Multiresolution Estimates of Classification Complexity

Sameer Singh, *Member*, *IEEE*

Abstract—In this paper, we study two measures of classification complexity based on feature space partitioning: "purity" and "neighborhood separability." The new measures of complexity are compared with probabilistic distance measures and a number of other nonparametric estimates of classification complexity on a total of 10 databases from the University of Calfornia, Irvine, (UCI) repository.

Index Terms—Pattern recognition, classification complexity, feature space partitioning.

1 Introduction

A number of approaches have been used to compute the classification complexity of data sets with two or more classes [18] as such analysis can be used for several pattern recognition applications. An example application includes feature selection, where the aim is to select a subset of features that minimizes the overall classification complexity. A bibliography of studies on the study of misclassification error estimation appears in Toussaint [42] and for the effects of dimensionality, sample size, and structure of classification algorithm on misclassification see [34], [40]. The approaches include

1. Bayes error-based parametric and nonparametric approaches [15] (use of probability distance measure bounds [7], [45], entropy measures [4], [47], nonparametric estimation including k nearest neighbor [6], [10], [25], [46], and Parzen estimation [28], interclass distance measures [9], [13], [41], [47], probability distances such as Bhattacharya [2], Chernoff [5], etc., for multiclass problems [1], [14]). See [13], [43] for a criticism of such approaches;
2. scatter matrices [7], [13];
3. information-theory-based approaches [26], [44];
4. boundary methods [30], [31], [32], [35], [36];
5. correlation-based approaches [33];
6. nonparametric methods [17], [20]; and
7. feature space partitioning methods [24].

The above approaches to estimating class separability are very different to each other in terms of their methodology, assumptions, and computational complexities. The main emphasis for practical purposes is on selecting those approaches that best correlate with classifier test errors [19]. It is important to note that it is not unreasonable to assume that more than one measure may be needed to fully quantify the true classification complexity of a problem and how to combine such information on a given data set remains a research problem.

2 Multiresolution Estimates of Data Separability

In this paper, we propose a novel set of classification complexity measures that complement the established set of measures. These measures have been discussed in greater detail in [39] under the acronym of Pattern Recognition using Information slicing Method (PRISM) that is based on the concept of feature space partitioning. Data partitioning has been of interest in pattern recognition, computer graphics, parallel computing, databases, and other fields for estimating data density [8], modeling 3D objects including texture [29], indexing [3], etc. A number of approaches for feature space partitioning have been suggested in the past including Octrees [11], [21], [22], [27], hyperboxes [24], specification of decision boundaries using decision trees and Simpson's min-max approach [37] and fuzzy ART/ARTMAP approaches, and pyramid strategy [3]. Our partitioning algorithm assigns each data point uniquely to a cell. The partitioning algorithm is based on a simple scheme of generating hypercuboids in d dimensional space. The data points are assigned to the bins by simple integer division (or rounding with the ceiling or floor functions). The main reasons for choosing this simple yet efficient scheme are: 1) The partitioning scheme does not need to be optimized for a given problem such that results are uniformly comparable across different problems and 2) The scheme does not require any parameter setting and does not introduce experimenter bias. A note on the difference between multiresolution partitioning used here and partitioning with decision trees is in order here. The multiresolution separability approach uses regular partitioning as opposed to decision trees that are based on adaptive partitioning. Decision tree based partitioning itself can be based on optimizing measures of impurity such as *entropy*, *variance*, *gini*, or *misclassification* [8]. In decision trees, one of the main research issues is how to control the complexity (size) of the tree. The stopping criteria for splitting may be based on cross-validation, change in

- *The author is with the Department of Computer Science, University of Exeter, Exeter EX4 4PT, United Kingdom. E-mail: s.singh@ex.ac.uk.*

Manuscript received 2 Oct. 2002; revised 1 Apr. 2003; accepted 26 June 2003. Recommended for acceptance by T.-K. Ho.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 117495.

impurity, or trading complexity for test accuracy. This would finally lead to a set of cells that differ in size and each would mostly contain pure data of only one class. It will be of benefit to study our multiresolution estimates of classification complexity with each resolution, which instead of being defined regularly, is defined by a decision tree split. For future work, we acknowledge that more sophisticated schemes including those based on kernel methods, meshes, and decision trees must be compared.

We introduce two new measures of classification complexity called "Purity" and "Neighborhood Separability." The basic idea is to compute these measures cumulatively by partitioning data space at various resolutions where each resolution is defined by the number of partitions per feature. If we partition the data space for resolutions $B = 0 \dots y$, where $y \geq 0$ is a user-set parameter, we find that as the resolution increases, the data space contains many more cells than at the lower resolution, and each cell has less amount of data than before. For most practical problems, we find that, at the highest resolution, most cells contain a single data point. In our experiments, computation is performed at different resolutions from $B = 0$ (no partitioning) to a higher resolution of up to $B = 31$ (up to 32 cells per axis).The basis of the proposed separability measures is that for each resolution we compute how separable data within each cell is. These cell measurements are then linearly weight summed for a single estimate. The weight applied to estimates of a cell is proportional to the number of elements in it. Also, the overall measurement across all cells at a given resolution is exponentially weighted to give more weight at less number of partitions. For each measure, the area under the curve (separability versus resolution) defines the overall data separability that is well bounded within the [0,1] range; for *purity* and *neighborhood separability*, the higher the measure the more separable the data is. The proposed measures are conceptually uncorrelated to the number of features, classes, or number of data points in a given data set. Therefore, two classification tasks with different values for these parameters can be directly compared.

The measure *purity* defines how pure the data is. If a cell contains data from only one class, then it is totally pure and if it contains data from a number of classes in equal amounts it is then totally impure. The basic idea is to determine the probability of all classes c_i, $1 \leq i \leq K_\ell$ in a cell H_ℓ, as

$$p_{i\ell} = \frac{\lambda_{i\ell}}{\sum_{i=1}^{K_\ell} \lambda_{i\ell}},$$

where $\lambda_{i\ell}$ is the number of data points of class c_i available in that cell. The purity of the cell is defined as

$$S_{H(\ell)} = \sqrt{\left(\frac{K_\ell}{K_\ell - 1}\right) \sum_{i=1}^{K_\ell} (p_{i\ell} - 1/K_\ell)^2},$$

where K_ℓ is the number of classes in cell H_ℓ. If cell H_ℓ contains N^ℓ points out of a total of N, then the overall *purity* of the classification problem can be estimated as a weighted average of the *purity* of different cells. This is given as:

$$S_H = \sum_{\ell=1}^{H_{total}} S_{H(\ell)} \cdot \frac{N^\ell}{N}.$$

If this measure is being computed at a resolution B, then it is weighted by a factor of $w = \frac{1}{2^{(B)}}$, for $B = (0, 1, \dots, 31)$. This is to give larger weights to lower resolutions. We can plot S_H versus normalized resolution curve and compute the area under the curve as the *purity* of the classification problem. A detailed algorithm and discussion on its properties is available in [39].

The main limitation of the *purity* measure is that it does not reflect classification complexity in the context of class boundaries. Singh [39] introduced another measure of estimating classification complexity called "nearest neighbor separability." The basic idea is to find, for each data point in a cell, the proportion of its nearest neighbors that come from the same class. We have for cell H_ℓ, $1 < \ell \leq H_{total}$, for class c_i, $1 \leq i \leq K_\ell$ a total of $\lambda_{i\ell}$ data points given by $(x_1, x_2, \dots, x_{\lambda_{i\ell}})$. Let the total number of data points be N^ℓ of all classes in the cell H_ℓ and N in total. Say, for example, if we wish to find the proportion p_{kj} of neighbors of data point $x_j \in c_i$, within a neighborhood of k data points. We find that p_{kj} will decrease with increasing k. We can estimate ϕ_j as the area under the curve that plots p_{kj} against k for sample x_j. An average proportion $\hat{\phi}$ for all data points is now computed and their weight averaged across cells to give an overall nearest neighborhood separability at a given resolution. This estimate at multiresolutions is then treated as earlier with purity (by defining the area under the curve of multiple resolution estimates) to obtain a final estimate of the complexity of the problem. An algorithm and discussion on its properties is available in [39].

The main intuitive appeal behind the proposed set of measures is the manner in which they are derived through recursive feature space partitioning. In addition to determining the boundary-based complexity of data using a nearest neighbor approach (as measured by boundary methods and minimum spanning trees), they also implicitly measure data compactness and distance between distributions (as measured by inter/intracluster distance ratios) under a unified framework. In addition, the proposed methods have a very few parameters setting and, therefore, form a more generic framework for different classes of data. Fig. 1 shows these measurements on random data. We create a total of 50 data sets containing two features and two classes. Each dataset has 100 observations; 50 for each class. The results show that for all data sets, the purity is quite low (< 0.2), and the nearest neighbor separability is around 0.5 which is as expected since data points are randomly assigned to one of the two classes. For a well-separated two-class classification problem, these measurements would take a value of 1.0.

3 Baseline Methods

In this paper, we compare the multiresolution separability-based classification complexity measures with well-known and established measures including the probability distance based measures including Bhattacharya B_D, Chernoff C_D, Divergence D_D, Mahalanobis M_D, Matusita m_D, Class Discriminability Measure (CDM) [24], and those suggested

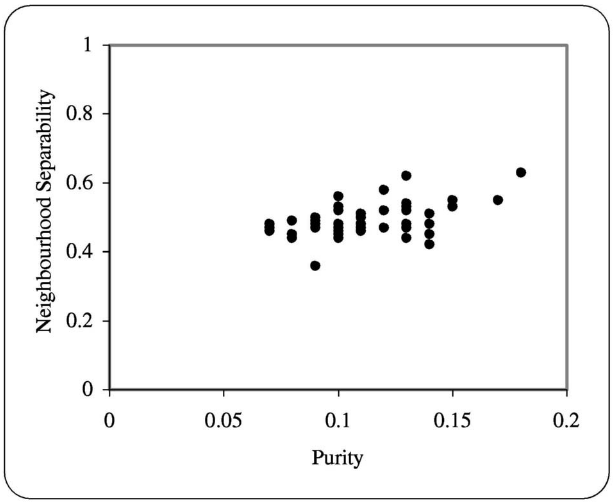

Fig. 1. The estimates of "purity" and "neighborhood separability" measures on 50 randomly generated data sets.

in [20] including the Fisher discriminant ratio ($F1$), Volume of Overlap ($F2$), Feature Efficiency ($F3$) [16], Minimum spanning Tree (MST) [12], Inter/Intra cluster distance (SW/SB), and Space Covered by Epsilon Neighborhoods ($SE_{\in}$) [18]. Pairwise estimates for the above techniques are weighted by class sizes and averaged.

4 Experimental Set-Up

We consider a total of 10 databases from the University of California, Irvine, (UCI) repository (http://www.ics.uci.edu/~mlearn/MLRepository.html)—see Table 1. For each data set, we compute the two proposed measures of "purity" and "neighborhood separability." Figs. 2 and 3 show their computation as the area beneath the curve after each point on the curve is weighted. It is important to note that there is no relationship between the proposed measures and the number of classes or features in the UCI data sets used. The correlation between the number of classes and AS_H is equal to .017 ($p = .96$) and between number of classes and AS_{NN} is equal to .20 ($p = .57$). The correlation between the number of features and AS_H is equal to .497 ($p = .14$) and between number of features and AS_{NN} is equal to .161 ($p = .66$).

For each data set, we also compute baseline estimates of classification complexity. In order to determine the relative efficiency of different classification complexity measures, we compute the correlation between the classification of four well-known classifiers with all of the above approaches. The classifiers considered in this study include the Least mean square Linear Discriminant Analysis (LDA), Quadratic Discriminant Classifier (QDC), k nearest neighbor classifier (kNN), and Decision Tree (C5.0) classifier. The separability measure with the highest correlation with these errors is best representative of the complexity of the classification task.

5 Results

In Table 1, we show the five probabilistic separability measurements on the 10 databases (columns 3-7). It should be remembered that the larger the value of the distance measure, the more separable data is. We also show seven more baseline measures for comparison (columns 8-14) and our proposed measures (columns 15-16). The interpretation of these measures is as follows: For measures $F1$, $F3$, MST, $SC_{\in}$, AS_H, and AS_{NN}, the higher the measure, the more separable the data is. On the other hand, for measures $F2$, SW/SB, and CDM, the lower the measure, the more separable the data is. These measures have been originally defined to show separability between two classes. For a multiclass problem, we use a weighted average across different class combinations as our estimate. We next perform leave-one-out cross validation with our chosen four classifiers. Linear and quadratic discriminant analyses are extended for multiclass problem by defining c discriminant functions for c classes [8]. The best reported rates are only available for the following data sets at the UCI repository webpage: Abalone: 65 percent, Wisconsin: 94 percent, Ecoli: 81 percent, Pima: 76 percent, Wine: 100 percent, and Yeast: 55 percent.

In Table 2, we show the correlation between the classification errors of these four classifiers with the five probability-based class separability measures, seven nonparametric baseline measures, and the proposed two measures of "purity" and "neighborhood separability." We only show the magnitude of the correlation and the

TABLE 1
The UCI Data Composition (Property Is Defined as Features (Classes)-Samples) and Nonparametric Classification Complexity Measures for UCI Data

Data	Property	B_D	C_D	D	m_D	M_D	$F1$	$F2$	$F3$	MST	SW/SB	$SC_{\in}$	CDM	AS_H	AS_{NN}
Abalone	7(3)-4176	.058	.034	.525	.165	.414	1.07	.061	.501	.567	.740	.004	.142	.287	.327
Balance	4(3)-625	0.302	0.246	5.57	0.325	2.07	.249	1.0	0.0	.792	.855	.064	0.00	.495	.792
Bupa	6(2)-345	0.059	0.043	0.518	0.16	0.159	.055	.260	.579	.623	1.00	.008	.243	.223	.629
Ecoli	7(8)-336	.063	.057	1.36	.137	.022	2.60	.694	.743	.812	.589	.102	0.00	.530	.807
Glass	5(7)-214	1.00	0.925	34.97	0.462	5.20	3.05	.117	.695	.705	.719	.106	0.00	.410	.677
Iris	4(3)-150	4.48	4.34	77.00	0.46	34.88	28.66	.114	.500	.960	.433	.538	.033	.384	.959
Pima	8(2)-768	2.77	2.12	55.22	0.675	13.79	.576	.251	.651	.679	.881	.008	.131	.430	.694
Wine	13(3)-178	0.088	0.054	0.756	0.182	0.428	8.13	.001	.564	.769	.674	.165	.117	.576	.862
Wisc	9(2)-683	2.66	1.71	52.94	0.464	17.06	3.56	.217	.350	.959	.300	.520	.015	.667	.970
Yeast	6(8)-1484	1.08	1.11	18.44	0.320	4.44	1.44	.487	.793	.528	.888	.005	.004	.380	.520

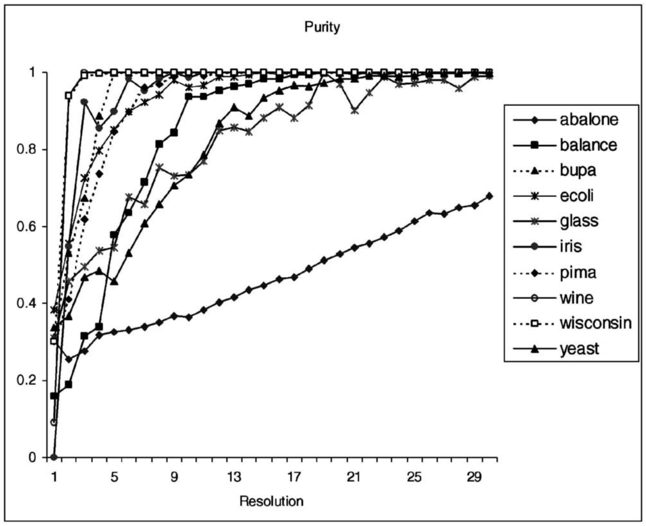

Fig. 2. The calculation of the measure “purity” as the area under the curve of S_H versus the resolution for the 10 UCI classification tasks.

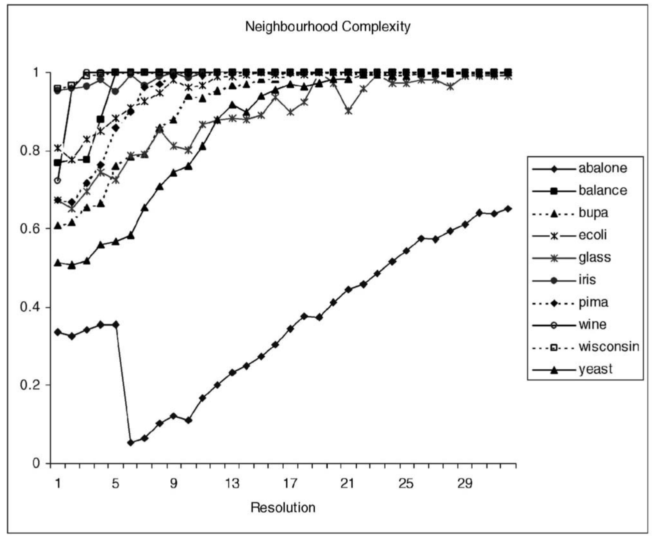

Fig. 3. The calculation of the measure “neighborhood separability” as the area under the curve of S_{NN} versus the resolution for the 10 UCI classification tasks.

statistical significance. All correlations found to be statistically significant at the 5-percent-level have been highlighted. The results show that the two proposed measures correlate best with linear and quadratic classifier's training and test errors compared to any other measure. For Gaussian classifier errors, AS_{NN} correlates the best followed by AS_H and MST. Finally, for the kNN generalization error, MST correlates the best followed by AS_{NN} and $SC_{\in}$. The MST method is likely to fail in those cases where data of two classes is separated by narrow margins which was not the case with our data sets. We also find that all of these correlations are statistically significant. Finally, for the decision tree C5.0 training error, AS_{NN} is best correlated followed by MST and $SC_{\in}$. For decision tree C5.0 test errors, the SW/SB metric best correlates followed by $SC_{\in}$,

TABLE 2
The Correlation Across the Different Measures of Classification Complexity

	Train error (LDA)	*Test error (LDA)*	*Train error (QDC)*	*Test error (QDC)*	*Test error (kNN)*	*Train error (C5.0)*	*Test error (C5.0)*
B_D	.130 (*p*=.36)	.139 (*p*=.35)	.274 (*p*=.22)	.282 (*p*=.21)	.438 (*p*=.10)	.312 (*p*=.38)	.540 (*p*=.11)
C_D	.040 (*p*=.44)	.077 (*p*=.41)	.225 (*p*=.26)	.233 (*p*=.25)	.381 (*p*=.14)	.295 (*p*=.41)	.491 (*p*=.15)
D	.074 (*p*=.42)	.071 (*p*=.42)	.218 (*p*=.27)	.200 (*p*=.29)	.411 (*p*=.12)	.331 (*p*=.35)	.543 (*p*=.11)
m_D	.012 (*p*=.48)	.044 (*p*=.45)	.070 (*p*=.42)	.041 (*p*=.45)	.213 (*p*=.28)	.157 (*p*=.66)	.325 (*p*=.36)
M_D	.170 (*p*=.32)	.206 (*p*=.28)	.335 (*p*=.17)	.346 (*p*=.16)	.507 (*p*=.07)	.363 (*p*=.30)	.559 (*p*=.093)
$F1$	.221 (*p*=.27)	.304 (*p*=.19)	.411 (*p*=.12)	.417 (*p*=.11)	.428 (*p*=.11)	.449 (*p*=.19)	.585 (*p*=.07)
$F2$	.091 (*p*=.39)	.082 (*p*=.41)	.161 (*p*=.33)	.161 (*p*=.33)	.261 (*p*=.23)	.008 (*p*=.98)	.393 (*p*=.26)
$F3$	.456 (*p*=.09)	.576 (*p*=.04)*	.454 (*p*=.09)	.500 (*p*=.07)	.533 (*p*=.05)*	.128 (*p*=.72)	.009 (*p*=.98)
MST	.743 (*p*=.007)*	.738 (*p*=.007)*	.856 (*p*=.001)*	.816 (*p*=.002)*	***.964*** (*p*=.000)*	.778 (*p*=.008)*	.762 (*p*=.01)*
SW / SB	.523 (*p*=.06)	.544 (*p*=.05)*	.637 (*p*=.02)*	.617 (*p*=.03)*	.717 (*p*=.01)*	.509 (*p*=.13)	***.825*** (*p*=.003)*
$SC_{\in}$	.511 (*p*=.06)	.564 (*p*=.04)*	.633 (*p*=.02)*	.626 (*p*=.02)*	.759 (*p*=.005)*	.638 (*p*=.047)*	.767 (*p*=.01)*
CDM	.103 (*p*=.38)	.161 (*p*=.33)	.337 (*p*=.17)	.282 (*p*=.21)	.424 (*p*=.11)	.305 (*p*=.39)	.357 (*p*=.31)
AS_H	.785 (*p*=.005)*	.758 (*p*=.004)*	.809 (*p*=.002)*	.770 (*p*=.005)*	.673 (*p*=.016)*	.594 (*p*=.07)	.733 (*p*=.016)*
AS_{NN}	***.794*** (*p*=.003)*	***.762*** (*p*=.000)*	***.877*** (*p*=.000)*	***.821*** (*p*=.002)*	.893 (*p*=.000)*	***.930*** (*p*=.000)*	.750 (*p*=.013)

TABLE 3
The Magnitude of Correlation Coefficient and Significance Levels between Classifier Errors and Classification Complexity Estimates

	B_D	C_D	D	m_D	M_D	$F1$	$F2$	$F3$	MST	SW/SB	$SC_\in$	CDM	AS_H	AS_{NN}
B_D	1.0	.98	.97	.77	.97	.69	-.28	-.02	.55	-.52	.77	-.23	.14	.49
C_D	.98	1.0	.95	.70	.98	.78	-.27	.02	.51	-.47	.68	-.24	.03	.45
D	.97	.94	1.0	.85	.93	.61	-.31	.01	.53	-.51	.68	-.27	.17	.48
m_D	.77	.70	.85	1.0	.64	.19	-.13	-.03	.27	-.16	.30	-.25	.19	.29
M_D	.97	.98	.93	.64	1.0	.80	-.28	-.09	.63	-.58	.79	-.23	.11	.54
$F1$	.69	.78	.61	.19	.80	1.0	-.34	-.01	.59	-.55	.73	-.18	.02	.52
$F2$	-.28	-.27	-.31	-.13	-.28	-.34	1.0	-.39	.03	.25	-.27	-.43	.17	.08
$F3$	-.02	.02	.01	-.03	-.09	-.01	-.39	1.0	-.41	.13	-.25	.11	-.26	-.29
MST	.55	.51	.53	.27	.63	.59	.03	-.41	1.0	-.82	.87	-.42	.65	.92
SW/SB	-.52	-.47	-.51	-.16	-.58	-.55	.25	.13	-.82	1.0	-.89	.48	-.63	-.64
$SC_\in$	.72	.68	.68	.30	.79	.73	-.27	-.25	.87	-.89	1.0	-.35	.48	.76
CDM	-.23	-.24	-.27	-.25	-.23	-.18	-.43	.11	-.42	.48	-.35	1.0	-.56	-.37
AS_H	.14	.03	.17	.19	.11	.02	.17	-.26	.65	-.63	.48	-.56	1.0	.70
AS_{NN}	.49	.45	.48	.29	.54	.52	.08	-.29	.92	-.64	.76	-.37	.70	1.0

MST, and AS_{NN}. In general, we find that the following baseline methods perform reasonably well: MST, SW/SB, and $SC_\in$. The results of the probabilistic distance measures is rather disappointing with those of $F1$, $F2$, $F3$, and CDM.

It is important to comment on the generalizability of the results given a small sample size of 10 samples for the correlation study. The high correlation coefficients with significance level below 5 percent show that the result is not a fluke. The results quoted in Table 2 are based on Pearson's correlation coefficient which is mostly used on normally distributed data and is a preferred choice in most studies due to its higher accuracy in measuring correlations. We also measured Spearman's rank correlation (preferred for small samples, i.e., data with little knowledge of its distribution and assumed nonparametric) and found that the correlations follow the same order as listed in Table 2 but tend to be more pessimistic (lower) across the board. For further studies, we recommend the use of parametric statistics if the number of samples is greater $(n > 100)$. From Table 2, the most important inference one can draw is that there is a strong relationship between some of the classification complexity measures and classifier errors. Also, our results show that the order of such correlations is preserved with nonparametric rank-order statistical correlation. For further work, a comprehensive analysis of a large number of synthetic data sets can reveal true differences in classification complexity estimates with parametric statistics—for smaller data sets one must rely on nonparametric statistics and have a cautious approach in interpreting results (e.g., more emphasis should be placed in rank differences rather than absolute measurement differences). Finally, in Table 3, we show how much each of the measurements correlate among themselves, i.e., measure similar characteristics of data. Undoubtedly, some of the measures adopt a similar methodology and, therefore, measure similar things. For example, those methods based on probabilistic distances correlate well among themselves. Similarly, measures based on the concept of using nearest neighbor information and data compactness, such as MST, SW/SB, CDM, SC_in, AS_H, and AS_{NN}, correlate well with each other. Measures such as $F2$ and $F3$ describe how much each feature contributes to the separation of two classes in a very gross manner but they do not consider the joint effects of features. These measures do not seem to correlate well with any of the other measures. Our recommendation would be to use measures such as MST when computational cheapness is required and, otherwise, use proposed multiresolution estimates.

6 Conclusion

Our suggested measures can be used for maximization in a feature selection task and used to study the behavior of classifiers. In addition, the previous study by Kishore et al. [23] suggests that feature space partitioning can also be used for using multiple classifiers in subspaces. The use of multiresolution schemes for data, therefore, present a framework within which one can perform localized learning, feature selection, and outlier removal [38].

References

[1] M. Ben-Bessat, "Use of Distance Measures, Information Measures and Error Bounds in Feature Evaluation," *Handbook of Statistics*, pp. 773-791, 1982.
[2] A. Bhattacharya, "On a Measure of Divergence between Two Statistical Populations Defined by Their Probability Distributions," *Bull. of Calcutta Soc.*, vol. 35, pp. 99-110, 1943.
[3] C. Böhm, S. Berchtold, and D.A. Keim, "Searching in High Dimensional Spaces: Index Structures for Improving the Performance of Multimedia Databases," *ACM Surveys*, 2001.
[4] C.H. Chen, "On Information and Distance Measures, Error Bounds, and Feature Selection," *Information Sciences*, pp. 159-173, 1976.
[5] A. Chernoff, "Estimation of a Multivariate Density," *Annals of the Inst. of Statistical Math.*, vol. 18, pp. 179-189, 1966.
[6] T.M. Cover and P.E. Hart, "Nearest Neighbor Pattern Classification," *IEEE Trans. Information Theory*, pp. 21-27, 1967.
[7] P.A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach.* Prentice Hall, 1982.
[8] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification.* John Wiley, 2000.
[9] A. Fisher, *The Mathematical Theory of Probabilities.* John Wiley, 1923.
[10] E. Fix and J.L. Hodges, "Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties," Technical Report 21-49-004, US Air Force School of Aviation and Medicine, 1949.

[11] J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes, *Computer Graphics: Principles and Practice,* second ed. Addison Wesley, 1987.

[12] J.H. Friedman and L.C. Rafsky, "Multivariate Generalisations of the Wald-Wolfowitz and Smirnov Two Sample Tests," *The Annals of Statistics,* vol. 7, no. 4, pp. 697-717, 1979.

[13] K. Fukunaga, *Introduction to Statistical Pattern Recognition.* Academic Press, 1990.

[14] N. Glick, "Separation and Probability of Correct Classification Among Two or More Distributions," *Annals of the Inst. of Statistical Math.,* vol. 25, pp. 373-383, 1973.

[15] T.K. Ho and H.S. Baird, "Estimating the Intrinsic Difficulty of a Recognition Problem," *Proc. 12th Int'l Conf. Pattern Recognition,* pp. 178-183, 1994.

[16] T.K. Ho and H.S. Baird, "Pattern Classification with Compact Distribution Maps," *Computer Vision and Image Understanding,* vol. 70, no. 1, pp. 101-110, 1998.

[17] T.K. Ho, "Complexity of Classification Problems and Comparative Advantages of Combined Classifiers," *Lecture Notes in Computer Science,* no. 1857, pp. 97-106, 2000.

[18] T.K. Ho and M. Basu, "Measuring the Complexity of Classification Problems," *Proc. 15th Int'l Conf. Pattern Recognition,* pp. 43-47, 2000.

[19] T.K. Ho, "Data Complexity Analysis for Classifier Combination," *Proc. Second Int'l Workshop Multiple Classifier Systems,* pp. 53-67, 2001.

[20] T.K. Ho and M. Basu, "Complexity Measures of Supervised Classification Problems," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 3, pp. 289-300, Mar. 2002.

[21] G.M. Hunter, "Efficient Computation and Data Structures for Graphics," PhD thesis, Dept. of Electrical Eng. and Computer Science, Princeton Univ., 1978.

[22] C. Jackins and S.L. Tanimoto, "Oct-Trees and Their Use in Representing Three Dimensional Objects," *Computer Graphics and Image Processing,* vol. 14, no. 3, pp. 249-270, 1980.

[23] J.K. Kishore, L.M. Patnaik, V. Mani, and V.K. Agrawal, "Genetic Programming Based Pattern Classification with Feature Space Partitioning," *Information Sciences,* vol. 131, pp. 65-86, 2001.

[24] A. Kohn, L.G.M. Nakano, and M.O. de Silvas, "A Class Discriminability Measure Based on Feature Space Partitioning," *Pattern Recognition,* vol. 29, no. 5, pp. 873-887, 1996.

[25] D.O. Loftsgaardne and C.P. Quesenberry, "A Nonparametric Estimate of Multivariate Density Function," *Annals of Math. Statistics,* pp. 1049-1051, 1965.

[26] J. Maddox, "Complicated Measures of Complexity," *Nature,* vol. 344, p. 705, Apr. 1990.

[27] D. Meagher, "Geometric Modelling Using Octree Encoding," *J. Computer Vision, Graphics, and Image Processing,* vol. 19, no. 2, pp. 129-147, 1982.

[28] E. Parzen, "On Estimation of a Probability Density Function and Mode," *Annals of Math. Statistics,* vol. 33, pp. 1065-1076, 1962.

[29] K. Perlin and E. Hoffert, "Hypertexture," *Computer Graphics,* vol. 23, pp. 253-261, 1989.

[30] W.E. Pierson, "Using Boundary Methods for Estimating Class Separability," PhD thesis, Dept. of Electrical Eng., Ohio State Univ., 1998.

[31] W.E. Pierson, B. Ulug, S.C. Ahalt, J.L. Sancho, and A. Figueiras-Vidal, "Theoretical and Complexity Issues for Feature Set Evaluation Using Boundary Methods," *Proc. SPIE Conf. Automatic Target Recognition VII,* 1998.

[32] W.E. Pierson, B. Ulug, and S.C. Ahalt, "On the Use of Boundary Methods for Measuring Class Separability," submitted for publication.

[33] A.F.R. Rahman and M. Fairhurst, "Measuring Classification Complexity of Image Databases: A Novel Approach," *Proc. Int'l Conf. Image Analysis and Processing,* pp. 893-897, 1998.

[34] S. Raudys and V. Pikelis, "On Dimensionality, Sample Size, Classification Error, and Complexity of Classification Algorithm in Pattern Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 2, no. 3, pp. 243-252, 1980.

[35] J.L. Sancho, B. Ulug, A.F. Vidal, W. Pierson, and S.C. Ahalt, "Boundary Methods for Distribution Analysis," *Signal Processing and Comm.,* 1996a.

[36] J.L. Sancho, A.R. Figueiras, B. Ulug, W. Pierson, and S.C. Ahalt, "Distribution Analysis Using Boundary Methods," *Proc. IEE Conf. Maths in Signal Processing,* 1996b.

[37] P.K. Simpson, "Fuzzy Min-Max Neural Networks—Part1: Classification," *IEEE Trans. Neural Networks,* vol. 3, pp. 776-786, 1992.

[38] S. Singh and A.P. Galton, "Pattern Recognition Using Information Slicing Model (PRISM)," *Proc. 15th Int'l Conf. Pattern Recognition,* vol. 2, pp. 144-147, 2002.

[39] S. Singh, "PRISM—A Novel Framework for Pattern Recognition," *Pattern Analysis and Applications,* vol. 6, 2003.

[40] S.Y. Sohn, "Meta Analysis of Classification Algorithms for Pattern Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 11, pp. 1137-1144, Nov. 1999.

[41] C.W. Therrien, *Decision Estimation and Classification.* John Wiley, 1989.

[42] G.T. Toussaint, "Bibliography on Estimation of Misclassification," *IEEE Trans. Information Theory,* vol. 20, pp. 472-479, 1974.

[43] V.N. Vapnik, *Statistical Learning Theory.* Wiley, 1998.

[44] C.S. Wallace and D.M. Boulton, "An Information Theoretic Measure for Classification," *Computer J.,* vol. 11, no. 2, pp. 185-194, 1968.

[45] A.C. Williams, "A Comparative Study of the Gaussian Form of the Bhattacharya Metric for ATR Performance Evaluation," *Proc. SPIE Int'l Symp. Optical Eng.,* 1997.

[46] Q. Xie, C.A. Laszlo, and R.K. Ward, "Vector Quantization Technique for Nonparametric Classifier Design," *IEEE Trans. Pattern Analysis Machine Intelligence,* pp. 1326-1330, 1993.

[47] T.Y. Young and K.S. Fu, *Handbook of Pattern Recognition and Image Processing.* Orlando, Fla.: Academic Press, 1986.

Sameer Singh received the Bachelor of Engineering degree with distinction in computer engineering from Birla Institute of Technology, India, the MS degree in information technology for manufacturing from the University of Warwick, United Kingdom, and the PhD degree in speech and language analysis from the University of the West of England, United Kingdom. His main research interests are in image processing, medical imaging, neural networks, and pattern recognition. He is the director of the Pattern Analysis and Neural Networks (PANN) group at Exeter University. He serves as the editor-in-chief of the *Pattern Analysis and Applications Journal* by Springer, editor-in-chief of the Springer book series on advances in pattern recognition, and chairman of the British Computer Society Specialist group on Pattern Analysis and Robotics. He is an associate editor of journals including *IEEE Transactions on Systems Man and Cybernetics B*, *Pattern Recognition*, *Real-Time Imaging*, *Knowledge and Information Systems*, and *Neural Computing and Applications*. He is a fellow of the Royal Statistical Society and a member of the BMVA-IAPR, IEE, and IEEE.