

## 基于 SOM-SVM 频谱感知算法\*

翟旭平 孟田 杨兵兵

(上海大学 特种光纤与光接入网重点实验室 上海 200072)

**摘要:** 频谱感知是认知无线电中的关键技术。本文提出了利用 SOM-SVM 模型进行频谱分类的方法。SOM-SVM 模型是利用 SOM 的聚类特点,将含有相同特征的输入样本聚集在一起,并把离聚类中心较远的输入样本舍去。经过 20% 的样本压缩后,将含有代表性的小样本再送入 SVM 进行训练。本文的样本集通过实验平台采集,验证了基于支持向量机的频谱感知方法在实际数据测试条件下也能取得很好的感知性能。仿真结果表明,SOM-SVM 模型在低信噪比下,频谱检测率接近 100%,检测错误率也得到了很好的改善。

**关键词:** 频谱感知; 认知无线电; 支持向量机; 自组织映射网;

**中图分类号:** TN929.5    **文献标识码:** A    **国家标准学科分类代码:** 510.5015

## Spectrum sensing method based on SOM-SVM

Zhai Xuping Meng Tian Yang Bingbing

(Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Shanghai University, Shanghai 200072, China)

**Abstract:** The technology of spectrum sensing plays a key part in cognitive radio(CR). This paper proposed the SOM-SVM spectrum sensing model. This model takes the advantage of the clustering of SOM which can contain the same input samples together and decrease the input samples near by the center of the other clustering. SVM is trained by the representative training set after compressed by 20% in SOM. In this paper, sample sets are generated by laboratory instruments, verifying that the SVM-based spectrum sensing model also can obtain good performance under the condition of actual communication environment. The simulation results show that in low SNR the probability of detection is 100% and the probability of error detection is improved by SOM-SVM spectrum sensing model.

**Keywords:** spectrum sensing; cognitive radio; support vector machine; self organizing maps

## 1 引言

无线通信技术的迅猛发展使得频谱资源日益紧张,然而 0~6 GHz 之间的频谱利用率只有 15~85%。为了改善固定频谱分配制度带来的频谱利用率低的情况,认知无线电技术得到了深入的研究。它可以感知频谱空洞<sup>[1]</sup>,并在主用户不进行通信时,将频谱资源分配给次用户使用,从而提高了频谱利用率。常见的频谱感知方法有,能量检测<sup>[2]</sup>、匹配滤波器检测<sup>[3]</sup>、周期平稳过程特征检测<sup>[4]</sup>等。其中,能量检测由于其方法较为简单而且不需要待检测信号的任何先验知识而被广泛应用于认知无线电的频谱感知技术中。但是,因其判决门限依赖噪声信号,而造成信噪比较低时,能量检测方法的检测错误率较高。近年来,支持向量机的应用得到了广泛的研究,例如,行人步态识别<sup>[5]</sup>,网页分类<sup>[6]</sup>等。常用的核函数有高斯径向基核函数(RBF)、Sigmoid 核函数、多项式核函数、Poly 核函数等。由于 RBF

核函数在频谱感知中有较好的性能<sup>[7]</sup>,所以本文选取 RBF 作为核函数。为了验证基于支持向量机的频谱感知方法在实际数据测试条件下也能取得很好的感知性能,通过搭建实验平台获取训练样本集和测试样本集。本文提出了利用 SOM-SVM 模型对频谱进行分类的方法。该模型能够充分利用 SOM 无监督聚类和 SVM 有导师分类的优点,在遥感图像分类<sup>[8]</sup>、道路类型识别<sup>[9]</sup>等得到了广泛研究。SOM-SVM 模型是利用 SOM 的聚类特点,将含有相同特征的输入样本聚集在一起,并按照 20% 的比例将距离非其聚类中心较近的训练样本舍去。经过样本压缩后,将含有代表性的小样本送入 SVM 进行训练。仿真结果表明,SOM-SVM 模型在低信噪比下,可正确无误的检测出主用户的存在。

## 2 SOM-SVM 模型

## 2.1 自组织映射网

自组织特征映射算法能够自动地找出输入数据之间的

收稿日期:2016-01

\* 基金项目:国家自然科学基金(61171085)资助项目

类相似度,将相似的输入在网络上就近配置,因此是一种可以构成对输入数据有选择地给予反应的神经网络<sup>[10]</sup>。SOM 网络包括两层:输入层,模拟外界输入信息的视网膜;输出层,也叫映射层,模拟做出响应的大脑皮层。该学习算法的步骤如下。

1) 初始化。对输入层到输出层之间权值向量  $W_j = (w_{j1}, w_{j2}, \dots, w_{jN})^T$  赋予较小的数值  $\alpha$  并进行归一化处理。一般情况下,  $0 < \alpha < 1$ 。其中  $N$  为权值向量的维数,  $j = 1, 2, \dots, p$ ,  $p$  为输出神经元的个数。

2) 输入向量。输入向量的维数为  $N$ , 将向量  $X = (x_1, x_2, \dots, x_N)^T$  输入给输入层。

3) 寻找获胜神经元。计算输出层的权值向量与输入向量之间的欧式距离:

$$d_j = \|X - W_j\| = \sqrt{\sum_{i=1}^N (x_i(t) - w_{ij}(t))^2} \quad (1)$$

通过计算,寻找出一个最小距离的神经元,为胜出神经元。记为  $j^*$ , 及确定出某个单元  $k$ , 使得对于任意  $j$ , 都有  $d_k = \min_j(d_j)$ 。

4) 确定优胜邻域。以  $j^*$  为中心确定  $t$  时刻的优胜邻域,一般初始值  $N_{j^*}(0)$  比较大,大约为总节点的 50%~80%。优胜邻域  $N_{j^*}(t)$  的大小随着学习之间的增大而缩小。

5) 调整权值。按照公式修正胜出神经元在  $t+1$  时刻的权值。

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)(x_i(t) - w_{ij}(t)) \quad (2)$$

式中:  $\eta$  是一个随着时间变化逐渐减小为 0 的变量,且  $0 < \eta < 1$ 。  $\eta$  的取值可以按照式(3)选取。

$$\eta(t) = \eta(0)(1 - \frac{t}{T}) \quad (3)$$

式中:  $T$  为学习步长,  $\eta(0)$  为初始值,一般数值为 0.01~0.03, 本文选取为 0.02。

6) 计算输出。按照式(4)将高维的输入样本映射到二维的输出层。

$$Y = f(\min_j \|X - W_j\|) \quad (4)$$

$f(*)$  一般为 0~1 的函数或其它非线性函数。

7) 是否达到预先设定的要求:如达到训练时间的要求则算法结束;否则,返回步骤 2) 进入下一轮学习。

## 2.2 支持向量机

最早应用的支持向量机是线性可分的,其训练集能用一个超平面完全正确的分开,例如图 1 所示二维空间  $R^2$  上的分类问题。两种形状不同的图形代表两种训练集,假设训练样本集为  $D = \{(x_1, y_1), \dots, (x_l, y_l)\}$ ,  $x_i \in R^2$ ,  $y_i \in \{1, -1\}$ , 其中  $i = 1, \dots, l$ ,  $x_i$  代表矢量训练点,  $y_i$  代表  $x_i$  的分类标签。

支持向量机进行训练时会产生一个分类超平面,该平面  $(\omega \cdot x) + b = 0$  可以最大化距离的将两类训练集分开。为了使间隔  $2/\|\omega\|$  最大化,最优化问题应为对变量  $\omega$  和  $b$

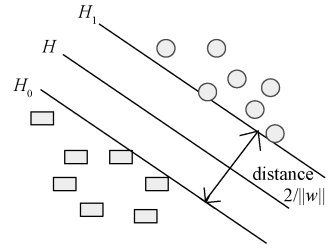


图 1 线性可分时的分类超平面

凸的二次规划问题。

$$\begin{cases} \min_{\omega} \frac{1}{2} \|\omega\|^2 \\ \text{s. t. } y_i(\omega \cdot x_i + b) - 1 > 0 \end{cases} \quad (5)$$

现实中,两类训练样本集会混合在一起,不能被完全分开的,即线性不可分。那么需要引入松弛变量  $\xi_i$  来解决此问题。具体是将出现混叠训练点  $x_i$  的位置稍作  $\xi_i$  大小的移动,使得原混叠的两类数据拉开,并且能够进行线性划分。为了控制  $\xi_i$  的大小,引入惩罚参数  $C$ 。其中,既要最大化间隔  $\frac{2}{\|\omega\|}$ , 也要最小化对训练数据的破坏程度  $\sum_{i=1}^l \xi_i$ 。那么优化问题转化为式(6)。

$$\begin{cases} \min_{\omega} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s. t. } y_i(\omega \cdot x_i + b) \geq 1 - \xi_i, i = 1, \dots, l \end{cases} \quad (6)$$

选取适当的从空间  $R^n$  到 Hilbert 空间的变换  $\Phi: x \rightarrow \Phi(x)$  以及惩罚参数  $C > 0$ , 利用 Lagrange 函数优化方法把上述二次规划问题转化为对偶问题:

$$\begin{cases} \min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j (\Phi(x_i) \cdot \Phi(x_j)) \alpha_i \alpha_j - \sum_{j=1}^l \alpha_j \\ \text{s. t. } \sum_{i=1}^l y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, i = 1, \dots, l \end{cases} \quad (7)$$

求得构造决策函数为:

$$y = \text{sgn} \left[ \sum_{i=1}^l y_i \alpha_i^* K(x_i, x) + b^* \right] \quad (8)$$

式中:  $K(x_i, x) = (\Phi(x_i) \cdot \Phi(x))$  称之为核函数。本文选择高斯径向基核函数,表达式为:

$$K(x_i, x) = \exp(-\|x_i - x\|^2 / 2\sigma^2) \quad (9)$$

基于支持向量机的频谱感知模型可看作二元假设型,  $H_1$  代表主用户信号存在,  $H_0$  代表主用户信号不存在。定义  $x(k)$  为信号在  $k$  时刻的采样点,  $n(k)$  为高斯白噪声,  $s(k)$  为待检测信号,检测模型可描述为:

$$x(k) = \begin{cases} n(k), & H_0 \\ s(k) + n(k), & H_1 \end{cases} \quad (10)$$

样本的采样点序列为:  $x(1), x(2), \dots, x(N)$ 。将  $N$  个采样点构成一个样本,  $M$  个样本构成一个样本集。可用一个  $M \times N$  的矩阵来表示:

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{pmatrix}, x_i \in R^N, i = 1, \dots, M \quad (11)$$

训练样本集中含有主用户信号的采样点的标签设置为  $y=1$ , 不含有主用户信号的采样点标签设置为  $y=-1$ 。将训练样本集及其标签送入支持向量机, 构造出一个分类超平面。分类超平面可将测试样本集进行分类, 判别出主用户信号是否存在。

### 3 实验平台搭建与数据采集

为了验证基于支持向量机的频谱感知模型在实际测试环境下也能取得很好的感知性能, 本文通过搭建实验平台采集出样本集。实验器材有: 双通道基带信号发射器 (AMU200A)、矢量信号发生器 (SMBV100A)、合路器和频谱分析仪。双通道基带信号发射器中的主用户信号类型有 PSK、QAM、FSK、ASK, 噪声信号的类型有 AWGN、单脉冲噪声 (IMP)。双通道基带信号发射器中还可以设置衰落场景, 类型有 CDMA、GSM、WLAN、WIMAX 等无线通信环境中常见的衰落。信号的带宽在双通道基带信号发生器中设置为 1 M, 采样率在频谱分析仪中设置为 10 M。

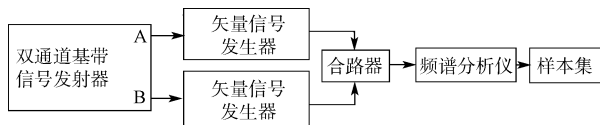


图 2 搭建实验平台

搭建的实验平台如图 2 所示, 双通道基带信号发生器的 A 通道产生主用户信号, B 通道产生噪声。噪声信号的功率设为  $-30$  dBm, 主用户信号的功率可根据信噪比的值进行设定。在主用户信号存在和主用户信号不存在的两种情况下进行数据库的采集, 每种情况下的采样点个数相同。在搭建实验平台过程中, 可根据不同类型的主用户信号、不同类型的噪声和不同的信噪比, 分别进行多次数据的采集。

## 4 仿真结果与性能分析

### 4.1 SOM-SVM 模型参数选取

为了提高训练样本集的适用性, 当测试样本集的信噪比不同时, 本文利用同一个训练样本集对 SVM 进行训练。该训练样本集中含有不同的信噪比,  $-20$  dB、 $-15$  dB、 $-10$  dB、 $-5$  dB、 $0$  dB、 $5$  dB 和  $10$  dB, 每种信噪比下的样本个数相同。由于训练样本集中含有不同信噪比的采样点, 在送入 SVM 之前, 先对其进行数值归一化处理。

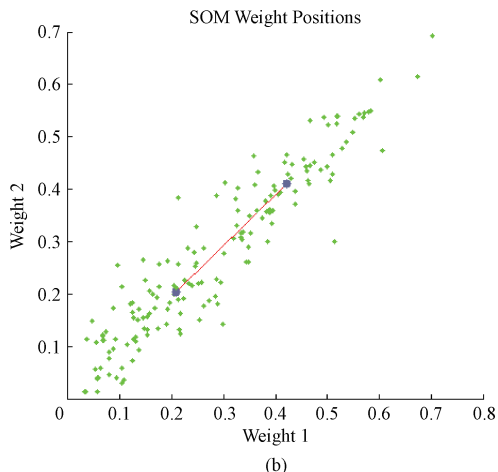
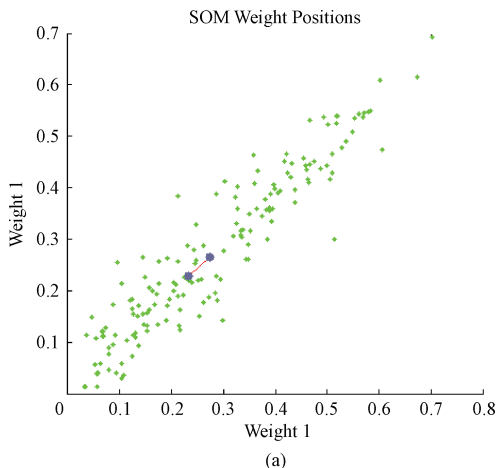
训练样本集中行向量的信噪比相同, 列向量的信噪比不同, 所以本文采用  $\min - \max$  标准化方法对样本集的列向量进行归一化处理, 它是对原始数据的线性变换。

本文利用 MATLAB 中的神经网络工具箱和 Libsvm 工具箱对 SOM-SVM 模型进行构建。对于 SOM 模型, 参数有聚类类别个数、距离函数、学习速率、学习步长、邻域函数等, 它们的选择方法及原则如下:

聚类类别个数: 频谱感知分类中, 按照主用户是否存在将输入样本集分为两类, 即输出神经元为两个。距离函数: 距离函数对 SOM 聚类效果影响很小, 本文的距离函数选取欧式距离函数。学习速率:  $\eta(t)$  取值可以按照式 (3) 选取。学习步长: 学习步长对 SOM 网络影响较大, 学习步长较小时自组织映射网络没有达到充分学习。过长时, 会出现“过学习”状态。由于本文是对频谱进行二分类, 输出神经元只有两个, 所以优胜邻域设置为 0。

对于 SVM 模型的参数选取, 核函数类型为 RBF、核参数选取为 0.9、训练样本集的信噪比为  $-5$  dB, 个数为 280、测试样本集的个数为 100、样本集的维数为 1 000。

学习步长对 SOM 网络影响较大, 下面对步长的选取进行仿真研究。图 3(a) 为学习步长等于 100 时的聚类结果图, 图 (b) 中的学习步长为 300, 图 (c) 中的学习步长为 600, 图 (d) 中的学习步长为 1 000。其中, 每幅图展示了在不同步长的情况下, 各个样本和两个聚类中心在二维空间的投影映射情况。



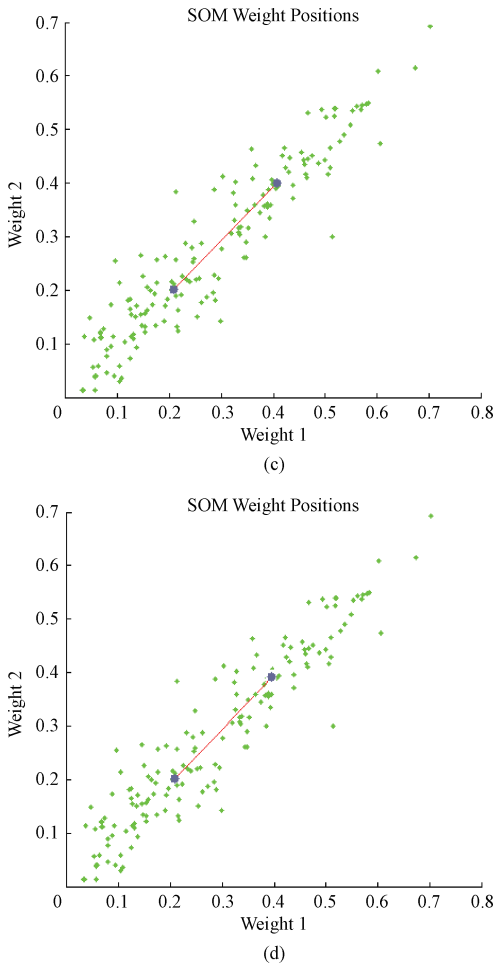


图 3 不同学习步数下的聚类结果

图 3 为二维平面图,其中横、纵坐标的“Weight1”和“Weight2”的含义是“特征分量”的意思,而不是权重。这两个新的特征分量可以理解为是高维空间中样本集的维数  $N$  个数值特征的“凝聚”,即  $N$  维空间向二维空间映射的结果。从四个图中可知,当学习步数为 100 时,两类聚类中心相距很近,表明 SOM 网还没有充分学习。当学习步长为 1 000 时,SOM 两个聚类中心相对于 300、600 时较近,表明 SOM 网出现了过学习状态。所以当学习步长为 300 时最佳。

SOM 网对样本集进行聚类之后,需要对训练样本集进行压缩。计算各个样本点与聚类中心之间的欧式距离,舍去非规整样本点,选择具有代表性的样本点对 SVM 进行训练。本文按照 20% 的比例去掉与对应聚类中心较远的那些样本点。

### 4.2 SOM-SVM 性能分析

本小节对 SOM-SVM 模型的频谱感知性能进行仿真研究。利用搭建的实验平台采集出样本集。SOM-SVM 模型的参数如下:聚类类别个数为 2 个、学习步长为 300、训练样本个数为 280、维数为 1 000、训练样本集压缩率为

20%。测试样本的个数为 100。核函数为 RBF、核参数等于 0.9、主用户类型为 BPSK、噪声类型为 AWGN。

图 4 中,  $p_{error}$ ,  $p_{false}$ ,  $p_{miss}$ ,  $p_{detect}$  分别代表 SOM-SVM 模型进行频谱感知的错误率,虚警率,漏检率和检测率。随着测试样本集信噪比的增加,频谱检测错误率也逐渐减小到 0。在低信噪比下,频谱检测率达到了 100%。即主用户存在时,SOM-SVM 模型能够正确检测到频谱的存在。

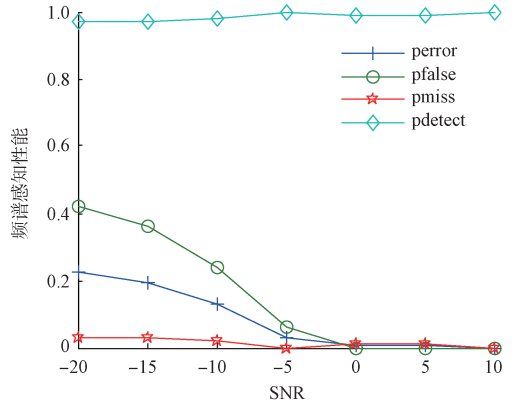


图 4 SOM-SVM 模型的频谱感知性能

为了验证 SOM-SVM 模型进行频谱感知的有效性和优越性,下面将在相同样本集下,同样进行样本集预处理,通过单独使用 SVM 进行频谱感知,对比 SOM-SVM 模型下的频谱感知性能。

图 5 中,随着信噪比的增大,SVM 模型和 SOM-SVM 模型中的频谱检测错误率逐渐减小。样本集维数为 1 000 时的频谱检测性能相对于其等于 500 时较高。SOM-SVM 模型与 SVM 模型相比,不仅减小了训练样本集的规模,而且频谱检测平均错误率也得到了很好的改善。

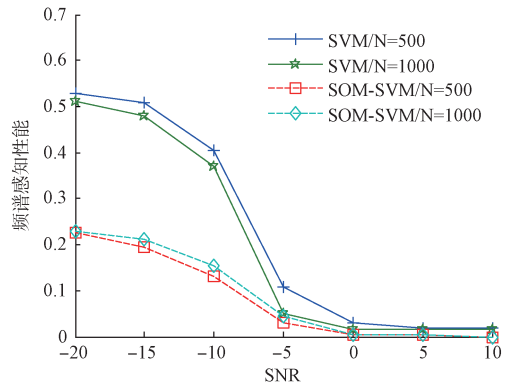


图 5 SVM 模型与 SOM-SVM 模型的频谱感知性能

## 5 结 论

本文将支持向量机应用到频谱感知技术中。通过搭建实验平台采集出训练样本集和测试样本集,验证了基于支

持向量机的频谱感知模型在实际的测试条件下也能得到很好的感知性能。为了压缩训练样本集,选取优良的样本点对 SVM 进行训练,本文提出了利用 SOM-SVM 模型进行频谱分类的方法。通过自组织映射网对训练样本集进行聚类 and 20% 的压缩处理,将具有代表性的样本集送入支持向量机进行训练。仿真结果表明,SOM-SVM 模型在低信噪比下,可正确检测出主用户的存在,并且其检测错误率也得到了很好的改善。

## 参考文献

- [1] AISSA I, FRIKHA M, TABBANE S. Dynamic spectrum hole management in cognitive radio [C]. 2011 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS), IEEE, 2011:1-4.
- [2] 张学军, 严金童, 田峰, 等. 基于差分能量检测的双门限协作频谱感知算法[J]. 仪器仪表学报, 2014, 35(6):1325-1330.
- [3] SABERALI S A, BEAULIEU N C. Matched-filter detection of the presence of MPSK signals[C]. 2014 International Symposium on Information Theory and its Applications (ISITA), IEEE, 2014:85-89.
- [4] ALI R, AHMAD A, HUSSAIN B, et al. Cooperative cognitive network: performance analysis of cyclostationary spectrum detection[C]. IEEE International Conference on Emerging Technologies (ICET), 2012:1-5.
- [5] 吴哲君, 赵忠华, 唐雷. 基于 SVM 的行人步态实时分类方法[J]. 电子测量技术, 2015, 38(7):41-44.
- [6] THAMRONGRAT P, PREECHAVEERAKUL L, WETTAYAPRASIT W. A novel voting algorithm of multi-class SVM for web page classification[C]. 2nd IEEE International Conference on Computer Science and Information Technology (ICCSIT 2009), 2009: 327-330.
- [7] YE Z, LI H. Based on radial basis Kernel function of support vector machines for speaker recognition[C]. 2012 5th International Congress on Image and Signal Processing (CISP), 2012:1584-1587.
- [8] PATRA S, BRUZZONE L. A novel SOM-SVM-based active learning technique for remote sensing image classification [J]. IEEE Transactions on Geoscience & Remote Sensing, 2012, 52 (11): 6899-6910.
- [9] LI Z, HOU J, WANG Q, et al. Road type recognition based on SOM and SVM [C]. 2011 International Conference on Consumer Electronics, Communications and Networks (CECNet), IEEE, 2011:2570-2573.
- [10] APINANTANAKON W, SUNAT K. A two-stage SOM for noise reduction in SOM map [C]. 2014 International Computer Science and Engineering Conference (ICSEC), IEEE, 2015:198-202.

## 作者简介

翟旭平, 副教授, 博士, 研究方向为认知无线电、宽带无线通信等。

E-mail: zhaixp@shu.edu.cn