

改进的 DBSCAN 算法在调制识别中的应用

吴月娴, 葛临东

(信息工程大学 信息工程学院, 河南 郑州 450002)

摘要: 提出了一种基于星座聚类的通信信号调制识别新方法。将基于密度思想的 M-DBSCAN 聚类算法运用于信号星座聚类, 通过自适应更新聚类参数, 利用其良好的抗噪声性能实现了较低信噪比高斯白噪环境下信号星座图的重构。将该方法应用到对 PSK/QAM 信号的调制识别, 实验结果表明该方法是实际有效的。

关键词: 调制识别 M-DBSCAN 算法 聚类 星座图

通信信号调制方式识别是软件无线电关键技术之一。在数字幅相调制信号的识别中, 可以对信号星座图进行处理, 将形状识别与调制识别相结合, 获取关键调制特征。在现有文献中, 对星座图的重构基本采用模糊 C 均值的聚类算法^{[1][2]}。该方法的不足在于所有噪声参与到重构过程中, 使得聚类中心受噪声影响较大而产生较大误差。另外, 模糊 C 均值在聚类初始值选定上存在着主观性以及对信号先验知识的缺乏, 使最终的聚类结果具有很强的随机性。此外, 星座图重构的准确性还与载波恢复密切相关。

本文将改进的 DBSCAN 算法 (Density Based Spatial Clustering of Applications with Noise) 用于被加性高斯白噪声干扰后信号的星座图重构, 即运用减法聚类算法^[3]确定聚类初始中心, 并通过对聚类结果的有效性分析, 自适应地调整输入参数, 提高了识别性能。该算法可以有效剔除部分噪声, 且对星座旋转不敏感, 减法聚类则有效降低了聚类时间开销。为进一步提高星座图重构的准确性, 将该识别模块嵌入一多模式数字锁相环中, 实现非数据辅助的载波恢复, 解决了载波频率偏移及相位抖动的问题。对 PSK/QAM 信号识别的仿真结果表明, 此算法具有较好的抗噪性, 在低信噪比下仍有较高的识别率。

1 改进的 DBSCAN 算法

1.1 DBSCAN 算法

DBSCAN 算法^[4]是一种基于类内密度挖掘聚类具体形状的方法, 适用于任意形状几何图形的聚类, 具有较强的抗噪性。首先需要介绍两个概念:

(1) 密度直达点: 在 Eps 和 $Minpts$ (门限值) 下, 点 p 为点 q 的密度直达点, 当且仅当: ① p 在 q 的 Eps 邻域内; ② q 为 Eps 邻域内点的个数 $|N_{Eps}(q)|$ 大于 $Minpts$;

(2) 密度可达点: 若存在一组点链 $p_1, p_2, \dots, p_n, p_1=q$, 其中 p_{i+1} 是 p_i 的密度直达点, 则称 p 为 q 的密度可达点。

由此可得出聚类 C 的定义, 即在 Eps 和 $Minpts$ 条件下, 若 p 为 C 中满足其 Eps 邻域内点的个数 $|N_{Eps}(p)| \geq Minpts$ 的点, 则聚类 C 为集合: $O=\{o | o \text{ 为在 } Eps \text{ 和 } Minpts \text{ 条件下 } p \text{ 的密度可达点}\}$ 。

DBSCAN 算法实现步骤如下所述:

步骤 1: 确定输入参数 Eps 、 $Minpts$, 令总点集为 G , 非核心点集合 H , 集合 $F=G-H$, 初始时 H 为空, 聚类数 $K=0$;

步骤 2: 若 F 非空, 则在 F 中选择初始点 p ; 否则, 转至步骤 6;

步骤 3: 判定 p 是否满足核心点条件。若满足, 则在 G 中找出 p 的所有密度直达点, 记作集合 Q ; 否则, p 归入 H , 返回步骤 2;

步骤 4: 判定 Q 中元素 q 是否为边界点, 是, 则将 p 点和 q 点归入当前类, 在 Q 中另取一点 p' 代替 p 点, 并从 G 中去除 p 点和 q 点, 返回至步骤 3; 否则, $K=K+1$, 执行下一步;

步骤 5: 判定 G 中元素个数是否大于 $Minpts$, 是, 则返回至步骤 2, 否则执行下一步;

步骤 6: 将 G 中剩余元素归入噪声类, 结束。

这可以看作一个森林遍历问题。令原始点 p 为根节点, 每个节点均为上一层节点的直达点, 而每个叶子则是边界点, 聚类的过程也就是对这森林中树的遍历过程。

1.2 改进算法——M-DBSCAN

DBSCAN 算法依赖于输入参数 Eps 、 $Minpts$, 当输入参数不同时, 对数据的聚类结果也不同。文献[3]中阐述了确定 Eps 和 $Minpts$ 的方法, 但它是在对数据的先验知识较为丰富的情况下得出的。对信号星座图的重构过程中, 并不满足这样的条件。此外, 虽然 DBSCAN 算法本身与输入数据顺序无关, 但初始聚类中心的选取仍然会影响聚类速度。为此, 本文用减法聚类确定初始聚类中心, 以缩短节点遍历过程, 并在每次聚类结束时根据有效性分析结果调整。这里将改进的算法称为 M-DBSCAN 算法。

1.2.1 减法聚类算法

该算法提出了数据点的密度指标概念, 每个数据点都有自己的密度指标, 等价于数据场中的势函数概念。数据集中, 每个点都是聚类中心的候选者, 数据点 D_i 处密度指标定义为:

$$D_i = \sum_{j=1}^n \exp \left\{ -\frac{\|x_i - x_j\|^2}{(\gamma_a/2)^2} \right\} \quad (1)$$

这里半径 γ_a 是一个正数, 它定义了该点的一个领域, 半径以外的点对该点的密度指标贡献甚微。在计算每个数据点密度指标后, 选择具有最大密度指标的数据点作为第一个聚类中心, 令 x_{c_1} 为选中的点, D_{c_1} 为其密度指标。那么每个数据点 x_i 的密度指标可用下式计算:

$$D_i = D_i - D_{c_1} \exp \left\{ -\frac{\|x_i - x_{c_1}\|^2}{(\gamma_b/2)^2} \right\} \quad (2)$$

其中 γ_b 是一个正数, 定义了一个密度指标显著减小的领域, 通常取 $\gamma_b = 1.25\gamma_a$, 以避免出现相距很近的聚类中心。修正每个数据点的密度指标后, 选定下一个聚类中心, 再次修正数据点的所有密度指标。该过程不断重复, 直至产生足够的聚类中心。

1.2.2 聚类有效性分析

聚类有效性在聚类分析中有重要的研究价值, 它用来度量各聚类之间的独立性和聚类中模式的紧密性, 文献[5]详细叙述了描述聚类性能的指标。在 Xie-Beni 方法的基础上, 假定聚类数为 C , 聚类有效性函数 S 定义为:

$$\begin{aligned} S_{sep} &= \frac{(d_{min})^2}{comp} = \frac{(d_{min})^2}{\frac{1}{C} \sum_{i=1}^C \sum_{j=1}^n U_{ij} \|x_i - x_j\|^2} \quad (3) \\ &= \frac{\min_{i,j} \|c_i - c_j\|^2}{\frac{1}{C} \sum_{i=1}^C \sum_{j=1}^n U_{ij} \|x_i - x_j\|^2} \end{aligned}$$

其中: sep 、 $comp$ 分别定义了聚类的分离性和紧密性, 体现了 C 个聚类的平均离散度。聚类越独立, sep 越大, S 也越大; 聚类越紧密, $comp$ 越小, S 也越大, 这样最大化 S 就代表了一个有效的最优划分。但当 C 变的很大, 接近 n 时, 有效性函数存在单调递减的趋势, 因此, 将函数 S 改写成如下形式:

$$S = \frac{C \cdot \min_{i,j} \|c_i - c_j\|^2}{\sum_{i=1}^C \sum_{k=1}^n \delta_{ik} \|x_i - c_i\|^2} \quad (4)$$

$$\delta_{ij} = \begin{cases} 1 & U_{ij} > U_{kj}, i \neq k \\ 0 & \text{其他} \end{cases}$$

引进函数 δ 的目的是将距离聚类中心较远、对聚类的紧密性贡献不大的数据剔除掉。当 C 未达到最优解时, S 随着 C 的增加而增大, 趋向于紧密; 当 C 到达最优值而继续增大时, S 将会由最大值而减小。因此取 S 随

C 的增加达到最大时的 C 值作为聚类数。

图 1 为 M-DBSCAN 算法流程图, 其中最大聚类数目 C 根据需要确定。

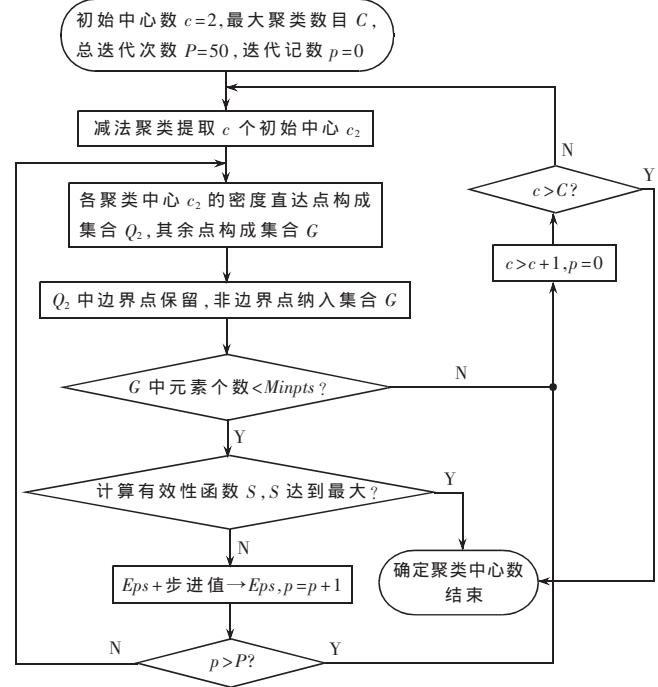


图 1 M-DBSCAN 算法流程图

2 PSK/QAM 信号的识别应用

对于数字正交调制信号, 发送信号 $S_m(t)$ 可表述为:

$$S_m(t) = A_m p(t) \cos(2\pi f_c t + \phi_m)$$

其中, A_m 为输入码元幅度, ϕ_m 为码元相位信息, f_c 为载频, $p(t)$ ($0 < t < T_s$) 为持续时间 T_s 的基带脉冲。PSK 信号的各星座点幅度相等, 而 QAM 信号的星座点幅度则有较大差异, 故可以通过星座图重构后星座点幅度的归一化方差 $R = \sigma^2 / \mu^2$ 识别 PSK 与 QAM。

对 PSK/QAM 信号调制识别简要过程如下:

- (1) 对仿真数据做归一化处理;
- (2) M-DBSCAN 算法星座聚类, 将聚类中心数目按 2 的幂次就近确定调制阶数;
- (3) 根据重构星座点的参数区分 PSK 与 QAM。

在星座图重构中, 接收信号的载波频率、初始相位、符号定时等参数的偏差都将产生影响。为提高信号星座图恢复的准确性, 考虑在载波已基本同步、载频偏差较小的情况下, 将调制类型识别模块嵌入一多模式数字锁相环结构中, 用调制识别结果控制载波恢复回路的开合, 实现非数据辅助的载波恢复^[6], 其系统框图如图 2 所示。

3 仿真结果及分析

以 QPSK 和 16QAM 信号识别为例验证 M-DBSCAN 算法的可行性。实验中假定已进行了载波同步, 剩余载波相位在小范围抖动。PSK 和 QAM 信号符号速率 $f_d = 1200$ 波特, 载波频率 $f_c = 1800\text{Hz}$ 。减法聚类算法中, $\gamma_a = 0.625$, $\gamma_b = 0.865$ 。

表 1 $SNR=5dB$ 时 QPSK 信号的星座聚类结果

C	V_1	V_2	V_3	V_4	Eps	S
4	(-1.0018, 0.9996)	(1.0558, -0.9933)	(-1.0380, -1.0026)	(1.0114, 1.0281)	0.275	0.0943
4	(-1.0147, 1.0138)	(1.0482, -0.9908)	(-0.9901, -0.9854)	(1.0120, 1.0172)	0.3	0.1144
4	(-1.0056, 1.0055)	(1.0535, -0.9833)	(-0.9970, -0.9861)	(1.0090, 1.0217)	0.325	0.1301
2	(-0.4232, 0.3034)	(1.0504, -0.9871)			0.35	0.6519

表 2 $SNR=5dB$ 时 16 QPSK 信号的星座聚类结果

C	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9
17	(-3.024, 1.042)	(3.057, -2.99)	(0.943, 1.075)	(-2.854, 3.066)	(3.042, 2.991)	(3.019, -1.002)	(0.949, -0.982)	(-0.951, -3.04)	(-0.953, 1.054)
16	(-3.025, 1.02)	(3.078, -2.989)	(0.945, 1.028)	(-2.845, 3.109)	(3.045, 2.993)	(3.005, -1.034)	(0.941, -0.945)	(-0.961, -3.013)	(-0.955, 1.011)
15	(-3.028, 1.037)	(3.030, -2.920)	(0.928, 1.054)	(-2.913, 3.003)	(3.031, 3.000)	(2.951, -1.012)	(0.950, -0.970)	(-0.995, -3.018)	(1.036, 2.993)

续表

V_{10}	V_{11}	V_{12}	V_{13}	V_{14}	V_{15}	V_{16}	V_{17}	Eps	S
(0.976, -3.103)	(2.950, 0.991)	(-0.912, 2.955)	(-2.881, -2.947)	(-1.123, -1.034)	(1.023, 2.879)	(0.811, 3.362)	(-3.050, -0.911)	0.325	0.0247
(0.992, -3.087)	(2.944, 0.998)	(-0.916, 2.967)	(-2.979, -2.975)	(-1.047, -1.038)	(0.966, 3.018)	(-3.042, -0.919)		0.35	0.0218
(-0.99, 0.094)	(1.005, -3.027)	(2.988, 0.981)	(-0.918, 3.018)	(-3.011, -2.977)	(-3.063, -0.890)			0.375	0.0502

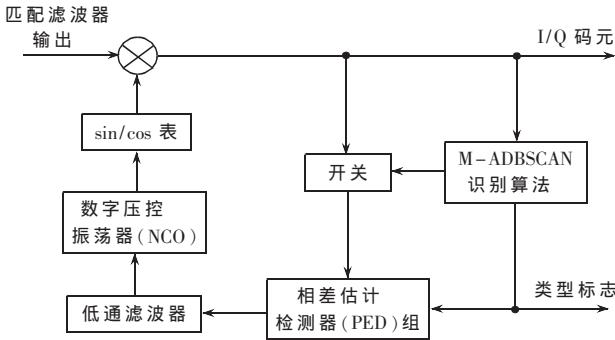


图 2 多模式数字锁相环 (PLL) 框图

表 1、表 2 分别给出了 $SNR=5dB$ 时 QPSK 和 16QAM 信号的星座聚类结果。由表 1、表 2 可以看出, 重构的 PSK 和 QAM 信号星座图在归一化幅度方差上区别非常明显, 同时通过有效性函数 S 的最大化求解, 可以自适应调整 Eps 参数, 从而达到最佳聚类结果。根据重构的星座图, 可以实现调制信号的识别并获得信号的幅度、相位信息。

表 1 中: $Minpts=7$ 条件下, 当 S 最小时, 聚类最优, 此时 $R=8.63\times10^{-5}$ 。

表 2 中: $Minpts=7$ 条件下, 当 S 最小时, 聚类最优, 此时, $R=0.12$ 。

图 3 给出了 AWGN 信道条件下, 1 000 个码元的 QPSK、8PSK 和 16QAM 随信噪比变化的正确识别率曲线。曲线表明, 识别率随信噪比的增大而增大并趋于平稳, 当信噪比达到 6dB 时, 三类信号的识别率均能达到 90% 以上。

本文将 M-DBSCAN 算法应用于 PSK/QAM 信号的调制识别, 聚类过程中运用减法聚类确定初始聚类中心, 缩短了算法的遍历过程, 根据聚类有效性自适应修改聚类参数, 利用 DBSCAN 算法的良好抗噪性, 实现了在信噪比较低情况下对信号调制方式的识别。此算法运算结

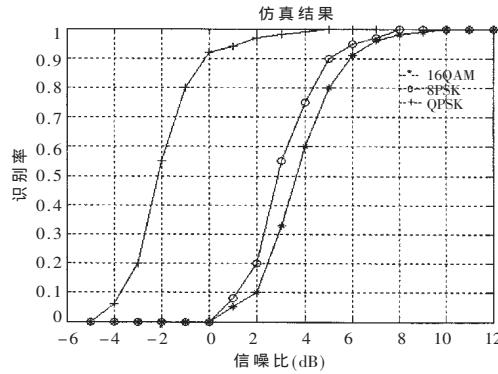


图 3 识别率与信噪比变化曲线

构简单, 适用于软件无线电接收平台。

参考文献

- Mobasseri B G. Digital modulation classification using constellation shape. Department of Electrical and Computer Engineering, Villanova, PA 19085, November 20, 1997
- Mobasseri B G. Constellation shape as a robust signature for digital modulation recognition, IEEE, 1999
- 边肇祺, 张学工. 模式识别(第二版). 北京: 清华大学出版社, 2000
- Martin Ester, Hans-Peter Kriegel, Jorg Sander et al. A Density-Based algorithm for discovering clusters in large spatial databases with noise. Institute for Computer Science, University of Munich Oettingenstr.67, D-80538 Munchen, Germany
- Maria Halkidi, Michalis Vazirgiannis. Clustering validity assessment: Finding the optimal partitioning of a data set. In Proceedings IEEE International Conference, 2001:187~194
- Kenta Umebayashi, Robert H. Morelos-Zaragoza. A method of Non-Data-Aided carrier recovery with modulation identification. In: Global Telecommunications Conference, 2001:3375~3379

(收稿日期: 2006-02-23)