

## 一种 K-means 改进算法的软扩频信号伪码序列盲估计

张天骐 杨 强\* 宋玉龙 熊 梅

(重庆邮电大学信号与信息处理重庆市重点实验室 重庆 400065)

**摘 要:** 针对软扩频信号因采用了编码技术使得伪码序列难以估计的问题, 该文提出一种基于 K-means 聚类改进的软扩频信号伪码序列盲估计方法。该方法首先以单倍伪码周期的窗长对接收信号进行数据分段以构造观测数据矩阵, 其次利用相似测度的理论从观测数据中寻找出 K-means 算法最优的初始聚类中心点, 然后通过搜索平均轮廓系数(Silhouette Coefficient, SC)最大的绝对值以完成伪码集合规模数的估计, 最后找到估计的伪码集合规模数所对应的聚类中心点集合, 进一步完成对软扩频信号伪码序列的盲估计。通过仿真实验表明, 在伪码序列估计错误概率低于 0.1 的情况下, 该文方法比未改进方法提高信噪比约 4 dB; 而且在同一条件下, 该文方法对信号的盲解扩性能优于未改进的方法。

**关键词:** 软扩频信号; 伪码序列; K-means 聚类; 盲估计

中图分类号: TN911.7

文献标识码: A

文章编号: 1009-5896(2018)01-0226-09

DOI: 10.11999/JEIT170306

## Blind Estimation PN Sequence in Soft Spread Spectrum Signal of Improved K-means Algorithm

ZHANG Tianqi YANG Qiang SONG Yulong XIONG Mei

(Chongqing Key Laboratory of Signal and Information Processing, Chongqing University of Posts and  
Telecommunications, Chongqing 400065, China)

**Abstract:** For the problem of the soft spread spectrum signal Pseudo-Noise (PN) sequence is difficult to estimate by using the coding technology, a blind estimation PN sequence method of soft spread spectrum signal is proposed based on improved K-means algorithm. Firstly, the received signal is divided into continuous non-overlapping temporal vectors according to one period of PN sequence to construct observation data matrix. Secondly, the similarity measure theory is applied to find out the optimal initial clustering center point of K-means algorithm from the observed matrix. Then the number of scale of PN sequence can be estimated by searching for the maximum absolute value of the average Silhouette Coefficient (SC). Finally, the estimated clustering center point corresponding to the number of scale of PN sequence is found, the blind estimation PN sequence of the soft spread spectrum signal is further completed. The simulation results show that the proposed method improves the Signal-to-Noise Ratio (SNR) about 4 dB compared to the traditional method under the condition of the estimation error probability of PN sequence is less than 0.1. Moreover, the blind despreading performance is also better than unmodified method under the same condition.

**Key words:** Soft spread spectrum; Pseudo Noise (PN) sequence; K-means clustering; Blind estimation

### 1 引言

在现代通信系统中, 直接序列扩频(Direct

Sequence Spread Spectrum, DSSS)作为扩展频谱通信的主要技术之一, 是通信技术的一次重大突破。由于 DSSS 信号采用了伪随机编码调制信息码, 使其频谱得以展宽, 较之常规的窄带通信信号具有隐蔽性好, 抗干扰、低截获概率、保密性强等优点被广泛地应用于军事通信和民用通信<sup>[1]</sup>。但是, 传统 DSSS 信号的抗多径、抗干扰和多址复用的能力都是以展宽信号频谱为代价的, 这使得在带宽较窄以及处理增益要求较高的信道中传输时, 传统的 DSSS 信号就无法达到传输的要求。为了解决传统的 DSSS 信号在上述信道中进行信号传输所存在的问题, 研

收稿日期: 2017-04-07; 改回日期: 2017-07-11; 网络出版: 2017-09-14

\*通信作者: 杨强 727491991@qq.com

基金项目: 国家自然科学基金(61671095, 61371164), 信号与信息处理重庆市市级重点实验室建设项目(CSTC2009CA2003), 重庆市教育委员会科研项目(KJ130524, KJ1600427, KJ1600429)

Foundation Items: The National Natural Science Foundation of China (61671095, 61371164), The Project of Key Laboratory of Signal and Information Processing of Chongqing (CSTC2009 CA2003), The Research Project of Chongqing Educational Commission (KJ130524, KJ1600427, KJ1600429)

究者们提出了一种高效的直扩通信技术，即采用编码的软扩频(或称多进制扩频)技术<sup>[2]</sup>。

软扩频通信系统与传统的扩频通信系统相比，其优点在于提高了传输效率、占用的带宽小，获得扩频增益的同时也获得了编码的增益。基于软扩频信号以上的优点，使得软扩频通信已成为一种高效率传输和通信的有效途径。当前软扩频信号已被广泛地使用，如美军使用的联合战术信息分配系统<sup>[3]</sup>(Joint Tactical Information Distribution System, JTIDS)就采用了(32, 5)的软扩频通信技术；挪威战地网使用的战术数字通信<sup>[3]</sup>(Tactical Digital Communication, TDC)也提出采用(256, 8)和(32, 7)的正交矩阵编码。因此，在非协作情况下，完成对软扩频信号的盲解扩以及信号中的各项参数进行有效的检测和估计是一个值得研究的重要课题。

由于软扩频信号引入了编码的调制方式，较之传统的 DSSS 信号采用乘性调制的扩频方式有所不同，所以对传统 DSSS 信号伪码序列盲估计的主模解扩方法，如对于长、短码 DSSS 信号伪码序列估计的奇异值分解法<sup>[4,5]</sup>、神经网络法<sup>[6]</sup>、最小二乘算法<sup>[7]</sup>、三阶相关函数法<sup>[8,9]</sup>等不适用于软扩频信号的伪码序列估计。目前，对于软扩频信号参数和序列估计研究甚少。文献[3]分析了 JTIDS 信号伪码序列的特性，对接收序列进行循环移位累加取集平均的方法实现了信号的伪码序列估计，但该方法对扩频码所使用的伪码序列类型具有局限性。文献[10]提出使用聚类的方法实现了对软扩频信号伪码序列的盲估计，该方法通过平均侧影度为聚类指标估计出了伪码序列集合规模和伪码序列，但是该方法对聚类的初始中心点进行了随机的选取，对伪码序列估计的准确性有一定的影响。为了提高 K-means 算法对软扩频信号伪码序列估计的准确性和算法自身的鲁棒性，本文提出一种改进的 K-means 算法。首先对聚类的初始中心点进行选取，然后通过搜索聚类评价指标平均轮廓系数<sup>[11]</sup>(Silhouette Coefficient, SC)最大的绝对值来完成伪码集合规模数的估计，最后以该伪码集合规模数进行聚类，并选择与其所对

应的聚类中心点集合取符号，即可完成软扩频信号伪码序列的盲估计。

## 2 软扩频信号模型

本文采用的软扩频信号模型如图 1 所示，其中信息码  $d_n(t)$  经过串并变换后得到  $k$  比特的并行数据，即将信息码以每  $k$  比特进行分组，每组信息码含有  $M = 2^k$  个状态，然后根据每组信息码的状态选取对应的  $M$  条伪码序列中的一条伪码序列进行输出(即  $M$  条伪码序列与  $k$  比特信息码的  $M$  个状态相对应)，最后经过平方根升余弦滚降滤波器完成码片成型。

设待传输的信息数据为

$$d(t) = \sum_{n=0}^{+\infty} d_n h_d(t - nT_d) \quad (1)$$

式中， $d_n$  为信息码元， $T_d$  为每一位信息码元的周期， $h_d(t)$  为矩形切普脉冲函数， $h_d(t) = \begin{cases} 1, & 0 \leq t \leq T_d \\ 0, & \text{其他} \end{cases}$ 。对信息数据  $d(t)$  按每  $k$  位一组进行

数据传输，则  $d(t)$  可进一步表示为

$$d(t) = \sum_{i=0}^{+\infty} d_k(t - iT_0) \quad (2)$$

式中， $d_k(t) = \sum_{l=0}^{k-1} d_l h_d(t - lT_d)$ ， $T_0 = kT_d$  为  $k$  位信息码元的周期。

若  $k$  位信息码元的加权值  $p = \sum_{l=0}^{k-1} d_l 2^{k-1-l}$ ， $p$  表示与传输  $k$  位信息码元相对应的  $2^k$  条伪码序列的编号。则所用的伪码序列为

$$c_j(t) = \sum_{n=0}^{N-1} c_{j,n} g_c(t - nT_c) \quad (3)$$

式中， $j = 0, 1, \dots, 2^k - 1$ ， $N$  为伪码序列的长度， $T_c$  为每一位伪码码元的周期且伪码序列的周期为  $T_0 = kT_d = NT_c$ ， $c_{j,n} \in \{\pm 1\}$  表示第  $j$  条伪码序列的第  $n$  个伪码码元， $g_c(t)$  也为矩形切普脉冲函数，

$g_c(t) = \begin{cases} 1, & 0 \leq t \leq T_c \\ 0, & \text{其他} \end{cases}$ 。所以扩展后的伪码序列为

$$x(t) = \sum_{i=0}^{+\infty} c_p(t - iT_0) \quad (4)$$

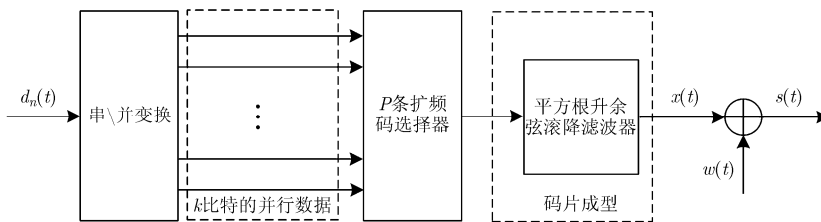


图1 软扩频信号模型

式中,  $c_p(t)$  的下标  $p$  即为  $k$  位信息码元  $d_k(t - iT_0)$  所对应的加权值。

不失一般性, 软扩频信号经过一理想的加性高斯白噪声信道, 在接收端接收到的基带软扩频信号可表示为

$$s(t) = A_x x(t - \tau) + w(t) \quad (5)$$

式中,  $A_x$  ( $A_x > 0$ ) 为信号的幅度,  $\tau$  为信号的传输时延,  $\tau \in [0, N]$  为了简便规定其为整数,  $w(t)$  是均值为零, 方差为  $\sigma_n^2$  的高斯白噪声, 而且  $w(t)$  和  $s(t)$  相互统计独立。假设软扩频信号的伪码周期和码片速率已由文献[12]中的方法估计得到, 则用估计出的伪码码片速率对式(5)进行采样, 进一步可得到离散基带信号模型为

$$s(n) = A_x x(t - \tau) + w(t) |_{t=nT_c} = x(n - \tau) + w(n) \quad (6)$$

把采样后的离散信号序列  $s(n)$  以伪码序列周期  $NT_c$  连续分段, 每段信号含有  $N$  个采样值, 即

$$\mathbf{s}_m = [s(mN - N) \ s(mN - N + 1) \ \cdots \ s(mN - 1)] \quad (7)$$

式中,  $m = 1, 2, \dots, M$  表示第  $m$  个数据,  $M$  表示总的数组数, 则采样后的数据矩阵为  $\mathbf{S} = [\mathbf{s}_1 \ \mathbf{s}_2 \ \cdots \ \mathbf{s}_M]^T$ 。

在非合作通信条件下, 想要知晓截获信号中的有用信息, 必须先要估计出用于通信传输的  $P$  ( $P = 2^k - 1$ ) 条伪码序列集合  $\mathbf{C} = \{\mathbf{c}_0(t) \ \mathbf{c}_1(t) \ \cdots \ \mathbf{c}_p(t)\}$ , 从而进一步实现对软扩频信号的盲解扩。而传统的直接序列扩频信号属于  $k = 1$  即调制所用的伪码序列  $\mathbf{c}_0 = \mathbf{c}_1 = \cdots = \mathbf{c}_p$  时软扩频信号的特殊形式, 所以对软扩频信号伪码序列的估计问题有着重要的研究意义。

### 3 改进 K-means 聚类算法的伪码序列盲估计

由于软扩频信号引入了编码的技术, 所以使得传统直扩信号伪码序列盲估计的方法不再适用于软扩频信号。根据软扩频信号的特性和扩频原理, 可以借鉴模式识别中用于无监督聚类的 K-means 方法<sup>[13]</sup>实现软扩频信号的伪码序列估计。聚类分析是数据挖掘和模式识别等领域的重要研究工具, 而 K-means 算法作为其中之一, 由于其算法简单、计算复杂度低, 处理速度快等优点被广泛地应用于数据分析和数字图像处理。K-means 算法通过将数据集按照某种测度准则把所有的数据集划分到若干个簇(或称聚类中心簇), 每一个簇中的数据称作一个对象, 聚类就是使每一个簇中对象之间有较高的相似度, 每一个簇间的对象有较高的相异度, 从而达到数据分类的目的。

文献[10]正是借鉴这种思想完成了软扩频信号

伪码序列的盲估计, 但是文中聚类时对聚类初始中心点随机地进行选取, 影响了伪码序列估计的准确性。针对此问题, 本文在文献[10]的基础上, 从以下两个方面对 K-means 算法进行改进, 一是基于相似测度的理论对聚类初始中心点进行选取; 二是使用平均轮廓系数的绝对值作为聚类结果的评价指标以估计伪码集合规模数(即最佳的聚类数)。

#### 3.1 K-means 聚类初始中心点的确定

传统的 K-means 算法由于初始聚类中心点选择的随机性, 导致聚类结果随着不同初始聚类中心点而改变, 影响算法的稳定性, 可能会造成局部最优值的问题<sup>[14,15]</sup>。为此, 结合模式识别中相似性测度(Similarity Measure, SM)的理论可得, 数据集中具有相似度大的对象会以较大的概率划分到同一个簇中, 而相似度小的对象会以较小的概率划分到同一个簇中。所以根据这一特性, 尽可能地选取相似度小(即距离较远)的数据对象作为聚类初始点, 避免了随机选取聚类初始中心点造成聚类结果不稳定的情况。

设待聚类的软扩频信号数据集为  $\mathbf{S}$ , 可表示为

$$\mathbf{S} = [\mathbf{s}_1 \ \cdots \ \mathbf{s}_i \ \cdots \ \mathbf{s}_M]^T, \quad i = 1, 2, \dots, M \quad (8)$$

式中,  $\mathbf{s}_i$  为数据集中的第  $i$  个分段数据。用于计算相似度的表达式为

$$\text{sm}(\mathbf{s}_i, \mathbf{s}_j) = \frac{\mathbf{s}_i^T \mathbf{s}_j}{\|\mathbf{s}_i\|^2 \|\mathbf{s}_j\|^2} \quad (9)$$

式中,  $\mathbf{s}_i$  和  $\mathbf{s}_j$  表示  $\mathbf{S}$  中不同的两个分段数据。

当已知聚类数  $k$  (即伪码集合规模数) 的搜索范围时, 确定聚类初始中心点的具体步骤如下:

步骤 1 不失一般性, 初始化聚类数为 2, 计算数据集  $\mathbf{S}$  中任意两个对象  $\mathbf{s}_i$  和  $\mathbf{s}_j$  之间的相似度  $\text{sm}(\mathbf{s}_i, \mathbf{s}_j)$ , 选取满足  $(\mathbf{s}_i, \mathbf{s}_j) = \arg \min_{i \neq j} [\text{sm}(\mathbf{s}_i, \mathbf{s}_j)]$ ,  $i, j = 1, 2, \dots, M$  中的两个对象  $\mathbf{s}_i$  和  $\mathbf{s}_j$ , 并将它们作为前两个初始聚类中心点  $\mathbf{c}_1$  和  $\mathbf{c}_2$ ;

步骤 2 当搜索的聚类数为 3 时, 在剩余的  $(M - 2)$  个对象中, 计算其与前两个聚类中心点各自相似度的乘积, 选取满足  $\mathbf{s}_i = \arg \min_{i \neq 1, i \neq 2} [\text{sm}(\mathbf{c}_1, \mathbf{s}_i) \times \text{sm}(\mathbf{c}_2, \mathbf{s}_i)]$  中的对象  $\mathbf{s}_i$  作为第 3 个初始聚类中心点  $\mathbf{c}_3$ ;

步骤 3 依次类推, 当搜索的聚类数为  $k$  ( $k < k_{\max}$ ) 时, 在剩余的  $(M - k + 1)$  个对象中, 计算其与前  $(k - 1)$  个聚类中心点各自相似度的乘积, 选取满足  $\mathbf{s}_i = \arg \min_{i \neq 1, \dots, (k-1)} [\text{sm}(\mathbf{c}_1, \mathbf{s}_i) \times \cdots \times \text{sm}(\mathbf{c}_{k-1}, \mathbf{s}_i)]$  中的对象  $\mathbf{s}_i$  作为第  $k$  个初始中心点  $\mathbf{c}_k$ 。

### 3.2 软扩频信号伪码集合规模数的估计

对于软扩频信号伪码序列的盲估计与传统直扩信号伪码序列的盲估计有所不同，在非合作情况下对于软扩频信号使用的伪码序列集合规模数(即  $k$  位信息码元所对应的  $2^k$  条伪码序列)也是不知道的，所以我们需要对伪码序列集合规模数进行估计，才能进一步完成对软扩频信号伪码序列的估计。

轮廓系数<sup>[11]</sup>，即 SC 指标，表示聚类后各类中样本之间的紧密程度和各类之间的离散程度。同一类中样本之间的距离越小，不同类间的距离越大，则  $SC(i, j)$  的值越大，此时的聚类效果越好。因此，常用 SC 作为评价聚类结果的性能指标。SC 指标<sup>[11]</sup>的表达式为

$$SC(i, j) = \frac{b(i, j) - d(i, j)}{\max[b(i, j), d(i, j)]} \quad (10)$$

式中， $SC(i, j) \in [0, 1]$ ， $b(i, j)$  表示第  $i$  类中的第  $j$  个样本与其它每一个类中样本平均距离的最小值， $d(i, j)$  表示第  $i$  类中的第  $j$  个样本与第  $i$  类中其它样本的平均距离，即

$$b(i, j) = \min_{1 \leq t \leq k, k \neq i} \left( \frac{1}{n_t} \sum_{p=1}^{n_t} \|c_p^{(t)} - c_j^{(i)}\|^2 \right) \quad (11)$$

$$d(i, j) = \frac{1}{n_i - 1} \sum_{q=1, q \neq j}^{n_i} \|c_q^{(i)} - c_j^{(i)}\|^2 \quad (12)$$

式中， $t$  和  $i$  分别表示不同的分类， $p$  和  $j$  分别表示第  $t$  类和第  $i$  类中的样本， $q$  和  $j$  表示第  $i$  类中的样本， $n_t$  和  $n_i$  分别表示第  $t$  类和第  $i$  类中样本的总数， $\|\cdot\|^2$  表示求两个样本之间欧氏距离的平方。

然而，式(10)中仅表示某一个聚类中单个样本的聚类情况，为了评价整个数据集所有样本的聚类情况，可以按照式(13)和式(14)求取数据矩阵  $\mathbf{S}$  中所有样本 SC 指标的平均值  $\overline{SC}(k)$ ，即可求出最佳的聚类数  $k_{\text{opt}}$ ，这里的  $k_{\text{opt}}$  即为信号伪码序列集合规模数的估计。

$$\overline{SC}(k) = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_j} SC(i, j) \quad (13)$$

$$k_{\text{opt}} = \arg \max_{2 \leq k \leq k_{\text{max}}} [\overline{SC}(k)] \quad (14)$$

式中， $\{k | 2 \leq k \leq k_{\text{max}}\}$  表示每一次搜索的聚类数， $n_j$  表示第  $i$  类中的样本总数， $n$  表示所有的样本总数。由式(14)可得，当使得  $\overline{SC}(k)$  最大的  $k$  即为最佳聚类数  $k_{\text{opt}}$ ，此时聚类的效果也越好。

### 3.3 软扩频信号伪码序列的估计

在 K-means 聚类算法中，称式(7)中的  $\mathbf{s}_m$  为一个聚类对象， $M$  个这样的聚类对象构成了接收信号数据集  $\mathbf{S} = [\mathbf{s}_1 \ \mathbf{s}_2 \ \cdots \ \mathbf{s}_M]^T$ ， $M$  为接收信号的数据

分组数。根据 3.1 节的分析可知，传统的 K-means 算法在确定最佳聚类数  $k_{\text{opt}}$  时，对于搜索的每一个  $k$  值，都是随机对聚类中心点进行选取，由于随机的选取初始聚类中心点，使得用于评价不同聚类数对应聚类结果的性能指标比较差，出现聚类结果不稳定的情况。所以，对传统的 K-means 算法进行改进，采用基于相似测度的理论确定初始中心点，随着  $k$  值的不断增加，已确定的聚类中心点不会改变，使得不同聚类数下的初始中心点之间有着继承关系，并根据评价聚类结果的  $\overline{SC}(k)$  指标确定伪码集合规模数，并进一步通过伪码集合规模数对数据进行聚类完成伪码序列的估计。具体的算法步骤如下：

步骤 1 输入接收信号的观测数据矩阵  $\mathbf{S}$ ，初始化迭代次数  $\alpha = 0$ ，聚类数  $k = 2$ ，即聚类算法中聚类中心点的数目；

步骤 2 按照 3.1 节中基于相似测度的方法对初始聚类中心点进行选取，得到聚类中心点的集合

$$\mathbf{C}_k = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k] \quad (15)$$

步骤 3 令  $\alpha = \alpha + 1$ ，将  $\mathbf{S}$  中的每一个对象按照最小类内距离划分到相应的簇中，这里使用的距离准则是 Tanimoto 距离<sup>[16]</sup>，即

$$s_T(\mathbf{s}_i, \mathbf{c}_k) = \frac{\mathbf{s}_i^T \mathbf{c}_k}{\|\mathbf{s}_i\|^2 + \|\mathbf{c}_k\|^2 - \mathbf{s}_i^T \mathbf{c}_k} \quad (16)$$

式中， $\mathbf{s}_i$  表示  $\mathbf{S}$  中的第  $i$  个数据对象， $\mathbf{c}_k$  表示第  $k$  个聚类中心点；

步骤 4 根据式(16)计算  $\mathbf{S}$  中所有数据对象到所有聚类中心簇的 Tanimoto 距离，并将所有数据对象划分到对应距离最小的中心簇中；然后计算每一个簇中所有对象的平均值  $\overline{\mathbf{c}}_k^{(\alpha)}$ ，更新所有的聚类中心点  $\overline{\mathbf{C}}_k^{(\alpha)} = [\overline{\mathbf{c}}_1^{(\alpha)}, \overline{\mathbf{c}}_2^{(\alpha)}, \dots, \overline{\mathbf{c}}_k^{(\alpha)}]$ ；

步骤 5 重复步骤 4，判断是否满足停止迭代的条件  $\|\overline{\mathbf{C}}_k^{(\alpha+1)} - \overline{\mathbf{C}}_k^{(\alpha)}\| \leq 0$ ，若满足，则进入步骤 6；若不满足，则返回步骤 3；

步骤 6 按照 3.2 节中的式(10)-式(14)计算当前聚类数  $k$  下的聚类性能指标 SC 的平均值  $\overline{SC}(k)$ ；

步骤 7 令聚类数  $k = k + 1$ ，重复步骤 2-步骤 7；并比较当前聚类数  $k$  与  $k_{\text{max}}$  的大小，若  $k < k_{\text{max}}$ ，则返回步骤 2；否则，结束整个算法。

最后，通过确定  $\overline{SC}(k)$  最大值的位置即可完成最佳聚类数的估计，即信号伪码序列集合规模数的估计，而且对最佳聚类数所对应的聚类中心点集合取符号就可完成信号伪码序列的估计，即

$$\left. \begin{aligned} \hat{P} &= k_{\text{opt}} \\ \hat{\mathbf{C}} &= \text{sign}(\bar{\mathbf{C}}_{k_{\text{opt}}}) = \text{sign}[(\bar{c}_1, \bar{c}_2, \dots, \bar{c}_{k_{\text{opt}}})] \end{aligned} \right\} \quad (17)$$

式(17)中,  $k_{\text{opt}}$  表示伪码序列集合规模数的估计,  $\hat{\mathbf{C}}$  表示原始伪码序列集合  $\mathbf{C}$  的估计,  $\text{sign}[\cdot]$  表示符号函数。本文算法的基本流程如图 2 所示。

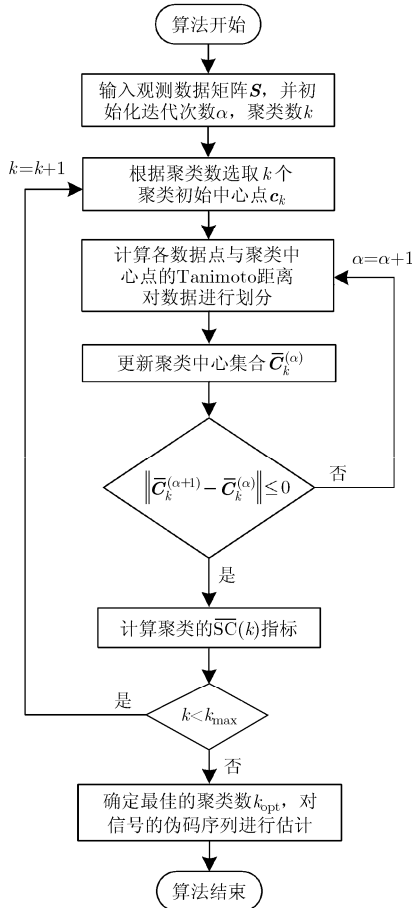


图 2 本文算法的基本流程

### 3.4 算法复杂度分析

针对某一聚类数  $k$  下的聚类情况, 传统的 K-means 算法计算复杂度约为  $O(ndk\alpha)$ , 这里的  $nd$  表示数据集的维数为  $n \times d$ ,  $k$  表示聚类数,  $\alpha$  表示算法收敛时的迭代次数, 而基于相似测度理论确定初始中心点的方法, 每寻找一次新的聚类初始中心需要进行约  $nd$  次相似度计算, 算法收敛时的迭代次数为  $\alpha_1$ , 则寻找  $k$  个聚类中心需要的计算复杂度为  $O(ndk\alpha_1)$ , 而本文算法由于对初始中心点进行了选取, 避免传统 K-means 算法随机选取初始中心点过于邻近或来自同一簇中的情况, 使得算法的迭代次数减少, 从而降低了计算复杂度, 而且确定了聚类初始中心点也使得聚类结果更加稳定。

## 4 仿真实验与结果分析

本文所有仿真实验主要是针对 BPSK 调制下的软扩频信号, 伪码序列集合为 gold 序列组成, 其他的实验参数设置和各项性能指标定义如下

(1) 伪码序列估计的错误概率

$$P_r = \frac{1}{L} \sum_{l=1}^L \frac{n_l}{2^k N} \quad (18)$$

式中,  $n_l$  表示第  $l$  次蒙特卡洛仿真实验中伪码序列估计的错误码元总数;  $N$  表示信号中伪码序列的长度;  $2^k$  表示  $k$  比特信息码元对应的伪码序列总数。

(2) 信噪比

$$\text{SNR} = 10 \lg \left( \frac{\sum_{n=0}^{N-1} s^2(n)}{\sum_{n=0}^{N-1} w^2(n)} \right) \quad (19)$$

式中,  $\sum_{n=0}^{N-1} s^2(n)$  和  $\sum_{n=0}^{N-1} w^2(n)$  分别表示信号的能量和噪声的能量。

(3) 实验 3 至实验 8 中信噪比的取值范围为  $\text{SNR} = -25 \sim 0$  dB, 且每一个信噪比下进行蒙特卡洛仿真实验的次数均为 300。

**实验 1** 下面将通过仿真实验来验证本文所提算法对软扩频信号伪码序列集合规模数的估计。仿真信号所用的伪码序列长度  $N=127$  chip, 接收信号分组数为  $M=1000$  组, 信噪比  $\text{SNR} = -5$  dB, 信息码元分别按  $k=2, 3, 4, 5$  分为一组, 则对应的伪码集合规模数 ( $P = 2^k$ ) 分别为 4, 8, 16, 32。设置伪码序列数据集合规模的变化范围为 2~40, 并计算每一个数据集合规模的归一化平均轮廓系数的绝对值, 则它随伪码序列集合规模数变化的结果如图 3 所示。

从图 3(a)-图 3(d) 可得, 归一化的平均轮廓系数的绝对值随着伪码序列集合规模数的增加呈现先递增后减小的变化趋势, 且在伪码序列集合规模数的正确估计值(最佳聚类数)处归一化平均轮廓系数的绝对值取得最大。因为此时聚类结果中, 类间距离远大于类内样本距离的平均值, 从而使得平均轮廓的绝对值最大。这与 3.2 节的理论分析相一致, 所以通过  $\overline{\text{SC}}_{\text{max}}(k)$  即可完成信号伪码序列集合规模数的估计。

**实验 2** 为了验证本文所提算法对软扩频信号伪码序列估计的有效性。取伪码序列长度  $N = 127$  chip, 信息码元按  $k = 2$  分为一组, 则伪码序列集中对应着 4 条不同的伪码序列, 其它参数与实验 1 相同。软扩频信号伪码序列的真实值和估计值如图 4(a)、图 4(b) 所示。

从图 4(a)、图 4(b) 可得, 软扩频信号伪码序列的估计值与真实值的符号完全一致。因此, 本文所提出的算法能够实现软扩频信号的伪码序列盲估计。

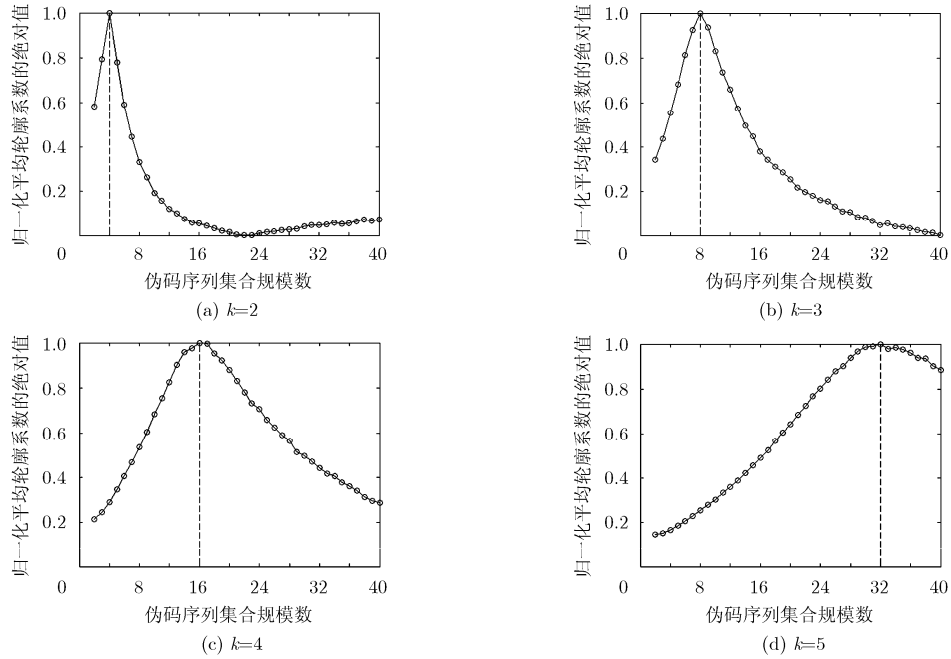


图 3 伪码序列集合规模数估计

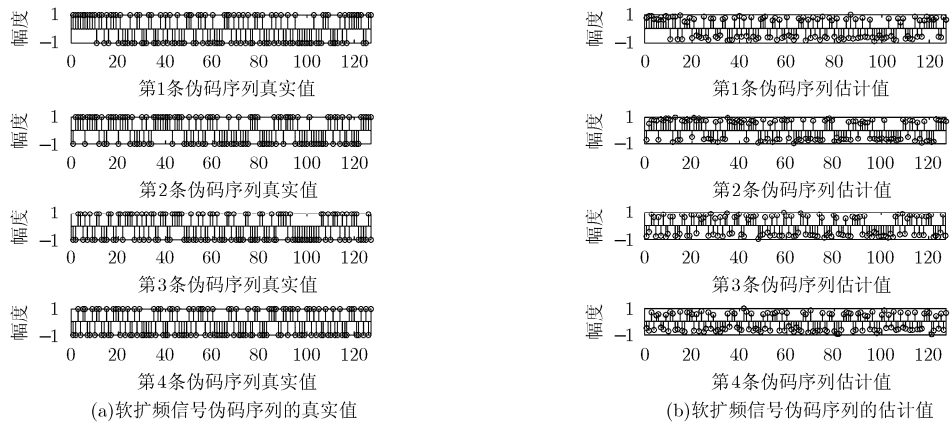


图 4 软扩频信号伪码序列的真实值与估计值对比

**实验 3** 为了比较本文所提算法在不同信息码元分组下对软扩频信号伪码序列估计性能的影响。取伪码序列长度  $N = 127$  chip，接收信号分组数为  $M = 1000$  组，信息码元分别按  $k = 2, 3, 4, 5$  分为一组进行仿真。软扩频信号伪码序列估计的错误概率如图 5 所示。

从图 5 可得，软扩频信号伪码序列估计的错误概率在不同  $k$  (伪码序列集合规模)下随着信噪比的增加而不断减小；在同一信噪比下，随着  $k$  值的不断减小，软扩频信号伪码序列估计的错误概率也随之减少；且对于  $k = 2, 3, 4, 5$  达到 0.1 以下的错误概率所需要的信噪比为  $-19$  dB,  $-16$  dB,  $-14$  dB,  $-12$  dB。这是因为随着伪码序列集合规模的增大，对伪码序列进行估计时相互之间的干扰也随之增

大、特征参量也难以检测，因此信号伪码序列估计的错误概率不断增加。

**实验 4** 为了比较本文所提算法在不同伪码序列长度下对软扩频信号伪码序列估计性能的影响。取伪码序列长度  $N = 127$  chip,  $255$  chip,  $511$  chip,  $1023$  chip，接收信号分组数为  $M=1000$  组，信息码元分别按  $k = 3$  分为一组进行仿真。软扩频信号伪码序列估计的错误概率如图 6 所示。

从图 6 可得，软扩频信号伪码序列估计的错误概率在不同伪码序列长度下，随着信噪比的增加而不断减小；在同一信噪比下，随着伪码长度的不断减小，软扩频信号伪码序列估计的错误概率也随之增加；且对于  $N = 511$  chip 达到 0.1 以下的误码率所需要的信噪比为  $-20$  dB。这是因为随着伪码序列

长度的增加,信号的扩频处理增益也随之增大、抗噪声能力进一步增强,因此信号伪码序列估计的错误概率不断减小。

**实验 5** 为了比较本文所提算法在不同接收信号分组数下对软扩频信号伪码序列估计性能的影响。取伪码序列长度  $N = 127$  chip,接收信号分组数分别为  $M = 300$  组,500 组,1000 组,2000 组,信息码元分别按  $k = 3$  分为一组进行仿真。软扩频信号伪码序列估计的错误概率如图 7 所示。

从图 7 可得,软扩频信号伪码序列估计的错误概率在不同  $M$ (接收信号的分组数)下随着信噪比的增加而不断减小;在同一信噪比下,随着  $M$  的不断增大,软扩频信号伪码序列估计的错误概率也随之减少;且对于  $M = 300, 500, 1000, 2000$  达到 0.1 以下的错误概率所需要的信噪比为  $-13$  dB,  $-15$  dB,  $-16$  dB,  $-18$  dB。这是因为随着接收信号长度的不断增大,用于伪码序列估计时的有用信息就越多,信号伪码序列估计的错误概率不断减少。因此,也可以得出本文算法信噪比容限的改变可以通过增加接收信号的长度来实现。

**实验 6** 为了验证本文所提算法对软扩频信号在不同码型下伪码序列估计的通用性。对 m 序列、gold 序列、哈达码矩阵的正交序列进行仿真实验,信号所用伪码序列的长度为  $N = 127$  chip,接收信号分组数为  $M = 1000$  组,信息码元分别按  $k = 3$  分为一组进行仿真实验。软扩频信号伪码序列估计的错误概率如图 8 所示。

从图 8 可得,软扩频信号伪码序列估计的错误概率在不同的伪码码型下随着信噪比的增加而不断减小,在同一信噪比下,且 3 种码型的估计性能几乎相一致。因此,可以得出本文算法对软扩频信号伪码序列进行估计时,不受伪码序列码型的限制且正交序列和准正交序列的估计性能相一致。

**实验 7** 为了验证本文改进的方法与文献[10]中的方法对软扩频信号伪码序列估计的性能。取信号伪码序列的长度分别为  $N = 127$  chip,信息码元分

别按  $k = 2, k = 4$  分为一组进行蒙特卡罗仿真实验。考察两种方法对软扩频信号伪码序列估计的错误概率结果如图 9 所示。

从图 9(a)、图 9(b)可得,本文提出的算法对信号伪码序列估计的性能优于文献[10]中的方法,当  $k = 2$ 、信号分组数为  $M = 500$  和  $M = 1000$ 、伪码序列估计的错误概率在 0.1 以下时,本文算法与文献[10]中的方法相比在性能上提高约 4 dB 和 5 dB;当  $k = 4$ 、信号分组数为  $M = 500$  和  $M = 1000$ 、伪码序列估计的错误概率在 0.1 以下时,本文算法与文献[10]中的算法相比在性能上提高约 9 dB 和 6 dB。这是由于本文在聚类开始时,并没有随机地进行聚类初始中心点的选取,而是在聚类时通过对聚类初始中心点按照 3.1 节的方法进行选取,提高了伪码序列估计的稳定性和准确性,避免了类似文献[10]中由于随机选取聚类初始中心点造成聚类结果不稳定。

**实验 8** 为了进一步比较本文改进的算法与文献[10]中的算法对软扩频信号盲解扩的性能。取信号伪码序列的长度分别为  $N = 127$  chip,信息码元分别按  $k = 2, k = 4$  分为一组进行蒙特卡罗仿真实验。考察两种算法对软扩频信号盲解扩的性能结果如图 10 所示。

从图 10(a)、图 10(b)可得,两种算法对软扩频信号盲解扩的性能(信息码元估计的错误概率)随着信噪比的增加而减小,本文提出的算法对软扩频信号盲解扩的性能不仅优于文献[10]中的方法,而且盲解扩性能与合作解扩性能相近,但是随着  $k$  的不断增大,对信号信息序列估计的错误概率也不断地增加。

## 5 结束语

对于软扩频信号的盲解扩问题,本文在传统 K-means 聚类的基础上,通过相似测度的理论对聚类的初始中心点进行选取,并使用平均 SC 指标对伪码序列集合规模数进行了估计,克服了传统

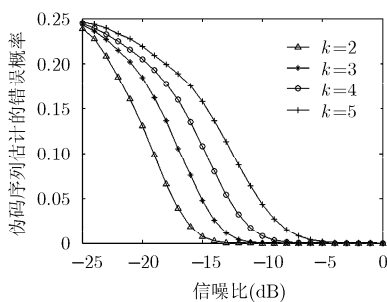


图 5 不同  $k$  值下软扩频信号伪码序列估计的错误概率

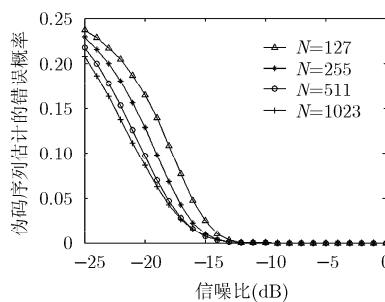


图 6 不同伪码序列长度下伪码序列估计的错误概率

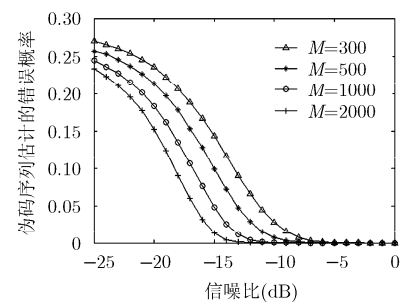


图 7 不同数据长度下伪码序列估计的错误概率

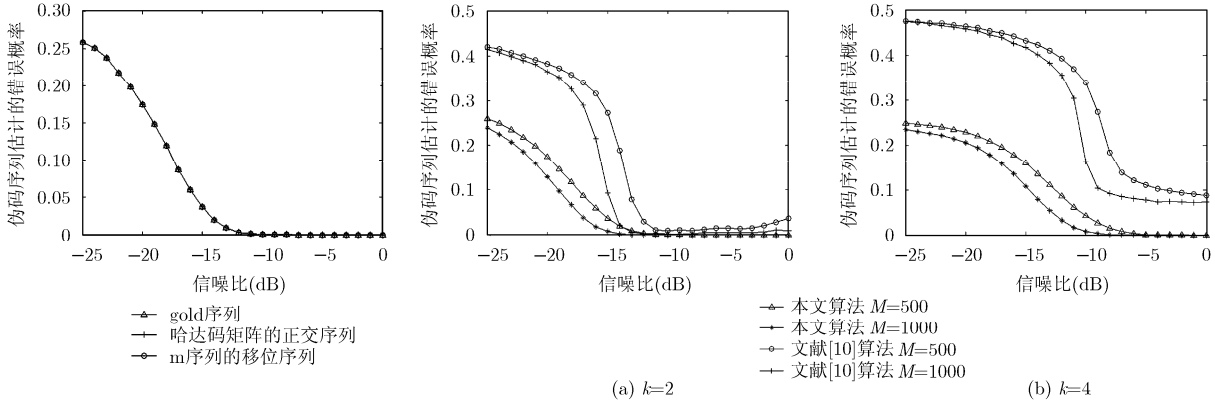


图 8 不同码型下伪码序列估计的性能对比

图 9 不同 k 和 M 下的伪码序列估计方法对比

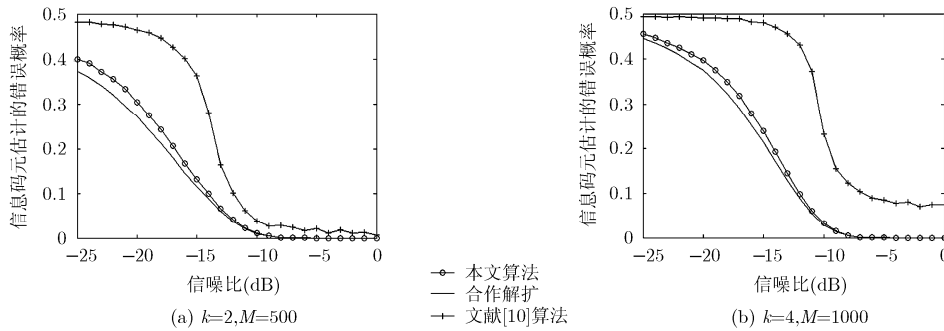


图 10 不同算法对软扩频信号盲解扩性能对比

K-means 算法因初始中心点的不确定性，造成聚类结果不稳定的问题。仿真实验结果论证了改进后的 K-means 聚类算法在不同的信息码元分组、不同的伪码长度、不同的伪码码型和不同的接收信号长度下对软扩频信号伪码序列进行估计时具有估计效率高、估计误差小和抗干扰能力强等特点，与未改进的 K-means 算法在伪码序列的估计上及信号盲解扩的性能上相比，明显优于后者。而且本文改进的软扩频信号盲解扩方法，可以在信噪比为 -15 dB 的条件下，使伪码序列估计的错误概率达到 0.1 以下，而且该方法可以对任何码型的伪码序列进行估计。由于本文只考虑了理想的高斯白噪声信道，而在实际通信中往往存在窄带干扰、多径衰落、多用户等复杂情况，针对以上干扰条件下的软扩频信号伪码序列盲估计将是我们下一步研究的重点。

参 考 文 献

[1] 田日才, 迟永钢. 扩频通信[M]. 第 2 版, 北京: 清华大学出版社, 2014: 1-5.  
TIAN Ricai and CHI Yonggang. Spread Spectrum Communication[M]. 2nd Ed. Beijing: Tsinghua University Press, 2014: 1-5.

[2] PURSLEY M B and ROYSTER T C. High-rate direct-sequence spread spectrum with error control coding[J]. *IEEE*

*Transactions on Communications*, 2006, 54(9): 1693-1702. doi: 10.1109/TCOMM.2006.881256.

[3] 周佳晶, 唐友喜. JTIDS 扩频序列的估计[J]. *电子科技大学学报*, 2007, 36(5): 1054-1056.  
ZHOU Jiajing and TANG Youxi. Spread spectrum sequence estimation for JTIDS[J]. *Journal of University of Electronic Science and Technology of China*, 2007, 36(5): 1054-1056.

[4] 沈斌, 王建新. 窄带干扰条件下含有未知载频的直扩信号的伪码序列估计[J]. *电子与信息学报*, 2015, 37(7): 1556-1561. doi: 10.11999/JEIT141322.  
SHEN Bin and WANG Jianxin. Estimation of PN sequence in DSSS signals with unknown carrier frequency under narrow band interferences[J]. *Journal of Electronics & Information Technology*, 2015, 37(7): 1556-1561. doi: 10.11999/JEIT141322.

[5] 任啸天, 徐晖, 黄知涛, 等. 短码 DS-SS 信号扩频序列及信息序列联合盲估计方法[J]. *通信学报*, 2012, 33(4): 169-175. doi: 10.3969/j.issn.1000-436X.2012.04.023.  
REN Xiaotian, XU Hui, HUANG Zhitao, et al. Joint blinding estimation of the spread-spectrum sequence and information sequence for short-code DS-SS signal[J]. *Journal on Communications*, 2012, 33(4): 169-175. doi: 10.3969/j.issn.1000-436X.2012.04.023.

[6] 张天骢, 赵军桃, 江晓磊. 基于多主分量神经网络的同步 DS-SS 伪码盲估计[J]. *系统工程与电子技术*, 2016, 38(11): 2638-2647. doi: 10.3969/j.issn.1001-506X.2016.11.27.



- ZHANG Tianqi, ZHAO Juntao, and JIANG Xiaolei. PN code sequence blind estimate of synchronous DS-CDMA based on multi-principal component neural network[J]. *Systems Engineering and Electronics*, 2016, 38(11): 2638–2647. doi: 10.3969/j.issn.1001-506X.2016.11.27.
- [7] 张天骐, 强幸子, 马宝泽, 等. 基于最小二乘的同步多用户非周期长码直扩信号扩频序列估计[J]. *电波科学学报*, 2016, 31(6): 1113–1123. doi: 10.13443/j.cjors.2016030201.
- ZHANG Tianqi, QIANG Xingzi, MA Baoze, *et al.* Estimation of the spread spectrum sequence for synchronous multi-user a periodic long-code DSSS signals based on least squares[J]. *Chinese Journal of Radio Science*, 2016, 31(6): 1113–1123. doi: 10.13443/j.cjors.2016030201.
- [8] GU Xiaolei, ZHAO Zhijin, and SHEN Lei. Blind estimation of pseudo-random codes in periodic long code direct sequence spread spectrum signals[J]. *IET Communications*, 2016, 10(11): 1273–1281. doi: 10.1049/iet-com.2015.0374.
- [9] 赵知劲, 李淼, 尚俊娜. 基于矩阵填充和三阶相关的长短码 DS-CDMA 信号多伪码盲估计[J]. *电子与信息学报*, 2016, 38(7): 1788–1793. doi: 10.11999/JEIT151087.
- ZHAO Zhijin, LI Miao, and SHANG Junna. Blind estimation of LSC-DS-CDMA signal based on matrix completion and triple correlation[J]. *Journal of Electronics & Information Technology*, 2016, 38(7): 1788–1793. doi: 10.11999/JEIT151087.
- [10] 王航, 郭静波, 王赞基. 基于聚类的软扩频信号盲解扩方法[J]. *电子与信息学报*, 2009, 31(2): 422–425.
- WANG Hang, GUO Jingbo, and WANG Zanji. Clustering based blind despread method of tamed direct sequence spread spectrum signals[J]. *Journal of Electronics & Information Technology*, 2009, 31(2): 422–425.
- [11] KISORE N R and KOTESWARAIAH C B. Improving ATM coverage area using density based clustering algorithm and voronoi diagrams[J]. *Information Sciences*, 2016, 376: 1–20. doi: 10.1016/j.ins.2016.09.058.
- [12] ZHANG Tianqi, QIAN Wenrui, ZHANG Gang, *et al.* Parameter estimation of MC-CDMA signals based on modified cyclic autocorrelation[J]. *Digital Signal Processing*, 2016, 54: 46–53. doi: 10.1016/j.dsp.2016.03.007
- [13] 李晓瑜, 俞丽颖, 雷航, 等. 一种 K-means 改进算法的并行化实现与应用[J]. *电子科技大学学报*, 2017, 46(1): 61–68. doi: 10.3969/j.issn.1001-0548.2017.01.010.
- LI Xiaoyu, YU Liying, LEI Hang, *et al.* The parallel implementation and application of an improved K-means algorithm[J]. *Journal of University of Electronic Science and Technology of China*, 2017, 46(1): 61–68. doi: 10.3969/j.issn.1001-0548.2017.01.010.
- [14] 翟东海, 鱼江, 高飞, 等. 最大距离法选取初始聚类中心的 K-means 文本聚类算法的研究[J]. *计算机应用研究*, 2014, 31(3): 713–715. doi: 10.3969/j.issn.1001-3695.2014.03.017.
- ZHAI Donghai, YU Jiang, GAO Fei, *et al.* K-means text clustering algorithm based on centers selection according to maximum distance[J]. *Application Research of Computers*, 2014, 31(3): 713–715. doi: 10.3969/j.issn.1001-3695.2014.03.017.
- [15] 张健沛, 杨悦, 杨静, 等. 基于最优划分的 K-Means 初始聚类中心选取算法[J]. *系统仿真学报*, 2009, 21(9): 2586–2589.
- ZHANG Jianpei, YANG Yue, YANG Jing, *et al.* Algorithm for initialization of K-means clustering center based on optimized-division[J]. *Journal of System Simulation*, 2009, 21(9): 2586–2589.
- [16] THEODORIDIS S and KOUTROUMBAS K. *Pattern Recognition*[M]. Fourth Ed. USA: Academic Press, 2010: 415–417.
- 张天骐: 男, 1971 年生, 博士后, 教授, 研究方向为通信信号的调制解调、盲处理、语音信号处理、神经网络实现以及 FPGA、VLSI 实现.
- 杨强: 男, 1991 年生, 硕士生, 研究方向为扩频信号盲处理.
- 宋玉龙: 男, 1992 年生, 硕士生, 研究方向为导航信号捕获与跟踪.
- 熊梅: 女, 1991 年生, 硕士生, 研究方向为语音信号处理.