UvA-DARE (Digital Academic Repository)

The classification of (cell) samples using the population function

Smeulders, A.W.M.; Bosveld, G.; Duin, R.P.W.

Publication date
1986

Published in
Medinfo 86

[Link to publication](#)

The Classification of (Cell) Samples Using the Population Function

A.W.M. Smeulders, G. Bosveld, R.P.W. Duin

The Department of Medical Informatics, Free University, 1007MB Amsterdam, and the Department of Pattern Recognition, University of Technology, Delft, The Netherlands

Population classification is a set of statistical techniques to classify populations (e.g. samples) on basis of observations of their members (of cells). This decision making involves a cascade of classifiers: one for the members, and, based on the resulting classifications, one for the populations. Also a rule is to be given how many members should be analysed before a statistically sound decision on the population can be given. In search for a more integral view of the problem the population function is introduced. Consideration of the population function implies that rather than fixing the member classifier a priori, it is better to use an ordinal member classifier. The concept of sequential classification is use to reach the population decision efficiently. From model studies, the population classification performance is favourably compared to other techniques.

1. Definition of the problem.

Many instances of medical decision making can be perceived as the result of a partitioning of a feature domain. The feature values represent the input knowledge on which the classifier is working, and the classifier transposes a particular set of feature values into the name of the corresponding partition. A wide range of these classifiers can be described as a partitioning operation $\underline{M}$ working on a multidimensional feature space $\underline{W}$ with features $\underline{w}$, yielding a decision symbolized by a vector $\underline{v}$ from space $\underline{V}$: $\underline{v} = \underline{M} \circ \underline{w}$. The decision $\underline{v}$ may be a code or name (nominal classification), possibly from an ordered list (ordinal classification). Typical for this kind of situation is the fact that the decision rule $\underline{M}$ is directly working on the input feature values, as a reduction of them. Decision rules optimized for these situations are found applying the appropriate statistical or hierarchical pattern recognition procedures.

In this paper, we describe a basically different type of decision procedure: population classification. A decision now is made at the level of the population (sample), whereas the observations are made on the level of the members of the population. To take a simple example, when analysing a sample of cells, each cell first is to be recognised as a member of the normal or the abnormal cell class. The

decision on a cell will contribute each own little weight to the decision of the entire sample. It is the combination of the decisions of many cells which determines the decision on the entire sample. Compared to standard pattern recognition procedures now three decision rules are needed in stead of just one: One rule specifying how to classify the members into two (or more) member classes given a set of features, one rule which says how many members are to be analysed before a statistically sound decision can be made on the underlying sample and one rule which combines the member classifications into a decision for the entire population. This is a cascade of classifiers [1], the performance of which is closely interrelated. They should be taken into consideration simultaneously when developing optimized decision procedures, rather than optimizing each of these classifiers independently.
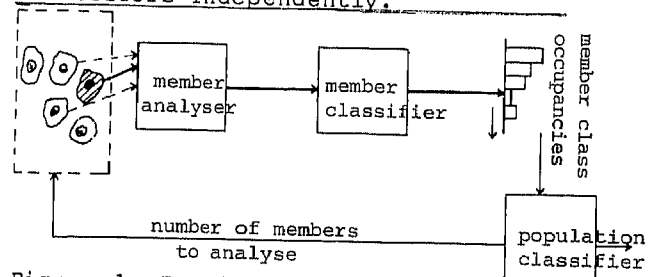

Figure 1: Population classification or a cell sample by a cascade of classifiers.

2. The population function.

Let the members of the I-th population be numbered by index i, with $i=1,\ldots,n_I$. To each member, there is a feature vector $\underline{w}_i$. The member decision rule makes an ordering in the (more) dimensional feature space by a (non-)linear projection $v = \underline{M} \circ \underline{w}$. Examples of such an ordering may be found in [2,3,4], though arbitrary other ones may be thought of. For v is a nominal decision, as is requested in the simplified case discussed in [1], much of the original information in the feature $\underline{w}$ is lost. It is preferable to take v ordinal, as is considered here. Let v have a range $v=1,\ldots,m$. Thus v_i is the ordered version of a feature vector $\underline{w}_i$. As described above, v_i can also be perceived as an ordinal class label assigned to member i. Classifying member by member of a population I, the resulting sum for each class is accumulated in the row $p_I(v)$. It is convenient to use the cumulative and normalised version of p_I by accumulating the sum of all classes $v \leq v'$ in v' and dividing each sum with n_I. $q_I(v)$ Is the cumulative density function of v on population I, and it is a representation unique for population I. Viewed upon otherwise, $q_I(v')$ indicates the proportion of members of the classes upto including class v' in that population I.

Populations of the same type form a population class and make up a distribution of rows q, called the population function $P(v,q(v))$, see [5]. Each class of populations will have its own population function $P_K(v,q(v))$, with $K=1,\ldots,M$ being the number of different population classes.

Operating on an unknown occupancy row $q_{I'}(v)$ of population I', the population decision rule will assign the underlying population into one of the population classes. For each member class upto including v', the Bayes-rule will give the optimal population decision rule $q_B(v')$, see figure 2. The shaded area is the remaining population classification error Q, making $q=q_B$ the population classifier with minimum error for that $v=v'$. For two populations ($M=2$) using the member bound v', the population decision rule assigns I' to $K=1$ for which $P_1(v',q(v')) > P_2(v',q(v'))$ and to $K=2$ for $P_1(v',q(v')) < P_2(v',q(v'))$ leaving the case where $P_1(v',q(v')) = P_2(v',q(v'))$ undecided. So varying v' over v, $q_B(v)$ is defined implicitly by: $P_1(v,q_B(v)) = P_2(v,q_B(v))$, giving for each v the optimal population decision criterion $q_B(v)$. As a sidestep, it should be noted that if one wishes to use only one member classifier, the same in the analysis of all populations, the optimal performance is obtained using member bound v_m with $v_m = v \mid min \{ Q(v,q_B(v)) \}$. This is the theoretical solution for the problem raised in [6]. Secondly it should be noted that for a given member bound v', the theoretical minimum population classification error can be predicted not to be less than $Q(v',q_B(v'))$. Thus one should not use the entire domain of v, but select a range v_t of v where $Q(v',q_B(v')) < E_t$, with E_t the tolerated minimum performance of the population classifier. The range v_t indicates for what feature values of the members P_1 may be discriminated from P_2 at all.

3. Operational rules.

In practice, analysing a population I', we cannot dispose of $q_{I'}(v)$ exactly but only an estimate $q_n(v)$ based on n observed members (for the moment omitting the index I'). Rather than comparing $q(v)$ with $q_B(v)$, we now are comparing $q_n(v)$ with $q_B(v)$. A very efficient way of doing so, using a minimal number of members to reach a decision, is to use a sequential procedure, similar to [7]. The null hypothesis is that $q_n(v)$ is a sample row drawn from $q_B(v)$. In that case no decision is taken on the population and a following member of I' is analysed, incrementing n. At the first n where $q_n(v)$ differs from $q_B(v)$ significantly, for some bound v', the population I' is classified to either class $K=1$ or class $K=2$, depending on which $P_K(v',q_B(v'))$
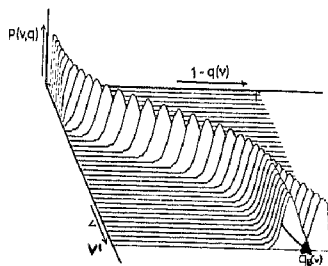

Figure 2: An example of two population functions $P_K(v,q(v))$, with the optimal specimen classifier $q_B(v)$ drawn in the frontmost value $v=v'$, and the residual error $Q(v',q_B(v'))$ indicated by shading. Note the limited range (v_t) where $P_1(v,q(v))$ differs sufficiently from $P_2(v,q(v))$, i.e. population 1 is distinguishable from population 2.

is maximum. For v is nominal, a x^2-test should be used. For v is ordinal a distribution test is used to demonstrate the significancy of the difference of $q_n(v)$ from $q_B(v)$. The Kolmogorov Smirnov test may be used, or the multiple nomial test [5]. To guarantee a minimum performance of the population classifier, the comparison of $q_n(v)$ with $q_B(v)$ should be limited to the range v_t, defined above.

Another, more simple procedure may be to fix v'=v" prior to the analysis [1]. The best possible member classifier for the discrimination of the populations, i.e. v_m defined above, is unlikely to be the best or most efficient member classifier in the analysis of each population. In the ordinal distribution comparisons, v' indicated the member classifier tailored to population I'. Therefore, fixing v' to v" is likely to lead to a larger classification error and a lesser efficiency of the population classifier.

Two types of population classification errors may occur. The first one is due to the nature of the problem, due to the overlap of $P_1(v,q(v))$ with $P_2(v,q(v))$. This is expressed by $Q(v,q(v))$ controlled by E_t. The other error comes from the fact that $q_n(v)$ only is an estimated for $q_T(v)$. This error may be reduced by analysing more members as is controlled by the level of confidence of the statistical test.

4. Learning phase.

To estimate an population function, there are good reasons to use the following Parzen estimator, [8]:

$$\hat{P}_K(v,q(v)) = \frac{1}{N_K} \sum_{I \in K} \frac{1}{s_I \sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

with $z = \frac{q(v)-q_n(v)/s_n}{\sqrt{q_n(v)(1-q_n(v))/n_I}}$ and $s_n = $ where $q_n(v)$ is the estimate of $q_T(v)$ based on the n members analysed from population I. N_K is the number of populations in population class K.

Note that doing so, nothing is to be known about the a priori class membership of the members of the learning set of populations.
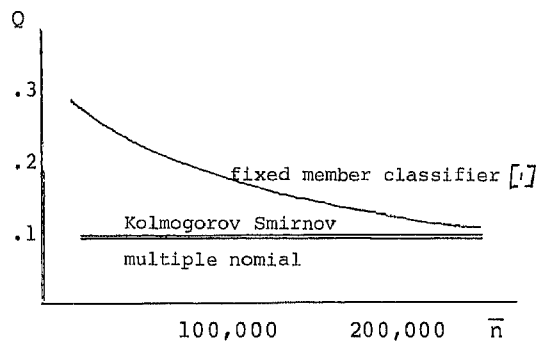
5. Experimental results.

To establish the difference in performance and efficiency of the above mentioned methods, Monte Carlo experiments have been performed [9] for the discrimination of two populations with a mixture of two member types. The

member types have overlapping feature distributions. One of the member types is rare in both population classes but with higher probability more present in K=2 than in K=1. The parametrisation of the experiment displayed in figure 3 was such that theoretically in experiment I an error of 8.4% was anticipated, and in experiment II of 3%.

Monte Carlo experiment I:
Two classes of populations. Each population has two types of members with overlapping feature values taken from uniform distributions. One member type is in majority (on average 99% of all members), one rare (on average 1% of all members). The probability of occurance of the two types is drawn from a uniform distributions. The circumstances are taken to simulate normal and abnormal cell samples [1,2,3,4,5].

Monte Carlo experiment II: Similar to experiment I, now using normal distributions.
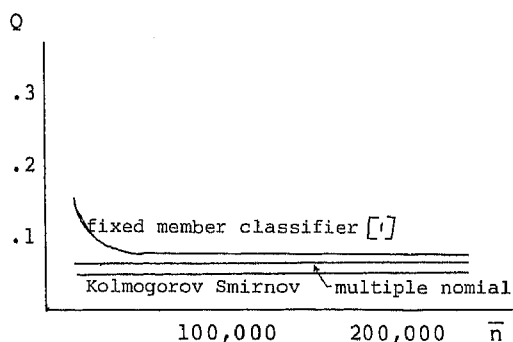

Figure 3: Experimental results from Monte Carlo procedures for two population classes with two member classes for three different methods.

The figure shows that, as exspected, while analysing more members (n) before coming to a conclusion will have a

positive effect on the classification error (Q). However, not all proposed methods are equally accurate in taking a decision on a limited number of members. The figure shows that the Kolmogorov Smirnov method and the multiple nomial test are superior for relatively small numbers of members compared to the procedure with fixed v'. This supports the theoretical considerations that specification of a member classifier a priori is inferior to not using a member classifier.

Alternative parametrisations of the Monte Carlo procedures are currently being investigated [9].

6. Conclusion.

This paper introduced the population function and its use in the classification of populations, such as samples of cells, leading to the following results.

1) From theoretical considerations, it is possible to formulate the best possible a priori member classifier v_m.

2) For common circumstances it is, however, better not to use an a priori member classifier (see figure 3). Rather, a distribution test is to be preferred as it were tailoring the member classifier to the population after the analysis.

3) Using the range v_t of Q, the error of the population classifier due to the overlap of the populations can be predicted and controlled by specification of E_t. The error due to the limited number of members analysed can be controlled by the confidence level of the statistical test.

4) It is not necessary to go through the elaborous work of assigning all members of the populations from the learning set used to formulate the decision rules. In stead, the crosssection of the population

function gives the best decision rule for discrimination of the population classes.

5) It should be noted that in the case of a more dimensional feature vector $\underline{w}$ for the members, the quality of the decision rule cannot be viewed apart from the projection $\underline{M}$.

7. References.

[1] Castleman KR, White BS: Optimizing cervical specimen classifiers, IEEE trans. PAMI 2, 451-457, 1980.

[2] Bartels P, Bibbo M, Richards DL, Sychra JJ, Wied GL: Patient classification based on cytologic sample profiles. I Basic measures for profile constructions. Acta Cytologica 22, 253-260, 1978.

[3] 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.

[4] Burger G, Juetting U: Specimen classification in cytometry: an intercomparison of various means of decision making. In: Pattern Recognition in Practice II (ES Gelsema and LN Kanal eds) pp. 509-519, Amsterdam, 1986.

[5] Smeulders AWM: Non parametric population classification. In: Pattern Recognition in Practice II (ES Gelsema and LN Kanal eds) pp. 497-507, Amsterdam, 1986.

[6] Castleman KR, White BS: The tradeoff of cell classifier error rates. Cytometry 1, 601-613, 1980.

[7] Wald A: Sequential analysis. Dover Publications, New York, 1947.

[8] Smeulders AWM: Pattern analysis of cervical specimens. PhD Thesis, Leyden State University, 1983.

[9] Bosveld G: Classificatie van populaties, MSc-thesis, Delft, 1986.