

分布式无线传感器网络的目标分类

陈瑞义, 席斌, 刘健

(厦门大学信息科学与技术学院, 福建 厦门 361005)

摘要: 研究了无线传感器网络中的运输车辆的检测问题, 并利用 AR - Burg 最大熵算法对不同车辆时间序列进行谱特征提取. 分类结果是使用支持向量分类器类得到的. 仿真结果表明, 该分类方法可以达到较高的分类率, 计算快且稳定, 生成数据量小.

关键词: 无线传感器网络; 目标分类; AR - Burg 算法; 支持向量机

中图分类号: TP274, TP212.9

文献标识码: A

Classification in distribute wireless sensor network

CHEN Rui - yi, XIB in, LIU Jian

(College of Information Science and Technology, Xiamen University, Xiamen, Fujian 361005, China)

Abstract: This paper investigates the detection theory of wireless sensor networks and extracts the spectral feature vectors of different type of moving vehicles by AR - Burg algorithm. We get the classification results by using the support vector machine classifier. The result shows that high classification rate and low compute complexity can be obtained by the method of this paper.

Keywords: wireless sensor networks; classification; AR - Burg; support vector machine

无线分布式传感器网络 (WDSN) 技术^[1]国内仍处于理论研究阶段, 且多集中在路由协议控制和拓扑控制和安全控制方面. 本文着眼于从探测到的目标事件信号中提取谱特征向量, 并进行高效的分类. 数据来源于威斯康星大学无线传感器研究数据^[2].

1 相关理论

1.1 AR - Burg 最大熵谱估计

AR - Burg 最大熵谱估计是一种性能优良的参数谱估计方法, 其估计方差小, 谱分辨率高, 计算量小且稳定^[3]. 设传感器节点探测到的目标事件时间序列 $\{x(0), x(1), \dots, x(N - 1)\}$ 为输入, N 个值的 m 阶估计器的前向和后向预报误差 $e_f^{(m)}$ 和 $e_b^{(m)}$ 为输出. 其递推公式为:

$$e_f^{(m)}(n) = e_f^{(m-1)}(n) + \alpha_m e_b^{(m-1)}(n-1) \quad e_b^{(m)}(n) = e_b^{(m-1)}(n) + \alpha_m^* e_f^{(m-1)}(n-1) \quad (1)$$

其中: α_i 为反射系数, $i = 1, \dots, m$, $(\cdot)^*$ 为共轭函数. 式 (1) 相当于 m 个递推的格型级联网络. 利用 Burg 提出的前、后向预报误差功率之和最小以及 Levinson - Durbin 公式来求反射系数和 AR 模型参数 a_k , 最后计算最大熵 PSD 值表达式:

$$\hat{S}_X(\omega) = \frac{1}{2\pi} \left| 1 + \sum_{k=1}^m a_k e^{-j\omega k} \right|^{-2} \quad (2)$$

1.2 支持向量机分类

将采用非线性软间隔支持向量 C - SVC 分类器^[4]进行车辆目标分类. 其原始问题可表述如下. 优化问题: $\min_{w, b} \left[\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \max\{0, 1 - y_i (w^T \phi(x_i) + b)\} \right]$, 约束为 $y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i$. 其中 x_i 为传感器探测到的事件时间信号的功率谱密度向量, 部分作为训练集. w 为权值, b 为 x_i 到超平面的距离, C 为惩罚参数, $\phi(x_i)$ 为映射

收稿日期: 2008 - 06 - 13

作者简介: 陈瑞义 (1982 -), 男, 硕士研究生; 通讯联系人: 席斌, 男, 副教授, 博士.

函数, $\lambda_i = 0$ for $i = 1, 2, \dots, l$ 为松弛变量, y_i 为类标记. 选取 RBF核函数, 得到最优解 $\hat{x} = (\hat{x}_1, \dots, \hat{x}_l)^T$. 决策函数为:

$$f(x) = \text{sign} \left(\sum_{i=1}^l y_i K(x_i, x) + b \right) \tag{3}$$

2 目标分类

2.1 目标探测

当网络中有移动目标出现时, 目标周围的传感节点采集并记录各种模式 (声音或振动) 的时间信号, 图 1 展示了在传感器节点 1 采集到 AAV3 车辆的声音信号. 目标是否出现可利用 CFAR (Constant False Alarm Rate) 算法来实现. 当信号呈现高能量, 表示车辆真实出现, 提取其对应周期的时间信号. 其具体理论和操作可参考文 [2].

2.2 目标频谱特征的提取

实验证明车辆信号的功率谱密度是一种很好的分类特征向量 [5]. 一次处理的数据长度 3 720 个数据点, 为一个 CFAR 探测周期. 在进行 PSD 估计之前须对数据进行预处理, 具体步骤如下:

- 1) 分割数据, 每个数据段长为 3 720 个点值, 如图 (2) 最上面的分图所示.
- 2) 去线性趋势处理, 以消除机器积累性零点偏离, 其处理效果如图 2 中间的分图所示.
- 3) 五点平滑平均平滑处理, 消除毛刺噪声干扰, 效果如图 2 最后下面的分图所示.
- 4) 利用 AR - Burg 方法估计其半边 PSD 值. 截取其前 500 个点后在进行奇偶平均光滑处理和去均值归一化处理产生 250 个点的声音或振动的 PSD 作为分类的特征向量, 如图 3.
- 5) 重复上述所有过程后, 提取传感节点上所有数据段的 PSD 向量, 加上类标记 label, AAV3 标记为 2.

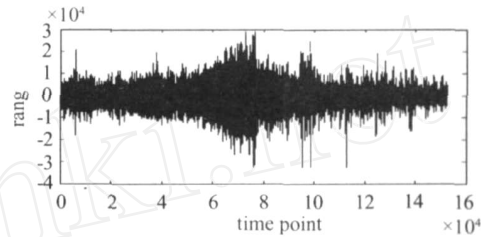
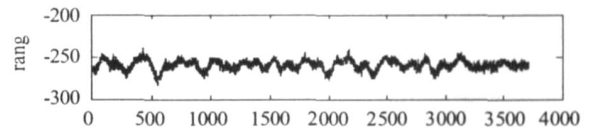
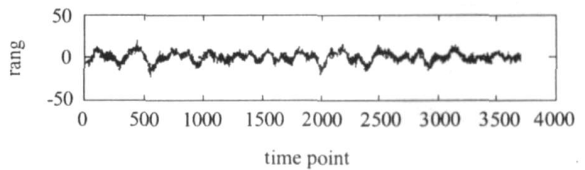


图 1 声音原始时间序列

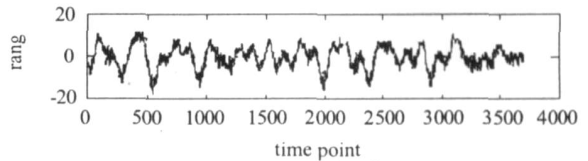
Fig. 1 Acoustic original time series



(a) Sample time series



(b) Remove linear trend



(c) Smooth handing

图 2 预处理结果 (振动)

Fig. 2 Results of pretreatment (seismic)

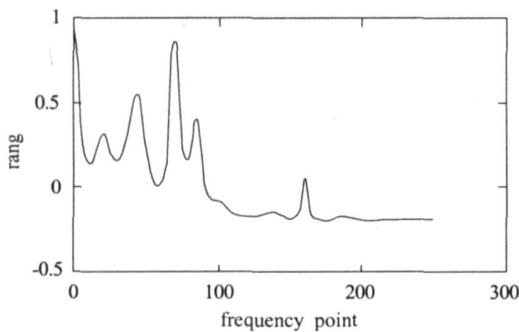


图 3 aav3 在节点 1 上声音频率谱密度 (单周期)

Fig 3 aav3 - notel - Acoustic PSD (single cycle)

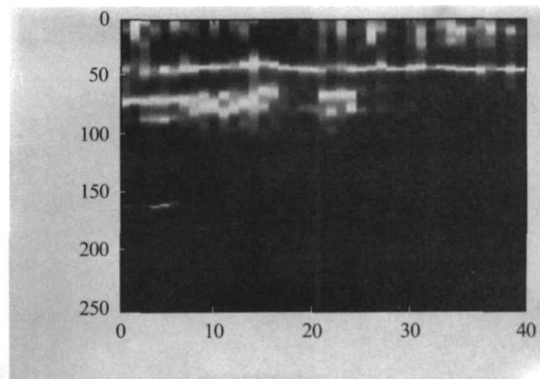


图 4 用于分类的声音特征

Fig. 4 Acoustic features for classification

DW 标记为 3, 噪声标记为 4. 这就构成了分类的特征向量. 其中第一列为车辆类别标记. 每一行作为一个分类的特征向量; 每一列为分类的属性. 利用 SVM 进行训练和分类. 图 4 展示了车辆在节点 1 上的声音功率谱密度特征矩阵图.

3 仿真结果分析

由于传感器固有采样频率为 4 960 Hz, 频率分辨为 1. 333 3 Hz, 声音和振动信号奇偶点平均后频率分辨率都为 2 666 6 Hz. AR - Burg 最大熵谱估计时, AR 模型为 100 阶的, 增加了零均值高斯白噪声类作为第三类, 其 PSD 如图 3 所示. 表 1 和表 2 展示了分类仿真的结果, 分别记为平均周期法和 AR - Burg 法进行谱特征提取并分类为方法 1 和方法 2.

表 1 用于分类的数据容量和特征集

Tab 1 Data volume and features set size for classification.

方法	内容	单事件声音		多事件声音		单事件振动		多事件振动	
		数据 /kB	特征数	数据 /kB	特征数	数据 /kB	特征数	数据 /kB	特征数
平均 周期	AAV3	186	297	803	1 280	187	297	811	1 282
	DW 3	214	341	815	1 296	215	341	819	1 521
	噪声	181	292	724	1 168	181	292	724	1 168
AR - Burg	AAV3	133	41	521	177	135	41	585	177
	DW 3	153	47	586	179	155	47	591	179
	噪声	128	40	512	160	128	40	512	160

注: kB 为数据文件大小单位; 特征的单位用“个”表示

表 2 分类结果

Tab 2 Results of classification

事件	项目	平均周期法			AR - Burg法		
单 事 件 声 音	正确分类率 /%	R11 = 75. 17	R22 = 78. 01	R33 = 99. 66	R11 = 95. 35	R22 = 100. 00	R33 = 100. 00
	错误分类率 /%	E12 = 24. 83	E13 = 0. 00	E21 = 21. 408	E12 = 100. 00	E13 = 4. 88	E21 = 100. 00
		E23 = 0. 59	E31 = 0. 00	E32 = 0. 34	E23 = 100. 00	E31 = 100. 00	E32 = 100. 00
	总分类率 /%	Accuracy = 93. 5484 (870/930)			Accuracy = 98. 442 (126/128)		
交叉验证 /%	Validation Accuracy = 92. 043			Validation Accuracy = 96. 875			
多 事 件 声 音	正确分类率 /%	R11 = 90. 25	R22 = 85. 14	R33 = 96. 88	R11 = 96. 67	R22 = 84. 54	R33 = 100. 00
	错误分类率 /%	E12 = 8. 47	E21 = 14. 43	E31 = 2. 20	E12 = 2. 67	E21 = 15. 46	E31 = 0. 00
		E13 = 1. 17	E23 = 0. 43	E32 = 0. 93	E13 = 0. 57	E23 = 0. 00	E32 = 0. 00
	总分类率 /%	Accuracy = 90. 4431 (3388/3746)			Accuracy = 92. 83 (479/516)		
交叉验证 /%	Validation Accuracy = 89. 6957			Validation Accuracy = 91. 4729			
单 事 件 振 动	正确分类率 /%	R11 = 75. 17	R22 = 78. 01	R33 = 99. 66	R11 = 81. 08	R22 = 78. 43	R33 = 100. 00
	错误分类率 /%	E12 = 24. 83	E21 = 21. 408	E31 = 0. 00	E12 = 18. 92	E21 = 21. 57	E31 = 0. 00
		E13 = 0. 00	E23 = 0. 59	E32 = 0. 34	E13 = 0. 00	E23 = 0. 00	E32 = 0. 00
	总分类率 /%	Accuracy = 83. 871 (780/930)			Accuracy = 85. 9375 (110/128)		
交叉验证 /%	Validation Accuracy = 83. 7634			Validation Accuracy = 86. 7188			
多 事 件 振 动	正确分类率 /%	R11 = 74. 35	R22 = 69. 28	R33 = 99. 914	R11 = 78. 03	R22 = 77. 05	R33 = 100. 00
	错误分类率 /%	E12 = 25. 65	E21 = 30. 66	E31 = 0. 00	E12 = 21. 97	E21 = 22. 95	E31 = 0. 00
		E13 = 0. 00	E23 = 0. 05	E32 = 0. 09	E13 = 0. 00	E23 = 0. 00	E32 = 0. 00
	总分类率 /%	Accuracy = 79. 5266 (3158/3971)			Accuracy = 84. 4961 (436/516)		
交叉验证 /%	Validation Accuracy = 79. 174			Validation Accuracy = 83. 7209			

注: R11 表示把 1 类识别为 1 类的正确率, 以此类推 R22、R33; E12 表示把 1 类识别为 2 类的错误分类率, 以此类推 E13、E21、E23、E31、E32

1) 从生成的特征维数和数据量来看: 方法 2 产生的特征数和数据量都小于方法 1, 所以方法 2 在 SVM 分类计算时比方法 1 快, 用于数据存储的空间小. 如果传感节点探测和分析的事件个数庞大的话, 这种节约计算代价和存储空间的优越性将能越来越明显.

2) 从分类的准确率来看: 方法 2 的声音单事件的 Total Accuracy 为 98.442%, 3 way 交叉验证平均分类率达到 96.875% 分别高于方法 1 对应的 93.5484% 和 92.043% 约 4.9 和 4.8 个百分点; 同样单事件振动信号的分类效果来看以及多事件声音和振动信号的分类效果方法 2 都有更好的准确率.

3) 从仿真的时间长度来看: 为减小谱估计方差方法 1 采用了分段平均, 虽然在产生 PSD 时利用 FFT 变化速度比较快, 但其分辨率降低了, 同时生成过多的特征向量, 在 SVM 分类时需要消耗较多的时间. 方法计算 PSD 时虽消耗的时间比较多, 但由于采用整个周期数据进行一次处理, 消除了分段产生的谱泄露, 提高了谱分辨率, 同时生成的特征向量也少了很多, 故在 SVM 分类阶段运行速度比较快.

4 结语

本文分类方法不管在分类正确率还是生成数据量大小上, 都优于威斯康星大学无线传感器研究小组曾采用的平均周期法进行谱特征提取和 C-SVC (多项式核) 进行分类的方法. 设计高效的节点决策融合为我们今后继续要进行的工作.

参考文献:

- [1] Estrin D, Culler D, Pister K, *et al* Connecting the physical world with pervasive networks[J]. IEEE Pervasive Computing, 2002(1): 50 - 69.
- [2] Duarte M F, Hu Y H. Vehicle classification in distributed sensor networks[J]. Journal of Parallel and Distributed Computing, 2004, 64(7): 826 - 838.
- [3] 肖先赐. 现代谱估计原理与应用 [M]. 哈尔滨: 哈尔滨工业大学出版社, 1991.
- [4] 邓乃扬, 田英杰. 数据挖掘的新方法支持向量机 [M]. 北京: 科学出版社, 2004: 193 - 196.
- [5] Li D, Wong K D, Hu Y H, *et al* Detection, classification and tracking of targets[J]. IEEE Signal Processing Magazine, 2002 (19): 16 - 30.
- [6] Wu H, Siegel M, Khosla P. Vehicle sound signature recognition by frequency vector principal component analysis[J]. IEEE Transactions on Instrumentation and Measurement, 1999, 48.

(责任编辑: 王阿军)