nodes of the graph, and decision rules – as directed branches of the flow graph, whereas strength of a decision rule is interpreted as flow of the corresponding branch.

### 3 Rough *Modus Ponens* and Rough *Modus Tollens*

Classical rules of inference used in logic are *Modus Ponens* and *Modus Tollens*, which have the form

$$\begin{array}{l}
 \text{if } \Phi \rightarrow \Psi \text{ is true} \\
 \text{and } \Phi \text{ is true} \\
 \hline
 \text{then } \Psi \text{ is true}
 \end{array}$$

and

$$\begin{array}{l}
 \text{if } \Phi \rightarrow \Psi \text{ is true} \\
 \text{and } \sim \Psi \text{ is true} \\
 \hline
 \text{then } \sim \Phi \text{ is true}
 \end{array}$$

respectively.

*Modus Ponens* allows us to obtain true consequences from true premises, whereas *Modus Tollens* yields true negation of premise from true negation of conclusion.

In reasoning about data (data analysis) the situation is different. Instead of true propositions we consider propositional functions, which are true to a "degree", i.e., they assume truth values which lie between 0 and 1, in other words, they are probable, not true.

Besides, instead of true inference rules we have now decision rules, which are neither true nor false. They are characterized by three coefficients, *strength*, *certainty* and *coverage factors*. Strength of a decision rule can be understood as a counterpart of truth value of the inference rule, and it represents frequency of the decision rule in a database.

Thus employing decision rules to discovering patterns in data boils down to computation probability of conclusion in terms of probability of the premise and strength of the decision rule, or – the probability of the premise from the probability of the conclusion and strength of the decision rule.

Hence, the role of decision rules in data analysis is somehow similar to classical inference patterns, as shown by the schemes below.

Two basic rules of inference for data analysis are as follows:

$$\frac{\begin{array}{l} \text{if } \Phi \rightarrow \Psi \text{ has } cer(\Phi, \Psi) \text{ and } cov(\Phi, \Psi) \\ \text{and } \Phi \text{ is true with the probability } val(\Phi) \end{array}}{\text{then } \Psi \text{ is true with the probability } val(\Psi) = \alpha val(\Phi)}.$$

Similarly

$$\frac{\begin{array}{l} \text{if } \Phi \rightarrow \Psi \text{ has } cer(\Phi, \Psi) \text{ and } cov(\Phi, \Psi) \\ \text{and } \Psi \text{ is true with the probability } val(\Psi) \end{array}}{\text{then } \Phi \text{ is true with the probability } val(\Phi) = \alpha^{-1} val(\Psi)}.$$

The above inference rules can be considered as counterparts of *Modus Ponens* and *Modus Tollens* for data analysis and will be called *Rough Modus Ponens* (RMP) and *Rough Modus Tollens* (RMT), respectively.

There are however essential differences between MP (MT) and RMP (RMT).

First, instead of truth values associated with inference rules we consider certainly and coverage factors (conditional probabilities) assigned to decision rules.

Second, in the case of decision rules, in contrast to inference rules, truth value of a conclusion (RMP) depends not only on a single premise but in fact depends on truth values of premises of all decision rules having the same conclusions. Similarly, for RMT.

Let us also notice that inference rules are transitive, i.e., *if*  $\Phi \rightarrow \Psi$  *and*  $\Psi \rightarrow \Theta$  *then*  $\Phi \rightarrow \Theta$  and decision rules are not. *If*  $\Phi \rightarrow \Psi$  *and*  $\Psi \rightarrow \Theta$ , then we have to compute the certainty, coverage and strength of the rule  $\Phi \rightarrow \Theta$ , employing formulas (9), (10), (12) and (13).

This shows clearly the difference between reasoning patterns using classical inference rules in logical reasoning and using decision rules in reasoning about data.

## 4 An Example

Suppose that three models of cars  $\Phi_1$ ,  $\Phi_2$  and  $\Phi_3$  are sold to three disjoint groups of customers  $\Theta_1$ ,  $\Theta_2$  and  $\Theta_3$  through four dealers  $\Psi_1$ ,  $\Psi_2$ ,  $\Psi_3$  and  $\Psi_4$ .

Moreover, let us assume that car models and dealers are distributed as shown in Fig. 1.

Applying RMP to data shown in Fig. 1 we get results shown in Fig. 2. In order to find how car models are distributed among customer groups we have to compute all connections among cars models and consumers groups, i.e., to apply RMP to data given in Fig. 2. The results are shown in Fig. 3.

For example, we can see from the decision network that consumer group  $\Theta_2$  bought 21% of car model  $\Phi_1$ , 35% of car model  $\Phi_2$  and 44% of car model  $\Phi_3$ . Conversely, for example, car model  $\Phi_1$  is distributed among customer groups as follows: 31% cars bought group  $\Theta_1$ , 57% group  $\Theta_2$  and 12% group  $\Theta_3$ .

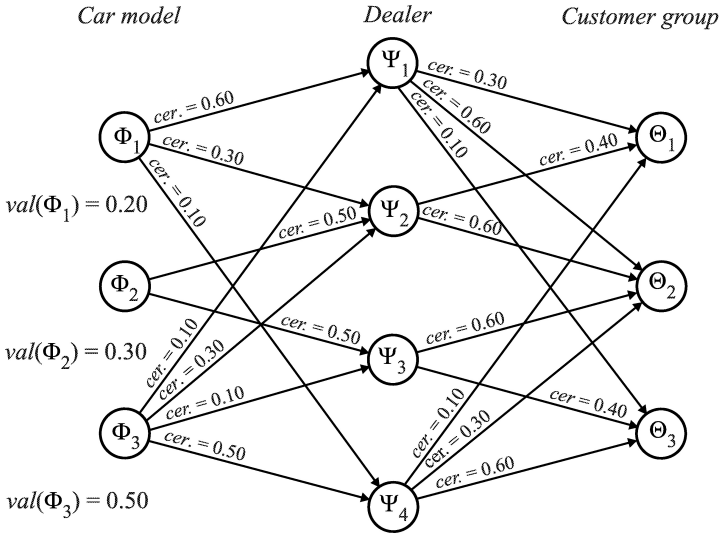


Fig. 1. Car and dealer distribution

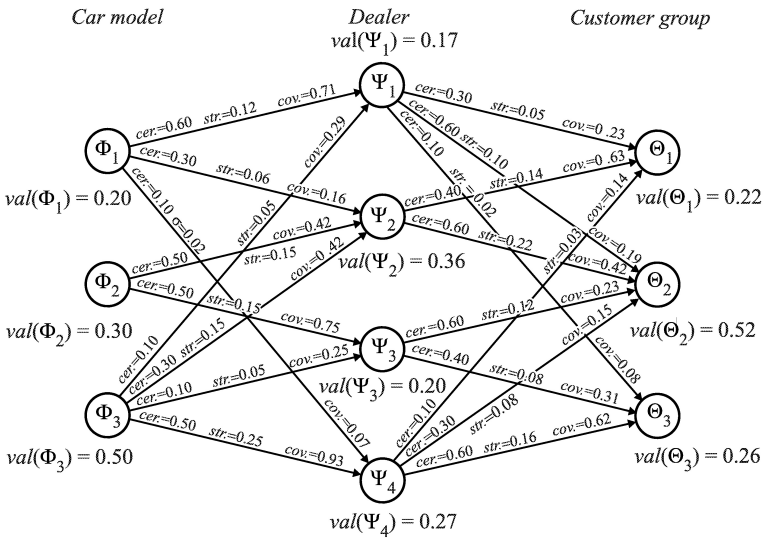


Fig. 2. Strength, certainty and coverage factors

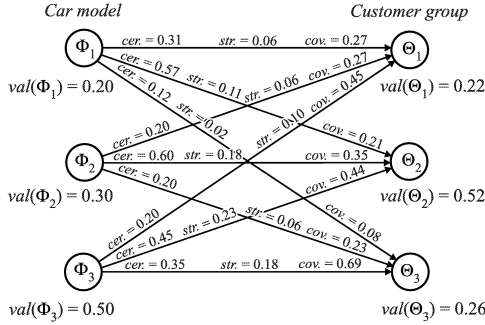


Fig. 3. Relation between car models and consumer groups

## 5 Conclusion

In this paper we compare inference rules and decision rules. Both are expressions in form "if  $\Phi$  then  $\Psi$ " but the meaning of these rules is different. We study the differences and show how they work in logical inference and data analysis, respectively.

## References

1. Łukasiewicz, J.: Die logischen Grundlagen der Wahrscheinlichkeitsrechnung. Kraków (1913), in: L. Borkowski (ed.), Jan Łukasiewicz – Selected Works, North Holland Publishing Company, Amsterdam, London, Polish Scientific Publishers, Warsaw (1970) 16-63
2. Pawlak, Z.: Decision networks, RSCTC2004 (to appear)
3. Pawlak, Z.: Flow graphs and decision algorithms, in: G. Wang, Q. Liu, Y. Y. Yao, A. Skowron (eds.), Proceedings of the Ninth International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing RSFDGrC'2003), Chongqing, China, May 26-29, 2003, LNAI 2639, Springer-Verlag, Berlin, Heidelberg, New York, 1-11