

印度洋水声环境特征自组织网络区划分型

王彦磊¹, 张 韧¹, 杨 波², 刘科峰¹, 余 鹏¹

(1 解放军理工大学气象学院, 江苏 南京 211101; 2 总参气象水文中心, 北京 100081)

摘 要: 利用 WOA05 世界大洋数据集, $1^{\circ} \times 1^{\circ}$ 分辨率、标准分层的气候态平均温、盐场资料, 提取一组表征印度洋海域水下声场环境的特征指标。借助一组聚类结果评估函数, 确定最佳区划分型数目, 利用自组织特征映射网络(SOFM)可以保持拓扑结构不变的特点, 对西太平洋海域水下环境进行区划分型。分类区划结果对该海域的声场背景分析和声纳探测具有参考应用意义。

关键词: 自组织网络; 水声场环境; 分类区划

中图分类号: P731 文献标识码: A 文章编号: 1003 - 0239 (2008) 2 - 0084 - 08

1 引言

印度洋海域是我国石油资源生命线的必经之途。该海域具有十分复杂的地形、地貌和水文气象特征, 开展其海洋环境研究, 特别是影响声纳探测的海洋中上层声场环境的特征结构和区划分类更具有重要性。海水的温、盐和声物理属性是各种水声传播模型和声纳探测模型的重要输入因子, 声场的分布及变化是水下导航、通讯的基础和影响声纳探测的重要环境因素。周燕遐^[1]等选用世界大洋数据集(WOD98)调查资料, 对海洋跃层进行分析, 揭示了大尺度的大洋水体中跃层分布规律和变化特征; 陈奕德^[2]等利用 Argo (Array for Real-time Geostrophic Oceanography) 资料对太平洋进行跃层分析, 得出一些有意义的结论; Huang^[3]等利用船舶报资料对海洋气象水文要素进行分类, 提出船舶航线风险评估区划。目前水声研究中较多讨论的是在某一假定的理想环境条件下, 水声的传播规律, 海洋区划研究中主要是对温、盐场及跃层分布进行区划, 而基于实际海洋资料的声场环境的分型区划较少。

针对印度洋声场环境特性及水声研究需要, 根据自组织特征映射网络(Self-organizing feature map, SOFM 网络)不仅能够学习输入分布情况, 而且可以学习输入向量的拓扑结构的特点^[4]。充分利用自组织特征映射网络无监督学习的优势, 对印度洋环境声场特性进行区划分型; 但自组织特征神经网络必须给定分型数目, 为解决此缺陷, 可借助一组评价分类效果优劣的函数指标, 综合确定最佳分类数目。本文利用 WOA05 数据资料, 采用 SOFM 网络对印度洋海域的声场环境进行了水声背景分析和特征区划分型。

收稿日期: 2007-10-22; 修订日期: 2008-03-25

基金资助: 中国博士后科学基金(No. 2004036012)与江苏省博士后科研资助计划(0401068B)。

作者简介: 王彦磊(1979-), 男, 博士在读, 从事海洋声场环境信息处理。

2 自组织特征映射神经网络

自组织特征映射网络是一种竞争式学习网络^[5],在学习中能无监督地进行自组织学习。该网络是由芬兰赫尔辛基大学网络专家 Kohonen 教授 1981 年提出,这种网络模拟大脑神经系统自组织特征映射的功能。SOFM 网络结构由输入层和竞争层组成。它是一种竞争学习网络,通过引入二维网格,在神经元之间建立了拓扑关系,神经元之间的联系是由它们在网格上的相互位置所决定的,并且这种联系模拟了人脑中的神经元之间的侧抑制功能,是网络实现竞争的基础;在学习过程中逐步缩小神经元之间的作用邻域,并用 Hebb 学习规则增强中心神经元的激活程度,则去掉各神经元之间侧向连接,实现“近兴奋远抑制”的效果,具有把高维输入映射到低维的能力。SOFM 网络另一个重要特点是拓扑保形特性,它能将任意输入模式在输出层映射成一维或二维离散图形,并保持其拓扑结构不变,即在无教师示教情况下,通过对输入模式的自组织学习,在竞争层将分类结果表示出来。此外,网络通过输入模式的反复学习,可以使连接权矢量空间分布密度与输入模式的概率分布区域一致,即连接权矢量空间分布能反映输入模式的统计特征。

自组织特征映射网络的算法,具体过程如下:对于输入模式 X ,首先确定中心神经元 M_c ,满足 $\|X - M_c\| = \min\{\|X_i - M_i\|\}$ 。然后,对以 M_c 为中心的周围的神元权值向量按下式进行调整:

$$M_i(k+1) = \begin{cases} M_i(k) + \alpha(k)[X - M_i(k)], & i \in N_c(k) \\ M_i(k), & i \notin N_c(k) \end{cases} \quad (1)$$

其中 N_c 表示由 M_c 为中心的周围神经元组成的领域。在学习过程中, $N_c(k)$ 的初始可选大些,然后逐步收缩,学习系数 $\alpha(k)$ 在初始时可取接近 1.0 的常数,然后逐渐变小。

3 最佳聚类数目确定

常规聚类算法中,需要预先知道对所分类的数据集的分类数目,而预先确定海洋声场环境的特征类数既很困难,也不符合客观实际。因此必须根据海洋环境数据集本身的拓扑结构,客观寻找其最佳的聚类数。当聚类中心数目未知时,一般有三种途径可以客观得到聚类数。第一种途径:设置一组阈值来控制新聚类类别的创建,对在线学习算法的应用比较合适,但它过分依赖数据加入的顺序;第二种途径:将数据集分成尽量细小的类别,类别数大于或远大于最佳聚类类别数,定义一组聚类效果评估函数(准则函数),设置函数的指标阈值,合并相似程度小于阈值的类别,重复以上步骤,直至达到最佳聚类数目,但该方法需要丰富的专家经验知识来确定指标阈值,具有一定局限性;第三种途径:尝试多种可能的聚类数目,并且根据定义的聚类效果评估函数指标,比较不同聚类数目下的准则函数值,选择聚类效果最优的聚类数目,即为最佳的聚类数。本文采取第三种途径以获取海洋声场的最佳分型数目。

在确定最佳聚类数过程中,用三组聚类结果评估函数来评估分类结果的优劣,第一类评估准则函数为 Partition Coefficient (PC)和 Classification Entropy (CE),表征分类结果的重叠和模糊程度。

$$PC(c) = \frac{1}{N} \sum_{j=1}^c \sum_{i=1}^N (\mu_{ij}^2) \quad CE(c) = -\frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N \mu_{ij} \log(\mu_{ij}) \quad (2)$$

第二类评估准则函数为 Partition Index (SC)、Separation Index(S) 和 Xie and Beni's Index (XB),它们是一个比率,指数最小值对应最佳分类数目,表征分类结果的致密性和可分离程度。

$$SC(c) = \frac{\sum_{i=1}^c \sum_{j=1}^N (\mu_{ij})^m \|x_j - v_i\|^2}{\sum_{i=1}^c N_i \sum_{k=1}^c \|v_k - v_i\|^2} \quad S(c) = \frac{\sum_{i=1}^c \sum_{j=1}^N (\mu_{ij})^m \|x_j - v_i\|^2}{N \min_{i,k} \|v_k - v_i\|^2} \quad XB(c) = \frac{\sum_{i=1}^c \sum_{j=1}^N (\mu_{ij})^m \|x_j - v_i\|^2}{N \min_{i,k} \|v_j - v_i\|^2} \quad (3)$$

第三类评估准则函数为 Dunn's Index (DI) 和 Alternative Dunn Index (ADI),表征聚类结果各类内紧致性及类间可分性的优劣。

$$DI(c) = \min_{i \in c} \left\{ \min_{j \in c, i \neq j} \left\{ \frac{\min_{x \in c_i, y \in c_j} d(x, y)}{\max_{k \in c} \{ \max_{x, y \in c_k} d(x, y) \}} \right\} \right\} \quad (4)$$

根据三角形三边定理 $d(x, y) \geq \kappa(y, v_j) - d(x, v_j)$ 可得:

$$ADI(c) = \min_{i \in c} \left\{ \min_{j \in c, i \neq j} \left\{ \frac{\min_{x \in c_i, y \in c_j} \kappa(y, c_j) - d(x, c_j)}{\max_{k \in c} \{ \max_{x, y \in c_k} d(x, y) \}} \right\} \right\} \quad (5)$$

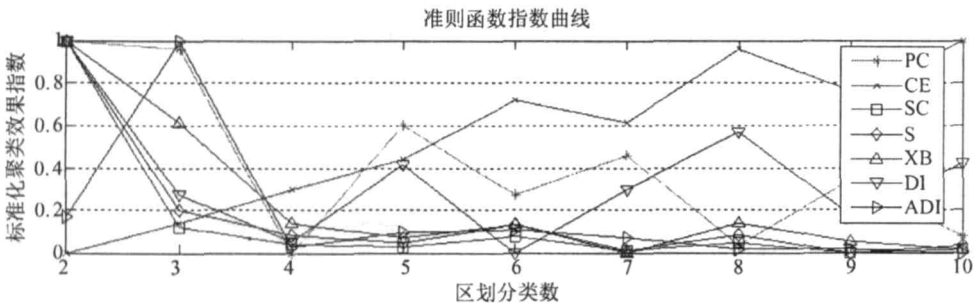


图 1 准则函数指数随聚类数目变化的曲线图

根据印度洋声场环境的复杂性,初步设定声场环境可能的最大分型数为 10,则可能的分类数目 c 为 2~10 中的 9 种可能,对其进行自组织网络分类,并利用三类分类效果的评估函数,计算 PC、CE、SC、S、XB、DI 和 ADI 评估指数(见图 1)。每一种指数只能表征一类有效性,不能够单凭单一指数值来确定分类数目。PC 和 CE 指数有随着分类

数单调变化的固有缺陷,不能单凭 PC 和 CE 指数确定分类数目;在聚类数为 4 时,更加有效的 SC 和 S 指数达到局域最小, XB 也达到比较小的值,同时 DI 和 ADI 指数也比较小;综合考虑三类评估指数,最佳分类结果对应得分分类数目为 4,即西太平洋声场环境的最佳区划分型数目为 4。

4 水声环境特征聚类分型

4.1 环境特征计算

本文采用世界大洋数据集 $1^\circ \times 1^\circ$ 分层的温、盐资料,并计算了海水声速和层结参数。借鉴水团分析中主要指标的计算方法^[6],提取了海水声速的均值、水平平均方差、垂直梯度和年变差的一组特征向量,用以刻画和表征海洋声场环境。

选取 $120^\circ \sim 160^\circ\text{E}$; $10^\circ \sim 50^\circ\text{N}$ 区域, 500m 以上的中上层水团作为研究对象,即选取 20、30、50、75、100、125、150、200、250、300、400、500m 12 层的温、盐数据,采用 Fofonoff^[11] 声速公式和 Fofonoff 绝热比密距平效正法(简称 F 法)计算每个温、盐点的声速值,运用 $5^\circ \times 5^\circ$ 滑动窗来求每一格点 (i, j, k) 的均值 $\bar{V}_{i,j}$ 和 \hat{V} 水平平均方差;利用上下两点的声速值之差除以两者之间的垂直距离得到垂直梯度 V'_k ;月声速距平绝对值的平均值即为该点声速年变差 \hat{V} 。由此得到每点的温度均值、温度垂直梯度、盐度均值、声速均值、声速水平方差、声速垂直梯度、声速年变差和层结参数 8 维特征向量,从而得到印度洋环境区划分型的特征矩阵 $P_{q,N}$ (其中 q 为格点数, N 为特征维数)。

$$\begin{aligned} \text{均值: } \bar{V}_{i,j} &= \sum_{m=i-2}^{i+2} \sum_{n=j-2}^{j+2} \frac{V_{m,n}}{5 \times 5}; \text{水平平均方差: } V_{ij} = \sum_{m=i-2}^{i+2} \sum_{n=j-2}^{j+2} \frac{(V_{m,n} - \bar{V}_{i,j})^2}{5 \times 5}; \\ \text{垂直梯度: } V'_k &= \frac{V_{k-1} - V_{k+1}}{d_{k+1} - d_{k-1}}; \text{年变差: } \hat{V} = \left(\sum_{y=1}^{12} (V_y - \bar{V})^2 / 12 \right)^{\frac{1}{2}}; \text{其中, } \bar{V} = \sum_{y=1}^{12} V_y / 12 \end{aligned}$$

4.2 自组织特征映射网络分类

特征矩阵 $P_{q,N}$ 作为 SOFM 网络的输入, q 为输入样本数, N 为竞争神经元数,即分类区划的类型数目,自组织特征映射网络分类步骤如下:

Step1 初始化:将网络的连接权值 $\{W_{ij}\}$ 赋予 $[0,1]$ 区间内的随机值,确定学习率 $\eta(0)=0.9$,邻域初始值 $N_g(0)=4$ 。

Step2 提供特征指标输入:将预报因子中的一个时次的指标 p_k 提供给网络的输入层,并进行归一化处理 $\bar{p}_k = p_k / \|p_k\|$ $k=1, 2, \dots, q$ 。

Step3 计算输入预报因子与每个输出神经元之间的距离:

$$\bar{w}_j = w_j / \|w_j\|, d_j = \left[\sum_{i=1}^N (\bar{p}_i^k - \bar{w}_{ji})^2 \right]^{\frac{1}{2}} \quad j=1, 2, \dots, M$$

Step4 找出最小距离 d_j 确定获胜神经元 g : $d_g = \min[d_j]$ 。

Step5 进行连接权值的调整,对竞争层邻域内的所有神经元与输入层神经元之间的连接权进行修正: $\bar{w}_{ji}(t+1) = \bar{w}_{ji}(t) + \eta(t)[p_i^k - \bar{w}_{ji}(t)]$, $j \in N_g(t)$, $j=1, 2, \dots, M$ 。

Step 6 选取另一个预报因子样本提供给网络的输入层, 返回步骤(3), 直至 q 个指数样本全部提供给网络。

Step 7 更新学习速率及邻域;

Step 8 令 $t=t+1$, 返回步骤(2), 直至 $t=S$ 为止。至此可以得到样本的隶属度矩阵。

5 区划分析

5.1 特征地理分布

基于 WOA05 世界大洋数据集温、盐格点资料, 计算温度垂直梯度、声速、声速水平变差、声速年变差等海洋环境特征。在 150m 深度层上, 其分布(见图 2), 温度垂直梯度高值区主要集中在 5°S 和 10°N 之间, 低值区在 30°S 以南; 声速的高值区主要有大洋西北部和 20°S 纬度带两个区域, 其间有一低值区, 30°S 以南也为低值区; 声速水平变差

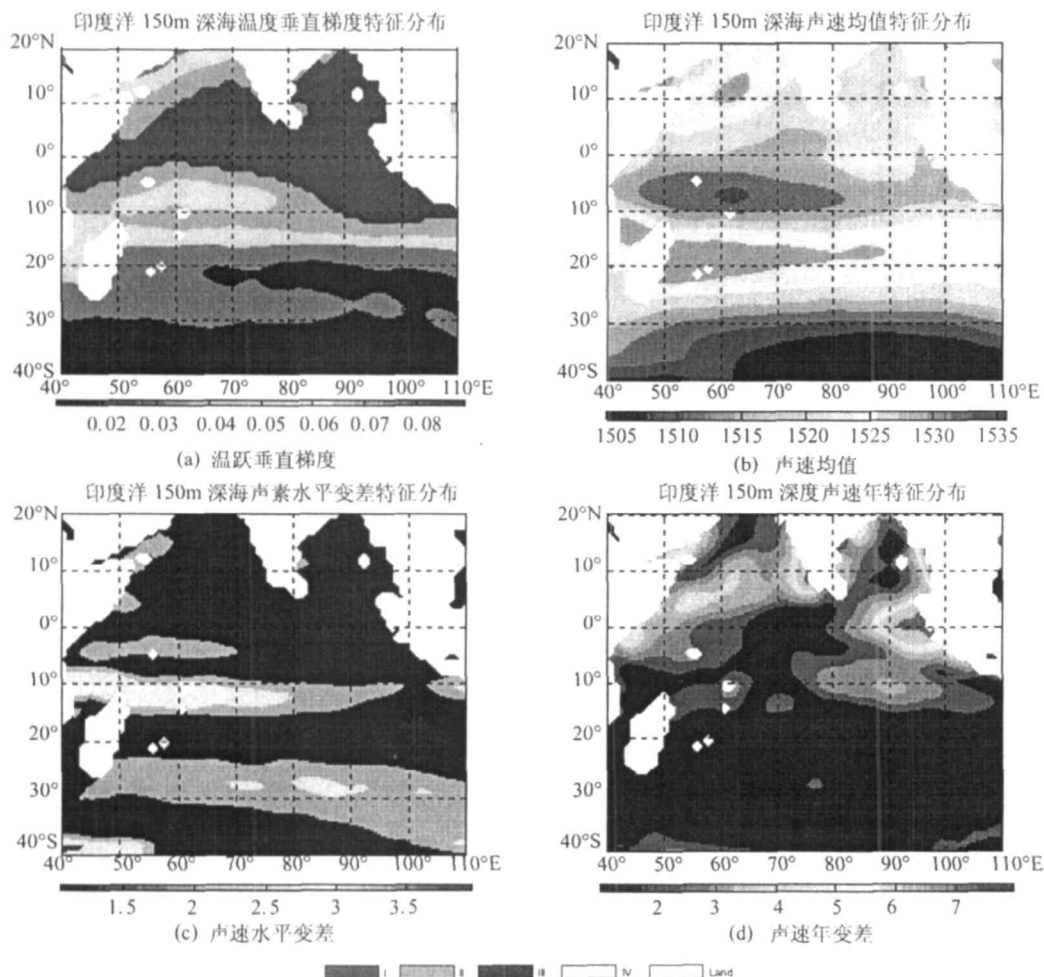


图 2 印度洋 150m 环境特征分布

最大的区域为 80°E 以西 10°S 纬度带;声速年变差较大的区域在大洋的北部。

5.2 特征分类与区划

运用上述自组织网络特征分类试验得到了印度洋水声环境区划分型结果,将其投影到印度洋环境特征空间上,其分布情况(见图 3)。水声环境区划分型在 150m 和 250m 深度层的分布见(见图 4),沿 20°S 纬向剖面 and 沿 60°N 经向剖面分布(见图 5)。从中分析可以得出,四类区划分型的基本特征如下:

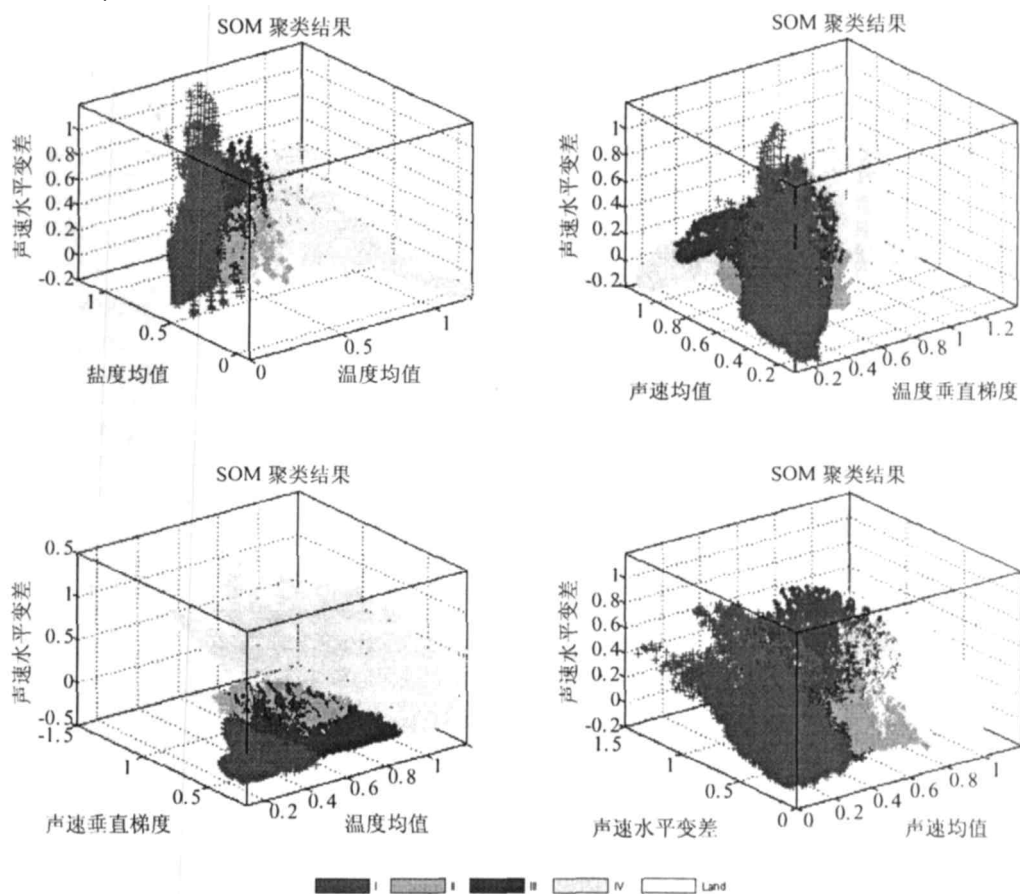


图 3 分类结果在特征空间投影

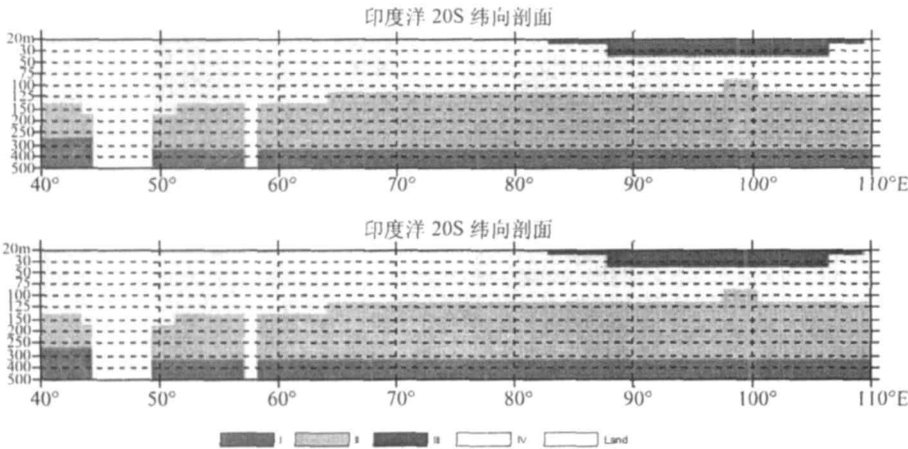
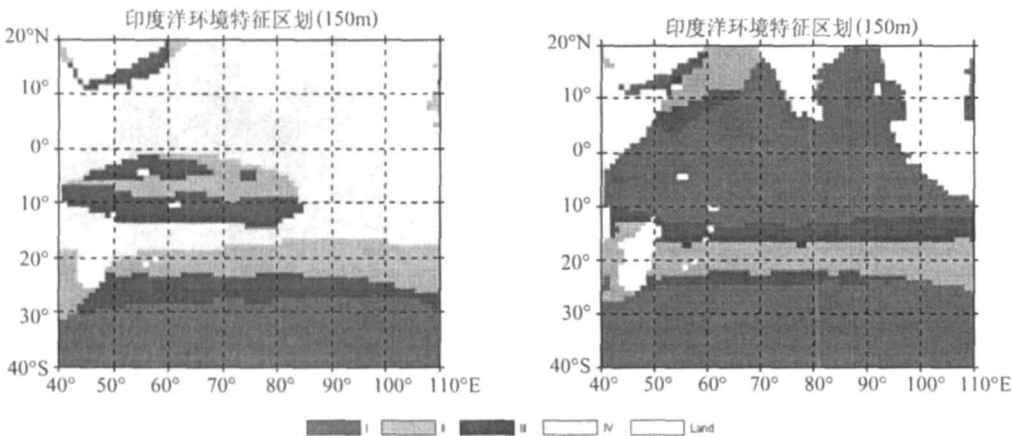
类型 I: 从其分布的区域情况来看,在表层 38°S 以南都属于该类型,随着深度的加深,该类型区域的北界逐渐向北扩展。到 300m 深度时占据整个大洋。该类型具有温度和声速比较小,其垂直梯度也较小的特性,声场环境比较单一,该类型区域跃层对潜艇活动和声纳探测的影响相对来说较小。

类型 II: 具有声速梯度较大,而声速年变差和水平变差较小的特性。从分布情况来看,在 100m 深度时才出现,在 125 ~ 300m 之间该类型主要在 20°S 纬度带和 5°S 纬度带 $40^{\circ} \sim 80^{\circ}\text{E}$ 之间。从属性特征值可知,该类型声速垂直梯度大,强声速梯度的存在可导

致声呐作用距离明显缩短。声跃层起伏使声传播产生明显的屏蔽效应，使声传播异常衰减，对潜艇活动和声呐探测的影响很大。

类型Ⅰ：从分布情况来看，该类型在表层占据了25°S 区域，随着深度增加，总体呈西南-东北走势的椭球型且逐渐缩小；150 ~ 250m 之间水层，在 80°E 以西 10°S 纬度带和大洋西北部也有被该类型占据，300m 以下消失。该类型特征与类型Ⅱ相反，声速垂直梯度较小，但声速年变差和水平变差较大，声速水平变差将使长距离声信号传播受到影响，声速年变差较大，在该类区划活动的潜艇应该注意不同季节对声速变化的影响。

类型Ⅲ：从分布情况来看，该类型占据了 150m 以上，15°S 以北的表层广阔水域。该类样本在特征空间分布较散，且温度和声速均值及垂直梯度较大，表层充分混合，层结稳定度大，易形成等温层(海面波导)，潜艇可以充分利用海面波导的特点。



6 结语

本文基于世界大洋数据集的温盐资料，从中提取一组海洋声场环境的特征向量，并

用聚类效果评估指数客观确定最佳聚类数目。发挥自组织特征映射网络能够保持样本数据空间拓扑结构不变的优势,运用自组织网络对西太平洋海洋环境进行了水声环境区划分型。该区划分型方案具有客观性和合理性,分类区划结果符合客观实际,对该海区的水声场分布特征和声纳探测应用有一定的参考意义。当然,上述分类方法对海洋环境特征的描述还比较简单,海洋区划分型还比较粗糙,还有待进一步的改进提高和不断完善。

参考文献:

- [1] 周燕遐,李炳兰. 世界大洋冬夏季温度跃层特征[J],海洋通报,2002,21(1):16~22.
- [2] 陈奕德,张 韧. 太平洋海域温跃层分类及其特征分析[A],Argo 应用研究论文集,北京:海洋出版社,2006.
- [3] Huang Zhisong, Wang Yanlei, Zhang Ren. Divisions of Hydrological Environment in the Northwest Pacific Ocean Sea Area and Risk Assessment on Maritime Navigation. Proceedings of the first International Conference on Risk Analysis and Crisis Response [M]. ATLANTIS PRESS, 2007.
- [4] 李宏东,姚天翔,等译. Richard O Duda, Peter E Hart, David G Stork 模式分类[M]. 北京:机械工业出版社,2003,9 25~28.
- [5] 叶世伟,史忠植,译. Simon Haykin 神经网络原理[M]. 北京:机械工业出版社,2004,321~225.
- [6] 李凤岐,苏育嵩. 海洋水团分析[M],青岛:青岛海洋大学出版社,2000,306~314.

Division of Underwater Acoustic Environment Character in the Indian Ocean Based on Self-organizing Map

WANG Yan-lei¹, ZHANG Ren¹, YANG Bo², LIU Ke-feng¹, YU Peng¹

(1. Institute of Meteorology, PLA University of Science and Technology, Nanjing 211101 China; 2. General Staff Meteorology and Hydrology Center, Beijing 100081 China)

Abstract: Basing on temperature-salinity data of the world oceans data sets (WOA05) possessed of one degree resolution and the standard layer, a group of character indicators were distilled from underwater acoustic environment of the Indian Ocean. The optimal number of clusters was determined by a group of function assessing the results clustering. Underwater acoustic environment of the Indian Ocean was divided using self-organizing feature map. Classification Division results have practical significance for underwater acoustic field analysis and sonar detection.

Keyword: Self-organizing Map; Underwater Acoustic Environment; Classification Division