

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

CSE Journal Articles

Computer Science and Engineering, Department  
of

---

3-1983

## Candide's Practical Principles of Experimental Pattern Recognition

George Nagy

*University of Nebraska-Lincoln*

Follow this and additional works at: <https://digitalcommons.unl.edu/csearticles>



Part of the [Computer Sciences Commons](#)

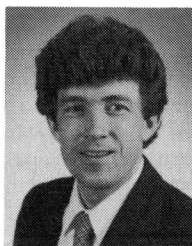
---

Nagy, George, "Candide's Practical Principles of Experimental Pattern Recognition" (1983). *CSE Journal Articles*. 3.

<https://digitalcommons.unl.edu/csearticles/3>

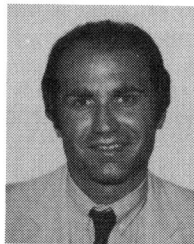
This Article is brought to you for free and open access by the Computer Science and Engineering, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in CSE Journal Articles by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

- [11] B. A. Dolan, "A generalized Mellin transform for the analysis of learning machines," Ph.D. dissertation, Stanford Univ., Aug. 1965.
- [12] M. McDonnell, "A clarification on the use of the Mellin transform in optical pattern recognition," *Opt. Commun.*, vol. 25, pp. 320-322, June 1978.
- [13] P. Baudelaire, "Linear stretch-invariant systems," *Proc. IEEE*, vol. 60, pp. 467-468, Apr. 1974.
- [14] G. Gambardella, "The Mellin transforms and constant-Q spectral analysis," *J. Acoust. Soc. Amer.*, vol. 66, pp. 913-915, Sept. 1979.
- [15] P. W. Hawkes, "A note on inverse filtering for anisoplanatic systems with coherent illumination," *Pattern Recognition*, vol. 7, pp. 59-60, 1975.
- [16] Anon., *Programs for Digital Signal Processing*. New York: IEEE Press and Wiley, 1979.
- [17] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
- [18] I. S. Gradshteyn and I. W. Ryzhik, *Tables of Integrals, Series and Products*. Associated Press, p. 30, eq. 1.342, 1965.
- [19] D. Casasent and D. Psaltis, "Multiple-invariant space-variant optical processors," *Appl. Opt.*, vol. 17, pp. 655-659, Feb. 1978.
- [20] D. Psaltis and D. Casasent, "Optical correlation of functions distorted as  $f(x') = f(x^a)$ ," *Opt. Commun.*, vol. 21, pp. 307-310, May 1977.
- [21] D. Casasent and D. Psaltis, "Space-bandwidth product and accuracy of the optical Mellin transform," *Appl. Opt.*, vol. 16, p. 1472, June 1977.
- [22] G. Kopec, "Relating algorithms and architectures," *Trends and Perspectives in Signal Processing* (published by Signal Processing Resources, Inc.), vol. 1, pp. 6-8, July 1981.



Philip E. Zwicke (S'70-M'77) was born in Sacramento, CA, on July 24, 1949. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from Virginia Polytechnic Institute and State University, Blacksburg, in 1971, 1973, and 1978, respectively.

From 1974 to 1975 he was a digital design engineer with NORD Instruments, Roanoke, VA. Since 1978 he has been with United Technologies Research Center, East Hartford, CT. His major interests are in digital signal processing, pattern recognition, and adaptive systems.



Imre Kiss, Jr. was born on July 11, 1948. He received the B.S. degree in chemical engineering from the University of Connecticut, Storrs, and the M.S. degree in electrical engineering from the University of Bridgeport, Bridgeport, CT, in 1974 and 1979, respectively.

He has worked as a chemical engineer at the American Cyanamid Research Center, Stamford, CT. From 1979 to 1982, he worked as a research engineer at Norden Systems Division of United Technologies, Norwalk, CT. In 1982 he joined the staff at United Technologies Research Center, East Hartford, CT. His current area of interest is automatic target classification.

## 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.

#### 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. "Throw-away" 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.

### A Step Towards Unification of Syntactic and Statistical Pattern Recognition

K. S. FU

**Abstract**—The problem of pattern recognition is discussed in terms of single-entity representation versus multiple-entity representation. A combined syntactic-semantic approach based on attributed grammars is suggested. Syntax-semantics tradeoff in pattern representation is demonstrated. This approach is intended to be an initial step toward unification of syntactic and statistical approaches to pattern recognition.

**Index Terms**—Attributed grammar, control diagram, semantics, statistical pattern recognition, syntactic pattern recognition, syntax-semantics tradeoff.

#### I. INTRODUCTION

Many mathematical methods have been proposed for solving pattern recognition problems [1]. They can be grouped into two major approaches, the decision-theoretic or statistical approach and the structural or syntactic approach [1]–[6]. From the point of view of pattern representation or description, we can discuss pattern recognition in terms of single-entity representation versus multiple-entity representation, and suggest a combined syntactic-semantic approach on the basis of using attributed languages.

Consider an *m*-class pattern recognition problem. When we consider each pattern as a single entity we can use a set of *n* characteristic measurements (features) to represent each pattern under study. In such a case, each pattern is represented by an *n*-dimensional feature vector and the recognition of patterns can be accomplished by applying various tech-

Manuscript received June 24, 1982; revised September 14, 1982. This work was supported by the National Science Foundation under Grant ECS 81-19886.

The author is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907.