

基于半监督聚类的 MQAM 信号盲识别算法研究

李苹苹 孙钢灿* 申金媛 赵海东

(郑州大学信息工程学院, 郑州 450001)

摘要 传统的聚类算法用在 MQAM(multilevel quadrature amplitude modulation, 多进制正交幅度调制)信号的调制识别中, 算法的迭代次数多, 特别对高阶调制信号运算时间长。针对此问题, 提出了一种半监督聚类重构星座图的方法, 由自适应减法聚类确定初始聚类中心, 在其周围标记部分样本点并赋予初始隶属度值 f_{ik} , 根据标记的样本点数目确定可信度参数 α 的值。用 f_{ik} 和 α 来监督隶属度和聚类中心的更新, 误差平方和函数迭代次数减少 1/2。接收端识别时, 提出基于星座图圆半径的调制识别方式, 该方法能很好应对初始聚类中心数目不准确的情况, 不需要进行聚类中心的合并与分裂。通过提取接收端星座图的特征参数 R 并与标准参数 R_s 进行比较, 实现对 MQAM 信号调制方式的识别。仿真结果表明运算时间是传统聚类算法的 1/3, 对 4~256QAM 信号的调制方式识别率在 93% 以上。

关键词 调制识别 半监督聚类 MQAM 信号

中图分类号 TN911.23; **文献标志码** A

通信信号调制方式的识别是信号分析的重要部分, 也是软件无线电的关键技术。在某些通信过程中, 接收端不知道接收信号的调制方式, 需要进行调制信号的识别以满足信号处理的需求。MQAM 信号同时进行了幅度和相位的调制, 而任何一种数字幅相调制信号都可以用星座图表示, 因此基于星座图重构的识别方法就成为数字幅相调制信号识别的主要方法之一^[1]。

星座图重构中最常用的聚类算法是均值聚类算法及其改进算法, 这些算法都是基于代价函数最优的聚类算法^[2], 可收敛于代价函数的最小值。均值聚类算法需要预先确定初始聚类中心的位置及其数目, 基于聚类重构星座图方法^[3], 给定大的初始中心数, 根据隶属度值合并聚类中心, 基于代价函数值变化情况来估计聚类中心个数。该方法可操作性差, 计算量大, 并且容易产生误判。为解决这个问题, 用基于信噪比估计的自适应减法聚类算法^[4]。但当信噪比较小时, 聚类结果和实际调制星座图的

误差较大, 因此本文用半监督聚类法对自适应减法聚类得到的初始星座图进行二次聚类, 在接收端对重构星座图提取特征参数 R 进行调制阶数判别。半监督聚类算法重构星座图, 在每个聚类中心周围标记部分样本点, 用标记的样本点指导隶属度及聚类中心的更新, 在很大程度上减少了算法的迭代次数、缩短运算时间、降低了算法复杂度。

1 基于信噪比的自适应减法聚类

针对任何一种调制方式, 在合适的“密度”半径下, 基于“密度”指标的减法聚类算法^[5, 6]能够高效的找到初始聚类中心。对不同阶数的调制信号, 根据信噪比估计值设定不同大小的“密度”半径值, 就可以完成星座图的初始聚类。下面给出基于信噪比估计的减法聚类算法。

考虑 N 个码元同步复信号 (r_1, r_2, \dots, r_N) ($k = 1, 2, \dots, N$), 不失一般性, 对其进行幅度归一化得 $r_{n,k}$ 。由于每个数据都是聚类中心的候选者, 因此 $r_{n,k}$ 处密度指标 D_k 定义为

$$D_k = \sum_{j=1}^N \exp\left(-\frac{\|r_{n,k} - r_{n,j}\|^2}{K_a (\theta_a(SNR))^2}\right) \quad (1)$$

$$\theta_a(SNR) = \left(\frac{\frac{1}{N} \sum_{k=1}^N (r_{n,k})^2}{SNR + 1}\right)^{1/2} \quad (2)$$

式中, SNR 表示信噪比值, $\theta_a(SNR)$ 定义了基于信噪比的密度指标领域, 是平均功率意义上的平均噪声幅度值, 表明了信号点的聚合程度; K_a 表示性能调

2014年3月13日收到 国家自然科学基金
(U1204604, 61172086)、中国博士后基金(2012M511587)、
河南省博士后基金(2011829)、河南省青年骨干教师
(2013GGJS-002)资助

第一作者简介: 李苹苹, 女, 硕士研究生。研究方向: 通信信号的调制识别。E-mail: 799757847@qq.com。

* 通信作者简介: 孙钢灿, 男, 副教授。研究方向: 无线通信信号盲处理, 无线通信网络安全。E-mail: iegcsun@zhu.edu.cn。

整系数,通过改变 K_a 的值找到聚类密度与平均噪声幅度的比例关系。实际应用中, K_a 的值通过多次实验得到。将式(2)代入式(1)可得:

$$D_k = \frac{\sum_{j=1}^N \exp\left(-\frac{\|r_{n,k} - r_{n,j}\|^2}{K_a \frac{1}{N} \sum_{k=1}^N (r_{n,k})^2}\right)}{\sum_{j=1}^N \exp\left(-\frac{\|r_{n,k} - r_{n,j}\|^2}{K_a P_n}\right)} \quad (3)$$

式(3)中, P_n 代表了平均每个符号的噪声功率。每个数据点的密度指标都计算完之后,选择密度最大的点作为第一个聚类中心,令 $r_{nc,1}$ 为选中的点, $D_{nc,1}$ 为其密度。每个数据点 $r_{n,k}$ 的密度将用式(4)进行更新:

$$D_k = D_k - D_{nc,1} \exp\left(-\frac{\|r_{n,k} - r_{nc,1}\|^2}{K_b P_n}\right) \quad (4)$$

式(4)中, K_b 是个常数,物理意义是密度减小的领域半径的调整系数。常数 K_b 通常大于 K_a , 为了避免出现相距很近的聚类中心,一般取 $K_b = 1.5K_a$ 。该过程不断重复,直至将所有数据点都包含在聚类中心辐射的范围内,找到所有的初始聚类中心。

2 半监督模糊聚类算法

传统的半监督模糊聚类算法^[7]是通过标记部分样本点来指导聚类中心的更新,将密度较大的点及其周围的点作为标记的样本点,并给这部分标记的样本点赋予初始的隶属度值进而指导隶属度及聚类中心的更新。设 N 个码元同步复信号组成的样本集合 $R = (r_1, r_2, \dots, r_N)$ 。 m_i 为减法聚类获得的初始聚类中心, C 表示初始聚类中心的数目, u_{ik} 是第 k 个样本对于第 i 个聚类中心的隶属度。用公式(5)选择密度较大点的周围的部分点作为标记样本点。

$$\|r_k - m_i\|^2 < K_u P_n^* ; \quad (5)$$

$$k = 1, 2, \dots, N, i = 1, 2, \dots, C$$

式(5)中, K_u 是常数用来调整密度较大点周围点数的选择^[8], P_n^* 是归一化后的信号噪声功率。由公式(6)计算标记样本点的初始隶属度值 f_{ik} 和 u_{ik} :

$$f_{ik} = \left[\sum_{j=1}^C \left(\frac{\|r_k - m_i\|}{\|r_k - m_j\|} \right)^{\frac{2}{b-1}} \right]^{-1} ; \quad (6)$$

$$k = 1, 2, \dots, h, j = 1, 2, \dots, C$$

式(6)中, h 是标记样本点的数目, C 是聚类中心的数目, b 是控制模糊程度的常数^[9], 实验中取值 2.3。按照公式(7)更新隶属度值:

$$u_{ik} = (1 - \alpha)u_{ik} + \alpha f_{ik}, \alpha = 1 - \frac{h}{N} \quad (7)$$

同时隶属度 u_{ik} 满足公式(8),即每个样本对各个聚类中心的隶属度之和为 $1^{[10]}$ 。

$$\sum_{i=1}^C u_{ik} = 1; k = 1, 2, \dots, h \quad (8)$$

更新完隶属度之后按公式(9)更新聚类中心

$$m_i = \frac{\sum_{k=1}^h (u_{ik})^2 r_k}{\sum_{k=1}^h (u_{ik})^2} \quad (9)$$

隶属度和聚类中心都更新完之后代入公式(10)计算误差平方和函数^[11], 计算公式如下:

$$J_e = (1 - \alpha) \sum_{i=1}^C \sum_{k=1}^h u_{ik}^b \|r_k - m_i\|^2 + \alpha \sum_{i=1}^C \sum_{k=1}^h (u_{ik} - f_{ik})^b \|r_k - m_i\|^2 \quad (10)$$

由公式(7)可知, α 与标记的样本点数成反比,在此可将 α 值理解为对标记样本点的可信度。标记的样本点占样本总数的比例越小,说明标记的点离密度最大点越近,从而标记的点成为聚类中心的可能性越大、可信度越高,相应 α 值就越大;相反标记的点越多可信度就越低, α 值就越小。在隶属度更新和误差平方和函数的求解中, α 的引入体现了标记样本点的监督和指导作用,是半监督思想的具体体现。通过迭代使代价函数逐渐收敛,收敛的判决条件如下:

$$J_e(k+1) \approx J_e(k) \quad (11)$$

传统的半监督聚类算法,用公式(5)选择标记点,利用标记点监督指导聚类中心和 J_e 的更新,在更新过程中所有的样本点都参与运算。这里存在两个问题:①标记点如何选择;传统算法中基于密度思想来标记样本点,密度大的点被选为标记样本点的可能性大,但是这种思想在信噪比较低时,噪声密度较大的点易被选为标记样本点从而错误指导聚类中心的更新。②更新过程中是否全部的样本点都要参与运算。在噪声功率和 K_u 一定的情况下,参与运算的样本点数目决定了算法的时间复杂度。参与运算的样本点数越多,算法花费的时间越多,所以如何减少参与运算的点数才是降低算法复杂度的关键。

基于此本文在传统聚类算法的基础上提出如下改进:在初始样本点数 N 一定、初始聚类中心确定的情况下,按公式(5)在每个聚类中心周围标记出大部分样本点并按距初始聚类中心的距离远近排序,然后对排过序的这部分点每隔一个点取一个与初始中心一起作为最终标记点,最后用标记的样本点指导隶属度和聚类中心的更新,从而在很大程度上缩短了运算时间,提高了运算效率。本文用等间

隔法取标记点,一方面保证了标记样本点的覆盖范围,确保聚类结果精度;另一方面也在一定程度上避免了相距过近的两个点重复参与运算,降低了算法的时间复杂度。

3 基于星座图圆半径的 MQAM 信号识别算法

通过对 MQAM 方形星座图的研究,星座图上的每个点可以按照到原点的距离不同划分到半径不同的圆上。为了实现对不同调制阶数的比较,定义一个相对的判别标准,即定义变量 R 使其等于星座图的圆中最大半径与最小半径之比:

$$R = \frac{r_{\max}}{r_{\min}} \quad (12)$$

首先根据标准星座图计算出标准参数值 R_s 。然后在得到 C 个聚类中心之后,计算每个中心所在圆的半径值并排序,取最大的 4 个半径值的均值作为最大半径 r_{\max} ,取最小的 4 个半径值的均值作为最小半径 r_{\min} ,从而求得参数 R 。将 R 和 R_s 比较实现 MQAM 信号不同调制阶数识别。基于半径的调制方式的识别方法,重构星座图的特征值提取简单,调制方式判决方便;能胜任星座图出现偏移的情形,星座图中部分中心偏移不影响正确判决;最大优点是当出现聚类的中心数比实际的星座图的中心数多或者少时,不需要进行中心的合并和分裂。

4 仿真结果

本文在 Gaussian 白噪声,误码率 15% 的情况下,对 4~256 阶 MQAM 信号分别进行 100 次实验。其中对 4QAM、8QAM、16QAM、32QAM 使用 2 000 点样本数据进行实验,对 64QAM、128QAM 信号使用 4 000 点样本数据进行实验,256QAM 使用 8 000 点样本数据进行实验。

4.1 聚类准确性和聚类时间

本文对传统聚类算法和半监督聚类算法分别进行 100 次的实验仿真,并统计了两种算法的运算点数、聚类中心数和运行时间如表 1。从表 1 可以看出,传统聚类算法对 4QAM、16QAM、...、256QAM 分别用 2 000、4 000、8 000 样本点进行运算更新,而半监督算法参与运算的样本点数近似为传统算法的 1/3。从聚类中心数目来看,改进后的算法在减少了大量参与运算的样本点数的情况下,并没有降低聚类中心的准确性。从算法的运算时间来看,传统的聚类算法运算所花费的时间近似是改进后算法的 3 倍。

表 1 算法改进前后聚类中心数和时间对比
Table 1 Algorithm of before and after cluster center numbers and time

阶数	样本点数		聚类中心数		时间/s	
	传统	改进	传统	改进	传统	改进
4	2 000	670	4	4	2.05	0.98
8	2 000	710	8	8	3.64	1.58
16	2 000	628	16	16	7.70	2.59
32	2 000	688	33	33	15.32	5.51
64	4 000	1 363	66	66	77.86	26.04
128	4 000	1 372	128	128	211.36	69.99
256	8 000	2 750	260	260	1 076.52	393.43

4.2 误差平方和函数曲线

图 1 和图 2 分别是 16QAM 信号迭代 50 次的误差平方和函数曲线,由图 1 看出传统的聚类算法要迭代 8 次时才逐渐收敛,而半监督聚类算法迭代 4 次就趋于收敛。迭代次数减少一半,算法运行所花费时间大大缩短。针对 128QAM、256QAM 等更高阶信号,参与运算的样本点更多,时间复杂度更高,运用半监督聚类算法优势会更加明显。因此通过半监督聚类算法,对高阶调制信号来说,时间缩短更明显、效率更高。

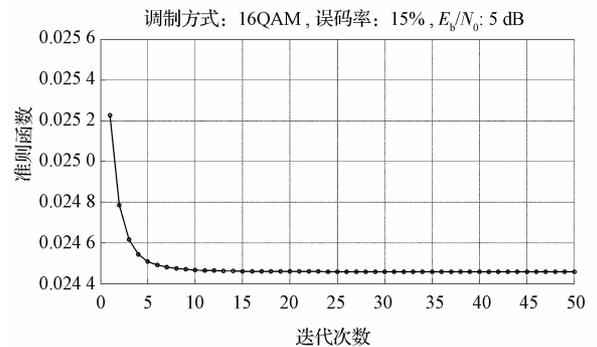


图 1 传统聚类算法的误差平方和曲线
Fig. 1 Error square curve before improvement

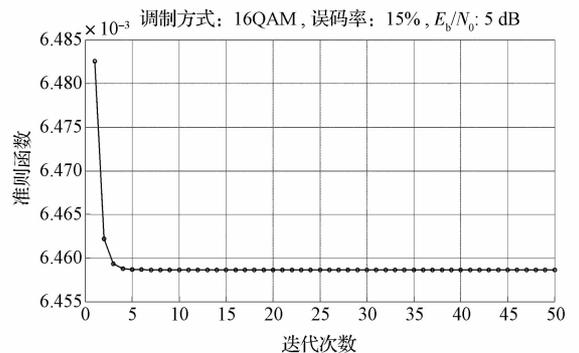


图 2 半监督聚类算法的误差平方和曲线
Fig. 2 Error square curve after improvement of the algorithm

4.3 基于半径的 MQAM 信号识别

根据标准的 R_s 值将不同阶数的调制信号的 R 值划分不同的范围如表 2 所示。对于不同阶数的调制信号将 R 落在相应范围内时即将信号的调制方式判别为相应的调制阶数。结合表 1 和表 2 可以看出对于低阶调制信号聚类中心数目准确,识别率达到 100%,对于高阶信号尽管聚类中心数目不太准确,但是用本文提出的基于半径的调制识别算法识别率仍然在 95% 左右。因此,重构星座图基于半径的识别算法在聚类中心数目不准确时能够正确识别,不需要聚类中心的合并和分裂,算法简单高效。

表 2 基于星座图圆半径的识别率

Table 2 Recognition rate based on constellation diagram circle radius

调制阶数	R_s	R	识别率
4QAM	1	0 ~ 1.6	100%
8QAM	2.2	1.6 ~ 2.6	100%
16QAM	3	2.6 ~ 3.5	100%
32QAM	4.1	3.5 ~ 5.6	100%
64QAM	7	5.6 ~ 8.1	97%
128QAM	9.2	8.1 ~ 12.1	93%
256QAM	15	> 12.1	95%

4.4 算法复杂度分析

文中 N 表示样本总数, C 表示聚类中心数目, h 表示标记样本点数目。改进前的聚类算法在计算初始聚类中心时的时间复杂度为 $O(N + C)$, 隶属度更新时的时间复杂度为 $O(NC)$, 聚类中心更新时的时间复杂度为 $O(C)$ 。即算法总的时间复杂度为 $O(N + C) + O(NC) + O(C)$ 。改进之后算法总的时间复杂度为 $O(N + C) + O(hC) + O(C)$, 而 $N \gg h$, 通过仿真表明改进之后算法的运算时间减少为原来的 $1/3$ 。

5 总结

本文用半监督聚类算法,由减法聚类得到初始聚类中心,用半监督思想标记每个聚类中心周围部分样本点参与隶属度和聚类中心的更新,与传统聚类算法的所有样本点都参与运算相比,在很大程度上减少了参与运算的样本点,降低了算法的时间复杂度,提高了运算效率。但是参与更新的样本点数减少,聚类中心的更新尺度减小,即初始聚类中心的权重变大,对初始聚类中心的依赖性大。因此聚类

中心的准确性和算法的复杂度需要折中处理,找到合适的分界点。

参 考 文 献

- 1 王华奎, 候健. 一种基于星座图聚类的 MQAM 识别方法. 信息传输与接入技术, 2009; 35(3): 35—38
Wang H K, Hou J. MQAM recognition based on research of constellation clustering. Information Transmission and Access Technology, 2009; 35(3): 35—38
- 2 刘方. 数据挖掘中半监督 K -均值聚类算法的研究与改进. 吉林: 吉林大学, 2010
Liu F. A semi-supervised K -average clustering algorithm in data mining research and improvement. Jilin: Jilin University, 2010
- 3 张路平, 王建新. MQAM 信号调制方式盲识别. 电子与信息学报, 2011; 22(2): 332—336
Zhang L P, Wang J X. Blind modulation recognition algorithm for MQAM signals. Journal of Electronics & Information Technology, 2011; 22(2): 332—336
- 4 孙刚灿. 非协作数字通信信号调制方式识别算法研究. 北京: 北京理工大学, 2008
Sun G C. Research on modulation recognition of digital communication signals in non-cooperative communication environment. Beijing: Beijing Institute of Technology, 2008
- 5 Yang Q, Zhang D, Tian F. An initialization method for fuzzy c -means algorithm using subtractive clustering, 2010 3rd International conference on Intelligent Networks and Intelligent Systems, Shenyang, 2010: 393—396
- 6 L Yanling, L Bingbing, Y Changyi. Modulation classification of MQAM signals using particle swarm optimization and subtractive clustering. 2010 the 10th IEEE International Conference on Signal Processing, Beijing, 2010: 1537—1540
- 7 曾山. 模糊聚类算法研究. 武汉: 华中科技大学, 2012
Zeng S. Research on fuzzy clustering algorithm. Wuhan: Huazhong University of Science and Technology, 2012
- 8 Chen C. A semi-supervised feature selection method using a non-parametric technique with pairwise instance constraints. Journal of Information Science, 2013; 39(3): 359—371
- 9 Coves T F, Hruschka E R, Ghosh J. A study of k -means-based algorithms for constrained clustering. Intelligent Data Analysis, 2013; 17(3): 485—505
- 10 Hashemi H, Javaherian A, Babuska R. A semi-supervised method to detect seismic random noise with fuzzy GK clustering. Journal of Geophysics and Engineering, 2008; 5(4): 457
- 11 de A T De Carvalho V M A F. An adaptive semi-supervised fuzzy clustering algorithm based on objective function optimization. IEEE World Congress on Computational Intelligence, Brisbane, Australia, 2012

Research on Blind Identification of MQAM Signal Based on Semi Supervised Clustering Algorithm

LI Ping-ping, SUN Gang-can^{*}, SHEN Jin-yuan, ZHAO Hai-dong

(School of Information Engineering, Zhengzhou University, Zhengzhou 450001, P. R. China)

[Abstract] Traditional clustering algorithms were used in MQAM (multilevel quadrature amplitude modulation) modulation recognition of signals; the number of iterations is more, particularly to the high-order modulation signals. To solve this problem, a method of semi-supervised clustering was presented to reconstruct constellation diagram. The initial clustering centers were determined by the subtractive clustering adaptive. Around them marking part of the sample point and given membership value f_{ik} , the value of parameter α of credibility labeled was determined according to the number of sample points. With f_{ik} and α to monitor and update the membership of the cluster center, iterations of squared error function reduced by 1/2. When identifying at the receiving end, the modulation recognition method was put forward based on constellation radius. The method can cope with the number of initial cluster centers inaccurate well, rather than merge with or split cluster centers. At the receiving end, extracting characteristic parameters R and compared with the standard parameters R_s , the recognition of MQAM signal modulation was achieved. Simulation results show that the computation time is 1/3 of the traditional clustering algorithms, the modulation recognition rate of 4 ~ 256QAM signals over 93%.

[Key words] modulation identification semi-supervised clustering MQAM signal

(上接第 192 页)

Wellbore Temperature Model Study and Application in Heavy Oil Thermal Recovery Well

LIU Guang-tian^{1,2}, SHAN Xue-jun²

(China University of Geosciences (Beijing)¹, Beijing 100083, P. R. China;

Sinopec International Petroleum Exploration & Production Corporation², Beijing 100029, P. R. China)

[Abstract] Exactly predicting wellbore temperature can guide choosing appropriate oil production technology in heavy oil production process, prevent oil wax and tacking. Using convection-diffusion model calculate annulus fluid heat transfer between tubing and rod, two-dimensional unsteady wellbore temperature distribution mathematical model is established. And the model realizes numerical solution through control volume method. The calculation results and the measured wellbore temperature coincide satisfactorily. Through analysis the influencing factors of wellbore temperature, the results show that the entrance temperature and flow rate has the most obvious influence on the wellbore temperature field, there exists an optimum range of hot fluid incorporation deep and the hollow rod cyclic heating method is better than that of casing. It is of guiding significance to improve the efficiency of oil production in heavy oil wells.

[Key words] heavy oil thermal recovery wellbore temperature model application