Chapter 7
The Concept of Good Classification (Diagnostic) Test

ABSTRACT

In this chapter, the definition of good diagnostic test and the characterization of good tests are introduced and the concepts of good maximally redundant and good irredundant tests are given. The algorithms for inferring all kinds of good diagnostic tests are described in detail.

INTRODUCTION

The definition of the good test is based on the partition model of classifications that is discussed in the previous chapter. Some characteristics of good tests allow proposing a strategy for inferring all kinds of these tests. We describe an algorithm called Background Algorithm based on the method of mathematical induction. This algorithm is applicable to inferring all kinds of good classification tests and, consequently, for inferring functional, implicative dependencies and association rules from a given data set. We discuss also, in this chapter, the possible ways of constructing an efficient algorithm for inferring good tests of any kind.

DEFINITION OF GOOD DIAGNOSTIC TEST

In chapter 6, we considered the set \( L(I(T)) \) of partitions produced by closing atomic partitions of \( I(T) \) with the use of operations + and * on partitions. This set is the algebraic lattice with constants over \( U \). Now we consider the table of examples \( T(U \cup K) \) with partition interpretation \( I(T \cup K) \) over \( U' = U \cup K \) and the algebraic lattice \( L(I(U')) \) with constants over \( U' \), where \( K \) is a given goal attribute.

DOI: 10.4018/978-1-60566-810-9.ch007
The Concept of Good Classification (Diagnostic) Test

Let \( X \subseteq U \) is a test for a goal attribute \( K \), and partitions \( P(X), P(K) \) are the interpretations of \( X \) and \( K \), respectively.

Denote by \( Q(K) \) the set of all diagnostic tests for \( K \): \( Q(K) = \{ X: X \subseteq U: P(X) \leq P(K) \} \).

Sub-lattice \( LK(I(U^*)) \) of \( L(I(U^*)) \): \( LK(I(U^*)) = \{ P: P \leq P(K) \} \) is the principal ideal generated by \( P(K) \) in \( L(I(U^*)) \) (see, please, definition 6.3 of principal ideal in Chapter 6).

For definition of good test, we use partition dependency: \( P(X) \subseteq P(K) \) (\( X \equiv K = X \equiv X + K = K \)).

If \( P(X) = P(K) \), then \( X \) is the ideal approximation of classification \( K \).

Of all tests \( X, X \subseteq U, P(X) \subseteq P(K) \), the good test will be \( X \) such that \( P(X) \) is the closest to \( P(K) \) element of \( LK(I(U^*)) \), i. e., for all \( P(Y), Y \subseteq U \) condition \( P(X) \subseteq P(Y) \subseteq P(K) \) implies \( P(X) = P(Y) \). Thus, we come to the following definition of a good diagnostic test.

**Definition 7.1.** A collection \( X \subseteq U \) is a good test or a good approximation of \( K \) of \( T \) if the following conditions are satisfied

a) \( X \in Q(K) \);  
b) there does not exist a collection of attributes \( Z, Z \subseteq U, X \neq Z \) such that \( Z \in Q(K) \) and \( P(X) < P(Z) \leq P(K) \).

We introduce the concept of the best diagnostic test as follows.

**Definition 7.2.** A good test \( X, X \subseteq U \) is the best one for a given classification \( K \) of \( T \) if the number of classes in partition \( P(X) \) is the smallest for all tests of \( Q(K) \).

Let us illustrate definitions of 7.1 and 7.2 based on the following example (see, please, Table 1).

In this table, \( P(K) = \{ \{ t_1, t_2, t_3, t_4 \}, \{ t_5, t_6 \} \} \).

We consider the set \( Q \) of tests containing not more than 2 attributes:

\[ Q = \{ AB, AC, AF, AG, FG, AE, BG, DG \} \]

But we have only three good tests in \( Q \):

\[ P(AE) = \{ \{ t_1, t_3, t_4 \}, \{ t_1 \}, \{ t_3, t_4 \} \} \text{ contains 4 classes;} \]

\[ P(BG) = \{ \{ t_1, t_2 \}, \{ t_3 \}, \{ t_4 \}, \{ t_5 \}, \{ t_6 \} \} \text{ contains 5 classes;} \]

\[ P(DG) = \{ \{ t_1 \}, \{ t_2, t_4 \}, \{ t_3 \}, \{ t_5 \}, \{ t_6 \} \} \text{ contains 5 classes} \]

**Table 1. The table of example for illustration good test definition**

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>t_1</td>
<td>a_1</td>
<td>b_1</td>
<td>c_1</td>
<td>d_1</td>
<td>e_1</td>
<td>f_1</td>
<td>g_1</td>
<td>k_1</td>
</tr>
<tr>
<td>2</td>
<td>t_2</td>
<td>a_2</td>
<td>b_2</td>
<td>c_2</td>
<td>d_2</td>
<td>e_2</td>
<td>f_2</td>
<td>g_1</td>
<td>k_1</td>
</tr>
<tr>
<td>3</td>
<td>t_3</td>
<td>a_1</td>
<td>b_1</td>
<td>c_2</td>
<td>d_1</td>
<td>e_1</td>
<td>f_1</td>
<td>g_2</td>
<td>k_1</td>
</tr>
<tr>
<td>4</td>
<td>t_4</td>
<td>a_1</td>
<td>b_2</td>
<td>c_1</td>
<td>d_2</td>
<td>e_1</td>
<td>f_1</td>
<td>g_1</td>
<td>k_1</td>
</tr>
<tr>
<td>5</td>
<td>t_5</td>
<td>a_2</td>
<td>b_1</td>
<td>c_1</td>
<td>d_2</td>
<td>e_1</td>
<td>f_1</td>
<td>g_2</td>
<td>k_2</td>
</tr>
<tr>
<td>6</td>
<td>t_6</td>
<td>a_1</td>
<td>b_1</td>
<td>c_1</td>
<td>d_1</td>
<td>e_1</td>
<td>f_1</td>
<td>g_1</td>
<td>k_2</td>
</tr>
</tbody>
</table>