

doi: 10.3969/j.issn.1006-1576.2011.01.016

## 基于支持向量机的地磁辅助导航匹配区域选取准则

吕云霄, 吴美平, 胡小平

(国防科学技术大学 机电工程与自动化学院, 湖南 长沙 410073)

**摘要:** 为综合考虑地磁标准差、地磁费歇信息量、地磁编码失真量等特征参数对选取结果的影响, 研究基于支持向量机的地磁匹配区域选取方法。先计算地磁特征参数, 准确有效地表示地磁信息; 再选择合适的训练样本集, 并通过选择合适的核函数及核函数参数, 实现了匹配区域及非匹配区域的正确分类; 最后, 将测试样本集作为检验核数据对分类器效果进行检验, 并根据总体精度和第二类错误率的大小选择合适的核函数参数, 通过测量的地磁数据进一步进行了验证。结果表明, 该方法分类简单, 适应性强, 能够进行匹配区域的正确选择。

**关键词:** 地磁辅助导航; 匹配区; 支持向量机; 特征参数

**中图分类号:** TP206 **文献标志码:** A

## Support Vector Machine for Geomagnetism Aided Navigation Matching Area Selection

Lu Yunxiao, Wu Meiping, Hu Xiaoping

(School of Mechatronics Engineering & Automation, National University of Defense Technology, Changsha 410073, China)

**Abstract:** Considering the impact on the selection of matching areas of geomagnetic standard variance, fisher information content and code distortion together, the paper studies a method for selecting suitable matching area by using support vector machine. First, reasonable measure parameters of geomagnetic field were calculated, which could represent geomagnetic information accurately and effectively. Then by coding the measured parameters as input feature vectors of training set, and choosing suitable kernel function and parameters of kernel function to classify the matching areas and non-matching areas. At last, by using testing set, the classified effect was tested. Based on classified precision and the second errors, the parameters of kernel function were picked, and by using geodesic geomagnetic data, the classified results were tested again. The experimental results show that this method holds the capability of flexibility and jamming resistance, and can choice correct geomagnetism aided navigation matching areas.

**Keywords:** geomagnetism aided navigation; matching areas; support vector machine; characteristic parameters

### 0 引言

由于地磁辅助导航技术具有全天候、无源性、隐蔽性、误差不随时间积累等明显优点, 近年来已逐渐成为导航界研究的热点, 并取得了一定进展。地磁辅助导航系统的核心技术是地磁匹配技术, 其主要思想是利用匹配算法将实时测量的地磁图与数据库内存储的基准地磁图进行匹配, 从而获得最佳匹配位置。在地磁匹配过程中, 定位精度与地磁匹配算法和地磁匹配区域的选择有关。在重力辅助导航研究中, 文献[1]提出了基于局部重力场标准差、局部重力场相关系数等参数特征的重力匹配区域选取准则。文献[2]提出了基于 SPSS 回归分析得重力匹配区域选择方法。在地磁匹配区域选择研究中, 文献[3]结合地磁图相关曲面的相关峰尖锐系数、独立性系数和抗噪声能力系数的定量关系, 对地磁图适配性进行了研究, 但这种方法只考虑了相关曲面特征, 没有对地磁图不同特征参数进行研究。故借鉴重力匹配区域选择相关知识, 提出了基于支持向

量机的地磁匹配区域选取方法, 能很好地实现地磁匹配区域的选择。

### 1 地磁匹配区域特征参数的选取

通常情况下, 地磁场数据是采用网格矩阵的形式存储在计算机内, 以离散点的形式表示地磁场空间的变化规律, 每一组离散点都包含了地磁场的位置信息和强度信息。设某区域地磁场的经纬跨度为  $M \times N$  的网格,  $f(i, j)$  为网格点  $(i, j)$  处的磁场强度值, 其中,  $(i, j)$  对应一组经纬度坐标  $(\phi, \lambda)$ ,  $(M, N)$  表示特征区域  $\Omega$  的大小。目前, 还没有统一的标准描述数字地磁图的特征。笔者运用数理统计、随机场理论及信息论的相关知识, 研究地磁图的特征参数, 并以此作为特征向量来研究匹配区域的选取准则。

#### 1.1 地磁标准差

通常情况下, 地磁标准差  $\delta_f$  主要反映了地磁场强度的离散程度和整个区域的起伏程度。从数学

收稿日期: 2010-08-02; 修回日期: 2010-09-15

作者简介: 吕云霄 (1985-), 女, 河北人, 硕士研究生, 从事导航技术研究。

意义上讲，样本标准差反映了一个样本集合总体上偏离其平均水平的程度，所以地磁标准差描述了地磁场强度的起伏状况是有效的。地磁标准差越大，说明地磁场所含的信息量越大。设  $M \times N$  区域的地磁场的平均值为  $\bar{f}$ ，则地磁标准差可定义为：

$$\delta_f = \sqrt{\frac{1}{MN} \sum_{i=0}^M \sum_{j=0}^N (f(i,j) - \bar{f})^2} \quad (1)$$

### 1.2 地磁费歇信息量

地磁的费歇信息量 (Fisher Information Content, FIC) 可以度量地磁场所包含的地磁信息量，地磁信息量是指在一定的地磁场区域内对任意义点的地磁特征值所能估计出该点在区域内位置的可能性度量。其公式如下 (推导过程参考[4-6])：

$$FIC = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \|\nabla h(x_{i,j})\|^2} \quad (2)$$

式中， $x_{i,j}$  是在地磁场区域内均匀分布的平面位置点坐标。理论上，地磁费歇信息量的取值范围是从 0 到无穷大，实际上不可能得到无穷大的地磁费歇信息量，它的上界是一个很大的数，其值越大，表明其含有的地磁信息越丰富。

### 1.3 地磁编码失真量

借鉴信息论的知识，载体位置信息可以看作是信源，信源由地球地磁场产生；地磁场强度值是信源经过编码后在信道中传输的信号；地磁传感器是信号传输的信道，通过它可以获得含有噪声的地磁场强度值；计算机内存储的数字地磁图具有解码器的功能，它利用实时测量的地磁值估计平面的位置信息。

在一定区域的地磁匹配中，所有位置点的地磁场值不可能都不相同，所以地磁匹配区不是一一对应的编码器。通常情况下，编码的输入与解码输出值之间存在随机误差，这里可采用概率平均作为编码器的失真<sup>[7]</sup>。地磁编码失真量的计算公式如下：

$$\bar{d}_f = \frac{1}{MN} \sum_{c(f_m)} \frac{1}{c(f_m)} \sum_{y(x_{i,j})=y(x_{k,l})=f_m} d(x_{i,j}, x_{k,l}) \quad (3)$$

其中， $c(f_m)$  表示在地磁匹配区域内被编码为同一个  $f_m$  的不同位置的个数， $d(x_{i,j}, x_{k,l})$  为失真函数，表示将位置  $(i, j)$  通过编码解码为位置  $(k, l)$  的失真量，通常用两个位置之间的平面直线距离表示，具体计算公式为：

$$d(x_{i,j}, x_{k,l}) = \sqrt{(x(x_{i,j}) - x(x_{k,l}))^2 + (y(x_{i,j}) - y(x_{k,l}))^2} \quad (4)$$

地磁编码失真量越小表示该区域内的地磁导航信息量丰富，越大表示含有的地磁导航信息越少，不适合作为地磁匹配区。

### 1.4 地磁粗糙度

地磁粗糙度是反映整个区域平均光滑程度，可以刻细微的局部起伏，通常用  $\sigma_f$  表示。

$$\sigma_f = \frac{Q_x + Q_y}{2} \quad (5)$$

其中， $Q_x = \sqrt{\frac{1}{M(N-1)} \sum_{i=1}^M \sum_{j=1}^{N-1} (f_{i,j} - f_{i,j+1})^2}$ ，

$$Q_y = \sqrt{\frac{1}{(M-1)N} \sum_{i=1}^{M-1} \sum_{j=1}^N (f_{i,j} - f_{i+1,j})^2}$$

粗糙度  $\sigma_f$  与标准差  $\delta_f$  之间没有必然的联系，但粗糙度与标准差之间的比值可以作为地磁起伏特征的度量<sup>[8]</sup>。 $\sigma_f / \delta_f$  值小，表示相邻采样点间变化比较小， $\sigma_f / \delta_f$  值大，表示相邻采样点间变化比整个区域起伏相对较大，含有的地磁信息相对丰富，更适合作为地磁匹配区。

## 2 支持向量机分类器的设计

支持向量机 (Support Vector Machine, SVM) 是 Vapnik 在 20 世纪 90 年代提出的一种基于统计学理论的机器学习方法，主要用于解决分类问题和回归问题。其中，用于解决分类问题的支持向量机称为支持向量分类机。

### 2.1 支持向量分类机的基本思想

在进行分类时，对于线性不可分的样本，支持向量机首先通过一个非线性映射将原空间的样本映射到一个高维的特征空间，又称核空间，使其在核空间内线性可分或近似线性可分，然后在核空间中进行线性分类，从而实现相对于原空间的非线性分类<sup>[9]</sup>。设已知训练集  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \in X \times Y$ ，其中， $x_i \in X = R^n, y_i \in Y = \{1, -1\}, i = 1, \dots, n$ 。为寻找结构风险最小的判决函数  $f(x)$ ，设分类超平面为  $w \cdot x + b = 0$ ，其中， $w$  和  $b$  分别为权重向量和偏置向量。为求解此超平面，当样本可分时，问题可以归结为二次规划问题<sup>[10]</sup>：

$$\begin{cases} \min \phi(w) = \frac{1}{2} (w \cdot w) \\ S.T. \quad y_i [x_i \cdot w - b] \geq 1 \quad i = 1, 2, \dots, n \end{cases} \quad (6)$$

若 2 类样本不可分, 则引入松弛变量  $\zeta$  和惩罚参数  $C$ , 问题可转化为:

$$\begin{cases} \min \phi(w, \zeta) = \frac{1}{2}(w \cdot w) + C \sum_{i=1}^n \zeta_i & (7) \\ \text{S.T. } y_i[(x_i \cdot w) - b] \geq 1 - \zeta_i & i = 1, 2, \dots, n \\ \zeta_i \geq 0 \end{cases}$$

利用其对偶形式得到线形分类器的决策函数为:  $f(x) = \text{sgn}(\sum_{i=1}^n a_i y_i (x_i \cdot x_i) + b)$ 。其中,  $a_i$  为拉格朗日因子, 每一个训练样本对应一个  $a_i$ , 为 0 时对应的训练样本对分类结果没有影响, 非零时产生影响, 这样的样本成为支持向量。

输入样本在原始空间内不可分时, 需要引入非线性映射函数将低维空间的线性不可分问题映射到高维特征空间, 转化为线性可分问题, 从而得到相应的非线性分类器:  $f(x) = \text{sgn}(\sum_{i=1}^n a_i y_i K(x \cdot x_i) + b)$ ,

其中称  $K(x \cdot x_i)$  为核函数。作为一种统计学习方法, 支持向量机有着很好的推广能力和理论基础<sup>[11]</sup>, 其关键在于核函数的选取。

## 2.2 支持向量机核函数的选取

核函数在支持向量机中具有重要的地位, 其中核参数的选取直接影响到核函数的推广能力。常用的核函数有: 线性核函数、多项式核函数、高斯径向基核函数等。在实际应用中, 每一种核函数都有各自的优点和缺点。核函数主要分为全局核函数和局部核函数。全局核函数具有整体特性, 允许相距较远的点对核函数有影响, 而局部核函数只允许相距较近的点对核函数值有影响。根据本文的解决分类问题, 通过多项测试发现选用径向基核函数时, 错误率较小, 径向函数表达式如式 (6):

$$K(x, x_i) = \text{Exp}\left(-\frac{|x - x_i|^2}{\sigma^2}\right) \quad (8)$$

## 2.3 训练样本集和测试样本集选取

基于利用支持向量机的分类包括 2 个重要过程: 一是利用训练样本集实现对分类器的训练学习; 二是利用测试样本集对训练好的分类器进行评价; 可见, 训练样本和测试样本的选取是需要解决的一个重要问题。笔者在网格精度为 100 m×100 m, 大小为 59.6 km×82.2 km 的某试验区地磁图上提取了 200 块 10×10 的地磁图 (网格精度为 100 m), 计算得到了这 200 块地磁图的地磁标准差  $\delta_f$ 、地磁费歇

信息量  $fic$ 、地磁编码失真量  $\bar{d}_f$ 、粗糙度与标准差的比值  $\sigma_f / \delta_f$  四个参量作为训练样本的输入, 即  $[\delta_f, fic, \bar{d}_f, \sigma_f / \delta_f]$ 。依据相关匹配算法<sup>[12]</sup>分别计算每块地磁图的匹配概率, 当匹配概率大于 0.95 时, 认为该样本可用, 利用  $Y=1$  表示; 当小于 0.95 时, 令  $Y=-1$ 。经计算, 在选取的 200 个训练样本中,  $Y=1$  的样本数为 134,  $Y=-1$  样本数为 66, 为加快训练速度, 随机选择 30 个误匹配样本。

测试样本也是在同一试验区地磁图上提取的不同于训练样本的 100 块 10×10 的地磁图 (网格精度为 100 m), 其中可匹配区为 52 块, 不可匹配区为 48 块。

## 2.4 样本的训练与分类

样本训练分类过程中会产生 2 类错误: 第一类错误是将可匹配区域判定为不可匹配区域; 第二类错误是将不可匹配区域判定为可匹配区域。对于地磁辅助导航系统来说, 第一类错误带来的后果是可以容忍的, 第二类错误带来的后果是致命的。为此, 在样本训练过程中, 通过调整核函数参数, 在保证整体精度的前提下, 尽量减小第二类错误。

将训练样本集的特征参数值作为输入, 使用径向基核函数进行训练, 不断调整核函数的参数, 最终统计出支持向量的个数。这些支持向量决定了可匹配区与不可匹配区之间的分界面, 实现分类<sup>[10]</sup>。

## 3 实验与分析

将选定的 164 个训练样本集的特征参数向量  $[\delta_f, fic, \bar{d}_f, \sigma_f / \delta_f]$  作为训练样本的输入, 将对应的  $[Y=1$  或  $-1]$  作为训练样本的输出。另外选定的 100 个训练样本的特征参数向量  $[\delta_f, fic, \bar{d}_f, \sigma_f / \delta_f]$  作为检核数据<sup>[13]</sup>。选择惩罚因子选择为无穷大, 径向基函数的宽度分别为 110 和 152.3 时, 支持向量机分类的结果如表 1。

表 1 实验分类结果

| 径向基函数宽度 | 第一类错误概率 | 第二类错误概率 | 总分类精度 |
|---------|---------|---------|-------|
| 110     | 0.15    | 0.11    | 0.87  |
| 152.3   | 0.24    | 0.03    | 0.86  |

在实际应用中, 第二类错误会辅助导航系统带来灾难性后果, 为此, 在总体精度相差不大的前提下, 选择径向基函数的宽度为 152.3。为进一步验证径向基函数宽度为 152.3 时的分类效果, 笔者从另外一试验区选择了 6 幅 100×100 (网格精度为 100 m) 的地磁图, 其中地磁图 1 和地磁图 6 的相关匹配结

果如图 1、图 2，其匹配结果的统计概率及利用支持向量机的分类结果如表 2（-1 代表不适合匹配区，1 代表适合匹配区）。

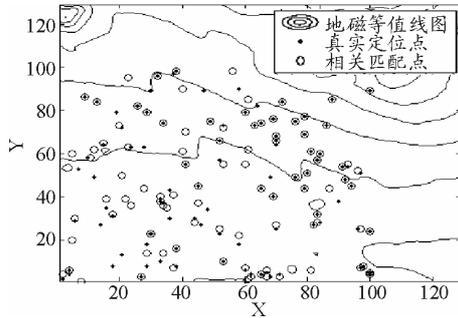


图 1 地磁图 1 的匹配定位结果

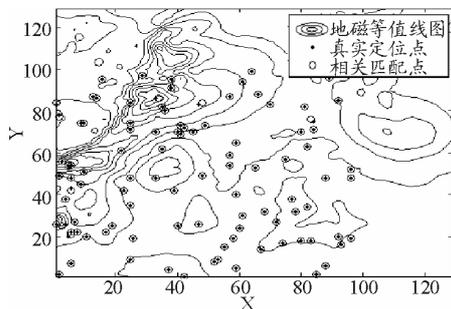


图 2 地磁图 6 的匹配定位结果

表 2 地磁图的匹配概率及分类结果

| 地磁图 | 匹配概率 | 分类结果 | 地磁图 | 匹配概率 | 分类结果 |
|-----|------|------|-----|------|------|
| 1   | 0.46 | -1   | 4   | 0.34 | -1   |
| 2   | 0.96 | 1    | 5   | 0.99 | 1    |
| 3   | 0.99 | 1    | 6   | 0.95 | 1    |

### 4 结束语

由实验结果可知，该方法充分考虑了不同特征参数之间的关系，具有简单、有效，适应性强的特点，为实现匹配区域的选择提供了参考。由于目前

的实验结果还存在第二类错误，下一步，将从数据处理、改进优化核函数等方面着手，严格杜绝第二类错误的出现，并提高总体定位精度。

### 参考文献：

- [1] 程力, 张雅杰, 蔡体菁. 重力辅助导航匹配区域选取[J]. 中国惯性技术学报, 2007, 15(5): 559-563.
- [2] 夏冰, 蔡体菁. 基于 SPSS 的重力匹配区域选择算法[J]. 中国惯性技术学报, 2010, 18(1): 81-84.
- [3] 杨亚鹏, 罗诗途, 吴美平, 等. 面向相关匹配算法应用的地磁图适应性分析[J]. 导航与控制, 2009(2): 9-15.
- [4] N. Bergman. Recursive Bayesian Estimation Navigation and Tracking Application[C]. Sweden: IEEE Proc. of Position Location and Navigation Conf. 1984: 286-293.
- [5] N. Bergman. A Bayesian approach to terrain-aided navigation[C]. In Proc. of SYSID'97, 11th IFAC Symposium on System Identification, 1997: 1531-1536.
- [6] 冯庆堂. 地形匹配新方法及其环境适应性研究[D]. 长沙: 国防科学技术大学, 2004.
- [7] 张宏基. 信源编码[M]. 北京: 人民邮电出版社, 1987.
- [8] Ratkovic J A. Estimation techniques and other work on image correlation[C]. //ICSP Proceedings. Feb. 1997: 54-60.
- [9] 齐志泉, 田英杰, 徐志浩. 支持向量机中的核参数选择问题[J]. 控制工程, 2005, 12(4): 379-381.
- [10] 杨朝辉, 陈映鹰. 基于支持向量机的景象匹配区选择方法[J]. 同济大学学报, 2009, 37(5): 690-695.
- [11] V ladimir N Vapnik. 张学工 译. 统计学习理论的本质 [M]. 北京: 清华大学出版社, 2000.
- [12] 朱永松, 国澄明. 基于相关系数的相关匹配算法的研究[J]. 信号处理, 2003, 19(6): 531-534.
- [13] Inglada J. Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2007, 62(3): 235-238.

\*\*\*\*\*

(上接第 40 页)

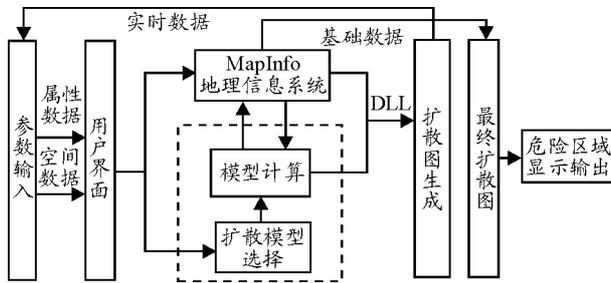


图 1 毒气扩散危险区域显示流程

### 3 结论

毒气模型和 GIS 的集成在实际应用中可以快速模拟有毒泄漏物的扩散过程，显示空间上受其影响的范围，而且有利于及时准确地掌握决策信息，最大限度地减少损失。但由于大气本身变化较为频繁，

尤其在近地面受到温度、地形等影响，下一步，还需要对模型参数及适用性不断修正，才能进行准确的过程模拟，提供可靠的决策服务。

### 参考文献：

- [1] 赵俊龙, 徐迎, 王英宇. 重要经济目标的次生灾害风险分析[J]. 中国人民防空, 2007(1): 36-37.
- [2] 杜建科. 毒气泄漏过程及其危险区域分析[J]. 中国安全科学学报, 2002, 12(6): 55-59.
- [3] 李骥. 化工储罐区空袭次生灾害危险性评价及减灾对策研究[D]. 合肥: 解放军理工大学, 2008.
- [4] 张明广, 蒋军成, 潘旭海. 基于 GIS 的重大危险源风险管理辅助决策系统[J]. 天然气工业, 2007, 27(7): 115-117.
- [5] 潘旭海. 事故性泄漏动力学过程的理论与实验研究[D]. 南京: 南京工业大学, 2004.
- [6] 汤国安, 赵牡丹. 地理信息系统[M]. 北京: 北京科学出版社, 2000.