Candide's Practical Principles of Experimental Pattern Recognition

George Nagy
University of Nebraska-Lincoln

Correspondence

Candide's Practical Principles of Experimental Pattern Recognition

GEORGE NAGY

Abstract—**This correspondence calls attention to several frequently used assumptions and techniques culled from the pattern recognition literature.**

Index Terms—**Classification, feature extraction, image processing, machine intelligence, pattern analysis, pattern recognition.**

The following items, which may be helpful to recently initiated acolytes of the art and science of pattern recognition, have been derived from an exhausting study of the pertinent literature of the past two decades. Specific references are omitted to avoid counterproductive arguments over priority, but the author waives any claim to originality of the following ideas.

I. Gaussian PDF's

According to the Central Limit Theorem, any feature may be presumed to be normally distributed if its mean and variance can be estimated from its empirically observed distribution.

Manuscript received June 17, 1982; revised August 9, 1982.
The author is with the Department of Computer Science, University of Nebraska, Lincoln, NE 68588.

Corollary: Noise is always additive, Gaussian, independent, and identically distributed. Variations in the patterns that do not fit this model, such as unwelcome data from a foreign population contaminating the pure Gaussian population, are acts of fate and can be ignored. Likewise, nonstationary phenomena should be attributed to *transducer artifacts* and may be ignored.

II. Statistical Dependence

Assume class-conditional independence between features. Generally this will minimize the predicted error rate. In image data, texture is the name of a defect in the samples which interferes with the assumption of statistical independence among neighboring pixels. Models of statistical dependence stronger than those required for texture are called structural or syntactic.

III. Classification Errors

Sampling, quantization, segmentation, and registration errors have no bearing on classification performance. Therefore, subtract their effect from the observed error rate to obtain the number of *real* errors.

Corollary: A recognition algorithm is considered successful if it demonstrates, by means of *highly promising* classification results (>50 percent), that the system would be practical provided *only* that the data acquisition method can be improved sufficiently. Such improvements fall, however, in the realm of engineering and are below the dignity of the dedicated practitioner of pattern recognition.

0162-8828/83/0300-0199$01.00

IV. Comparison of Classification Accuracies

Comparisons against algorithms proposed by others are distasteful and should be avoided. When this is not possible, the following *Theorem of Ethical Data Selection* may prove useful.

Theorem: There exists a set of data for which a candidate algorithm is superior to any given rival algorithm. This set may be constructed by omitting from the test set any pattern which is misclassified by the candidate algorithm.

Toussaint's Corollary: For every classification algorithm, there exists an optimal probability distribution function for generating the data to be classified.

V. Replication of Experiments

Since pattern recognition is a mature discipline, the replication of experiments on new data by independent research groups, a fetish in the physical and biological sciences, is unnecessary. Concentrate instead on the accumulation of novel, universally applicable algorithms.

Casey's Caution: Do not ever make your experimental data available to others; someone may find an obvious solution that you missed.

VI. Representative Training/Test Sets

To estimate the expected classification accuracy in the field, construct an appropriate training set by extracting random samples from a suitably selected *homogeneous* test set. In remote sensing classification problems, systematically eliminate artifacts such as border pixels from both the training set and the test set. If small-sample estimation gives rise to problems, use the same samples for training and testing.

VII. Asymptotic Error Rate

The asymptotic error rate provides a firm upper bound on the experimentally observed error rate. In *real* problems the empirically estimated Bayes risk is zero unless identical samples are sometimes labeled A and sometimes B. Therefore, nearest neighbor algorithms, which have a lower bound on the asymptotic error rate proportional to the Bayes risk, are optimal. Upper bounds on error probabilities, even with values greatly exceeding unity, may be readily constructed by multiple applications of standard inequalities.

VIII. Computational Complexity of Classification

1) The only acceptable criteria for *concrete* computational complexity are the wall-clocktime and the number of statements in your program.

2) To a first approximation, all classification algorithms run at a speed proportional to N, the number of patterns to be classified. Therefore, to increase the speed, reduce the experimental sample size.

3) Estimate the speed of the candidate algorithm under the assumption that it will be reprogrammed in language X on parallel processor Y. Assume that rival algorithms have been fully optimized already.

Corollary: The rival algorithms may be reprogrammed, if necessary, to run slower.

IX. Multiclass Generalization

Since all multiclass tasks consist of a set of pairwise decisions, arbitrary dichotomies may be used to estimate the overall error rate. To do so, always use the most easily separable class pair: in OCR, "A" versus "B"; in remote sensing, "emerging corn" versus "sea ice." It is also helpful to collapse similar categories, such as "O" and "Q" in OCR.

X. Clustering

Clustering the training patterns guarantees significant performance improvement without additional cost. Choose the criterion and algorithm for clustering completely independently from the method used in subsequent classification. Omit reporting unnecessary details, such as initial conditions.

XI. Adaptive Classification

For proper statistical design, do not alter the training set and the test set between experiments. Feedback from the errors on the test set may be used to adjust higher order parameters of both training and classification algorithms. This technique of adapting the classifier to the test set is superior to adaptation on the training set, particularly with small test sets. "Throwaway" test sets are an affectation of statistical sophisticates.

Corollary: Report the improvements resulting from introducing specialized procedures for coping with significant misclassifications observed on the test set by considering the earlier results as due to a *rival* (inferior) algorithm.

Acknowledgment

The author is greatly indebted to a number of colleagues, and to the Editor and referees of this Transactions for helping to clarify several of the points addressed in this correspondence.