Pattern Recognition, Vol. 20, No. 4, pp. 413–418, 1987.
Printed in Great Britain.

0031-3203/87 $3.00 + .00
Pergamon Journals Ltd.

A VALUATION OF STATE OF OBJECT BASED ON WEIGHTED MAHALANOBIS DISTANCE

E. KRUSIŃSKA

Institute of Computer Science, Wrocław University, 51-151 Wrocław, Przesmyckiego 20, Poland

(*Received* 26 *April* 1985; *in revised form* 14 *January* 1986)

Abstract—A method of valuation of the state of a multidimensional object is presented. The considered object may be described by continuous and discrete variables as well. The method consists in evaluating the weighted Mahalanobis distance between the object and a multidimensional norm estimate. This distance can be standardized to take the values from the interval [0, 1].

Weighted Mahalanobis distance Continuous and discrete variables Linguistic variables
Norm estimate Difference degree

1. INTRODUCTION

The so-called Mahalanobis distance[1] plays an important role nowadays in multivariate statistics. The Mahalanobis distance underlies the theory of discriminant analysis and it is also often used in cluster analysis.[2] Both discriminant and cluster analysis may be treated as a part of pattern recognition.

The problem, for example, of obtaining the linear discriminant function on the basis of a training sample is a common one in pattern recognition.[3,4] The obtained discriminant function gives the possibility to assign $\mathbf{x} = [x_1, x_2, \ldots, x_s]^T$ of an unknown origin to one of two (or generally more) distinct classes on the basis of the values of s predictor variables. Many statistical and nonstatistical procedures for obtaining the linear discriminant function have been proposed.[3–5] From the statistical point of view the classification rule equivalent to the classical Fisherian linear discriminant function[5] with equal *a priori* probabilities in all classes consists in comparing the Mahalanobis distance[1] between $\mathbf{x}$ and $\mathbf{m}_1$ and between $\mathbf{x}$ and $\mathbf{m}_2$ (where $\mathbf{m}_i$ is the mean vector (centre) of s predictor variables for the ith class; $i = 1, 2$). The individual $\mathbf{x}$ is assigned to this class where the calculated distance is the smallest one. Next in cluster analysis the Mahalanobis distance may be treated as a dissimilarity measure between two objects and it can be a basis for finding clusters of similar points (objects). Classically the linear discriminant function and the Mahalanobis distance are defined for continuous variables. The linear discrimination does not work for discrete variables because the assumption of normality of the multivariate distribution of $\mathbf{x}$ is not fulfilled. Generally an object $\mathbf{x}$ can be described by continuous and discrete variables as well. Therefore in discriminant analysis some special approaches have been elaborated to handle mixtures of continuous and discrete variables.[6] However Mahalanobis-like distances have been defined in cluster analysis.[2]

In this paper a new application of the Mahalanobis distance is shown. The Mahalanobis distance is not used to perform the discrimination between several classes or to cluster not distant objects. It is used for a valuation which shows how an abnormal object differs from a norm. A norm estimate is obtained on the basis of a training sample drawn out of the class of normal objects as a point or solid estimate. The distance between the considered abnormal object and the obtained norm estimate is treated as the searched measure of the difference degree. To extend an analysis to discrete variables a generalization of Mahalanobis distance[7] is used or a conversion of the groups of discrete variables into linguistic ones[8] is performed. To emphasize that each variable has its own weight in the problem under consideration the weighted Mahalanobis distance as defined by the author[9] instead of the classical unweighted Mahalanobis distance is used in this paper to a valuation of the state of an object.

2. THE STATEMENT OF THE PROBLEM

Let us assume that we have a sample $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ of n individuals drawn out of the class C_0 of normal objects (e.g. healthy persons) and a sample drawn out of the class C_1 of abnormal individuals (e.g. ill persons). The aim of this work is to evaluate the state of an object $\mathbf{x} \in C_1$, this means to decide how $\mathbf{x}$ differs from the norm estimate obtained on the basis of the sample $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$. As a measure of the state of an object characterized by an observational vector $\mathbf{x} = [x_1, x_2, \ldots, x_s]^T$ of s predictor variables we take the

weighted squared Mahalanobis distance d_a^2 between $\mathbf{x}$ and the norm estimate. It measures the dissimilarity between an object $\mathbf{x}$ and the norm estimate. Therefore in our application we use it to decide how an object differs from the norm. We assume that from the s considered predictor variables p of them are continuous, the remaining l are discrete. The r sets of discrete variables (each set describes one characteristical feature of the object) are transformed into r linguistic ones using the method of Saitta and Torasso.[8] This method allows for missing values. For this conversion we use j from l discrete variables, so $k = l - j$ variables remain. The details of the performed conversion are given in the next section. In the problem under consideration the obtained linguistic variables are treated as the continuous ones.

The weighted Mahalanobis distance d_a^2 for both continuous and discrete variables is defined as the sum

$$d_a^2 = d_c^2 + d_d^2, \tag{1}$$

where

d_c^2 is the weighted Mahalanobis distance for the continuous and linguistic variables together,
d_d^2 is the weighted Mahalanobis distance for the remaining k ($k = l - j$) discrete variables.

The distance d_c^2 equals

$$d_c^2 = [(\tilde{\mathbf{x}} - \bar{\mathbf{y}})W]^T S^{-1} (\tilde{\mathbf{x}} - \bar{\mathbf{y}}) W, \tag{2}$$

where
$\tilde{\mathbf{x}} = [x_1, x_2, \ldots, x_p, x_{p+1}, \ldots, x_{p+r}]^T$—the vector of observed values of continuous and linguistic variables for the considered individual $\mathbf{x} \in C_1$,
$\bar{\mathbf{y}} = [\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_p, \bar{y}_{p+1}, \ldots, \bar{y}_{p+r}]^T$—the mean vector (the centre) of the sample $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ drawn out of C_0 (the standard point estimate for continuous and linguistic variables),
S—the sample within-group covariance matrix (obtained for the groups (samples) drawn out of C_0 and C_1),
$W = \operatorname{diag}(w_1, w_2, \ldots, w_p, w_{p+1}, \ldots, w_{p+r})$.
This definition was given by the author.[9]
The vectors $\tilde{\mathbf{x}}$ and $\bar{\mathbf{y}}$ comprise continuous and linguistic variables. The elements of the weight vector $\mathbf{w} = [w_1, w_2, \ldots, w_p, w_{p+1}, \ldots, w_{p+r}]^T$ express the degree of importance of continuous and linguistic variables in the valuation of the object's state. A condition that $w_i \neq 0$ ($i = 1, 2, \ldots, p + r$) have to be fulfilled because the assumption for a distance in the metric space for d_c^2 must be kept.

The classical unweighted Mahalanobis distance calculated only for continuous variables equals $D^2 = (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$, where $\mathbf{x}$—the considered individual (comprised the values of continuous variables), $\boldsymbol{\mu}$, Σ are the parameters of the multivariate normal distribution assumed for the norm population and for the vector $\mathbf{x}$, this means they are the vector of expected values and the covariance matrix, respectively. In practice the parameters $\boldsymbol{\mu}$ and Σ are not known, therefore they ought to be estimated on the basis of the training sample. The sample equivalent of the Mahalanobis distance equals $D^2 = (\mathbf{x} - \bar{\mathbf{y}})^T S^{-1} (\mathbf{x} - \bar{\mathbf{y}})$, where $\bar{\mathbf{y}}$ is the sample mean vector—the estimate of $\boldsymbol{\mu}$ and S is the estimate of Σ. This is commonly the sample within-group covariance matrix obtained on the basis of the whole sample drawn out of C_0 and C_1. It is done similarly in formula (2). The new extension of the Mahalanobis distance consists in transforming the unweighted distance by the weight matrix W. This enables to emphasize different importance of variables in evaluating the state of the considered object (e.g. the state of a patient) and gives a possibility of a more adequate valuation of the difference degree between the object and the norm. Another extension consists in calculating the Mahalanobis distance together for continuous and linguistic variables obtained as a conversion of the groups of the original discrete ones.

Now let us define the distance d_d^2. Under assumption of independency of $k = l - j$ remaining (not transformed) discrete variables we can write the total distance for discrete variables as

$$d_d^2 = \sum_{i=1}^{k'} v_i^2 d_i^2, \tag{3}$$

where
v_i ($i = 1, 2, \ldots, k'$) is the weight for the ith discrete variable, d_i^2 ($i = 1, 2, \ldots, k'$) is the Mahalanobis distance (as defined by Kurczyński[7] and described in Section 4) between the observed value of the ith discrete variable and the norm estimate for it, $k - k'$ is the number of missing observations for the considered object among k remaining discrete variables.

This means that the distances d_i^2 are summed only for these single discrete variables whose values are known for the considered individual $\mathbf{x}$.

3. LINGUISTIC VARIABLES

It has been assumed that each object is characterized by an observational vector $\mathbf{x} = [x_1, x_2, \ldots, x_s]^T$ of s predictor variables. p of them are continuous, the remaining l are discrete. The j discrete variables are transformed into linguistic ones using the method of Saitta and Torasso.[8] The linguistic variables are then treated as continuous for calculating the Mahalanobis distance. The variables used for transformation are grouped in r sets. Each set describes one characteristical feature of an object (e.g. in medical applications—one disease symptom). A conversion of the groups of discrete variables into linguistic ones ought to be done only if basing on practical premisses it is sensible to join together the discrete variables in groups, and if the phenomena described by these groups have a fuzzy character.

Let us assume that the linguistic variable L_i ($i = 1, 2, \ldots, r$) is described by the set $Q^{(i)}$. A weight $\gamma_{gh}^{(i)} \in [-1, 1]$ is associated with each state $s_{gh}^{(i)}$ of the variable $q_g^{(i)} \in Q^{(i)}$ ($i = 1, 2, \ldots, r$; $g = 1, 2, \ldots, M_i$, where M_i is the number

of variables in the set $Q^{(i)}$; $h = 1, 2, \ldots, A_{ig}$, where A_{ig} is the number of states of the gth variable from $Q^{(i)}$) $\cdot\ \gamma_{gh}^{(i)}$ with the value near -1 tells us that there is no agreement between the state $s_{gh}^{(i)}$ and the variable L_i. It means that the state $s_{gh}^{(i)}$ is not characteristic for the phenomenon described by L_i. On the contrary when $\gamma_{gh}^{(i)}$ has the value near $+1$ the agreement degree between the state $s_{gh}^{(i)}$ and the variable L_i is very high. When $\gamma_{gh}^{(i)}$ equals 0 there is no information (positive or negative) about this agreement. So $\gamma_{gh}^{(i)}$ equals 0 for each missing value. The weights $\gamma_{gh}^{(i)}$ are fixed *a priori* by a specialist (e.g. by a physician) basing on practical premisses.

Next the vector $\boldsymbol{\alpha}^{(i)} = (\alpha_1^{(i)}, \alpha_2^{(i)}, \ldots, \alpha_{M_i}^{(i)})$ is defined. The weights $\alpha_g^{(i)}$ are proportional to the discriminative power of the variables from the set $Q^{(i)}$ in differentiating between classes C_0 and C_1. We assume that

$$\alpha_g^{(i)} = \begin{cases} t_g^{(i)}, & \text{when } t_g^{(i)} \text{ is greater or of the order of } t_{0.1} \\ 0, & \text{in the remaining cases,} \end{cases}$$

where

$t_g^{(i)}$ is the computed value of the Student t-statistic for the variable $q_g^{(i)}$ (calculated to compare the states of $q_g^{(i)}$ in C_0 and C_1),
$t_{0.1}$ is the critical value of the Student t-statistic for the significance level 0.1.

The total weight for the state $s_{gh}^{(i)}$ is given as

$$u_{gh}^{(i)} = \alpha_g^{(i)} \cdot \gamma_{gh}^{(i)}.$$

Further Saitta and Torasso[8] defined for each object from C_1 the variable m_i associated with the linguistic variable L_i as

$$m_i = \sum_{g=1}^{M_i} u_{g^*}^{(i)}, \tag{4}$$

where
$u_{g^*}^{(i)}$ is the weight of this state of $q_g^{(i)}$ which has been reached by the considered object.

On the basis of variable m_i a membership function of L_i which takes values from the interval [0, 1] may be calculated.

But having in mind the assumption of normality the variables m_i associated with L_i were used to obtain the weighted Mahalanobis distance.

The distance d_c^2 was calculated for $\tilde{\mathbf{x}} = [\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_p, \tilde{x}_{p+1}, \ldots, \tilde{x}_{p+r}]^T$, where $\tilde{x}_t$ $(t = 1, 2, \ldots, p)$ equals the value x_t of the original tth continuous variable and $\tilde{x}_{p+i}$ $(i = 1, 2, \ldots, r)$ equals m_i.

The method of Saitta and Torasso[8] was used by Krusińska and Liebhart[10] to describe the symptoms of the so called chronic obturative lung disease (this means: bronchial asthma, chronic bronchitis and lung emphysema). For each patient a questionnaire of 146 items was filled out to collect the results of a complex examination. The 81 variables from this set were divided into 14 groups basing on medical premisses. Each group described one phenomenon with the fuzzy character such as, e.g. cough, character of dyspnoea, intensity of dyspnoea, etc. The weights for discrete variables, this means $\alpha_g^{(i)}$ were calculated basing on the Student t-statistic for comparing the values of these variables in the group of ill persons and in the control group (healthy persons). The weights for states of discrete variables, i.e. $\gamma_{gh}^{(i)}$ were fixed *a priori* by a physician. The same linguistic variables as in Ref. (10) were used to calculate the Mahalanobis distance for a valuation of the state of patients.

4. THE GENERALIZED MAHALANOBIS DISTANCE FOR DISCRETE VARIABLES

Only the part of discrete variables can be used for the conversion into linguistic ones. Some variables (e.g. sex, coded: 0 — male, 1 — female) which do not describe phenomena with fuzzy character ought to rest as not transformed discrete ones. The Mahalanobis-like distance must be used to calculate the dissimilarity between an object and a norm estimate for them.

Such a generalization of Mahalanobis distance to discrete variables[7] between two classes, say C_1 and C_2 is defined for the ith discrete variable $(i = 1, 2, \ldots, k)$ as

$$d_i^2 = (\hat{\mathbf{p}}_1 - \hat{\mathbf{p}}_2)^T S^{-1} (\hat{\mathbf{p}}_1 - \hat{\mathbf{p}}_2), \tag{5}$$

where

k is the number of discrete variables (not transformed into linguistic variables),
$\hat{\mathbf{p}}_g = [\hat{p}_{g1}, \hat{p}_{g2}, \ldots, \hat{p}_{gz_i}]^T$ $(g = 1, 2)$ is the vector of probability estimates,
p_{gh} is the probability of obtaining the hth state of the considered ith discrete variable in the gth class,
z_i is the number of states of the ith discrete variable,
S is the sample covariance matrix for the ith discrete variable. It means that the elements of the matrix S equal

$$s_{ht} = \hat{p}_h(1 - \hat{p}_h) \quad h = t$$
$$s_{ht} = -\hat{p}_h\hat{p}_t \quad h \neq t,$$

where

$$\hat{p}_h = \sum_{g=1}^{2} n_g \hat{p}_{gh} \Big/ \sum_{g=1}^{2} n_g,$$

n_g is the number of individuals in the gth class $(g = 1, 2)$.

As the estimates $\hat{p}_{gh}$ the fractions of observing the hth state in the gth class are used.

In the problem of a valuation of the state of the considered object we ought to calculate the distance between class C_0 of normal objects and an object from the class C_1. For this reason the estimate $\hat{\mathbf{p}}_0 = [\hat{p}_{01}, \hat{p}_{02}, \ldots, \hat{p}_{0z_i}]^T$ which bases on the sample drawn out of C_0 and the common covariance matrix S which bases on the whole sample (drawn out of C_0 and C_1) ought to be estimated. The Mahalanobis distance for the ith considered discrete variable equals in this case

$$d_i^2 = (\hat{\mathbf{p}}_0 - \mathbf{e})^T S^{-1} (\hat{\mathbf{p}}_0 - \mathbf{e}), \quad (i = 1, 2, \ldots, k), \tag{6}$$

$\mathbf{e} = [0, \ldots, 0, 1, 0, \ldots, 0]^T$, where 1 appears on the uth place when the uth state of the ith discrete variable is observed for the considered individual from C_1.

Under assumption of independency of k discrete variables the total weighted Mahalanobis distance for discrete variables is given by formula (3) where d_i^2 $(i = 1, 2, \ldots, k)$ are defined as in equation (6).

5. THE STANDARDIZED DIFFERENCE DEGREE

The measure of the state of an object defined as d_a^2 [formula (1)] may be standardized to take the values from the interval [0, 1].

Let us discuss the distribution of the Mahalanobis distance. The unweighted Mahalanobis distance for continuous and linguistic variables has asymptotically χ^2_{p+r} distribution (where $p + r$ is the number of the degrees of freedom). The linear transformation by a matrix W does not change the distribution and the number of the degrees of freedom.[11] So the weighted Mahalanobis distance d_c^2 given by (2) has also asymptotically χ^2_{p+r} distribution.

The generalized Mahalanobis distance d_i^2 between two classes for the ith discrete variable [given by (5)] has asymptotically $\chi^2_{z_i-1}$ distribution (z_i—the number of states of the ith discrete variable). Under assumption of independency of discrete variables the Mahalanobis distance for k' variables defined as the sum of distances for succeeding variables [formula (3)] has $\chi^2_{\sum z_i - k'}$ distribution.

For the distance d_i^2 between class C_0 and an individual $\mathbf{e} \in C_1$ given by (6) this distribution is not kept but on the analogy we propose:

Definition

The standardized difference degree (DD) equals

$$DD = \frac{\min\left(d_a^2, \chi^2_{df,\alpha}\right)}{\chi^2_{df,\alpha}}, \tag{7}$$

where

$\chi^2_{df,\alpha}$ is the quantile of χ^2 distribution,

α is the significance level,

$$df = p + r + \sum_{i=1}^{k} z_i - k',$$

$k - k'$ is the number of missing observations for the considered object among k discrete variables, so the distances d_i^2 are summed in reality only for k' discrete variables.

6. THE WEIGHT MATRIX

The weight matrix $W = \operatorname{diag}(w_1, w_2, \ldots, w_{p+r})$ and the weights $v_i\,(i = 1, 2, \ldots, k)$ can be defined in various ways. Now we describe one feasible proposal.

Let us assume that we have *a priori* partition of the class C_1 into G distinct classes $C_{1g}\,(g = 1, 2, \ldots, G)$. The class C_{1g} groups the objects at the same state. The object from the class C_{1g} is closer to the norm than the object from the class $C_{1,g+1}$. The weights express the discriminative power of the variables in differentiating between classes $C_{11}, C_{12}, \ldots, C_{1G}$. As a measure of discriminative power of the ith variable $(i = 1, 2, \ldots, p + r + k)$ the Wilks Λ statistic[11] is used

$$\Lambda_i = \frac{\sum_{g=1}^{G}\sum_{h=1}^{n_g}\left[x_{hi}^{(g)2} - \frac{\left(\sum_{h=1}^{n_g} x_{hi}^{(g)}\right)^2}{n_g}\right]}{\sum_{g=1}^{G}\sum_{h=1}^{n_g} x_{hi}^{(g)2} - \frac{\left(\sum_{g=1}^{G}\sum_{h=1}^{n_g} x_{hi}^{(g)}\right)^2}{\sum_{g=1}^{G} n_g}}, \tag{8}$$

where

n_g is the number of individuals drawn out of C_{1g},

$\mathbf{x}_h^{(h)} = [x_{h1}^{(g)}, x_{h2}^{(g)}, \ldots, x_{h,p+r+k}^{(g)}]^T$ is the vector of observed values of continuous, linguistic and discrete variables for the hth individual (object) from the class C_{1g}.

Now the weights equal

$$w_i = 1 - \Lambda_i\,(i = 1, 2, \ldots, p + r)$$

and $\tag{9}$

$$v_t = 1 - \Lambda_t\,(t = 1, 2, \ldots, k),$$

where

Λ_i defined as in the formula (8).

Such a choice of weights is easily understood in medical applications. The class C_1 of abnormal objects (ill persons) can be easily divided based on medical premisses into subclasses of patients with increasing severity of disease from which they suffer. For evaluating the state of persons suffering from the chronic obturative lung disease (mentioned in Section 3), these were groups of patients with bronchial asthma or chronic bronchitis without complications (i.e. lung emphysema or chronic cor pulmonale) and with complications. The weights emphasized the importance of these variables which differentiated in the best way between illness without and with complications, i.e. between less and more severe state of disease.

7. CONCLUDING REMARKS

The presented method enables to evaluate the state of the considered multidimensional object on the basis of observed values of predictor variables. The method works for both continuous and discrete variables. It is not only possible to assign an object $\mathbf{x}$ to one of two classes C_0 or C_1 but it is also possible to answer the question of how $\mathbf{x}$ differs from C_0. So the method enables to observe fluently the changes of the state of an object during the considered period of time or to compare comprehensively the state of several objects. As a measure of the state of an object $\mathbf{x}$ a generalized weighted Mahalanobis distance evaluated for continuous, linguistic and discrete variables is used. This distance is calculated between an object $\mathbf{x} \in C_1$ and a norm estimate obtained on the basis of a sample drawn out of the class C_0 of normal objects. As it is the simplest norm estimate a point estimate (the sample mean vector of predictor variables) is used. It is also

possible to use solid estimates such as a confidence elipsoid or a multidimensional cube with the sides equal to confidence intervals. The Mahalanobis distance has good properties. It does not depend on the chosen units,[1] so the observed values of predictor variables need not be standardized. The generalized weighted Mahalanobis distance between $\mathbf{x}$ and the norm estimate may be transformed to take values from the interval [0, 1] [formula (7)]. This standardization simply enables to compare the states of the same kind of objects described by different sets of predictor variables.

The method was applied to a comparison of 171 patients suffering from bronchial asthma and 59 patients suffering from chronic bronchitis ($p = 31, r = 14, k = 12$). The details of this application are presented by Krusińska and Liebhart.[12] The control group used to obtain the norm estimates contained only healthy persons. On the basis of this sample the so called 'centre of health' was calculated. The standardized Mahalanobis distance (the difference degree DD) was treated as the dissimilarity measure between a patient and the norm, and described synthetically his actual state. The analysis was done separately for bronchial asthma and chronic bronchitis. Linguistic variables ($r = 14$) described the disease symptoms and were obtained as a conversion of the groups of original discrete variables chosen out the questionnaire of over 140 items. The weights w_i ($i = 1, 2, \ldots, p + r$) and v_t ($t = 1, 2, \ldots, k$) were calculated as in formula (9) for differentiating between disease without and with complications. The evaluations of the patient's state obtained with the difference degree DD [formula (7)] were compared with ranked physician's evaluations (in the scale: light, medium, severe). The rank correlation coefficients were statistically significant for the significance level $\alpha = 0.001$ for both bronchial asthma and chronic bronchitis. This means that it is a high correlation between the automatical and physician's valuation and it confirms the usefulness of the new technique of the difference degree in medical applications.

SUMMARY

The aim of this work is to present a new statistical method for a valuation of the state of a multidimensional object. The proposed method works for both continuous and discrete variables and enables simple and comprehensive comparison of the state of an object in the considered period of time, or a comparison of a group of several objects. As a measure of this how an abnormal object differs from a norm population the generalized weighted Mahalanobis distance is taken. The state of an object $\mathbf{x} = [x_1, x_2, \ldots, x_s]^T \in C_1$ (the class of abnormal objects) is described by the weighted Mahalanobis distance d_a^2 between $\mathbf{x}$ and a norm estimate. The norm estimate is obtained on the basis of a training sample drawn out of the class C_0 of normal objects (as a point or solid estimate). This estimate can be chosen as the sample mean vector or the confidence elipsoid or the multidimensional cube with sides equal to confidence intervals. Let us assume that from the s considered predictor variables p of them are continuous, the remaining l are discrete. The r sets of discrete variables (each set describes one phenomenon with fuzzy character) are transformed into r linguistic ones using the Saitta and Torasso method (*Fuzzy Sets and Systems* **5**, 245–258 (1981)). For this conversion we use j from among l discrete variables.

The weighted Mahalanobis distance calculated for both continuous and discrete variables is defined as a sum $d_a^2 = d_c^2 + d_d^2$, where d_c^2 is the weighted Mahalanobis distance for the continuous and linguistic variables, d_d^2 is the weighted Mahalanobis distance for the remaining k ($k = l - j$) not transformed discrete variables. The distance d_a^2 measures the dissimilarity between the considered object $\mathbf{x}$ and the norm estimate. The distance d_c^2 is a weighted equivalent of the classical Mahalanobis distance[1] transformed by the diagonal weight matrix W. The distance d_d^2 is obtained as a sum of the Mahalanobis-like distances.[7]

The weights for continuous, linguistic and discrete variables express their degree of importance in the description of the state of an object. One proposal for obtaining these weights has been presented in the paper. The usefulness of these kind of weights was proved in a practical medical application.

After a simple transformation of the distance d_a^2 the standardized difference degree with the values from the interval [0, 1] is obtained. Using the difference degree DD it is not only possible to assign $\mathbf{x}$ to one of two classes C_0 or C_1, but it is also possible to answer the question of how $\mathbf{x}$ differs from C_0. Therefore the difference degree DD is a new tool in statistical diagnostics.

The proposed method was applied for a comparison of the state of 171 patients suffering from bronchial asthma and 59 patients suffering from chronic bronchitis ($p = 31$, $r = 14$, $k = 12$). The rank correlation coefficients between the statistical evaluation of the patients' state and the ranked physician evaluation (in the scale: light, medium, severe) were statistically significant ($\alpha = 0.001$).

REFERENCES

1. K. V. Mardia, Mahalanobis distances and angles. *Proceedings of the 4th International Symposium on Multivariate Analysis*, pp. 495–512. North Holland, Amsterdam (1977).
2. K. V. Mardia, J. T. Kent and J. M. Bibby, *Multivariate Analysis*. Academic Press, London (1979).
3. O. R. Duda and P. E. Hart, *Pattern Classifications and Scene Analysis*. John Wiley, New York (1973).

4. N. J. Nilsson, *Learning Machines: Foundations of Trainable Pattern Classifying System.* McGraw-Hill, New York (1965).
5. P. A. Lachenbruch, *Discriminant Analysis.* Hafner Press, London (1975).
6. G. A. F. Seber, *Multivariate Observations.* Wiley, New York (1984).
7. T. W. Kurczyński, Generalized distance and discrete variables. *Biometrics* **26**, 525–534 (1970).
8. L. Saitta and P. Torasso, Fuzzy characterization of coronary disease. *Fuzzy Sets and Systems* **5**, 245–258 (1981).
9. E. Krusińska, A weighted Mahalanobis distance as a measure of the degree of illness. *COMPSTAT 84, Summaries of Short Communications and Posters,* Prague (1984).
10. E. Krusińska and J. Liebhart, A note on the usefulness of linguistic variables in differentiating between some respiratory diseases, *Fuzzy Sets and System* **18**, 131–142 (1986).
11. C. R. Rao, *Linear Statistical Inference and Its Application.* Wiley, New York (1965).
12. E. Krusińska and J. Liebhart, Objective valuation of degree of illness with weighted Mahalanobis distance. A study for patients suffering from chronic obturative lung disease. *Computers in Biology and Medicine* (submitted).

About the Author—EWA KRUSIŃSKA was born in Poland in 1957. She received the M. S. degree in Computer Science from the University of Wrocław in 1981 and the Ph.D. degree in Electrical Metrology from the Technical University of Wrocław in 1986. Now she is employed at the Institute of Computer Science, Wrocław University and works in the field of computational statistics. Her current interests are concerned mainly with Discriminant Analysis, Statistical Pattern Recognition and their Applications to Metrology and Medicine.