

基于 KHM 聚类算法的跳频信号分选

齐 昶, 王 斌, 丁海军

(信息工程大学信息工程学院, 郑州 450002)

摘要: 针对跳频信号分选, 主要研究了聚类算法及利用直方图来预估计聚类数目及初始中心的方法。首先对直方图方法进行改进, 得到了对跳频信号参数估计值误差不敏感的方法, 其次对初始化中心不敏感的 KHM 聚类算法进行改进并聚类, 最后提出了通过定义类内距类间距的方法来确定最佳聚类数的算法。通过改进的 KHM 算法和估计聚类个数方法, 利用跳频信号参数对跳频信号进行分选, 仿真结果表明, 分选效果良好。

关键词: 聚类; 跳频; 信号分选

中图分类号: TN911

文献标识码: A

文章编号: 1000-3630(2011)-06-0547-05

DOI 编码: 10.3969/j.issn1000-3630.2011.06.017

Identification of frequency hopping signals based on clustering

QI Chang, WANG Bin, DING Hai-jun

(Institute of Information Engineering, Information Engineering University, Zhengzhou 450002, China)

Abstract: In order to identify frequency hopping signals, clustering algorithms and searching parameter' histogram peak values are studied. The histogram algorithm is firstly modified to achieve insensitivity to the estimating error of frequency hopping signal parameters. Secondly, the K Harmonic-Means (KHM) algorithm is modified to achieve insensitivity to the random initialization of clustering centers. And finally, a new method of estimating the exact optimum clustering number is proposed. Combining the improved KHM algorithm with the method of estimating the clustering number, the frequency hopping signals can be identified with easily known signals' durations, directions and powers. The proposed algorithm is shown good in performance.

Key words: clustering; frequency hopping; signal sorting

0 引言

跳频通信由于具有优良的抗干扰和多址组网性能, 在通信领域中得到了广泛的应用^[1,2]。跳频通信常以跳频组网方式出现, 进一步加强了跳频通信的抗干扰能力, 并且由于短波电磁环境复杂, 定频、突发、跳频等信号可能同时存在, 增加了跳频信号识别和分选的难度。对此, 文献[3]将跳频通信信号看成是等间隔的跳频脉冲序列, 将应用于周期脉冲序列去交错的算法用于跳频信号的分选, 但是这种方法只利用了时间信息, 并且对脉冲丢失也比较敏感。文献[4]试探性地利用了独立分量分析理论对重叠的跳频信号进行了分离, 算法尚不实用。文献[5]基于阵列天线和二维谐波恢复, 提出了一种联合估计到达方向、跳频时间和跳变频率的多用户检测算法, 但运算量较大。本文主要考虑利用聚类的方法

实现跳频信号分选。文献[6]提出利用直方图方法进行最优划分的 K-Means 初始聚类中心选取算法, 减少了算法结果对参数的依赖; 文献[7]提出了一种 KHM 聚类方法, 在一定程度上解决了对初始化中心敏感的问题。文献[8]提出了一种优化初始聚类中心的方法预估计类的数目, 但不够完善。本文通过对文献[8]直方图估计类数目方法的改进, 得到了对样本参数估计值误差不明显的类数目方法, 并提出了在此估计值下确定最佳聚类数的算法, 利用文献[8]提出的聚类算法对非合作方能得到的持续时间、方位信息和功率对信号进行聚类, 完善了跳频信号的分选。

1 信号模型

通过前期预处理, 可以得到各个信号的方向及各段信号的跳周期、功率等信息。本文即是在此基础上, 对获得的方向信息、跳周期、功率等进行聚类。因为信号功率受信道传输的影响较大, 参数不稳定, 因此主要利用获得的方位信息、跳周期参数来进行分选。

收稿日期: 2010-12-10; 修回日期: 2011-03-03

作者简介: 齐昶(1982—), 男, 河南新乡人, 硕士研究生, 研究方向为通信中的现代信号处理。

通讯作者: 齐昶, E-mail: qifeiyang412829@163.com

信号模型为 $\mathbf{S}=\{s_{ij}\}$, $i=1,2,\dots,N$, $j=1,2,\dots,L$, 表示 N 个待分类的 L 维样本。记第 i 个行向量为 \mathbf{x}_i , 则待分类样本也可表示为 $\mathbf{S}=\{\mathbf{x}_i|i=1,2,\dots,N\}$ 。

经常应用的聚类算法如 K-Means 算法对随机初始化中心敏感, 且需要预先确定类的数目, 而 KHM 算法虽在一定程度上解决了对初始化中心敏感的问题, 但是不够彻底, 也需预设聚类数目。这 2 个问题亟待解决。

2 初始聚类中心的预选取和预估聚类数目

文献[8]利用搜索参数统计直方图来预估类数目并优化初始类中心。

通过前期处理, 将跳频信号的各维参数(持续时间、方位信息、功率等)估计值按列构造矩阵 $\mathbf{T}_{N \times M}$, 如果矩阵 \mathbf{T} 列数为 1, 直方图为二维图像; 矩阵列数为 2, 直方图为三维图像。将直方图推广到矩阵多列情况下, 即列数为 ($M \geq 3$) 的直方图将在 $M+1$ 维空间显示。可知矩阵 $\mathbf{T}_{N \times M}$ 为二维空间, 其直方图表示即张成了 $M+1$ 维空间。对矩阵进行直方图统计, 得区间集合 $Ce=\{ce_{i_1 \dots i_M}\}$, $i_1=1,2,\dots,p_1$; $i_2=1,2,\dots,p_2$; \dots ; $i_M=1,2,\dots,p_M$; p_M (p_M 为预估箱数), 区间对应的直条高度(即区间内所含对象数目)为 $Nu=\{nu_{i_1 \dots i_M}\}$, $i_1=1,2,\dots,p_1$; $i_2=1,2,\dots,p_2$; \dots ; $i_M=1,2,\dots,p_M$; 继而搜寻直方图的峰值数目和峰值位置, 峰值数目即为初始聚类中心数目, 峰值位置对应的区间质心作为聚类中心。

直方图峰值搜索时, 文献[8]提出的方法是: 将每个区间的对象数目与相邻的区间比较, 如果某个区间的对象数目均大于各邻区间数目, 则该区间即为峰值所在区间。任何一个非边界区间的相邻区间为 $2M$ 个, 边界上的区间的相邻区间数目范围为 $[M, 2M-1]$ 。由于跳频信号参数估计值进行聚类时, 功率参数不稳定, 一般利用持续时间和方位信息的估计值进行聚类, 因此, 本文只考虑 2 维参数以下的情况。仿真发现, 在 $M=2$ 时, 由于该方法未考虑相邻对角区间, 因此通过上述方法搜索时, 误差较大。由此, 本文提出, 任何一个非边界区间的相邻区间为 $4M$ 个, 边界上的区间的相邻区间数目范围为 $[2M-1, 2M+1]$ 。

直方图预估箱数 p_m ($m=1,2,\dots,M$) 决定了区间的范围及区间内的对象数目, 因此, 在该方法中非常关键。文献[8]中提出 p_m 值根据经验得到, 或为实际类数目的 3 倍左右。当类数目未知, 只能选取多的预估箱数。但在实际中, 类数目一般未知, 且在

仿真中依类数目 3 倍左右设置 p_m 值, 取得的效果也不理想。但由于在前期预处理时, 参数估计值的误差范围根据信噪比可确定, 因此, 本文提出利用参数估计值的最大误差范围来确定预估箱数。

3 基于改进的 KHM 算法跳频网台分选

3.1 KHM 聚类算法

KHM(K Harmonic Means)算法^[7], 利用调和平均数 HA(Harmonic Average)代替 K-Means 算法中数据点到各个中心的最小距离, 聚类时对初始化中心敏感的问题得到了一定程度上的解决, 其算法主要思想是:

$\mathbf{S}=\{\mathbf{x}_i|i=1,2,\dots,N\}$ 代表 N 个待分类的样本, $C=\{c_l|i=1,2,\dots,K\}$ 代表 K 个初始类中心, 代价函数定义为

$$J = \sum_{l=1}^K HA \{ \|\mathbf{x}_i - \mathbf{m}_l\|^p | \mathbf{m}_l \in \mathbf{M}, l=1, \dots, K \} = \sum_{l=1}^K \frac{K}{\sum_{i=1}^N \frac{1}{\|\mathbf{x}_i - \mathbf{m}_l\|^p}} \quad (1)$$

要使式(1)达到最小, 依据最优化准则, 可对各中心点 c_l , $l=1, \dots, K$ 求偏导数, 并令其为零得到。

令 $d_{il} = \|\mathbf{x}_i - \mathbf{m}_l\|$, 则

$$\mathbf{m}_k = \sum_{i=1}^N \frac{\mathbf{x}_i}{\left(\sum_{l=1}^K \frac{1}{d_{il}^p} \right)^2} \frac{1}{d_{ik}^{p+2}} \bigg/ \sum_{l=1}^K \frac{1}{\left(\sum_{i=1}^N \frac{1}{d_{il}^p} \right)^2} \frac{1}{d_{ik}^{p+2}} \quad (2)$$

本文取 $p=2$ 。从式(2)可以看出当 d_{il} 为零时, 出现分母为零的情况, 本文用一个足够小的常数 ε 代替。根据第 2 节得到初始化中心, 通过 KHM 算法聚类的流程图如图 1 所示。

3.2 确定聚类数目及类心

如何确定最佳聚类数 k_{opt} 、聚类中心及其算法的合理性, 被称为聚类有效性问题。文献[9]提出利用聚类有效性指标来确定最佳聚类数, 而本文算法通过直方图搜索预估出聚类数目, 由于各维参数估计值的误差影响, 聚类数目的取值有时会出现大于实际值的情况。

如图 2 所示的一组样本, 图中横纵坐标为样本点数所在位置, 很自然地将其分为 4 类, 其判断的依据是利用类内距紧密程度及类间距的分离程度。因此, 本文提出利用待分类样本通过第 4.1 节聚类算法后, 所得到的聚类中心、各类样本个数等参数, 通过定义类内类间距, 来最终确定最佳聚类数及各类的样本个数。

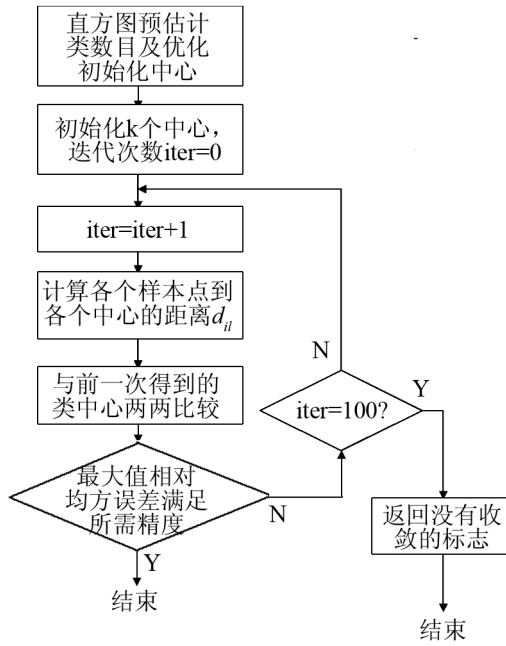


图1 改进后的KHM算法流程图
Fig.1 Flow chart of the modified KHM algorithm

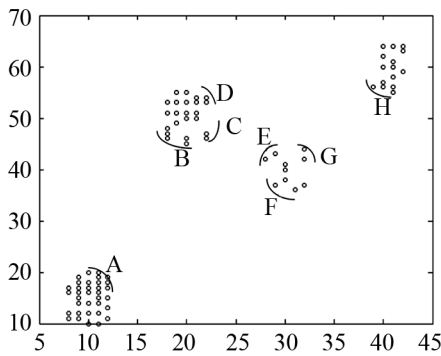


图2 分类示意图
Fig.2 Sketch map of classification

定义1 类内距:同一类内的互不相同样本点之间距离的平均。第 k 类的类内距记为 d_k 。类内的紧密程度由类内距由决定。

定义2 类间连接对:若类 j 中距离样本点 x_{ij} 最近的样本点为 x_{jk} ,且类 i 中距离 x_{jk} 最近的点也为 x_{il} ,称 (x_{il}, x_{jk}) 为类 i 和 j 之间一个连接对。记为 $d_{(x_{il}, x_{jk})}$ 。其中 $x_{il} \in \text{类}i, l=1, \dots, n(i); x_{jk} \in \text{类}j, j=1, \dots, n(j); n(\cdot)$ 为第 \cdot 类的样本点数。

定义3 类间距:类 i 和 j 之间的所有连接对的最小距离,即为 $D_{(i,j)}$ 。则 $D_{(i,j)} = \min(d_{(x_{il}, x_{jk})})$ 。类间分离程度由类间距决定。

如果类间距 D_{ij} 小于类内距,即

$$D_{(i,j)} < \min(d_i, d_j) \quad (3)$$

则两类应合并为一类,相反则认为两类聚类正确不用合并。逐次对各类内类间距进行比较,直至所有类间距大于类内距时,即

$$D_{(i,j)} > \max(d_i, d_j) \quad (4)$$

迭代结束。

合并后的类的聚类中心由合并前类中心的平均值所替代。表述为:两类的类间距小于其两类的类内距的最小值时,其两类的样本个数合并,样本中心由两类中心的平均值替代,聚类数降1,逐次对各类依据类内距类间距的关系进行判断,直至聚类数不再减少,重新计算经合并后各类的类内距及类间距,如类内类间距满足式(4),则对应的聚类个数即为最佳聚类数。

确定聚类数目及类心的具体算法步骤为:

(1) 初始化 $i=1, K$ 为利用直方图方法预估出来的聚类数, c_i 为第 i 类的样本中心, CLA_i 为第 i 类的样本个数。

(2) 计算所有的类内类间距,并根据式(3)逐次进行比较,满足的则样本个数合并,聚类数降1,合并后的样本中心为合并前两类中心的平均值。

(3) 重新计算合并后各类的类内距和类间距,如满足式(4),则最佳聚类数即为合并后得到的聚类数;否则,执行步骤(4)。

(4) 初始化 $i=1, K$ 为利用类内类间距方法各类合并后的聚类数, c_i 为第 i 类的样本中心, CLA_i 为第 i 类的样本个数。

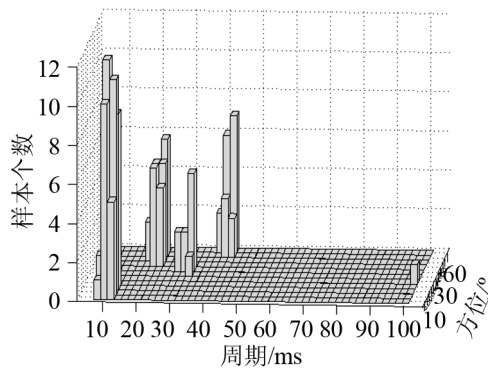
(5) 继续执行步骤(2)。

4 实验仿真

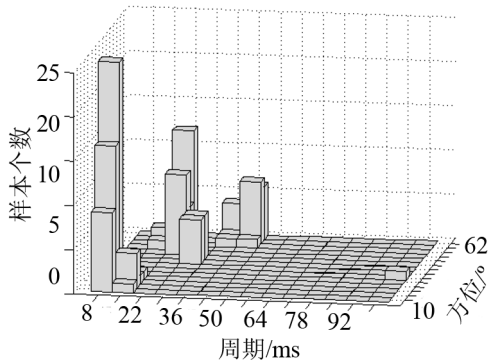
4.1 预分箱数选择

实验1:利用直方图来预估类数目,其中最重要的是对预分箱数的选择。仿真条件:信号1为跳频信号,60跳,跳周期、方位角为(10, 15°);信号2为跳频信号,30跳,参数为(20, 50°);信号3为跳频信号,10跳,参数为(30, 40°);信号4为跳频信号,20跳,参数为(40, 60°),将各维的参数随机叠加偏差,周期偏差在 $\{-2, 1, 0, 1, 2\}$ 均匀分布;角度偏差在 $[-5^\circ, 5^\circ]$ 均匀分布。信号5为定频信号,参数为(100, 30°)。图3(a)为本文方法利用各维参数的最大误差选择预分箱数统计的直方图。图3(b)是文献[8]方法统计的直方图。

由图3(a)、3(b)对比可知,本文提出的选择预分箱数的方法比较合理,直方图的峰值个数为5,与实际相同;而文献[8]的方法的直方图的峰值个数为4,没有搜索出正确的聚类数,误差较大。且在实际中,实际类数一般是未知的,因此,文献[8]的方法的预分箱数只能靠经验取得,但本文提出的方法不需要预知实际类数,只需获知在前期处理的



(a) 本文方法跳频参数统计的直方图



(b) 文献[8]方法跳频参数统计直方图

图3 跳频参数统计直方图

Fig.3 The histograms of FH parameters

跳频信号参数估计时的各维最大误差值即可。

4.2 聚类算法的性能分析

实验 2: 仿真条件同实验 1。

文献[8]估计的聚类个数为 4。文献[10]估计的聚类个数为 6。本文所提算法估计的聚类个数为 5。从仿真实验可知, 本文算法能够正确估计出最佳聚类出聚类数。

从仿真试验可知, 本文算法估计准确, 而其它两种方法没有准确估计出聚类数, 究其原因, 文献[8]算法受直方图预设箱数影响, 容易漏掉某些峰值。文献[10]算法需预设聚类数范围, 且受聚类算法影响较大。

实验 3: 仿真条件中信号 1 及信号 2 与实验 1 中相同, 进行 100 次蒙特卡罗仿真实验, 样本聚类正确率见表 1, 聚类中心和相应的样本个数如表 2 所示。

实验 4: 仿真条件同实验 1, 进行 100 次蒙特卡罗仿真实验, 样本聚类正确率见表 3, 聚类中心和相应的样本个数如表 4 所示。

表 1 信号类数为 2 时聚类正确率(%)

Table 1 Exactitude probability of two signal classifying(%)

信号类数/个	文献[8]方法	文献[10]方法	本文方法
2	56	100	100

表 2 信号类数为 2 时估计的聚类中心及相应样本数

Table 2 Clustering center and sample numbers for two signal classifying

实际中心	聚类中心	实际样本个数/个	聚类样本个数/个
(10,15°)	(9.9824,14.5973°)	60	60
(20,50°)	(20.3340,50.7160°)	30	30

表 3 信号类数为 5 时聚类正确率(%)

Table 3 Exactitude probability of five signal classifying(%)

信号类数/个	文献[8]方法	文献[10]方法	本文方法
5	30	65	94

表 4 信号类数为 5 时估计的聚类中心及相应样本数

Table 4 Clustering center and sample numbers for five signal classifying

实际中心	聚类中心	实际样本个数/个	聚类样本个数/个
(10,15°)	(9.9966,14.9340°)	60	60
(20,50°)	(20.1032,50.4958°)	30	10
(30,40°)	(30.2781,40.3196°)	10	60
(40,60°)	(39.9506,60.6226°)	20	20
(100,30°)	(9.9966,14.9340°)	1	1

从表 1 和表 3 可以看出, 文献[8、10]方法在聚类数较高时, 聚类正确率不高; 而本文改进方法在 2 种情况下均能得到较高的聚类正确率。从表 2 可以看出, 在方位角为 14.5973° 附近存在较多持续时间约为 9.9824 时间单元的信号, 在 50.7160° 方向附近存在较多持续时间约为 20.3340 时间单元的信号, 可以初步判断为 2 个跳频电台。对于表 4, 首先将个数较少的 1 类信号排除, 剩下的根据同 1 个跳频电台的每 1 跳持续时间相同、方向相同的特点, 可判断出存在 4 个跳频电台。

4.3 算法的收敛性能分析

实验 5: 仿真条件同实验 1, 进行 100 次蒙特卡罗仿真实验, 聚类中心的收敛曲线如图 4 所示。

由图 4 可知, 文献[10]的方法需迭代 6、7 次后

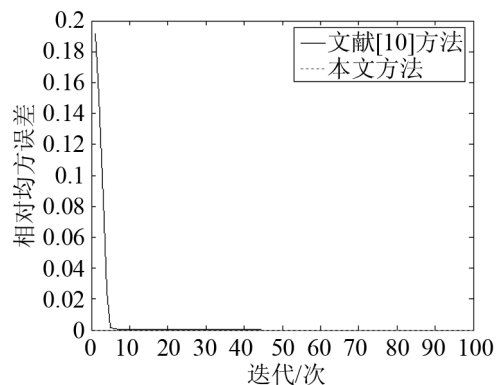


图 4 聚类中心收敛曲线

Fig.4 The convergence curve of clustering center

才能收敛,而本文算法只需要迭代1次即可收敛,表明该算法具有较高的收敛效率。究其原因,是因为文献[10]的方法初始化中心具有随机性,而本文算法通过直方图预估聚类的初始中心与实际类心较接近。

5 结语

本文通过对搜索参数统计直方图方法的改进,优化了聚类数及初始化中心,然后利用KHM聚类算法对跳频信号参数进行聚类分选,最后根据类内距及类间距的定义,提出了寻找最佳聚类个数的办法,仿真实验表明,该算法效果良好。

参 考 文 献

- [1] 梅文华,王淑波,邱永红,杜兴民著.跳频通信[M].北京:国防工业出版社,2005:15-22.
MEI Wenhua, WANG Shubo, QIU Yonghong, DU Xingmin. Frequency Hopping Communications[M]. Beijing: National Defense Industry Press.
- [2] 赵宝珍,申晓红.水声通信中的混沌跳频技术研究[J].声学技术,2005,24(3):137-139.
ZHAO Baozhen, SHEN Xiaohong. Chaotic frequency hopping in underwater communication[J]. Technical Acoustics, 2005, 24(3): 137-139.
- [3] 雷迎科,钟子发,郑大炜.一种短波非正交跳频网台信号分选方法研究[J].舰船电子工程,2006,26(5):135-139.
LEI Yingke, ZHONG Zifa, ZHENG Dawei. Study on the method of sorting the shortwave non-orthodox frequency-hopped network-station signal[J]. Ship Electronic Engineering, 2006, 26(5): 135-139.
- [4] 陈超,高宪军,李德鑫.基于独立分量分析的混叠跳频信号分离算法[J].吉林大学学报(信息科学版),2008,4(26):47-51.
CHEN Chao, GAO Xianjun, LI Dexin. Overlapped frequency-hopping communication signals separation algorithm based on Independent Component Analysis[J]. Journal of Jilin University (Information Science Edition), 2008, 4(26): 47-51.
- [5] LIU Xiangqian, Sidiropoulos N D, Swami A. Code-blind reception of frequency hopped signals over multipath fading channels[C]//IEEE ICASSP'03. 2003, 4: IV-592-5.
- [6] 张健沛,杨锐,杨静,张泽宝.基于最优划分的K-Means初始聚类中心选取算法[J].系统仿真学报,2009,9(21):86-90.
ZHANG Jianpei, YANG Rui, YANG Jing, ZHANG Zebao. Algorithm for initialization of K-Means clustering center based on optimized-division[J]. Journal of System Simulation, 2009, 9(21): 86-90.
- [7] ZHANG Bin. Generalized k-harmonic means boosting in unsupervised learning [EB/OL]. (2000-10-12). <http://www.hpl.hp.com/techreports/2000/HPL-2000-137.pdf>.
- [8] 陈利虎,张尔杨,沈荣骏.基于优化初始聚类质心K-Means算法的跳频信号分选[J].国防科技大学学报,2009,31(2):70-75.
CHEN Lihu, ZHANG Eryang, SHEN Rongjun. Algorithm with Optimal Initial Clustering Centers[J]. Journal of National University of Defense Technology, 2009, 31(2): 70-75.
- [9] 李双虎,张风海.一个新的聚类有效性指标[J].计算机工程与设计,2007,28(4):1772-1774.
LI Shuanghu, ZHANG Fenghai. New index for clustering validation[J]. Computer Engineering and Design, 2007, 28(4): 1772-1774.
- [10] 王斌,陈秋华,王翠柏.基于聚类的跳频信号分选[J].北京邮电大学学报,2009,32(2):80-84.
WANG Bin, CHEN Qiuhua, WANG Cuibai. Identification of frequency hopping signals based on clustering[J]. Journal of Beijing University of Posts and Telecommunications, 2009, 32(2): 80-84.

《噪声控制与建筑声学设备和材料选用手册》一书出版发行

由吕玉恒、燕翔等主编的《噪声控制与建筑声学设备和材料选用手册》(第3版)于2011年10月由化学工业出版社出版发行。全书共850页,130万字,精装,定价178元/本。

该手册荟萃了本世纪以来国内噪声振动控制以及建筑声学所用设备、材料的最新成果,可供设计、研究、施工、安装、教学等工程技术人员参考。主要章节包括概述、基础知识、消声器、吸声降噪、隔声构件、隔振器与阻尼材料、声振测量仪器、典型案例以及4个附录等。内容丰富、图文并茂,是一本综合性的工具书。

若有需要者,可直接与中国船舶重工集团公司第九设计研究院工程有限公司吕玉恒联系(地址:上海市武宁路303号,邮编200063,电话021-62549700-729),也可与清华大学建筑学院燕翔联系(地址:北京清华大学中央主楼104室,邮编100084,电话010-62781010-800)。