Non parametric population classification

Smeulders, A.W.M.

Published in:
Pattern Recognition in Practice II: proceedings of an International Workshop held in Amsterdam, June 19-21, 1985

PATTERN RECOGNITION IN PRACTICE II
E.S. Gelsema and L.N. Kanal (Editors)
Elsevier Science Publishers B.V. (North-Holland), 1986

NON PARAMETRIC POPULATION CLASSIFICATION

A.W.M. Smeulders

Pathological Institute
Free University Hospital
De Boelelaan 1117
1081 HV Amsterdam
The Netherlands

Population classification is a set of statistical techniques to classify populations on basis of observations of their constituing members. This decision making involves a cascade of classifiers: one for objects, and, based on the resulting classifications, one for populations. Also a rule is to be given how many objects should be analysed before the population is decidable. In search for a more integral view of the problem the population function is introduced. Consideration of the population function implies that rather then fixing the object classifier a priori, it is more efficient and more accurate to extend the classifier to a distinct range of the feature vector. The concept of sequential classification is defined as well. Population classification performance is favourably compared to [1,2,3,5,6].

1. INTRODUCTION

Population classification is the classification of a population on basis of observations on its members. Population classification deviates from usual decision making procedures in the two step classification procedure that is employed. In the first step, members of a population each are classified into one of the member classes. After the analysis of a certain number of members, in the second step the population itself is classified into one of the population classes. Population classification is a statistical technique with applications in many different fields such as industrial inspection, quality control and quality assessment of lots. For conceptual convenience, we will adopt here the terminology from the cytology field, where a specimen may be conceived as a population of cells.

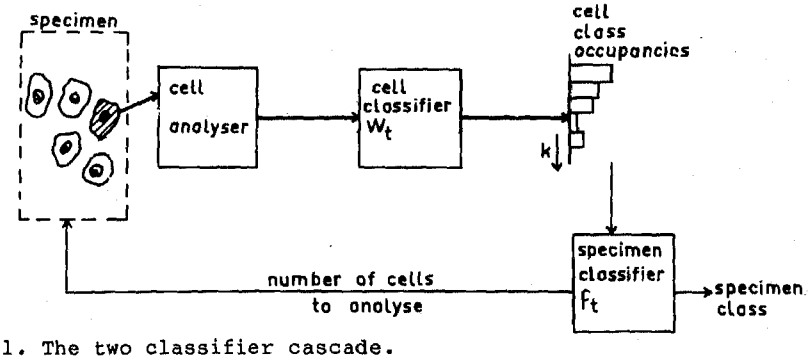

Fig. 1. The two classifier cascade.

In figure 1, the two step classification procedure is depicted. It is crucial in this cascade of classifiers that the accumulated outcome of the cell classification is the input of the specimen classification. So, an alteration in the cell classifier may influence the specimen classification result. In the analysis of specimens, the prime interest is not in the performance of the cell classifier on its own, but in its performance in respect to the specimen classification.

Let the cell of the I-th specimen be numbered by index i with i=1,...,n_I. A feature vector $\bar{w}_i$ is computed per cell, giving a list of n_I observations. The cascade of classifiers is composed of three tunable entities:

1. The cell classifier ($\bar{w}_t$) provides a set of rules to classify a cell characterized by $\bar{w}_i$. The cell is classified into one or more cell classes k with k=1,...,m.
   The cell classifications are accumulated in the cell class occupancy vector $\bar{f}$. It should be noted, that the results of this paper will suggest that it may be profitably not to use a cell classifier.
2. The number rule (n_t) controls the minimum number of cells to be classified before classifying the specimen.
3. The specimen classifier ($\bar{f}_t$) provides a set of rules to classify a specimen based on $\bar{f}$ after the analysis of n_t cells. The specimen is classified in specimen class K with K=1,...,M.

The objective of statistical population analysis is to design a procedure such that the specimen classification error is acceptable and the number of cells to classify is minimized by controlling these three entities. It is important to note that a specimen may either be erroneously classified by the fact that only a limited number of cells are analysed or by the fact that the occupancy vector $\bar{f}$ of a specimen, when compared to $\bar{f}_t$, leads to erroneous classification. The first error, the sample error, is related to the system efficiency and may be avoided by analysing more cells. The second error is related to accuracy of the system and is intrinsic to the nature of the problem. For a fixed $\bar{w}_t$ and $\bar{f}_t$, it cannot be cured by analysing more cells.

Castleman and White were the first to design a procedure for cascaded classification. In a series of papers, using a simplified model, they have given guidelines how to address the cell classifier [1], how to address the specimen classifier [2], and how the proportionality of different classes of cells influences the performance of the specimen classifier [3], once the cell classifier has been fixed. All references are restricted to M=2 with K=1 : 'normal specimens' and K=2 : 'abnormal specimens'. In [1] and [3], the number of cell classes is restricted to m=2. In that case, we will call cells classified into class k=1 'non-events', and cells into class k=2, 'events'. In [2] there is no restriction to the number of cell classes.

The procedure of Castleman and White is discussed here, without loss of generality, for m=2. They consider a previously settled cell classifier such that a fraction (ε_1) of normal cells (k=1) is erroneously classified into class k=2, and fraction (ε_2) of abnormal cells erroneously into k=1. Secondly, the presumption is made that specimens of the normal specimen class (K=1) entirely consists of k=1-cells. Also, it is presumed that abnormal specimens (K=2) all have the same, fixed, fraction p of k=2-cells, and a fraction 1-p of k=1-cells. Let q_K for K = 1,2 be the proportion of events from a specimen of class K, then we have $q_1=\varepsilon_1$ and $q_2=\varepsilon_1+p(1-\varepsilon_1-\varepsilon_2)$. Both q_1 and q_2 are thus constants for all specimens of class K.

It should be noted that the discrimination of ε_1 and ε_2 from p is not necessary. The same result would have been obtained by the presumption that the cell classifier produces a proportion of q_1 events for normal specimens and q_2 events for abnormal ones.

Under these presumptions, the classification procedure reduces to the recognition whether an observed proportion $\hat{q}$ of an unknown specimen originates from a specimen with q_1 or q_2 events. As $\hat{q}$ is based on n cells, $\hat{q}$ is distributed binomially around q_K. The binomial distribution is, with a very good approximation, replaced by a Gaussian distribution with mean q_K and standard deviation $s_K = \sqrt{q_K(1-q_K)/n}$. Let $f_e(x)$ be the error function

$$f_e(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{1}{2}t^2} dt \qquad \text{with inverse } f_e^{-1}(x), \tag{1}$$

and E_K^S, K=1,2, be the tolerated fractions of specimens classified wrong. The minimum number of cells to be analysed (n_t) and the specimen classification threshold (q_t) may be solved now as follows:

$$n_t = \left(\frac{f_e^{-1}(E_2^S)\sqrt{q_2(1-q_2)} + f_e^{-1}(E_1^S)\sqrt{q_1(1-q_1)}}{q_2 - q_1} \right)^2 \tag{2}$$

$$q_t = \frac{f_e^{-1}(E_1^S)q_2\sqrt{q_1(1-q_1)} + f_e^{-1}(E_2^S)q_1\sqrt{q_2(1-q_2)}}{f_e^{-1}(E_1^S)\sqrt{q_1(1-q_1)} + f_e^{-1}(E_2^S)\sqrt{q_2(1-q_2)}} \tag{3}$$

In summary, the proposed procedure is:

1. Select cell features $\bar{w}$ and a cell classifier $\bar{w}_t$, see [1].
2. Select a value for q_1 representative for normal specimens on basis of experience with this cell classifier. Similarly, choose a q_2. Compute n_t and q_t.
3. In the operational phase, analyse n_t cells for each specimen and find $\hat{q}$. If $\hat{q} < q_t$ then decide "normal specimen" else $\hat{q} \geq q_t$ and decide "abnormal specimen".

Although the procedure is attractive in the sense that it provided a relationship between the cell clasifier and the specimen classifier, the presumptions tend to oversimplify reality. As a consequence all information in $\bar{w}_i$ is reduced two parts above and below $\bar{w}_t$. All other information in $\bar{w}$ is lost. Secondly, in many practical cases, it is in conflict with reality to presume that a class of specimens can be represented by a fixed proportion of events. Thirdly, it is not necessary to select n a priori. Wald [4] has demonstrated that n_t is most efficiently computed during the analysis of the specimen, as was also recognised in [2]. Finally, chosing a cell classifier prior to and separately from the establishment of the specimen classifier is not optimal, as will be demonstrated later on.

An alternative approach was followed by Tanaka [5], acquiring the values of one cell feature in a cumulative histogram $\hat{q}(v)$. The procedure that was used, was as follows:

1. One cell feature, v, is selected.
2. In the learning phase, derive histograms $q_K(v)$ representing the K-th specimen class.
3. In the operational phase, establish after the analysis of n cells, the specimen histogram $\hat{q}_n(v)$, and apply the non-parametric Kolmogorov Smirnov test for the difference of two histograms.
   For K=1,2:
   $$v_t = v \text{ where max } | \hat{q}_n(v) - q_K(v) | \tag{4a}$$

$$n_t = \frac{(f_{KS}(E_K^S))^2}{(\hat{q}_n(v_t) - q_K(v_t))^2} \qquad (4b)$$

$$\text{if } n > n_t \text{ assign I to class K} \qquad (4c)$$

else assign I not to a class.

In this decision rule $f_{KS}(E_K^S)$ is a factor based on the confidence level E_K^S of the Kolmogorov Smirnov test. The values of f_{KS} may be found in standard statistical tables [11].

The analysis with a non-parametric test, only permits the use of a one dimensional cell feature. Non-parametric tests on the difference of two distributions, do not exist for more than one dimension, as the more dimensional space cannot be uniquely ordered.

More dimensional cell features ($\bar{w}$) can be used still, by using non-linear mapping (M) of the cell features onto what is called the axis of cell atypia. The index (v)

$$v = M * \bar{w} \qquad (5)$$

winds through the feature space, e.g. according to [6] or [7]. In reference [6], the axis is a result of a curve fitting procedure through a number of ordered cell classes represented by their covariance matrices, derived from a learning phase. The ordering of the cell classes permits the definition of a scale, called index of atypia, on the axis. De facto, this axis can be regarded as a way to order the cell feature space and classes, giving a meaning to the feature values in the context of the classification problem. The index of atypicality is an ordinal scale, and can on its own be regarded as a one dimensional cell feature. In [6], no hint is given how to use the index in a specimen classification procedure.

A conceptually different way to reduce the more dimensional cell feature space into a one dimensional cell scale, is making use of a posteriori probability of a cell classifier. Let M now denote a (set of) cell classifier(s), and let v now denote the resulting a posteriori probability of a cell to belong to cell class k=1, then v also is specified by equation 5. The feature v is an ordinal scale cell feature. For some problems, for M the Mahalanobis distance to the class of normal cells may be taken as a measure for the deviation of the cell from the normal [9]. In this concept and the previous one, a cell feature v is the residue from the more dimensional feature $\bar{w}$ by an operation M.

Ott [8] also applies a mapping from the more dimensional cell space onto a triangulated cell class space. Specimen features are defined on this cell class space as input to the specimen classifier. No hints are, however, given how to generally define specimen features or how to relate the number of cells to be analysed on a specimen to the classification error.

In the present paper, it is pointed out how the distribution of v relates to the specimen classification error by the introduction of the population function, see also [10].

2. THE POPULATION FUNCTION

Consider an arbitrary specimen I of specimen class K. For a one dimensional cell feature v, which may be the result of a mapping ($v = M * \bar{w}$), call the probability density function p(v). A convenient form to consider is the cumulative density function

$$q(v) = \int_{v}^{\infty} p(x)dx. \qquad (6)$$

It is interesting to relate q(v) to a cell classifier v' of the following type, see figure 2.

$$\begin{aligned} &\text{if } v_i < v' \text{ assign cell i to class k=1,} \\ &\text{else } v_i \geq v' \text{ and assign cell i to class k=2.} \end{aligned} \qquad (7)$$

In this context, q(v') denotes the fraction of cells that will be classified into class k=2 (events), when v' would have been used as the cell classifier. The fraction non-events accordingly is 1-q(v'). When v' is varied over the feature scale, q(v) denotes the performance for any cell classifier v on the given specimen.

For one cell classifier v', the fraction of events may vary from specimen to specimen, even within one class of specimen. This fraction q(v') thus is a variable itself, for which within one class of specimens K, a pdf may be formed: $P_K(q(v'))$. This function indicates the probability to observe a fraction q(v') of events when analysing a specimen of class K with cell classifier v'. Obviously the shape of $P_K(q(v'))$ is completely dependent on the nature of problem and is independent of q(v). Variation of v' over the feature

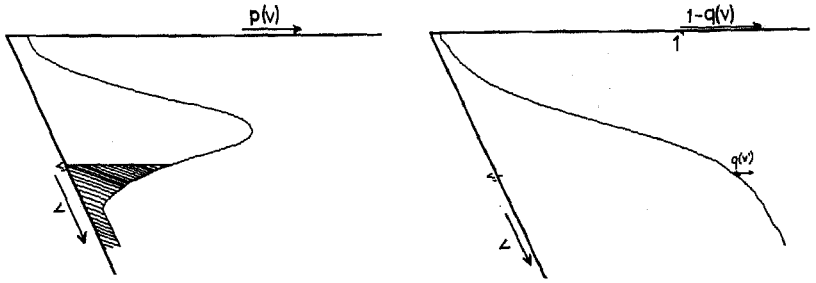

Figure 2: The proportion of events (cells classified into k=2) resulting from a cell classifier v' indicated by shading in the graph of p(v) and the cumulative function q(v).

scale yields the result for any feature: $P_K(v,q(v))$. This is the population function, giving a relationship between v and q(v) for population K. The integral form of P_K over q:

$$Q_K(v,q(v)) = \int_{0}^{q} P_K(v,x(v))dx \qquad (8)$$

relates an arbitrary cell classifier to the obtainable specimen classification error as follows. Consider a specimen classifier q' such that:

$$\begin{aligned} &\text{if } q_I < q' \text{ assign specimen I to specimen class K=1,} \\ &\text{else } q_I > q' \text{ and assign specimen I to class K=2.} \end{aligned} \qquad (9)$$

For a given cell classifier v' and a given specimen classifier q', $Q_K(v',q'(v'))$ denotes the probability of a specimen to be classified into class K=1. The function $1-Q_1(v,q(v))$ thus is the specimen classification error function for K=1-specimens, and the function $Q_2(v,q(v))$ is the specimen error-function for K=2-specimens.

The population function can be used in specimen classification procedures by considering for specimen class K=1,2, the population

functions $P_K(v,q(v))$ as illustrated in figure 3a,b. For each value v' of v, the Bayes-rule implies that the specimen classifier with smallest total error is $q_B(v')$ at the point where $P_1(v',q(v')) = P_2(v',q(v'))$. Varying v' again over the feature scale gives a curve $q_B(v)$, indicating for each v the associated best specimen classifier. The cumulative population function is displayed in figure 3b. For each v', $q_B(v')$ now is positioned at the location of maximum vertical distance between $Q_1(v',q(v'))$ and $Q_2(v',q(v'))$. In figure 3c, the intersection of $q_B(v)$ with $Q_K(v,q(v))$ is indicated. The remaining specimen classification error, is indicated by $1-Q_1(v,q\ (v))$ and $Q_2(v,q\ (v))$. In the example of figure 3c, for the larger part of the feature v the population 1 is indistinguishable from population 2, and consequently on that part both errorfunctions have a value .5.

Several conclusions may be drawn from figure 3c. Firstly, there may be a part of the feature scale v where no or little discrimination is possible between class-K=1 and class-K=2 specimens. Secondly, for the part of the scale where discrimation is possible, in figure 3c denoted by the region v_{test}, the specimen classifiction error is $1-Q_1(v,q\ (v))$ for K=1 and $Q_2(v,q\ (v))$ for K=2. This error is due to the nature of the problem as expressed in the shapes of $Q_K(v,q(v))$ and can for this feature not be avoided. Thirdly, if desired, the single best cell classifier v_o can be selected by taking

$$v_o = \max_{v,q} \mid Q_1(v,q_B(v)) - Q_2(v,q_B(v)) \mid . \qquad (10)$$

As we will see later on, it may still be advantageous not to select a cell classifier a priori.

3. LEARNING PHASE

Until thus far, we have considered cell classifiers and specimen classifiers for the true population functions. In principle, we cannot dispose of the true $P_K(v,q(v))$. Instead we have to rely on estimations or simplifications, based on a limited series of observations $\hat{q}_I(v)$, $I=1,\ldots,N_K$ for K=1,2. These observations constitute the learning set of specimens.

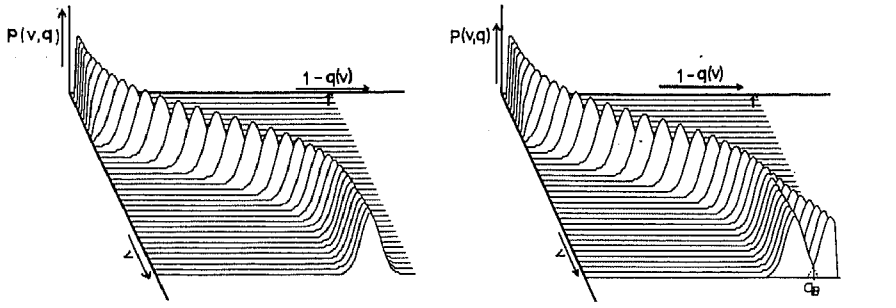

Figure 3a,b: The population function $P_K(v,q(v))$ shown left for one specimen class. For the ease of display $P(v,q(v))$ has been chosen here unimodal, Gaussian shaped for all cross-sections over v. Note that, in general, not all cross-sections of $P(v,q(v))$ are necessarily Gaussian shaped or even unimodal, nor have a similar shape. On the right side a pair of joint probability functions $P_K(v,q(v))$ is given. In the cross-section at the front, the Bayes specimen classifier for that cross-section is indicated by $q_B(v)$.

A rigorous simplification of the population function as used in [1-3] has already been discussed in the introduction. They use two independent simplifications. First, one cell classifier is selected, reducing $P_K(v,q(v))$ to a one dimensional function. Secondly, as $q(v)$ is assumed to be equal for all specimens of the same class, $P_K(v,q(v))$ then further reduces to a delta-pulse. Especially the second simplification is in conflict with common classification problems. Note, that in the context of this model no overlap occurs between the approximations for the population functions. As a consequence, there is no preference for a specimen classifier qt on basis of the assumed properties of the two populations at hand other than $q_1<q_t<q_2$. The expression for q_t given in eq.(3) is derived on basis of the finite number of cells only; a subject we will come to discuss in the next section.

Another, much better estimation, $\hat{P}_K(v,q(v))$ for the population function is the following Parzen-estimation procedure [10]:

$$\hat{P}_K(v,q(v)) = \frac{1}{N_K} \sum_{I \in K} \frac{1}{S_I\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{q(v)-\hat{q}_I(v)}{S_I}\right)^2} \quad \text{with} \qquad (11)$$

$$S_I = \sqrt{\frac{\hat{q}_I(v)[1-\hat{q}_I(v)]}{N_I}} \qquad (12)$$

Having established $\hat{P}_K(v,q(v))$ at the end of the learning phase, $q_t(v)$ can be determined as an estimate for $q_B(v)$. $q_t(v)$ is implicitly defined by $\hat{P}_1(v,q_t(v))=\hat{P}_2(v,q_t(v))$. Analogously to the cumulative population function, $Q(v,q(v))$ gives error estimates for this specimen classifier. If E_K^B is the tolerated specimen classifier error fraction for class K due to overlap of $\hat{P}_1(v,q(v))$ and $\hat{P}_2(v,q(v))$, then a region v_{test} of v may be selected as follows:

$$v_{test} = v \text{ where } 1-Q_1(v,q_t(v))<E_1^B \text{ .and. } Q_2(v,q_t(v))<E_2^B \qquad (13)$$

If we could dispose of the cumulative probability density function $q_I(v)$ of an unknown specimen I, we could classify the specimen by comparing $q_t(v)$ to $q_I(v)$ for the region v_{test}. In that case, the error may be expected to be bounded by E_K^B. In reality, this specimen classification error is not the only source of error as shall be discussed in the next section.

4. SEQUENTIAL CLASSIFICATION

In the analysis of an unknown specimen I, a finite number n of cells

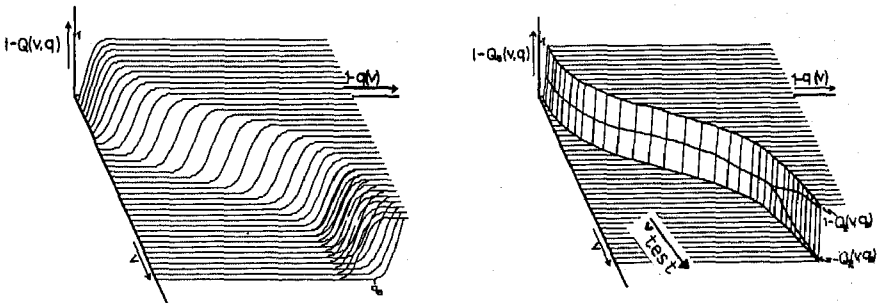

Figure 3c,d: The cumulative population fucntions $Q_K(v,q(v))$ on the left. The specimen classifier error as the intersection of $Q_K(v,q(v))$ with $q_B(v)$ on the right.

is analysed, resulting in a cumulative histogram of feature values $\hat{q}_n(v)$. The histogram is a sample of an unknown distribution $q_I(v)$. A specimen is said to be undecidable when $\hat{q}_n(v)$ may be a sample of $q_t(v)$ over the region v_{test}. In order to test the decidability a null hypothesis is formed stating that $\hat{q}_n(v)$ is a sample of $q_t(v)$. A hypothesis test on the difference of the two distributions is used. Two non-parametric tests are considered here.

The first to consider is the Kolmogorov Smirnov test. Let E_K^S be the tolerated specimen fraction to be classified wrong due to finite cell sample size. This factor then serves as the level of confidence in the test, see eq.(4). The specimen decision rule based on the KS-test is:

$$v_t = v \text{ where } \max_{v_{test}} | \hat{q}_n(v) - q_t(v) | \tag{14a}$$

$$n_t = \frac{(f_{KS}(E_K^S))^2}{(\hat{q}_n(v_t) - q_t(v_t))^2} \tag{14b}$$

if $\hat{q}_n(v) < q_t(v_t)$ .and. $n > n_t$ assign I to class K=1

if $\hat{q}_n(v) \geq q_t(v_t)$ .and. $n > n_t$ assign I to class K=2 (14c)

else specimen I is undecidable.

Note the difference between eq.(4) and eq.(14). The null hypothesis used in [5] was that $\hat{q}_n(v)$ originated from $q_1(v)$ or $q_2(v)$ which are example histograms of specimen class K=1 and K=2. In many problems, a class of specimens cannot reliably be represented by one example histogram. Doing so may lead to large classification errors. Alternatively, in eq. (14), $\hat{q}_n(v)$ is tested against $q_t(v)$, the critical function. When $\hat{q}_n(v)$ deviates significantly from $q_t(v)$, a decision can be taken with a known, bounded error probability.

A drawback of the KS-test is its aspecificity. Many samples are needed to demonstrate a significant difference. As a solution, the following test is proposed.

Consider one specimen class v'. Let the tolerated fraction of specimen to be classified wrong again be E_K^S. As $q_t(v')$ is a fraction, samples of $q_t(v')$ based on n observations (cells) are distributed binomially with mean $q_t(v')$ and standard deviation similar to eq.(12). With good approximation, the binomial distribution may be replaced by the Gaussian one with same mean and standard deviation. A relation now may be found between the value of $\hat{q}_n(v')$ needed to reject the hypothesis given $q_t(v')$, n and $f_e^{-1}(E_K^S)$, see eq.(15).

The same test holds, varying v' over v. In that case $q_I(v)$ is said to be decidable if one value v' is found for which $\hat{q}_n(v')$ significantly differs from $q_t(v')$.

The proposed test becomes:

For K=1,2

$$v_t = v \text{ where } \min_{v_{test}} n_t(v), \text{ with} \tag{15a}$$

$$n_t(v) = \frac{q_t(v)\,(1-q_t(v))\,(f_e^{-1}(E_K^S))^2}{(q_t(v) - \hat{q}_n(v))^2} \tag{15b}$$

if $n > n_t(v_t)$ assign specimen I to class K=1 if $\hat{q}_n(v_t) < q_t(v_t)$

if $n > n_t(v_t)$ assign specimen I to class K=2 if $\hat{q}_n(v_t) > q_t(v_t)$
else $n \leq n_t(v_t)$ and specimen I is undecidable.

The comparison of eq.(4) with eq.(14), can also be made for this test, comparing eq.(2) and (3) with eq.(15). Again, rather than comparing $\hat{q}_n(v)$ with one particular value $q_1(v)$ or $q_2(v)$ as was done in [1-3], $\hat{q}_n(v)$ here is compared with $q_t(v)$. In practice, it is likely to lead to smaller classification errors.

It should be noted that the above tests could be used to compute a value for n to analyze on all unknown specimens. It is more profitable, however, to apply eq.(14) or eq.(15) sequentially. In that case, the entire population classification becomes:

1. Define an operation M reducing the more dimensional cell feature space $\bar{w}$ to a one dimensional feature [6,9].
2. In the learning phase, estimate the population functions by eq.(11) and compute $q_t(v)$. Specify the tolerated error fractions due to specimen overlap and find v_{test}, eq.(13).
3. In the test phase, specify the tolerated error fractions due to the finite cell sample size. Update $\hat{q}_n(v)$ after the analysis of each cell and apply the following decision rule:
   Evaluate eq.(14) or eq.(15)
   If the specimen is decidable, stop the analysis,
   else proceed with the next cell.

It has been made plausible [10], that the procedure based on eq.(15) is both more efficient, in the sense that it needs fewer cells to arrive at a conclusion, and more accurate in the sense that it makes fewer errors in the classification, than the method of eq.(14). Also, this procedure is likely to be more efficient and accurate than the method based on eq.(2) and (3) [1-3], and more accurate than the method of eq.(4) [5]. Model studies are currently being done to investigate these conjectures.

5. CONCLUSION

In the foregoing, the population function $P_K(v,q(v))$ was introduced as a theoretical framework to describe the performance of cascade classifiers. Three items are important here, the cell (primary) classifier, the specimen (secondary) classifier and a rule prescribing how many cells need to be analysed before a specimen is decidable. The performance of a cascade classifier is established by its accuracy (specimen classification error) and its efficiency (number of cells to be analysed).

The population function is the equivalent of the probability density function for one-classifier problems. The equivalent of the Bayes-rule, optimizing overall accuracy, is for population classification given by eq.(10). Equation 10 will give the theoretical most accurate cell classifier selected prior to the analysis and the associated specimen classifier. This gives a theoretical solution to the problem raised in [1] and [2]. From eq.(10) it is also clear that the cell and specimen classifier are best optimized in combination, not separately, as was done in [1,2].

In practice, eq.(10) can only be approximated by a learning phase for which we propose the computation of the Parzen-estimation given in eq.(11), and the selection of a part of the cell feature as specified in eq.(13). The selection of this part of the feature implies that the specimen classification error is restricted to a bound.

In spite of eq.(10), it is not necessary (even unprofitable) to

select a cell classifier prior to the analysis, as was done in [1,2,3]. Bounding the tolerated specimen classification error to a specified fraction as proposed in eq.(13), fewer cells need to be analysed when using sequential classification techniques. The technique differs from previous ones in that the difference is demonstrated with the threshold function $q_t(v)$ between the two populations rather then the concordance with the examplatory population functions [3,6,5]. This is expected to lead to a higher accuracy. Presently, model studies are carried out to investigate the accuracy and effiency of the proposed models in comparison to the existing ones.

ACKNOWLEDGEMENT
Prof. A.R. Bakker is gratefully acknowledged for raising an intriguing question, and L. Dorst and G. Bosveld for critically reviewing the paper. The manuscript was carefully typed by D.I.M. de Jong.

REFERENCES

1. Castleman KR, White BS: The tradeoff of cell classifier error rates. Cytometry 1, 601-613, 1980.
2. Castleman KR, White BS: Optimizing cervical specimen classifiers. IEEE trans. PAMI 2, 451-457, 1980.
3. Castleman KR, White BS: The effect of abnormal cell proportion on specimen classifier performance. Cytometry 2, 155-158, 1981.
4. Wald A: Sequential analysis. Dover Publications, New York, 1947.
5. Tanaka N, Ikeda H, Ueno T, Mukawa A, Kamitsuma K: Field test and experimental use of CYBEST model 2 for practical gynecologic mass screening. J. Anal. Quant. Cytol. 1, 122-126, 1979.
6. Bartels P, Bibbo M, Richards DL, Sychra JJ, Wied GL: Patient classification based on cytologic sample profiles. I Basic measures for profile contructions. Acta Cytol. 22, 253-260, 1978.
7. Gelsema ES, Hunink M, Queiros CE, Timmers T: The use of correspondence analysis in the assessment of morphological changes during carcinogenesis. Cytometry 5, 463-468, 1984.
Submitted for publication in J. Anal. Quant. Cytol. 1985.
8. Ott R, Schurman J, Reinhardt ER, Bloss WH: Automated classification of cytological specimens based on features extracted from nuclei images. Pattern Recognition 13, 83-87, 1981.
9. Mayall B, Burger G: Personal communications.
10. Smeulders AWM: Pattern analysis of cervical specimens. Ph.D.-thesis, University of Leyden, The Netherlands, 1983.
11. Beyer WH et al: Handbook of probability and statistics. The Chemical Rubber co. Cleveland, 1968.

DISCUSSION

Timmers:
Do you see a way of going around the restriction that you work with only one feature?

Smeulders:
It should be noted that there is no possible way to define a Kolgomorov-Smirnov test in more than one dimension because you cannot order uniquely a two-dimensional space. This holds not only for this test but for any parameter-free distribution statistic. So basically the answer is no.

Kanal:
There was some work by Tukey on multivariant tolerance regions for nonparametric rank ordering.

Smeulders:
Basically you cannot order a two-dimensional space uniquely but I indicated that you use a multidimensional cell classifier to reduce the N-dimensional cell feature space to one dimension. This gives order in space and this is what you need. So you may either take an a posteriori probability of an arbitrary cell classifier or you may look at a (non)linear mapping of the multidimensional space. These are the ways to escape from this dilemma but essentially it cannot be solved.

Timmers:
Yes, but then you don't know whether it is optimal at the specimen level, isn't it?

Smeulders:
I believe it is optimal at least for the one-dimensional cell feature case. For the N-dimensional case it depends on the quality of the non-linear mapping.

Burger:
The reason why we think there is no escape in practice of single cell classifiers which precede the specimen classifier is that we would like to define the region of interest in the feature space. In many cases in cytometry a vast amount of the population is identical and is not worthwhile to be looked at with the same accuracy. So we apply an economic multivariate cell classifier in order to get rid of that part of the population with cheap features and then concentrate on that region in the feature space where we expect differences.

Smeulders:
I agree with you. In my terms this is the region where the two P-functions are different.