DISCUSSION / DISCUSSION

Comment on "Acoustic seabed classification: improved statistical method"[1]

J.M. Preston and R.L. Kirlin

In a discussion of methods for acoustic seabed classification, Legendre et al. (2002) claim to offer improvements over existing techniques and assert that their method "produces statistically better results than the classification method implemented in the QTC [Quester Tangent Corporation] software". Reasons for this assertion are not given in that paper but are given in an unpublished document. In this paper, we examine the basis for the assertion and discuss whether it should be accepted.

The method of Legendre et al. (2002) implements a *K*-means partitioning by an iterative process. They claim this as an advantage over QTC; however, QTC also implements a *K*-means partitioning with an iterative process (see, e.g., Preston et al. 2001), so this cannot be the explanation. Choosing the optimal number of clusters can be problematic, but Legendre et al. report differences between their methods and QTC even when both are computing with the same number of clusters, so the explanation cannot lie entirely here. What is left?

The answer, it appears, is that Legendre et al. (2002) wish to minimize the within-group sums of squares using a homogeneous measure of distance to the centre of a cluster (squared Euclidean distances), whereas QTC uses a likelihood-based measure in which the distance to the centre of any cluster is scaled by the variance of that cluster. A one-dimensional example illustrates the point at issue. Suppose we have a population that is an equal mixture of two Gaussian distributions: A, which is distributed as $N(0,(0.5)^2)$, and B ~ $N(2,(0.1)^2)$. How do we assign an observation at 1.5? Legendre et al. would say that it is a distance of 0.5 from the centre of B and 1.5 from the centre of A and therefore should be assigned to B. QTC would say that is it 5 standard deviations from the centre of B and 3 standard deviations from the centre of A and therefore should be assigned to A. In the example that we have sketched, the statement "x is from A and x = 1.5" has a higher probability density than "x is from B and x = 1.5"; that is, our method has a higher probability of making a correct assignment to a component of the mixture. Therefore, one cannot unambiguously identify Legendre et al.'s "best in the [homogeneous] least-squares sense" with best in the statistical sense. Attempts to prove the latter fail because any statistical test is also based on either homogeneous or variance-based measures; if the clustering method and the test use the same measure, the test scores will often be higher for that reason alone.

The paragraphs above summarize our comment. Legendre et al.'s (2002) claim to "statistically better results" is without support, arising as it did from applying a test that used squared Euclidean distances to two clustering methods, one using squared Euclidean distances and one using a likelihood-based measure. What remains is to provide a justification for using non-Euclidean measures for clustering and to extend this example to more dimensions.

A reasonable principle for choosing the class assignment is the maximum a posteriori probability (MAP). In other words, a vector $\mathbf{x}$ is observed and is to be assigned to one of q classes $\{\omega_1,\omega_2,\ldots,\omega_q\}$. MAP selects the class of maximum a posteriori probability from the candidates as

$$(1)\qquad \omega_k = \arg\max_{k=1,2,\cdots,q} P(\omega_k|\mathbf{x})$$

where $P(\omega_k|\mathbf{x})$ is the a posteriori probability for class ω_k. From Bayes' theorem, the a posteriori probability can be written as

$$(2)\qquad P(\omega_k|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_k)P(w_k)}{p(\mathbf{x})}$$

where $p(\mathbf{x}|\omega_k)$ is the density of the data given that they are drawn from ω_k, $P(\omega_k)$ is the a priori probability for ω_k, and $p(\mathbf{x})$ is the marginal density of observed data $\mathbf{x}$. If the a priori probability $P(\omega_k)$ is uniform, the MAP rule becomes

$$(3)\qquad \omega_k = \arg\max_{k=1,2,\cdots,q} p(\mathbf{x}|\omega_k)$$

Received 10 December 2002. Accepted 15 May 2003. Published on the NRC Research Press Web site at http://cjfas.nrc.ca on 12 November 2003.
J17236

J.M. Preston.[2] Quester Tangent Corporation (QTC), 201 – 9865 West Saanich Road, Sidney, BC V8L 5Y8, Canada.
R.L. Kirlin.[3] Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC V8W 3P6, Canada.

[1]Appears in Can. J. Fish. Aquat. Sci. **59**: 1085–1089 (2002).
[2]Corresponding author (e-mail: jpreston@questertangent.com).
[3]Retired.

 doi: 10.1139/F03-131

which is the maximum likelihood (ML) selection. Because $p(\mathbf{x})$ does not depend on class assignments, it drops out of the selection computations and each vector is assigned to cluster ω_i if (Cover and Hart 1967)

$$(4) \qquad P(\omega_i)p(\mathbf{x}|\omega_i) > P(\omega_j)p(\mathbf{x}|\omega_j), \ \forall j \neq i$$

It is usual to assume a multivariate normal density (Fraley and Raftery 1998):

$$(5) \qquad p(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{M/2}|\mathbf{C}_i|^{1/2}} \exp[-1/2(\mathbf{x}-\mathbf{m}_i)^T\mathbf{C}_i^{-1}(\mathbf{x}-\mathbf{m}_i)]$$

where M is the dimensionality, and $\mathbf{m}_i$ and $\mathbf{C}_i$ are the estimated mean and covariance of cluster w_i. Legendre et al. (2002) and QTC both cluster in three-dimensional space called Q space, thus $M = 3$.

We can straightforwardly estimate the mean, covariance, and prior probability of the ith class if the marginal density of x,

$$(6) \qquad p(\mathbf{x}) = \sum_{i=1}^{q} P(\omega_i) \frac{1}{(2\pi)^{M/2}|\mathbf{C}_i|^{1/2}} \exp[-1/2(\mathbf{x}-\mathbf{m}_i)^T\mathbf{C}_i^{-1}(\mathbf{x}-\mathbf{m}_i)]$$

can be appropriately simplified. The required simplification results when the clusters or modes of the Gaussian mixture are well separated. In that case $p(\mathbf{x})$ is well approximated in the region of the ith cluster by

$$(7) \qquad p(\mathbf{x}) = \frac{P(\omega_i)}{(2\pi)^{M/2}|\mathbf{C}_i|^{1/2}} \exp[-1/2(\mathbf{x}-\mathbf{m}_i)^T\mathbf{C}_i^{-1}(\mathbf{x}-\mathbf{m}_i)], \quad \mathbf{x} \in \omega_i$$

and we can assume that all samples of $\mathbf{x}$ assigned to cluster i have been drawn only from the approximate distribution in eq. 7. In this case, we can obtain ML estimates of the required parameters using only samples of $\mathbf{x}$ assigned to cluster i.

When selecting the maximum under eq. 4, it is convenient to minimize the logarithm of the reciprocal of the left-hand side, log being a monotonic function. This gives what may be called a Bayesian metric:

$$(8) \qquad d_i(\mathbf{x}) = -\log P(\omega_i) - \log(p(\mathbf{x}|\omega_i)) = -\log P(\omega_i) + ½\log|\mathbf{C}_i| + ½(\mathbf{x} - \mathbf{m}_i)^T \mathbf{C}_i^{-1} (\mathbf{x} - \mathbf{m}_i)$$

Two simplifications of eq. 8 deserve mention. The Euclidean metric, as used by Legendre et al. (2002), is the simplification of eq. 8 that assumes equal $P(\omega_i)$ and that the covariance matrices are equal, constant, and diagonal. Because assignments are based on minimization across classes, class-independent terms and constant factors are dropped, giving

$$(9) \qquad d_i(\mathbf{x}) = (\mathbf{x} - \mathbf{m}_i)^T(\mathbf{x} - \mathbf{m}_i) \quad \text{(Euclidean)}$$

Secondly, if we continue to use equal priors $P(\omega_i)$, but now allow the covariance matrices to differ but with approximately equal determinants, we have the weighted sum of squares metric

$$(10) \qquad d_i(\mathbf{x}) = (\mathbf{x} - \mathbf{m}_i)^T \mathbf{C}_i^{-1} (\mathbf{x} - \mathbf{m}_i) \quad \text{(Mahalanobis)}$$

The general case, using all terms of the Bayesian metric (eq. 8), requires estimates of the a priori probabilities, as well as the means and covariances of each cluster. QTC has implemented these estimates iteratively, with an outer loop for estimating priors and covariances and an inner K-means loop that adjusts assignments and cluster means. This is an optimal process under the MAP criterion, the only approximation being that in eq. 7 which approximates the whole density of $\mathbf{x}$ by its component due to class i alone, which is valid under the assumption of well-separated classes or mixture components.

For classifying regions of acoustic similarity, as part of acoustic sediment classification, we have described three metrics: the general case and two simplifications. Which gives optimal results? This will likely remain an open question. We have found very few statistical tests or clustering processes that work well with both simulated and real data sets and thus have come to believe that comparison with ground truth is the most meaningful basis for comparisons. However, it is unusual to have ground truth that is adequate in quantity and scope. For example, surface roughness can affect acoustic character but is destroyed when sampling with a grab. Over the years, the QTC implementation described above has repeatedly been found to give practical, useful, and accurate classes. Some recent examples are described by Morrison et al. (2001), Anderson (2001, 2002), and Ellingsen et al. (2002).

References

Anderson, J.T. 2001. Classification of marine habitats using submersible and acoustic seabed techniques. *In* Spatial processes and management of marine populations. Alaska Sea Grant College Program, AK-SG-01-02.

Anderson, J.T. 2002. Acoustic classification of marine habitats in coastal Newfoundland. ICES J. Mar. Sci. **59**: 156–167.

Cover, T.M., and Hart, P.E. 1967. Nearest neighbour pattern classification. IEEE Trans. Information Theory No. IT-13. pp. 21–27.

Ellingsen, K.E., Gray, J.S., and Bjornbom, E. 2002. Acoustic classification of seabed habitats using the QTC VIEW™ system. ICES J. Mar. Sci. **59**: 825–835.

Fraley, C., and Raftery, A.E. 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. Computer J. **41**: 578–588.

Legendre, P., Ellingsen, K.E., Bjornbom, E., and Casgrain, P. 2002. Acoustic seabed classification: improved statistical method. Can. J. Fish. Aquat. Sci. **59**: 1085–1089.

Morrison, M.A., Thrush, S.F., and Budd, R. 2001. Detection of acoustic class boundaries in soft sediment systems using the seafloor acoustic discrimination system QTC VIEW. J. Sea Res. **46**: 233–243.

Preston, J.M., Christney, A.C., Bloomer, S.F., and Beaudet, I.L. 2001. Seabed classification of multibeam sonar images. *In* Proceedings of MTS/IEEE Oceans 2001: An Ocean Odyssey, Honolulu, Hawaii, November 2001. Holland Publications, Escondido, Calif.