

统计学习理论算法在跳频信号分选中的应用

王 锐, 徐 祎

(电子工程学院, 安徽 合肥 230037)

摘 要:在第 3 方截获并检测到跳频通信信号后,由于无先验知识,所以对其进行网台分选不便采用传统有监督学习算法。即便采用无监督的算法,目前多数算法对分类个数等相关消息也必须有所借鉴并在多分选参数的逐级分选中通过先验知识进行有效性判断和筛选。针对电子支援中探测到的跳频信号分选所遇到的困难,利用统计学习理论在小样本学习及非线性分类上较其它传统分类算法更好的性能,提出基于统计学习理论的无监督及半监督学习算法,对第 3 方得到的跳频网台分选进行应用,取得理想结果。为跳频通信侦察过程中的分选工作,提供一种应用鲁棒性好,分选准确度高的方法。

关键词:跳频信号;网台分选;统计学习理论;半监督学习

DOI:10.3969/j.issn.1009-3516.2010.02.015

中图分类号: TN97 **文献标识码:** A **文章编号:** 1009-3516(2010)02-0067-06

跳频通信作为扩谱通信的一个基本分支,以其抗干扰性强,抗截获性高,稳定性好,安全性高等优点,在军事领域得到广泛应用。逐步成为战时无线通信的重要手段,在现代战争中发挥着越来越重要的作用。通信对抗侦察(即电子支援 Electronic Support, ES)^[1]的主要任务是检测和截获信号、分选、参数估计,从而获取信息。其中在对截获的跳频信号进行分选后,才是对感兴趣目标进行精确参数估计,为最终最佳干扰提供信息。它是电子支援(ES)关键一步,最终为电子对抗(Electronic Attack, EA)提供了保障。虽然它只是 ES 中的一个步骤,但是它的效果直接影响到整个通信电子战的有效性。目前在对跳频信号的分选研究中,一般基于有先验知识(或部分先验知识)条件下对跳频信号^[2-3]进行分选,然而实际条件往往无法满足,使得这些算法实际应用效果不理想。在通信对抗侦察中研究无任何先验知识或少量先验知识的跳频信号分选问题,具有较高的实际应用价值。

作为一种新的学习理论体系,近些年来统计学习理论在模式识别问题中的应用越来越广泛,其中在具有多维特征参数的数据集的分类中体现出它在非线性分类和小样本学习问题上的优势。而将统计学习理论应用在没有先验知识或训练集样本个数不足的分类问题上,同样具有优于其它聚类算法的特点。

目前,在高密度、复杂的脉冲信号环境下,对雷达信号进行分选应用中,采用统计学习理论的方法进行分段聚类的分选效果要优于传统信号分选方法^[4]。同时,在对于先验知识不足的恶劣信号环境下,进行跳频信号的分选研究,可以通过一些改进的 K-means 聚类一维直方图法,分层进行跳频信号分选^[5]。另外,为了对信号分选的准确率有进一步提高,半监督的统计学方法也被应用在卫星红外遥感等领域^[6-7]。

本文就是利用统计学习理论在模式识别中的优势,根据信号环境的未知性,采用基于统计学习理论的分类算法对上述恶劣背景环境下的跳频网台进行分选研究。

1 支持向量聚类算法(SVClustering)

支持向量聚类算法^[8-9]是利用一类支持向量机分类,再进行聚类的无监督分类方法。

* 收稿日期:2009-06-03

作者简介:王 锐(1979-),男,安徽安庆人,讲师,主要从事智能信号处理、信号检测与通信研究。

E-mail:wangrui20060908@126.com

SV Clustering 算法分为 2 部分:形成 SV 的聚类边界的 SVM 训练部分和聚类形成部分。其中 SVM 训练部分负责新知识模型的训练,包括 Gauss 核函数宽度系数的优化、Lagrange 乘子的计算、Hilbert 空间最小包络超球体半径的计算和有界支持向量(Bounded Support Vectors, BSVs)与支持向量(Support Vectors, SVs)的选取。聚类形成部分首先生成聚类标识关联矩阵,再通过 DFS(Depth First Search)算法根据关联矩阵进行聚类。

SV Clustering 推广了 SVM 的思想,通过解二次规划问题,得到全局最优解;能处理任意形状的聚类,对噪声能有效处理;无需指定聚类数目,需调参数少;容易处理高维数据。

该算法解决了 K-means 聚类的聚类中心数和迭代最大次数等参数设定的问题,在实际仿真中发现对数据对象的几何分布要求不高,不仅能更好地解决 K-means 聚类算法解决的问题,而且还在 K-means 聚类算法不宜进行的地方有很好地应用。

2 二维迭代转导支持向量分类机

为了进一步提高 SVClustering 算法的分类准确度,利用转导推理提出一种半监督学习分类方法 TDITS-VC。

一般性假设:在 t_1, t_2 时刻观测的样本集为 $\mathbf{R}_1, \mathbf{R}_2$, 训练集 $\mathbf{X}_1 = \{x_1^1, x_2^1, \dots, x_M^1\} \subset \mathbf{R}_1$, $\mathbf{X}_2 = \{x_1^2, x_2^2, \dots, x_N^2\} \subset \mathbf{R}_2$, $A = |\mathbf{X}_1|$, ($|\cdot|$ 为集合的势)。 $\Omega = \{\omega_1, \omega_2, \dots, \omega_L\}$ L 种类刻画 t_1, t_2 时刻的样本集的标号向量, $y_i^1, y_i^2 \in \Omega$ 分别代表 \mathbf{X}_1 和 \mathbf{X}_2 第 i 个样本对应的标号。假设: t_1 时刻 \mathbf{X}_1 对应的标号 $\mathbf{Y}_1 = \{y_1^1, y_2^1, \dots, y_M^1\}$ 是有效的,而 t_2 时刻 \mathbf{X}_2 对应的标号 $\mathbf{Y}_2 = \{y_1^2, y_2^2, \dots, y_N^2\}$ 是难以获得的, \mathbf{R}_1 和 \mathbf{R}_2 的统计分布几乎一致(最差的情况是不一致), \mathbf{X}_1 为训练集, \mathbf{X}_2 为转导集。

二维迭代转导支持向量机(Two Dimensional Iteration - Transductive Support Vector Classification, TDITSVC)利用 SVC,在无法获得的样本标号情况下,结合转导推理方法得到较好的结果。TDITSVC 主要思想是进行 y 维迭代转导,即在逐渐减少 \mathbf{X}_1 样本个数同时增加 \mathbf{X}_2 的样本数从而对 \mathbf{X}_2 的分类面形成过程减少 \mathbf{X}_1 对其影响;进行 x 维迭代转导,即在逐渐增大 \mathbf{X}_2 样本的松弛变量 ξ^* 的惩罚因子 C^* ,同时减少 \mathbf{X}_1 样本的松弛变量 ξ 的惩罚因子 C ,转导 \mathbf{X}_1 的信息用在 \mathbf{X}_2 的分类面建立上。

图 1 为 TDITSVC 算法流程图,其中步骤 1 表示对原始训练集 \mathbf{X}_1 进行 SVM 分类, $C^{(0)}$ (预设)训练集 \mathbf{X}_1 的初始惩罚因子生成分类面;

步骤 2 表示上面提过的 y 维迭代,生成分类面;

步骤 3 表示第 n 次 x 维迭代,有标号转导集的 x 维分组迭代的第一次迭代的惩罚因子是 C_1^* (预设, $C_1^* < C^{(0)}$),生成分类面。迭代总次数为 G (预设),由分类面(随 y 维迭代不断变化)决定标号的转导集被均分为 G 组。表示对经过第 m 次 y 维迭代后的训练集 $\mathbf{X}_1(m)$ 进行 SVM 分类,惩罚因子为 $C^{(0)}/(2G)$,生成分类面。将①的 x 维 G 次迭代转导后产生的分类面与②产生的分类面结合成一个总分类面 $\alpha(m)$ 。当训练集中样本 $\mathbf{X}_1(m)$ 为空时,得到的 $\alpha(m)$ 即为最终分类面。

这里以 2 类模式识别为例说明算法,再利用 1 against all 转化为多类分类器^[10]。

3 分选跳频信号

在第 3 方通信下,对不同组网方式的多个跳频(Frequency Hopping, FH)电台的分选是属于无监督或半

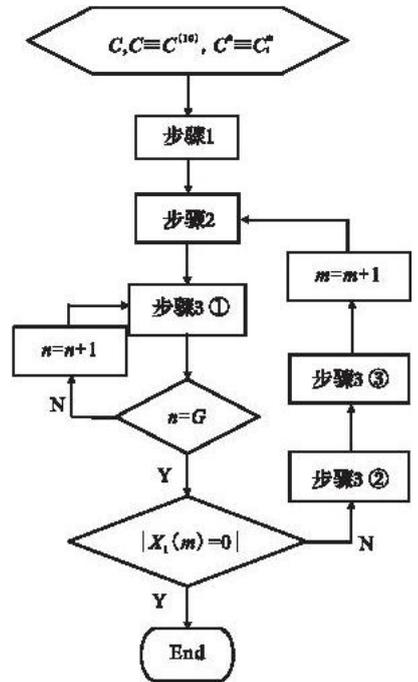


图 1 TDITSVC 算法流程图

Fig. 1 TDITSVC algorithm flow chart

监督学习的范畴,通常是利用 K-means 聚类算法进行分类,本文对 FH 电台的分选分别使用 SVClustering 算法和 TDITSVC 算法分类,取得较为满意的效果。

通过对跳频信号参数分析,提取 5 种特征参数进行跳频信号 hop 的网台分选:第 1 类是单天线测得的数据,驻留时间(跳速)、频点和跳时(第 1 个 hop 相对第 1 次扫描时间)。第 2 类是多天线测得的数据,DOA 包括仰角和方向角。第 1 类可以用短时傅里叶变换等高等信号处理的方法得到;而第 2 类通过 ESPRIT 算法等得到^[11]。再在对具有相应参数的 hop 信号进行预处理得到进行模式识别的有效 hop 信号。

实验信号参数集和噪声与干扰说明见表 1 和表 2。

表 1 FH 信号参数集

Tab.1 FH signal parameter set

跳速/(hop·s ⁻¹)	仰角, 方向角/(°)	跳时 /10 ⁻¹ ms	起始频点, 频率间隔/Hz	信噪比/dB
网络 I 同步正交网				
发射机 1	100, 30, 45	-50	9.51 M, 5k	10
发射机 2	100, 10, 87	-49		10
网络 II 异步非正交网				
发射机 3	200, 70, 350	-25	10.3 M, 5k	10
发射机 4	200, 68, 353	-15		10
发射机 5	200, 50, 210	-11		10
网络 III 异步非正交网				
发射机 6	500, 63, 147	-10	8.1 M, 20k	5
发射机 7	500, 27, 217	-5		10

表 2 噪声与干扰说明

Tab.2 Noise and interference notification

噪声	具有单位功率的白色噪声
定频干扰	存在 76 个定频干扰源,随机数生成其载频、相位、带宽、功率、仰角、方向角等参数。 干扰源 1 是一个载频 10 MHz、带宽 300 kHz、信噪比 10 dB 的宽带干扰源。 定频干扰源的基带信号都是具有相应带宽的白色噪声。

由于多径干扰的存在,检测和截获的差错率会有所提高。文中的仿真实验中提供的多径延时参数位于 0.5 ms 和 4.5 ms 之间。

为了减少伪 hop 对分类过程的影响,需要进行预处理对 hop 缓冲池中的 hop 数据结构进行整理。例如:删除驻留时间大于最大驻留时间的 hop,这些多数对应定频电台;删除驻留时间小于最小驻留时间的 hop,这些是由短促的噪声起伏导致的。文中从收集来的近 20 000 个 hop 整理出 141 个有效 hop 进行网台分选仿真实验。

3.1 SVClustering 分选

步骤 1 对具有 5 个特征参数(驻留时间、频点、跳时、仰视角和方向角)的数据点,求输入空间中数据点对应的二次优化参数 β 值,见图 2。

步骤 2 通过 β 值求出核函数 $K = \exp(-q \|x_i - x_j\|^2)$ 对应的输出特征空间的点距离所在超球面的球心的距离 R 的值,由于计算造成 R 值在实际半径上下允许范围内波动,见图 3。

步骤 3 利用等高线可以表明输入空间与输出特征空间的关系,图 4 可以看出以仰视角、方向角构成的输入空间与特征空间的关系。图中用数字标出图中 hop 数据点对应的发射机种类。

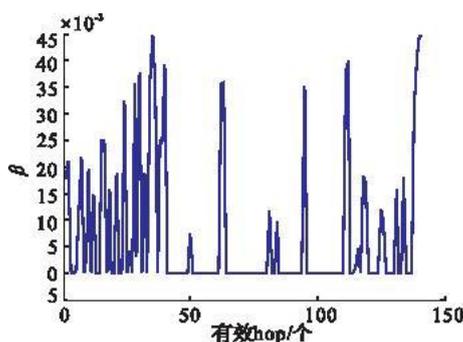


图 2 FH 信号对应的 β 值

Fig.2 The β corresponding to FH signal

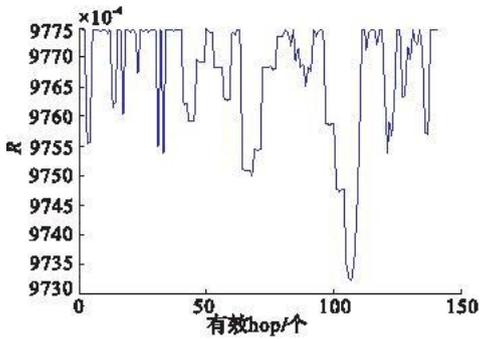
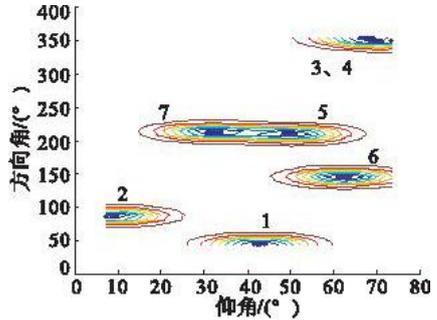
图3 $q = 169.64 \times 10^{-6}$ 时的 R 分布Fig. 3 The distribution of R at $q = 169.64 \times 10^{-6}$ 

图4 DOA 输入空间等高线分布图

Fig. 4 The contour distribution plot of the DOA input space

支持向量聚类 SVclustering 根据 hop 信号集的在驻留时间 - 频点 - 跳时 - 仰视角 - 方向角特征空间的分布情况, 形成良好的聚类边界, 再利用 DFS 的方法找出这些 hop 数据点之间的关联关系构成关联矩阵。通过关联矩阵, 将相关联进行聚类。

3.2 TDITSVC 分选

在训练集中有发射机 1-7, 共 7 类数据。

仿真 1: 训练样本为 (各类训练数据个数 - 10) 原始有效标号训练样本 + 10 个转导集样本。通过 TDITSVC 仿真, 得到发射机 6 分选正确率 = 93.33%, 发射机 7 分选正确率 = 100%, 发射机 3 分选正确率 = 88%, 发射机 5 分选正确率 = 96.55%, 发射机 4 分选正确率 = 100%, 发射机 1 分选正确率 = 93.33%, 发射机 2 分选正确率 = 86.67%。在仰角和方向角的输入空间中, 标出分类情况, 见图 5。图 6 给出发射机 3, 4 的分类图, 其中阴影部分为发射机 4。

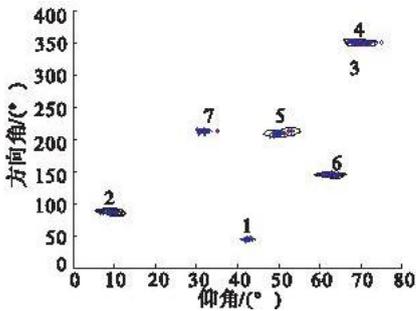


图5 仿真 1 得到的分类结果

ig. 5 The classification result gotten by the simulation

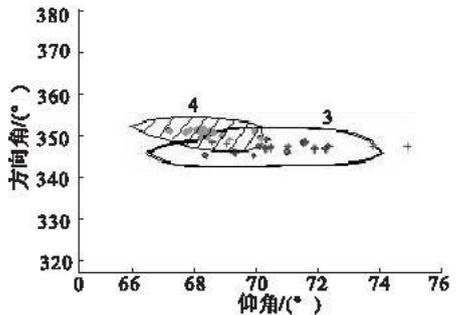


图6 发射机 3, 4 分类图

ig. 6 The classification plot of the transmitter 3 and 4

仿真 2: 在训练样本中, 增加转导集样本的占有比例。原有标号样本减少 12 个, 补充增加 12 个无标号样本, 组成训练样本进行半监督学习, 实现 TDITSVC 仿真。验证各个发射机的分类准确度为: 发射机 6 分选正确率 = 86.67%, 发射机 7 分选正确率 = 100%, 发射机 3 分选正确率 = 88%, 发射机 5 分选正确率 = 96.55%, 发射机 4 分选正确率 = 100%, 发射机 1 分选正确率 = 86.67%, 发射机 2 分选正确率 = 66.67%。

仿真 3: 在训练样本中, 增加转导集样本的占有比例。原有标号样本减少 14 个, 补充增加 14 个无标号样本, 组成训练样本进行半监督学习, 实现 TDITSVC 仿真。验证各个发射机的分类准确度为: 发射机 6 分选正确率 = 46.67%, 发射机 7 分选正确率 = 100%, 发射机 3 分选正确率 = 88%, 发射机 5 分选正确率 = 100%, 发射机 4 分选正确率 = 100%, 发射机 1 分选正确率 = 46.67%, 发射机 2 分选正确率 = 20%。

4 性能比较及方法论证

通过 3.1 节所述的 SVClustering 仿真和 3.2 节所述的 TDITSVC 仿真, 可将各发射机分选准确率的结果进行比较, 见表 3。

表 3 仿真试验数据分析

Tab.3 The analysis of the experimental data

发射机	有效 hop 数	分选准确率(%)			
		SVClustering	TDITSVC		
			仿真 1	仿真 2	仿真 3
6	15	75	93.33	86.67	46.67
7	22	90.48	100	100	100
3	25	68	88	88	88
5	29	93.1	96.55	96.55	100
4	20	58.62	100	100	100
1	15	73.33	93.33	86.67	46.67
2	15	86.62	86.67	66.67	20

可以得到以下结论:

1)除了 TDITSVC 仿真 2 的发射机 2 和仿真 3 中的发射机 6、1、2 这 4 个这样本的分选准确率低于无监督学习 SVClustering 的分选准确率,其他的数据均优于无监督学习。其中特例是由于在训练集中无标号样本占有率过高所造成的,分别为 80.00% 和 93.33%,93.33%,93.33%,此时有限的有标号样本无法给半监督学习 TDITSVC 提供有效帮助。总体来说,“分割”比“聚类”分类效果更好。

2)TDITSVC 中仿真 1、2、3 的分选性能,随着训练样本中有标号样本占有率的减少而下降。发射机 6 如图 5 所示,仿真 1 有标号占有率 33.33%,仿真 2 有标号占有率为 20%,仿真 3 有标号占有率为 6.67%。而分选准确率相应减小,仿真 1 为 93.33%,仿真 2 为 86.67%,仿真 3 为 46.67%。而发射机 7、3、5、4 的有标号样本占有率都为 50%左右,所以分选准确率并未有太大变化。

通过 2 种模式识别算法在短波电台分选中的仿真试验,可以看出 2 种算法的应用范围都很广泛。比较分选结果得出 TDITSVC 分选性能优于 SVClustering,所以条件允许的情况下,即有相应标号训练样本,并且占训练样本总数在 50%左右时,多采用 TDITSVC 进行半监督分类;如果无法满足条件则利用 SVClustering 进行无监督分类。实际应用中,无法分辨一组相关训练集中哪些在当前分类任务下有用,哪些无用,可以把它们一起拿来当成“有标号样本”使用,在迭代中逐渐消减无用样本的影响。

5 结束语

本文对完全(或部分)不知跳频信息的跳频电台分选,通过应用 SVClustering 和 TDITSVC 算法,得出一种多参数较原来单参数多层次分选更快的算法,分选性能也比较理想。但对于 100 hops/s - 2 000 hops/s 的中高速跳频,在 TDITSVC 算法中,二次优化会生成 Hessian 矩阵^[12-13],使得内存开销是样本数目的平方,因此,如何减小内存开销是提高 TDITSVC 算法速度的关键,也是下一步的任务。

参考文献:

- [1] Richard A. Poisel; Introduction to Communication Electronic Warfare Systems [M]. New York: Artech House, 2002.
- [2] Eric M, Dukic M L, Obradovic M. Frequency Hopping Signal Separation by Spatio - frequency Analysis Based on the MUSIC Method [C] // IEEE Sixth International Symposium on Spread Spectrum Techniques and Applications. New Jersey: IEEE Press, 2000: 78 - 82.
- [3] Liu Xiangqian, Sidiropoulos N D, Swami A. Blind Separation of FHSS Signals Using PARAFAC Analysis and Quadrilinear Least Squares [C] // IEEE MILCOM2001, Communications for Network - Centric Operations; Creating the Information Force. Vienna: IEEE Press, 2001: 1340 - 1344.
- [4] 国强,王常虹,郭利民,等. 分段聚类在雷达信号分选中的应用[J]. 北京邮电大学学报, 2008, 31(4): 132 - 136. GUO Qiang, WANG Changhong, GUO Limin, et al. Application of Segment Clustering in Radar Signal Sorting[J]. Journal of Beijing University of Posts and Telecommunications, 2008, 31(4): 132 - 136. (in Chinese)
- [5] 范伟. 跳频信号的盲检测和参数盲估计[D]. 合肥: 中国科学技术大学, 2005. FAN Wei. Blind Detection and Parameter Blind Estimation of FH Signals[D]. Hefei: University of Science and Technology of China, 2005. (in Chinese)

- [6] 邱磊. 基于内容的遥感图像挖掘方法研究[D]. 长沙:国防科学技术大学,2005.
QIU Lei. Research on Content – based Image Mining Methods for Remote Sensing Images[D]. Changsha:National University of Defense Technology,2005. (in Chinese)
- [7] Bruzzone L,Cossu R. A Multiple – Cascade – Classifier System for A Robust and Partially Unsupervised Updating of Land – cover Maps[J]. IEEE Transactions on Geoscience and Remote Sensing,2002,40(9):1984 – 1996.
- [8] Asa Ben Hur,David Horn,Siegelmann Hava T,et al. Support Vector Clustering[J]. Journal of Machine Learning Research, 2001, 2: 125 – 137.
- [9] Asa Ben Hur,David Horn,Hava T Siegelmann,et al. A Support Vector Method for Hierarchical Clustering[J]. Advances in Neural Information Processing Systems,2001,13:367 – 373.
- [10] XU Yi, WANG Rui. The Classification Method Study of Two – Dimension Iteration Transductive Support Vector Machine [C]//ICIE 09 WASE International Conference on Information Engineering. Taiyuan:IEEE Press,2009:32 – 35.
- [11] Richard Roy, Thomas Kailath. ESPRIT – Estimation of Signal Parameters via Rotational Invariance Techniques[J]. IEEE Trans on Acoustics, Speech, and Signal Processing,1989,37(7):984 – 995.
- [12] Ran El – Yaniv, Dmitry Pechyony, Vladimir Vapnik. Large Margin vs Large Volume in Transductive Learning [J]. Machine Learning,2008,72: 173 – 88.
- [13] Wang L,Jia Huading,Sun Shixin. A Fast and Accurate Progressive Algorithm for Training Transductive SVMs[C]//Proceeding of the 4th International Symposium on Neural Networks:Advances in Neural Networks,Part III. Berlin:Spring – Verlag,2007: 497 – 505.

(编辑:徐楠楠)

The Research on the Application of the Statistical Learning Theory Algorithm to the Hopping Frequency Signals' Separation

WANG Rui , XU Yi

(Electronics Engineering Institute, Hefei 230037, China)

Abstract: After the third party has captured and detected the frequency hopping (FH) communication signals, the lack of the priori knowledge leads to that the classic supervised learning algorithm is infeasible for separating HF signals. Even if, nowadays, many unsupervised algorithms are adopted there are still that the cluster number needs referring and multiple classification parameters separation by steps carries on the validity judging and filtering according to the priori knowledge. For the difficulties met in the course of the separation of FH signals detected in electronic support measures, we make use of the performance of the statistical learning theory (SLT), which is higher than that of the others in the small sample learning and the nonlinear classification to put forward unsupervised and semi – supervised learning algorithms based on SLT, by using which FH signals got by the third party are well classified. This research provides a kind of classification method that is of higher applicable robustness and higher accuracy rate for signals separation in FH communication reconnaissance.

Key words: frequency hopping signals; network separation; statistical learning theory; semi – supervised learning