

# 基于 Fisher 和核 Fisher 组合的磁暴识别方法<sup>\* 1</sup>

康瑞清 梁 盈

(北京科技大学自动化学院,北京 100083)

**摘 要** 为从地磁水平分量序列中自动识别出磁暴,提出了一种基于 Fisher 和核 Fisher 组合的磁暴识别方法。使用北京十三陵地磁台的数据进行测试,在数据样本较少的情况下,该方法对于磁暴与非磁暴的平均识别率达到了85%以上,并随着样本数据的增多,识别率还将有所提高。

**关键词** 小波包;磁暴;Fisher;核 Fisher;识别率

**中图分类号**:P207 **文献识别码**:A

## A COMBINED METHOD OF FDA AND KFDA FOR MAGNETIC STORMS RECOGNITION

Kang Ruiqing and Liang Ying

(Beijing University of Science and Technology, Beijing 100083)

**Abstract** To recognize the magnetic storms automatically from the horizontal component of geomagnetism, a combined method of magnetic storms recognition based on Fisher and kernel Fisher discriminant analysis is presented. The data used in the study are derived from Ming tombs of Beijing geomagnetic station. In the case of the little number of the data samples, the average rate of this combined method for magnetic storms recognition is above 85%, and the recognition rate will be improved by the increasing amount of data samples.

**Key words**: wavelet packets transform; magnetic storms; Fisher; kernel Fisher; recognition rate

## 1 引言

磁暴是一种剧烈的全球性的地磁扰动现象,对通信系统、电力系统、输油管道、空间飞行器等有严重影响,所以,磁暴研究有重要的实际应用价值<sup>[1]</sup>。国内外对磁暴监测预报已展开了大量的研究。周晓燕等<sup>[2]</sup>研究了地磁扰动及其预报的系统辨识建模方法;Akasofu<sup>[3]</sup>提出以日地系统中各环节的物理模型为基础进行物理预报;Valdivia 等<sup>[4]</sup>基于磁层响应存在非线性,提出一种基于 Dst 数据的非线性预报模型,能预报磁暴演化并从中度类型的磁暴中识别出强磁暴。上述方法对磁暴的预报都以太阳活动

的卫星观测数据以及日地系统中各环节的物理模型为基础,模型建立难度大且数据难以获取。因此本文尝试从较容易获得的地磁观测数据出发,来自动识别磁暴的发生<sup>[5]</sup>。

磁暴发生时,所有地磁要素都会发生剧烈变化,其中,水平分量  $H$  变化最大,最能代表磁暴过程的特点。本文采用北京十三陵地磁台 2003、2004、2006 年地磁观测得到的  $H$  分量绝对值的分均值数据,运用小波包分析进行信号处理和特征提取,并提出一种基于 Fisher 和核 Fisher 组合模式的模式识别

\* 收稿日期:2012-10-24

基金项目:中央高校基本科研基金(FRF-TP-09-017B);教育部新世纪优秀人才支持计划资助项目(NCET-09-0216);国家自然科学基金(61273082)

作者简介:康瑞清,女,副教授,主要研究方向:地磁匹配. E-mail: krq@ustb.edu.cn

通讯作者:梁盈,女,1987年生,硕士研究生,研究方向:磁暴识别研究. E-mail: liangying0831@sina.com

技术来尽量准确地识别出磁暴的发生。

2 Fisher 分类与核 Fisher 分类

2.1 Fisher 分类 (FDA)<sup>[6]</sup>

Fisher 方法 (FDA; Fisher Discriminant Analysis) 的基本思想是将  $m$  类  $n$  维数据集尽可能地投影到一个方向,使得类与类之间尽量分开。图 1 是一个以二维模式为例来表示多维空间向一维空间投影的示意图。其中,  $W'$  方向为一个不利的投影方向,无法区分两类样本;  $W^*$  方向可以将两类样本清楚地区分开来。因此,如何找到最好的直线方向正是 Fisher 算法要解决的基本问题。根据使“类间离散度最大化,类内离散度最小化”的 Fisher 准则,找到最佳投影向量  $W^*$  及分界点,判断待测样品向  $W^*$  方向的投影点与分界点的关系,可将其归类至  $\omega_1$  或  $\omega_2$ 。

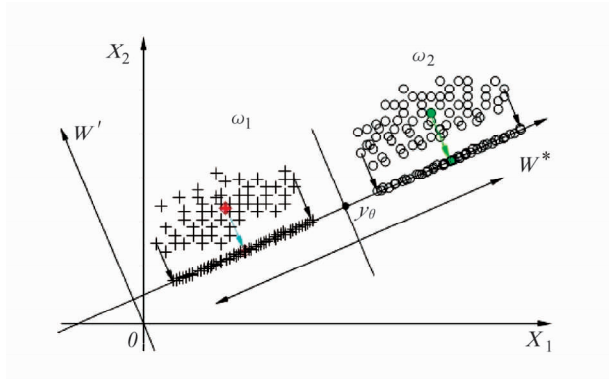


图 1 二维模式向一维空间投影示意图  
Fig. 1 Projection of two-dimensional mode to one-dimensional space

2.2 核 Fisher 分类 (KFDA)<sup>[7,8]</sup>

核 Fisher 判别分析首先把数据非线性的投影到某个特征空间  $F$ ,然后在这个特征空间中进行 Fisher 线性判别,这样在高维空间中的线性判别就相当于原输入空间的非线性判别。核 Fisher 方法的基础是实现向量的内积变换  $\langle X_i, X_j \rangle \rightarrow k(x_i, x_j) = \Phi(X_i) \Phi(X_j)$ 。

3 组合方法的提出

3.1 数据预处理

在地磁台提供的 3 年地磁数据中,总共发生了 43 次磁暴。为了消除太阳静日变化对磁暴识别的影响,采用 FMI 方法<sup>[7]</sup>得到磁暴当日的规则日变化  $S_R$  曲线( $S_R$  的主要成分是太阳静日变化  $S_q$ ,以及幅度很小的太阴日变化  $L$ ),并从当天的地磁台  $H$  分量的绝对值数据中减去,得到地磁中的磁扰值,然后从中提取每次磁暴前和磁暴开始后各 1 小时的序列。其中,为了消除直流分量的影响,先计算数据的序列均值,并从提取序列中减去,最后得到非磁暴和磁暴

分别 43 个零均值序列用于小波包分析。

3.2 用小波包分析进行特征提取

3.2.1 小波包分解与重构

采用 db1 小波对前面得到的零均值序列分别做小波包分解,其中  $(i, j)$  表示第  $i$  层的第  $j$  个结点,  $i = 0, 1, 2, 3, j = 0, 1, 2, \dots, 7$ 。

对小波包各个分解系数进行重构,提取第三层各频带范围的重构信号。分别以  $S_{30}$ 、 $S_{31}$ 、 $S_{32}$ 、 $S_{33}$ 、 $S_{34}$ 、 $S_{35}$ 、 $S_{36}$ 、 $S_{37}$  表示从低频到高频 8 个频带范围的重构信号,总信号表示为  $S = S_{30} + S_{31} + S_{32} + S_{33} + S_{34} + S_{35} + S_{36} + S_{37}$ 。

3.2.2 计算各频带信号的总能量

设  $S_{3j}(j = 0, 1, \dots, 7)$  对应的能量为  $E_{3j}(j = 0, 1, \dots, 7)$ , 有:

$$E_{3j} = \int |S_{3j}(t)|^2 dt = \sum_{k=1}^n |x_{jk}|^2$$

其中  $x_{jk}(j = 0, 1, \dots, 7, k = 1, 2, \dots, n)$  表示重构信号  $S_{3j}$  的离散点幅值。

3.2.3 构造特征向量

由于磁暴发生时会使各频带内信号的能量有较大的变化,因此我们以能量为元素构造特征向量  $T$ :

$$T = [E_{30}, E_{31}, E_{32}, E_{33}, E_{34}, E_{35}, E_{36}, E_{37}]$$

由于这 8 个频带范围的能量数值较大,并且有时会有较大差异,这会造成特征向量中各元素取值范围彼此差异很大,因此需要对特征向量进行改进。取  $E_{3j}(j = 0, 1, \dots, 7)$  的 4 次方根为元素,构造一个新的特征向量  $X'$ <sup>[6]</sup>, 即

$$E'_{3j} = (E_{3j})^{1/4} \quad (j = 0, 1, \dots, 7)$$
$$X' = [E'_{30}, E'_{31}, E'_{32}, E'_{33}, E'_{34}, E'_{35}, E'_{36}, E'_{37}]$$

经过以上步骤的处理,我们得到了磁暴发生前 43 个非磁暴的特征向量,以及磁暴发生时 43 个磁暴的特征向量。

3.3 用分类器进行磁暴与非磁暴的识别

用 Fisher 分类和核 Fisher 分类方法识别磁暴与非磁暴的步骤如下:在用小波包分解提取的特征向量中每类中分别随机选取 50% 作为训练集,用来训练分类器判别式;其余的 50% 作为测试集,用来测试识别率。我们用两种方法各进行了 200 次试验,对照实际磁暴发生情况,得到 Fisher 方法和核 Fisher 分类的方法对磁暴和非磁暴的识别率,其 200 次测试识别率的结果分别如图 2、3 所示。

从图 2、3 可以看出, Fisher 分类方法对非磁暴、核 Fisher 方法对磁暴的识别率均较高,200 次试验平均识别率分别达到了 85% 与 82% 以上;而 Fisher 方法用于识别磁暴、核 Fisher 方法用于识别非磁暴时,识别率很低。

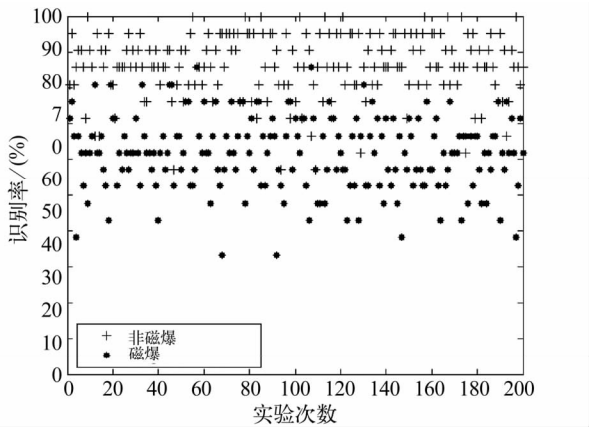


图2 Fisher 方法得到的识别率  
Fig.2 Recognition rate with FDA

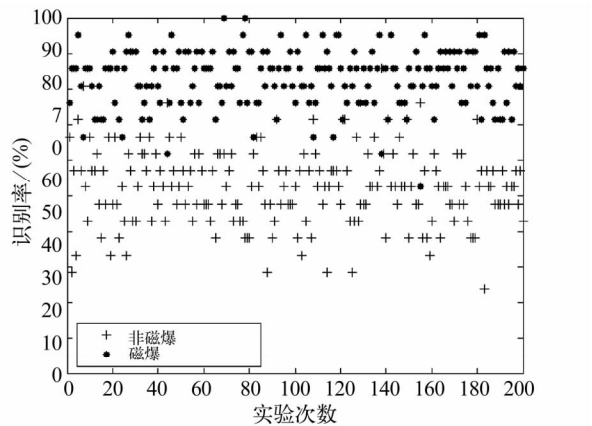


图3 核 Fisher 方法得到的识别率  
Fig.3 Recognition rate with KFDA

对比识别结果,Fisher 和核 Fisher 方法在样本较少的情况下并不能做到在识别两种情况时同时具有较高的识别率。Fisher 分类器在识别非磁暴时性能优,核 Fisher 分类器在识别磁暴时性能优。虽然各个分类器的分类性能有所差异,而且对某一类的分类精度不尽相同,但不同的分类器误判的样本并不完全重合。即对于某个分类器误判的样本,使用其他子分类器可能得到正确的分类结果,这表明不同的分类器对于分类信息可以互补,因此可以将两种分类器组合,即“基于 Fisher 和核 Fisher 组合分类方法”,以获得比单一子分类器更高的分类精度<sup>[8]</sup>。

### 4 组合方法的实施

在特征向量集的两类中分别随机选取 50% 作为训练集,用来训练 Fisher 线性判别式,其余的 50% 作为测试集,记录 Fisher 方法所具有较高识别率的非磁暴情况。再将此次选取的数据用来训练核 Fisher 分类器判别式,并记录其所擅长识别的磁暴发生情况。把记录的结果与实际磁暴发生与否的情况进行比对,分别得到该次识别的非磁暴和磁暴识别率。这样的

过程重复 200 次,每一次选取作为训练集和测试集的数据都是不同的,200 次分类的平均结果作为组合方法的最终识别率。图 4 是 Fisher 和核 Fisher 组合方法 200 次重复试验的识别率结果。

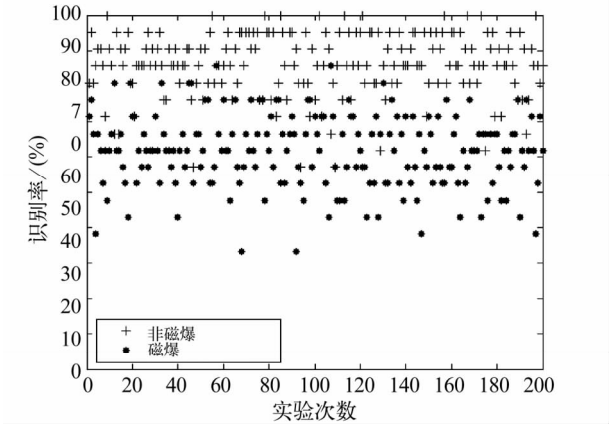


图4 Fisher 和核 Fisher 组合方法得到的识别率  
Fig.4 Recognition rate with combined method of FDA and KFDA

从图 4 可以看出,使用 Fisher 和核 Fisher 组合方法进行非磁暴和磁暴的识别,其效果都得到了明显的改善。经计算,组合方法对非磁暴和磁暴的平均识别率分别达到了 86.57% 和 83.76% (表 1)。

表 1 三种方法识别磁暴结果的比较(单位:%)  
Tab.1 Comparison among recognition rates with 3 recognition methods (unit: %)

	非磁暴	磁暴	平均
FDA	85.71	61.64 (失败)	73.68
KFDA	52.81 (失败)	82.33	67.57
组合法	86.57	83.76	85.17

由表 1 可知,在样本数相同的情况下,Fisher 和核 Fisher 组合方法在识别非磁暴时的识别率相比于核 Fisher 方法提高了 33.76%,在识别磁暴时的识别率相比于 Fisher 方法提高了 22.12%,并且组合方法在识别非磁暴方面与 FDA 方法、在识别磁暴方面与 KFDA 方法的识别率相比均有所提高,对非磁暴和磁暴的识别率分别达到了 86% 和 83% 以上,两类平均识别率达到了 85% 以上,识别效果得到了有效的改善。

此外,将 Fisher 方法、核 Fisher 方法、Fisher 和核 Fisher 组合方法这三种方法分别执行 10 次,计算耗时如表 2。

从表 2 可以看出,由于核 Fisher 分类加入了核方法,运算量大于单独使用 Fisher 分类方法,所以运算时间也相对较长。而 Fisher 和核 Fisher 组合的运算时间与只使用核 Fisher 方法的时间几乎相同,因此可以得出结论,两种方法的组合并没有增加整体

的运算时间。

表 2 三种方法的计算耗时比较(单位:s)

Tab.2 Comparison among computing time with 3 recognition methods( unit:s)

序号	FDA	KFDA	组合法
1	1.094	1.329	1.328
2	1.062	1.328	1.328
3	1.078	1.313	1.297
4	1.079	1.297	1.328
5	1.078	1.313	1.296
6	1.078	1.297	1.329
7	1.063	1.297	1.312
8	1.078	1.328	1.313
9	1.078	1.312	1.313
10	1.078	1.329	1.312
平均	1.076 6	1.314 3	1.315 6

## 5 结论

本文提出了一种基于 Fisher 和核 Fisher 的组合分类方法,即从地磁水平分量序列中提取能量元素构成特征向量,并实时监测其变化,磁暴识别率达到了 85% 以上。值得说明的是,由 Fisher 分类器和核 Fisher 分类器的原理可知,训练所使用的样本越多,勾画出各类事物在特征空间分布的规律性就越好,即样本数据越多,可供分类器学习的特征就会越多,那么分类效果就会越好。试验所用数据仅来自于 43 次磁暴,我们可以认为,如果可供训练的样本数据更多,其识别率还会有所提高。另外,Fisher 和核 Fisher 组合方法的运算时间与只使用核 Fisher 方法的运算时间几乎相同,因此两种方法的组合并没有对整体的运算速度造成影响,在计算速度上也有一定优势。

## 参 考 文 献

1 徐文耀. 地球电磁现象物理学[M]. 合肥:中国科学技术

(上接第 150 页)

4 中国计量科学研究院. 测量不确定度评定与表示[S]. 1999. (National institute of metrology P R China, Evaluation and expression of uncertainty in measurement[S]. 1999)

5 费业泰. 误差理论与数据处理[M]. 北京:机械工业出版社,2005. (Fei Yetai. Theory of error and data processing [M],Beijing: China Machine Press,2005 )

6 常青. 重复性测量在测量不确定度评定中的作用[J]. 中国计量,2006,(10):83-84. (Chang Qing. The role of re-

大学出版社,2009. (Xu Wenyao. Physics of electromagnetic phenomena of the earth [M]. Hefei:University of Science and Technology of China Press, 2009)

2 周晓燕,朱岗. 太阳风引起的地磁扰动的辨识建模[J]. 地球物理学报,1994,(03):291-297. (Zhou Xiaoyan andZhu Gang. Modeling research for solar wind magnetosphere coupling[J]. Chinese J Geophys. ,1994,(03):291-297)

3 Akasofu S Iand Fry C F. A first generation numerical geomagnetic storm pre-diction scheme [J]. Planet Space Sci. , 1986,34(1):77-92.

4 Valdivia J A, Sharma A S and Papadopoulos K. Prediction of magnetic storms by nonlinear models [J]. Geophys Res Lett. ,1996,(23):2899.

5 Ali G Hafez and Essam Ghamry. Automatic detection of geomagnetic sudden commencement via time-frequency clusters [J]. Advances in Space Research, 2011, 48 :1537-1544.

6 杨淑莹. 模式识别与智能计算——MATLAB 技术实现[M]. 北京:电子工业出版社,2008. (Yang Shuying. Pattern recognition and intelligent computing——Technical realization of Matlab [M]. Beijing: Electronics Industry Press, 2008)

7 杨马陵,等. FMI 方法计算地磁 K 指数的初步研究[J]. 地震地磁观测与研究,2004,(2):55-61. (Yang Malin, et al. Preliminary study of computing K indices with FMI method[J]. Seismological and Geomagnetic Observation and Research,2004,(2):55-61)

8 彭正林,等. 基于多分类器组合的遥感影像分类方法研究[J]. 国土资源遥感,2011,(2):19-25. (Peng Zhenglin, et al. Method for classification of remote sensing images based on multiple classifiers combination[J]. Remote Sensing for Land & Resources, 2011,(2):19-25)

peatability measurements in the evaluation of measurement uncertainty[J], China Metrology, 2006,(10):83-84)

7 李慎安. JJF 1059-1999《测量不确定度评定与表示》讨论之三十二关于不确定度合成中较小分量的忽略问题[J]. 工业计量,2011,(1):34-35. (Li Shenan. NO. 32 discussion on uncertainty synthesis of smaller component of the ignored problems of JJF 1059—1999 “evaluation and expression of uncertainty in measurement” [J]. Industrial Measurement, 2011(1):34-35)