Table 3.9 A Decision Table

	Attributes			Decision
	a	***b***	***c***	***d***
	terrain_familiarity	gasoline_level	distance	speed [m.p.h.]
x_1	poor	low	short	< 30
x_2	poor	low	short	< 30
x_3	good	low	medium	< 30
x_4	good	medium	short	30..50
x_5	poor	low	short	< 30
x_6	poor	high	long	> 50

3.3.2 Indiscernibility Relations and Partitions

An approach for rule induction from examples is presented here. The approach uses some of Z. Pawlak's concepts, presented in many papers (e.g., Pawlak, 1984). The presented algorithm was implemented as a Franz Lisp program LEM for VAX 11/780 (Dean and Grzymala-Busse, 1988).

An example of the decision table is presented in Table 3.9. Observed situations $x_1, x_2, \ldots, x_6$ are described in terms of attributes: terrain_familiarity, gasoline_level, and distance, and driver's decision: speed of a car.

3.3.2.1 Indiscernibility Relations

Let Q denote the set of all attributes and decisions. In Table 3.9, $Q = \{a, b, c, d\}$. Let P be an arbitrary nonempty subset of Q.

Let U be the set of all entities. In the example $U = \{x_1, x_2, \ldots, x_6\}$. Let x and y be arbitrary entities. Entities x and y are said to be *indiscernible by* P, denoted

$$x \underset{P}{\sim} y,$$

if and only if x and y have the same value on all elements in P. Thus x and y are indiscernible by P if and only if the rows of the table, labeled by x and y and restricted to columns, labeled by elements from P, have, pairwise, the same values. In the example,

$$x_3 \underset{\{a\}}{\sim} x_4,$$

$$x_2 \underset{\{b,d\}}{\sim} x_3,$$

and

$$x_1 \underset{Q}{\sim} x_2.$$

3.3.2.2 Partitions

Obviously, the indiscernibility relation, associated with *P*, is an equivalence relation on *U*. As such, it induces a *partition of* U *generated by* P, denoted *P**. For simplicity, the partition *P** of *U* generated by *P* is also called a partition *P** of *U*, or yet simpler, if *U* is known, as partition *P**. As follows from its definition, partition *P** is the set of all equivalence classes (also called *blocks*) of the indiscernibility relation. Thus in the example

$$\{a\}^* = \{\{x_1, x_2, x_5, x_6\}, \{x_3, x_4\}\},$$

$$\{b\}^* = \{\{x_1, x_2, x_3, x_5\}, \{x_4\}, \{x_6\}\},$$

$$\{c\}^* = \{\{x_1, x_2, x_4, x_5\}, \{x_3\}, \{x_6\}\},$$

$$\{d\}^* = \{\{x_1, x_2, x_3, x_5\}, \{x_4\}, \{x_6\}\},$$

$$\{a, b\}^* = \{\{x_1, x_2, x_5\}, \{x_3\}, \{x_4\}, \{x_6\}\},$$

$$\{a, c\}^* = \{\{x_1, x_2, x_5\}, \{x_3\}, \{x_4\}, \{x_6\}\},$$

$$\{a, b, c\}^* = \{\{x_1, x_2, x_5\}, \{x_3\}, \{x_4\}, \{x_6\}\},$$

$$Q^* = \{\{x_1, x_2, x_5\}, \{x_3\}, \{x_4\}, \{x_6\}\}.$$

Partition $\{a\}^*$ has two blocks, $\{x_1, x_2, x_5, x_6\}$ and $\{x_3, x_4\}$.

Note that the concept of a partition, generated by *P*, makes specific values of attributes or decisions unimportant. These values are necessary to define a partition, but whether a domain of an attribute or decision is the set {0, 1, 2} of integers, or the set {*H*, *L*} of characters, or strings, does not make any difference because all partitions are defined on the same set *U*. Also, the fact that domains for different attributes or decisions are not disjoint is irrelevant.

3.3.3 Attribute Dependency and Rule Induction

In this section the concept of a covering, based on yet another concept, attribute dependency, is introduced. In the process of rule induction, redundant attributes in rules may be avoided, provided that these rules are constructed from coverings. This justifies the importance of concepts of attribute dependency and covering.

3.3.3.1 Attribute Dependency Inequality

Let P and R be nonempty subsets of set Q of all attributes and decisions. Set R is said to *depend on* set P if and only if

$$\widetilde{P} \subseteq \widetilde{R} .$$

The fact that R depends on P is denoted by $P \rightarrow R$. Note that $P \rightarrow R$ if and only if

$$P^* \leq R^*.$$

The preceding inequality is called *attribute dependency inequality*. Partition P^* is smaller than or equal to partition R^* if and only if for each block B of P^* there exists a block B' of R^* such that

$$B \subseteq B'.$$

The statement "set R depends on set P" may be characterized by the following: If a pair of entities cannot be distinguished by means of elements from P, then it cannot be distinguished by elements from R.

In the example, let $P = \{a, b\}$ and $R = \{d\}$. Then

$$P^* = \{\{x_1, x_2, x_5\}, \{x_3\}, \{x_4\}, \{x_6\}\} \leq R^* = \{\{x_1, x_2, x_3, x_5\}, \{x_4\}, \{x_6\}\},$$

so $\{d\}$ depends on $\{a, b\}$.

Similarly,

$$\{a, c\}^* \leq \{d\}^*,$$

so $\{d\}$ depends on $\{a, c\}$.

Moreover, $\{d\}$ does not depend on $\{a\}$ because partition $\{a\}^*$ is not smaller than or equal to partition $\{d\}^*$.

However, $\{d\}$ depends on $\{b\}$ because the dependency inequality is fulfilled:

$$\{b\}^* \leq \{d\}^*.$$