doi:10.11959/j.issn.1000-436x.2017043

基于软信息的扰码盲识别方法

陈泽亮，彭华，巩克现，于沛东
（解放军信息工程大学信息系统工程学院，河南 郑州 450001）

摘 要：针对非合作接收的加扰信号，提出 2 种基于软信息的扰码盲识别方法。方法 1 利用软信息建立了扰码系数的代价函数，采用实数域的优化理论进行正向求解，不再需要对多项式测试闭集进行遍历；方法 2 利用软信息建立了符合度的概念，以每个测试扰码多项式符合度的大小作为判别的标准，相比硬判决识别算法，其对接收信息得到了更充分的利用。仿真结果表明，方法 1 相比 Cluzeau 提出的遍历方法，其自同步扰码多项式的识别时间可从 5 min 18 s 缩短为 8 s；方法 2 与现有硬判决算法相比，达到较高正确率时，具有约 2 dB 的信噪比增益。
关键词：软信息；扰码盲识别；代价函数；符合度；抗噪性能
中图分类号：TN911.7 **文献标识码**：A

Scrambler blind recognition method based on soft information

CHEN Ze-liang, PENG Hua, GONG Ke-xian, YU Pei-dong
(School of Information Systems Engineering, PLA Information Engineering University, Zhengzhou 450001, China)

Abstract: Two scrambler blind recognition methods based on soft information were proposed for received signal in non-cooperative ways. The first method established a cost function of the scrambler coefficients by using the soft information, and adopted the optimization theory of real number field for positive solution. So it didn't need to traverse the closed set of test polynomial any more. The second method built conformity degree concept with the soft information, and used the size of conformity of each test scrambler polynomial as the discriminant criteria. So it made more full use of the received information compared to the hard sentence recognition algorithm. Simulation results show that the first method can shorten the recognition time of a synchronous scrambler polynomial from 5 min 18 s to 8 s compared with the traversal method put forward by Cluzeau, and the second method has 2 dB *SNR* gain when to achieve the relatively high accuracy compared with the hard sentence recognition algorithm.
Key words: soft information, scrambler blind recognition, cost function, conformity degree, anti-noise performance

1 引言

在实际的数字通信系统中，为了提高数据传输的随机性，通常会对信息进行加扰处理，使加扰后的数据尽量达到(0,1)平衡，这样更加有利于接收端进行定时恢复。同时，伪随机扰码在密码系统中也有着广泛的应用，对数据进行加扰处理可增强信息传输的安全性。然而，在信号截获领域，侦察方对于扰码的相关参数是未知的，这就需要对其进行盲识别。

扰码可分为自同步扰码和同步扰码 2 种。对于自同步扰码的分析主要是对其扰码多项式进行识别，对于同步扰码还包括其初态的识别。实际通信中，信源信息的(0,1)并不平衡，一般信息中 0 的概率要大于$\frac{1}{2}$，即有$\Pr(x_k=0)=\frac{1}{2}+\varepsilon$，$\varepsilon\neq 0$，其中，$\varepsilon$的典型值一般为 0.1 和 0.05。近年来，很多识别算法利用信息序列的这种有偏性，对扰码参数的盲识别取得了较好的识别效果。文献[1]利用解扰序列构造一个随机变量，并说明了在测试多项式和实际的扰码多项式（或其倍式）相同和不相同 2 种情况下，随机变量服从不同的高斯分布，然后利用该判别准则，从低阶到高阶遍历寻找加扰多项式的 2 个

收稿日期：2016-04-26；**修回日期**：2016-09-01
基金项目：国家自然科学基金资助项目（No.61401511）
Foundation Item: The National Natural Science Foundation of China (No.61401511)

三项倍式，最后通过求该 2 个三项倍式的最大公约数完成对扰码参数的识别。文献[2]仍采用文献[1]的识别思想，但对文献[1]的方法进行了改进，更加准确地分析了扰码识别所需数据量，优化了利用扰码多项式的 2 个倍式求解扰码多项式的过程，并考虑了存在噪声情况下如何对扰码参数进行识别。文献[3]在高误码率下利用组合枚举求优势的方法对伪随机扰码（同步扰码）的生成多项式进行识别，并通过基于卷积码快速相关攻击的方法实现对其初态的识别过程。文献[4]针对自同步扰码的识别问题，巧妙地建立了抽头位置的比特状态概率分布与均匀分布之间的修正平方欧几里德距离的概念，很好地实现了自同步扰码的识别。但是上面的这些算法主要是基于硬判决序列对扰码参数进行识别，在接收系统中，所获得的软判决信息中不仅含有发送比特的符号信息，还包含了该符号的可靠度信息，使用硬判决序列进行编码参数的分析，会损失可靠度信息，从而一定程度上限制参数分析性能提升的空间。文献[5~7]提出的基于软信息的分析方法在低信噪比下取得了很好的识别效果，且具有相对低的复杂度。

本文提出了 2 种利用软信息对扰码参数进行识别的方法。方法 1 从信息序列的有偏性以及扰码多项式内部的约束关系出发，利用接收软信息，建立扰码生成多项式系数的代价函数，在此基础上，利用实数域中的优化理论和方法，求得扰码系数的最优解。方法 2 建立了符合度的概念，利用软信息对所有可能的扰码生成多项式进行遍历，取符合度最大的测试多项式为扰码多项式。这 2 种算法都利用软信息对扰码多项式进行识别，提高了算法的抗噪性能，取得了更优的识别效果。

2　扰码介绍及问题模型

自同步扰码[8]的加扰和解扰过程以线性反馈移位寄存器(LFSR)为基础，在 LFSR 的输出端和信息的输入端之间引入异或逻辑，并将得到的结果作为 LFSR 的输入。如图 1 所示，信息序列 $\{x_k\}$ 与 LFSR 的输出序列 $\{s_k\}$ 做异或处理，最终得到扰码序列 $\{y_k\}$，同时 $\{y_k\}$ 也作为 LFSR 的输入，加扰过程的数学表达式为

$$y_k = x_k \oplus \sum_{i=1}^{L} \oplus c_i y_{k-i} \tag{1}$$

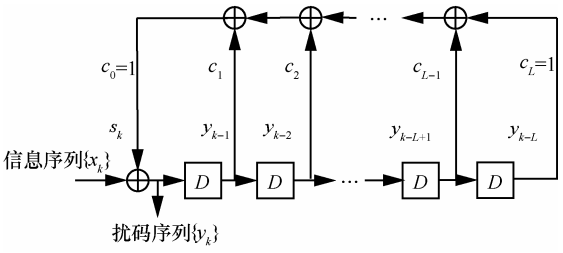


图 1　自同步扰码器

式(1)中的运算都在二元域中进行，其中，$\sum\oplus$ 表示模二求和，$c_i \in \{0,1\}, i = 0,1,\cdots,L$ 为扰码多项式的抽头系数。则扰码生成多项式可表示为

$$P(x) = c_0 + c_1 x + \cdots + c_{L-1} x^{L-1} + c_L x^L \tag{2}$$

对式(1)稍作变换，可得到自同步扰码的解扰过程

$$x_k = \sum_{i=0}^{L} \oplus c_i y_{k-i} \tag{3}$$

其原理如图 2 所示。

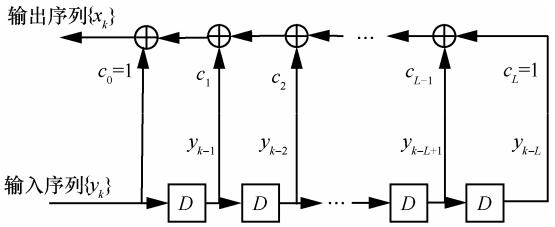


图 2　自同步解扰器

同步扰码[8]通过把独立的 m 序列加到信息序列上，从而产生扰码序列。扰码序列减去一个相同的 m 序列，被扰乱的数据得到恢复。同步扰码的加扰器和解扰器具有相同的结构，只需将信息序列 $\{x_k\}$ 和扰码序列 $\{y_k\}$ 调换位置即可，其加扰的数学表达式如式(4)所示，解扰的数学表达式如式(5)所示，原理如图 3 所示。

$$y_k = x_k \oplus \sum_{i=1}^{L} \oplus c_i s_{k-i} \tag{4}$$

$$x_k = y_k \oplus \sum_{i=1}^{L} \oplus c_i s_{k-i} \tag{5}$$

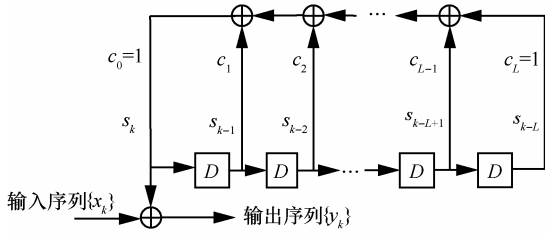


图 3　同步扰码器

实际上，加扰序列需要经过信道传输，在传输中扰码序列会叠加噪声。自同步扰码和同步扰码的问题模型分别如图 4 和图 5 所示。

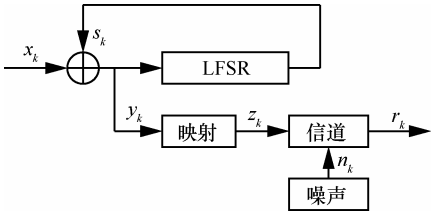


图 4 自同步扰码的问题模型

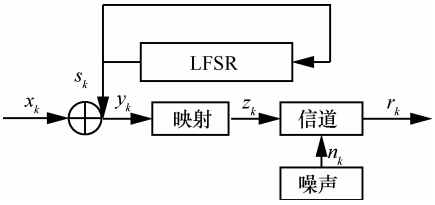


图 5 同步扰码的问题模型

上述模型是实际信号传输的简化模型，模型中省去了编码过程和调制过程的成型等。对于接收方来说，已知的是序列 $\{r_k\}$。根据具体通信协议，加扰和编码的前后顺序不定，所获得的软信息可能是解调输出的软判决信息或软译码后输出的软信息（一些软输入、软输出译码），这 2 种情况下的软信息主要是误码率不同，在此同等对待。最终需要解决的问题就是利用已知的数据序列 $\{r_k\}$ 对扰码多项式进行盲识别。

3 基于扰码系数代价函数的参数估计

3.1 自同步扰码的估计原理

对于自同步扰码器，由式(3)和信息序列的(0,1)不平衡性 $\Pr(x_k=0)=\frac{1}{2}+\varepsilon,(\varepsilon\neq 0)$，可知

$$\Pr(\sum_{i=0}^{L}\oplus c_i y_{k-i}=1)=\Pr(x_k=1)=\frac{1}{2}-\varepsilon, k\geqslant L \quad (6)$$

对式(6)稍作变换，可得

$$1-2\Pr(\sum_{i=0}^{L}\oplus c_i y_{k-i}=1)=2\varepsilon,\quad k\geqslant L \quad (7)$$

在继续推导算法之前，先给出一个概率运算的相关结论[9]。设 a、b 为二元域上 2 个相互独立的随机变量，则容易证明

$$1-2\Pr(a\oplus b=1)=(1-2\Pr(a=1))(1-2\Pr(b=1)) \quad (8)$$

显然，这一关系可以推广到多个随机变量的情况，即

$$1-2\Pr(\sum_{k=1}^{m}\oplus a_k=1)=\prod_{k=1}^{m}(1-2\Pr(a_k=1)) \quad (9)$$

其中，二元域随机变量 $a_1,\cdots,a_m$ 相互独立。

将式(7)中的 y_k 看成已知的常量，将系数 $c_0,c_1,\cdots,c_L$ 看成未知的随机变量，又由于各系数之间可视为相互独立，故式(7)可以结合式(9)改写为

$$\prod_{i=0}^{L}(1-2\Pr(c_i=1)\Pr(y_{k-i}=1))=2\varepsilon, k\geqslant L \quad (10)$$

对于每一个时刻 k 都有上面的关系式成立，则为了提高其识别的准确性，对于 M 个时刻，有

$$\sum_{k=L+jM}^{L+(j+1)M-1}\prod_{i=0}^{L}(1-2\Pr(c_i=1)\Pr(y_{k-i}=1))=M(2\varepsilon), j\geqslant 0 \quad (11)$$

接下来，引入软信息的概念，如图 4 所示，假设映射为简单的 BPSK 映射，则与扰码序列 y_k 对应的映射序列为 z_k $(0\leftrightarrow -1,1\leftrightarrow +1)$，对应的接收序列为 $r_k=z_k+n_k$，n_k 为高斯白噪声，则 y_k 的后验概率 $p_k=\Pr(y_k=1\mid r_k)$ 为

$$p_k=\frac{\mathrm{e}^{\frac{2r_k}{\sigma^2}}}{\mathrm{e}^{\frac{2r_k}{\sigma^2}}+1} \quad (12)$$

其中，σ^2 为 AWGN 信道的噪声功率。同时，为了便于算法推导，令

$$q_i=\Pr(c_i=1) \quad (13)$$

在实际应用中，y_k 的值是不得而知的，接收端得到的只有序列 r_k。于是，在式(11)中只能用 y_k 的后验概率 p_k 取代 $\Pr(y_k=1)$。但是取代后很难再存在一组系数 c_i 严格满足关系式(11)，为了尽量满足式(11)，对任意的 j，式(11)应尽可能成立，也就是使如下代价函数取得最小值。

$$D(n)=\sum_{j=0}^{n-1}\left|\sum_{k=L+jM}^{L+(j+1)M-1}\prod_{i=0}^{L}(1-2p_{k-i}q_i)-M(2\varepsilon)\right|^2 \quad (14)$$

为了降低计算复杂度，利用文献[10]中单变量搜索的方法将式(14)中的多维优化问题转变为一维优化问题，其求解过程如图 6 所示，以二维平面上最优解的搜索过程为例，从 $x^{(1)}$ 出发，分别沿着 2 个坐标轴方向找到极值点 $x^{(2)},x^{(3)},x^{(4)},\cdots$，最终找到

最优解 x_0。

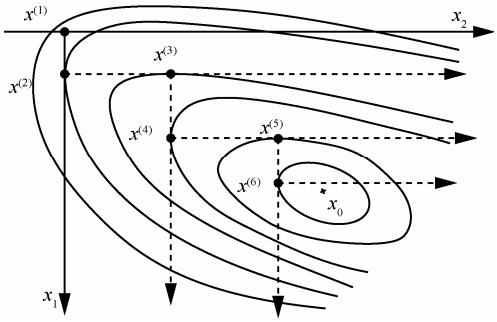


图 6　单变量搜索最优解的过程

根据上述所提单变量搜索法，对式(14)进行优化求解，令 $\frac{\partial D(n)}{\partial q_u}=0,(u=0,1,\cdots,L)$，得

$$q_u=\frac{1}{2}\frac{\sum_{j=0}^{n-1}\left[K_j(u)-M(2\varepsilon)\right]H_j(u)}{\sum_{j=0}^{n-1}H_j(u)^2} \tag{15}$$

其中，

$$\begin{aligned}K_j(u)&=\sum_{k=L+jM}^{L+(j+1)M-1}E_k(u)\\ H_j(u)&=\sum_{k=L+jM}^{L+(j+1)M-1}p_{k-u}E_k(u)\\ E_k(u)&=\prod_{\substack{i=0\\ i\neq u}}^{L}(1-2p_{k-i}q_i)\end{aligned} \tag{16}$$

类似于图 6 的求解过程，对于本文代价函数的优化求解，首先计算得到某个 q_u 的新值，然后用它取代原来的值，用于下次计算，这样经过多次迭代后使局部最优解趋于全局最优解。

由于 $q_u=\Pr(c_u=1),u=0,1,\cdots,L$，最后通过对 q_u 进行判决，得到自同步扰码系数 c_u。

$$c_u=\begin{cases}1, & q_u>0.5\\ 0, & q_u<0.5\end{cases}\quad(u=0,1,\cdots,L) \tag{17}$$

3.2　同步扰码的估计原理

同步扰码多项式的求解过程与自同步扰码的识别过程基本相同。由图 3 可知

$$s_k\oplus s_k=s_k\oplus\sum_{i=1}^{L}\oplus c_i s_{k-i}=\sum_{i=0}^{L}\oplus c_i s_{k-i}=0 \tag{18}$$

又因为 $y_k=x_k\oplus s_k$，可知

$$\begin{aligned}\Pr(\sum_{i=0}^{L}\oplus c_i y_{k-i}=1)&=\Pr((\sum_{i=0}^{L}\oplus c_i x_{k-i})\oplus(\sum_{i=0}^{L}\oplus c_i s_{k-i})=1)\\ &=\Pr(\sum_{i=0}^{L}\oplus c_i x_{k-i}=1)\end{aligned} \tag{19}$$

假设扰码多项式 $P(x)$ 为一个 L 阶 d 项式，即 $P(x)=1+\sum_{j=1}^{d-1}x^{i_j}$，$0<i_1<i_2<\cdots<i_{d-1}=L$，则可推导得

$$\begin{aligned}&\Pr(\sum_{i=0}^{L}\oplus c_i x_{k-i}=1)=\Pr(x_k\oplus\sum_{j=1}^{d-1}x_{k-i_j}=1)\\ &=\sum_{\substack{i=0\\ i=odd}}^{d}\mathrm{C}_d^i(\frac{1}{2}-\varepsilon)^i(\frac{1}{2}+\varepsilon)^{d-i}\\ &=\frac{1}{2}\left[\sum_{i=0}^{d}\mathrm{C}_d^i(\frac{1}{2}-\varepsilon)^i(\frac{1}{2}+\varepsilon)^{d-i}-\sum_{i=0}^{d}\mathrm{C}_d^i(\varepsilon-\frac{1}{2})^i(\frac{1}{2}+\varepsilon)^{d-i}\right]\\ &=\frac{1}{2}[1-(2\varepsilon)^d]\end{aligned} \tag{20}$$

由式(19)和式(20)可知

$$\Pr(\sum_{i=0}^{L}\oplus c_i y_{k-i}=1)=\frac{1}{2}[1-(2\varepsilon)^d],\quad k\geqslant L \tag{21}$$

对式(21)稍作变换，可得

$$1-2\Pr(\sum_{i=0}^{L}\oplus c_i y_{k-i}=1)=(2\varepsilon)^d,\quad k\geqslant L \tag{22}$$

接下来的推导过程和自同步扰码的相同，最后可建立如下所示的目标函数

$$D(n)=\sum_{j=0}^{n-1}\left|\sum_{k=L+jM}^{L+(j+1)M-1}\prod_{i=0}^{L}(1-2p_{k-i}q_i)-M(2\varepsilon)^d\right|^2 \tag{23}$$

对其进行优化求解，可得

$$q_u=\frac{1}{2}\frac{\sum_{j=0}^{n-1}\left[K_j(u)-M(2\varepsilon)^d\right]H_j(u)}{\sum_{j=0}^{n-1}H_j(u)^2} \tag{24}$$

其中，$K_j(u)$、$H_j(u)$ 与式(16)中的相同。对于其初态的识别过程，可以采用文献[3]中提出的基于卷积码快速相关攻击的方法对其进行识别。

4　基于符合度的扰码参数估计

由第 3 节中的式(6)和式(21)可知，不管是自同步扰码还是同步扰码，当利用正确的扰码多项式对其进行解扰时，得到的抽头位置上扰码序列的模二求和值是(0,1)不平衡的，由式(6)可知，自同步扰码的不平衡度为 $\varepsilon,\varepsilon\neq 0$，由式(21)可知，同步扰码的

不平衡度为 $\frac{(2\varepsilon)^d}{2}, \varepsilon \neq 0$。但是，当利用的解扰多项式并不是其实际的扰码多项式时，进行解扰后，得不到其信息序列，同时抽头位置上扰码序列的模二求和值仍处于(0,1)平衡状态，也就是说式(6)和式(21)中的 $\varepsilon = 0$。

若采用遍历的方法对扰码多项式进行识别，对于每一个测试扰码多项式的系数而言有 $\Pr(c_i = 1) = c_i, i = 0,1,\cdots,L$，则由式(11)可知，自同步扰码满足的关系式为

$$\sum_{k=L+jM}^{L+(j+1)M-1} \prod_{i=0}^{L} (1 - 2c_i \Pr(y_{k-i} = 1)) = M(2\varepsilon), j \geqslant 0 \quad (25)$$

对于同步扰码有

$$\sum_{k=L+jM}^{L+(j+1)M-1} \prod_{i=0}^{L} (1 - 2c_i \Pr(y_{k-i} = 1)) = M(2\varepsilon)^d, j \geqslant 0 \quad (26)$$

利用上面分析的特性，并用 y_k 的后验概率 p_k 取代 $\Pr(y_k = 1)$，建立符合度的概念如下。

$$F(j) = \sum_{k=L+jM}^{L+(j+1)M-1} \prod_{i=0}^{L} (1 - 2p_{k-i}c_i), \quad j \geqslant 0 \quad (27)$$

由式(25)可知，对于自同步扰码，当利用正确的测试多项式对其进行解扰时，有 $F(j) \approx M(2\varepsilon)$，$\varepsilon \neq 0$；否则有 $F(j) \approx 0$。由式(26)可知，对于同步扰码也有类似的结论，当测试扰码多项式正确时，有 $F(j) \approx M(2\varepsilon)^d$，$\varepsilon \neq 0$，否则有 $F(j) \approx 0$。则利用这种区别作为判别的标准，可以对扰码多项式进行识别。综上所述，对于给定的扰码序列，可以利用式(27)求得每个测试扰码多项式的符合度 $F(j)$，对应符合度最大的扰码多项式就是待识别的加扰多项式。

5 仿真实验及性能分析

对于自同步扰码和同步扰码，其问题模型分别如图 4 和图 5 所示，本节将分别采用基于扰码系数代价函数的参数估计方法（方法 1）和基于符合度的扰码参数估计方法（方法 2）对其进行仿真。

5.1 基于扰码系数代价函数参数估计方法的仿真

5.1.1 自同步扰码的仿真

对于自同步扰码，其问题模型如图 4 所示。仿真实验中首先利用扰码多项式为 $P(x) = x^{10} + x^6 + x^5 + x^3 + x^2 + x + 1$ 的 LFSR 对 $\varepsilon = 0.1$ 的信息序列 $\{x_k\}$ 进行加扰得到扰码序列 $\{y_k\}$，然后对扰码序列 $\{y_k\}$ 进行 BPSK 映射得到映射序列 $\{z_k\}$ $(0 \leftrightarrow -1, 1 \leftrightarrow +1)$，最后通过信噪比 $SNR = 6$ dB 的 AWGN 信道进行加噪得到已知的接收序列 $\{r_k\}$。由于事先并不知道其扰码多项式的级数 L，所以可以将其适当设大一些，本仿真实验中 L 设为 15，得到的仿真结果为 $[c_0, c_1, \cdots, c_{15}] = [1,1,1,1,0,1,1,0,0,0,1,0,0,0,0,0]$ 时，认为识别正确。利用方法 1 对接收序列 $\{r_k\}$ 进行仿真实验，在不同迭代次数的情况下，进行 500 次蒙特卡罗仿真实验其扰码序列长度和正确识别率的关系如图 7 所示。从图 7 可以看出采用方法 1 对自同步扰码进行识别，其识别正确率随着数据量的增加而增加。且当扰码序列的长度在 0～100 000 bit 时，其识别正确率增长较快；当数据量达到一定饱和度后，其识别正确率随数据量的增加明显变缓，此时数据量对其识别正确率的影响减小。同时从图 7 可以看出，当数据量相对较少时，迭代 2 次的效果要明显优于迭代 1 次的效果，这是因为数据量较少时进行迭代可以更加充分地利用所获得的有限信息；而当数据量充足时，进行 1 次迭代就可以得到正确率较高的识别结果。

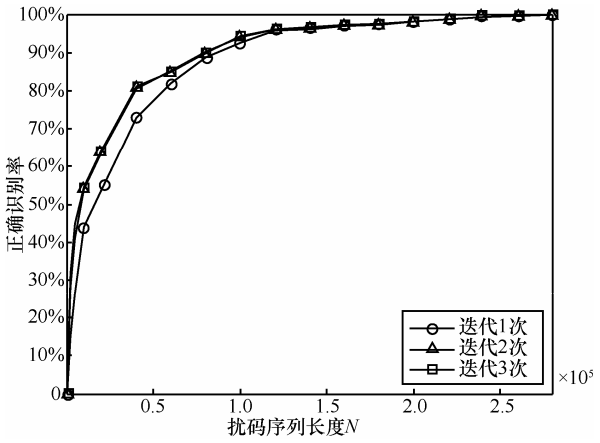


图 7 方法 1 自同步扰码数据量与正确识别率的关系

保持上面其他仿真参数不变，固定仿真所需的数据量 $N = 300\,000$ bit 和 $N = 50\,000$ bit，对于不同信噪比下得到的接收序列 $\{r_k\}$，进行 500 次蒙特卡罗仿真实验得到的仿真结果如图 8 所示。从图 8 可以看出，其正确识别率随着信噪比的增加而增加。且当数据量相对充足时（$N = 300\,000$ bit），迭代次数对其影响较小；当数据量相对较少时（$N = 50\,000$ bit），迭代 2 次的效果要明显优于迭代 1 次的识别效果。

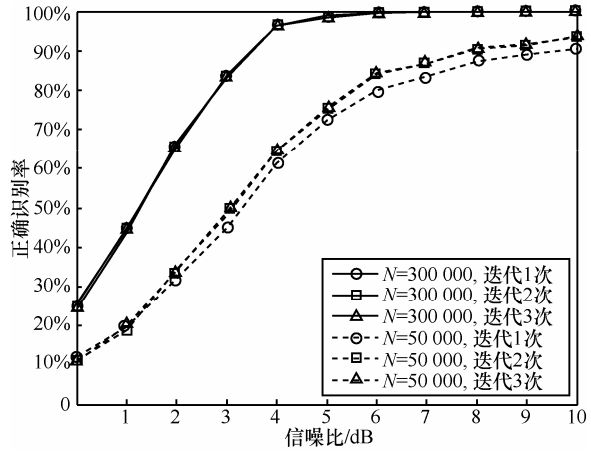


图 8　方法 1 自同步扰码信噪比与正确识别率的关系

方法 1 与文献[1]中 Cluzeau 提出的遍历方法相比，其所需要的数据量明显增加。但是当加扰多项式的三项式倍数阶数较高时，采用本文的方法 1 可以大大缩短其识别的时间，如 $P(x)=x^{10}+x^6+x^5+x^3+x^2+x+1$ 的扰码多项式，在 $SNR=6$ dB，数据量 $N=300\,000$ bit，有偏性 $\varepsilon=0.1$ 的实验条件下迭代 3 次（由图 8 可知正确识别率为 100%），在主频为 3.2 GHz 的 Pentium 4 机器上其识别的时间为 8 s；而利用文献[1]中的方法所用时间需要 5 min 18 s。可以看出对于自同步扰码采用本文的方法 1 在时间（复杂度）方面的优势较明显，分析其原因可以知道，文献[1]中的算法采用遍历的方法对扰码多项式进行识别，其求解过程比较盲目，当其正确的扰码多项式较难找到时，所需要的时间会很长，而本文的方法 1 不再需要遍历，进行一次求解过程就可以对扰码参数识别，所以所需时间较少。且对于自同步扰码，其项数 d 不会对算法产生影响，所以可对任意的多项式进行识别，当 L 一定时，其识别时间相当，不随多项式的特性变化，如对于扰码多项式 $P(x)=x^8+x^4+x^3+x^2+1$，在 $L=15$，$\varepsilon=0.1$，$SNR=6$ dB，$N=300\,000$ bit 的实验条件下，迭代 3 次其识别时间也为 8 s。

5.1.2　同步扰码的仿真

对于同步扰码，其相关问题模型如图 5 所示。实验中采用的扰码多项式为 $P(x)=x^7+x^4+1$，在 $SNR=15$ dB，$\varepsilon=0.1$，$L=15$ 的实验条件下，进行 500 次蒙特卡罗仿真，其扰码序列的长度和正确识别率的关系如图 9 所示。从图 9 可以看出，与图 7 所示的自同步扰码仿真实验结果相比，同步扰码需要的数据量更多，这是因为抽头位置上自同步扰码序列模二求和值的不平衡度为 $\frac{2\varepsilon}{2},\varepsilon\neq 0$，而同步扰码的不平衡度为 $\frac{(2\varepsilon)^d}{2},\varepsilon\neq 0$，所以在对同步扰码进行识别时，其有偏性不明显，需要更多的数据来进行识别。正是因为同步扰码需要的数据量很多，进行仿真实验的时间太长，图 9 和图 10 对仿真的数据量都进行了适当限制。从图 9 可以看出，当数据量有限时，迭代次数对其识别正确率的影响较大，迭代次数越多效果越好。

同时，同步扰码的不平衡度 $\frac{(2\varepsilon)^d}{2}$ 会随着项数 d 的增加而减小，所以对于同步扰码采用本文的方法 1 主要是对三项式和五项式进行识别，对于 d 值较大的多项式，采用本文的方法 1，所需数据量太大，这也是在对同步扰码进行识别时，本文方法 1 的不足之处，但是对于三项式和五项式仍有其在识别时间方面的优势。例如对于扰码多项式 $P(x)=x^{15}+x^5+x^4+x^2+1$，在 $SNR=18$ dB，数据量 $N=140\,000\,000$ bit，有偏性 $\varepsilon=0.1$ 的实验条件下迭代 4 次，在主频为 3.2 GHz 的 Pentium 4 机器上其识别的时间为 49 min，而利用文献[1]中的方法所用时间需要 58 min。

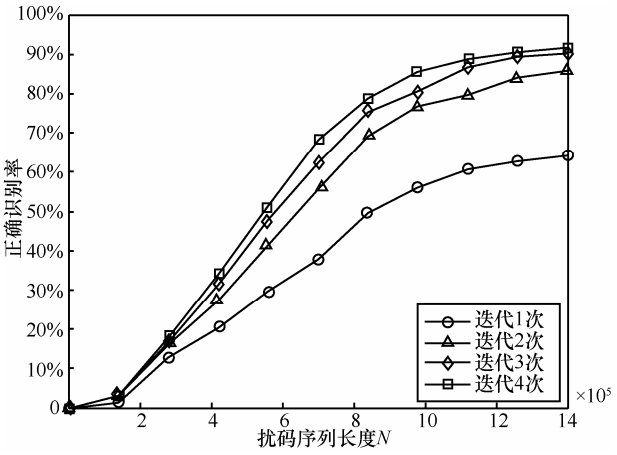


图 9　方法 1 同步扰码数据量与正确识别率的关系

其他仿真参数不变，固定仿真所需的扰码序列长度 $N=1\,400\,000$ bit，进行 500 次蒙特卡罗仿真实验其信噪比和正确识别率的关系如图 10 所示。从图 10 可以看出，当信噪比在 0～8 dB 时，正确识别率随信噪比的增加幅度较明显，但当信噪比继续增加时，其正确识别率趋于稳定，这是

因为此时正确识别率受数据量的限制，当数据量不够时，即使继续增加信噪比，对其正确识别率的影响也不大。且由于迭代处理能使有限的信息得到充分利用，从图 10 可以看出迭代次数越多，识别性能越好。

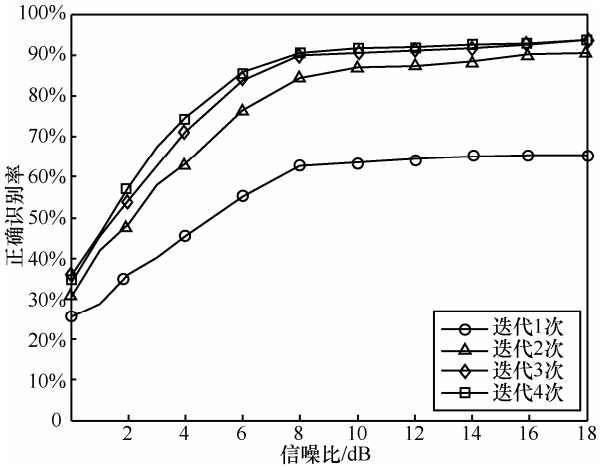


图 10 方法 1 同步扰码信噪比与正确识别率的关系

5.2 基于符合度扰码参数估计方法的仿真

5.2.1 自同步扰码的仿真

对于自同步扰码，为了更好地展示其性能，将与文献[4]中的算法做比较。其测试扰码多项式遍历的集合引用的是文献[4]中的测试集合，所采用的扰码多项式为文献[4]中的三项式 $P(x)=x^{23}+x^{18}+1$，信噪比为 $SNR=6$ dB。对于本文的方法 2 所利用的信息为软信息 $\{r_k\}$，对于文献[4]由于要统计抽头位置比特状态的概率，所以需要使用硬判决信息，在此对已知的软信息序列 $\{r_k\}$ 进行硬判决后，再利用文献[4]的方法进行识别。对于不同长度的信息序列 $\{r_k\}$，进行 1 000 次蒙特卡罗仿真实验，在 $\varepsilon=0.1$ 和 $\varepsilon=0.05$ 这 2 种情况下得到的仿真结果分别如图 11 和图 12 所示。从图 11 和图 12 的仿真结果可以看出，本文的方法 2 和文献[4]的方法，其正确识别率都随着扰码序列的长度增加而增加；同时当 $\varepsilon=0.05$ 时，正确识别需要的扰码序列长度都更长，因为当 $\varepsilon=0.05$ 时，正确的测试扰码多项式和错误的测试扰码多项式之间的差别更小。由于本文方法利用了软信息，其对所获得的接收信息进行了更充分的利用，所以不管在 $\varepsilon=0.1$，还是 $\varepsilon=0.05$ 的情况下，其识别性能都要比文献[4]中方法的性能更好。

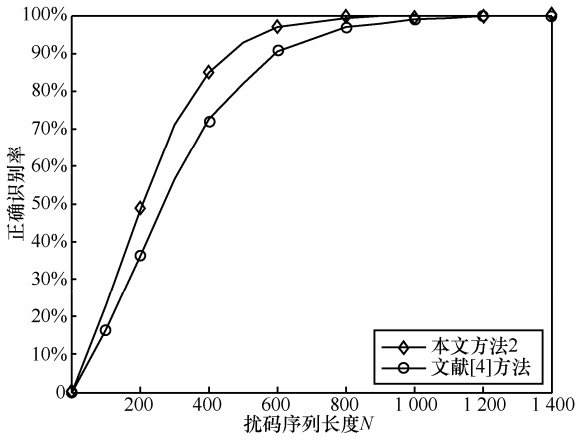


图 11 本文方法 2 和文献[4]方法自同步扰码数据量与正确识别率的关系(ε=0.1)

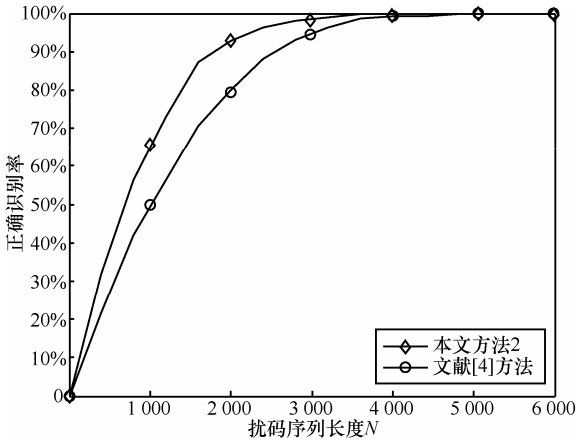


图 12 本文方法 2 和文献[4]方法自同步扰码数据量与正确识别率的关系（ε=0.05）

当 $\varepsilon=0.1$ 时，固定仿真所需的数据量 $N=800$ bit；当 $\varepsilon=0.05$ 时，固定仿真所需的数据量 $N=3\,500$ bit，其他的仿真参数不变。对于不同信噪比下得到的接收序列 $\{r_k\}$，进行 1 000 次蒙特卡罗仿真实验得到的仿真结果分别如图 13 和图 14 所示。

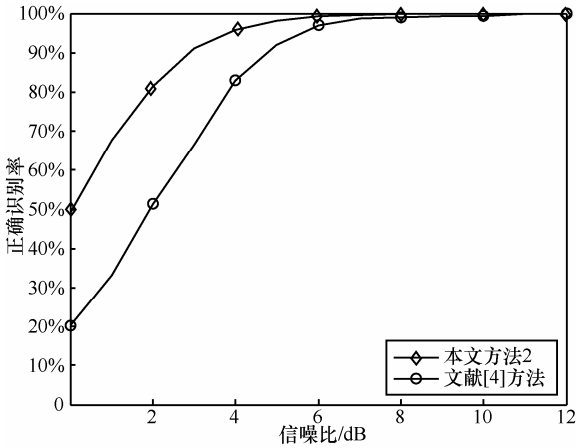


图 13 本文方法 2 和文献[4]方法自同步扰码信噪比与正确识别率的关系（ε=0.1）

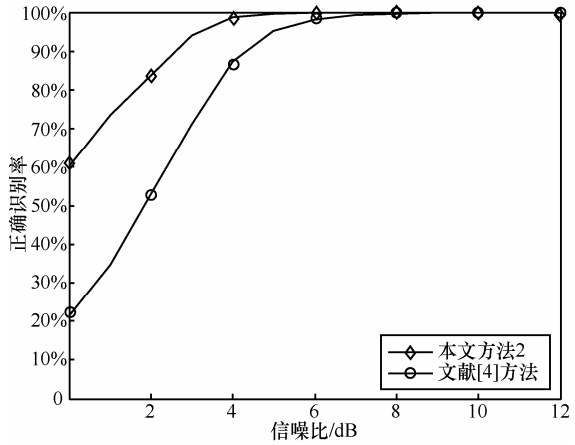


图 14　本文方法 2 和文献[4]方法自同步扰码信噪比与正确识别率的关系（ε=0.05）

由图 13 和图 14 的仿真结果可以看出，本文的方法 2 和文献[4]的方法，其正确识别率都随着信噪比的增加而增加。由于本文方法引入了软信息，其抗噪性能明显要优于文献[4]的方法，特别是当信噪比较低时，其差距较大。如从图 13 和图 14 的仿真结果可以看出，当 $SNR=0$ dB 时，在 $\varepsilon=0.1$ 的情况下，本文方法的估计正确率要比文献[4]的算法高 30%左右；在 $\varepsilon=0.05$ 的情况下，本文方法的估计正确率要高 40%左右。同时从图 13 和图 14 可看出，在 95%正确率处，本文方法具有约 2 dB 的信噪比增益，这是使用软信息所带来的优势。

最后对本文方法和文献[4]方法的复杂度进行简单的分析。假设扰码序列的长度为 N，扰码多项式的阶数和项数分别为 L 和 d，则对于一个确定的扰码多项式，文献[4]方法的计算量约为 $O(N-L)$（比特操作为主），本文的计算量约为 $O((N-L)(d-1))$（乘法操作为主）。在计算复杂度方面，本文方法确实要高于文献[4]方法。但本文方法的复杂度完全在可以承受的范围里，且从上述实验结果来看，虽然本文方法牺牲了部分计算量，但其性能得到了较大的提高。

5.2.2　同步扰码的仿真

对于同步扰码，下面的仿真实验将与文献[3]中的组合枚举求优势法做比较。对于组合枚举求优势法，由于利用的是硬判决信息，所以仍对软信息序列 $\{r_k\}$ 进行硬判决后，再利用文献[3]中的方法对扰码参数进行识别。仿真实验采用的扰码多项式为 $P(x)=x^{15}+x^{14}+1$，在 $SNR=10$ dB，$\varepsilon=0.1$ 的实验条件下，对于不同长度的接收序列，进行 500 次蒙特卡罗仿真实验得到的结果如图 15 所示。从图 15 可知，对于本文方法和文献[3]中的方法，其正确识别率随着数据量的增加而增加，且在同样的扰码序列长度下，由于本文利用了所获得软信息中的可靠度信息，对接收信息实现了更充分的利用，其识别性能更好。

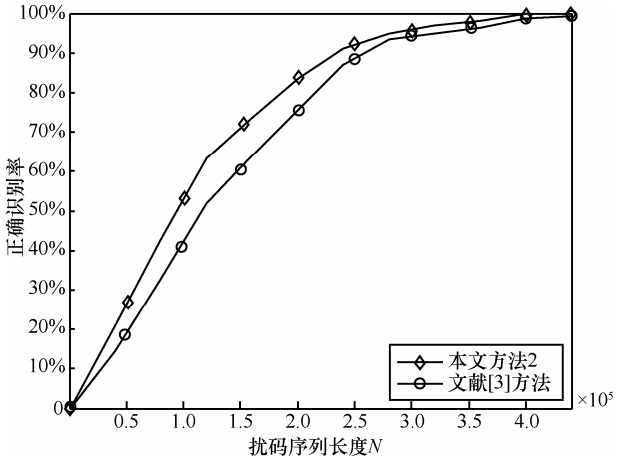


图 15　本文方法 2 和文献[3]方法同步扰码数据量与正确识别率的关系

当 $\varepsilon=0.1$ 时，固定仿真所需的数据量 $N=350\,000$ bit。在不同信噪比的情况下，进行 500 次蒙特卡罗仿真实验得到的仿真结果如图 16 所示。从图 16 中可以看出，本文的方法 2 和文献[3]的方法，其正确识别率都随着信噪比的增加而增加，但本文方法的抗噪性能更好。当 SNR=0 dB 时，本文方法的正确识别率要比文献[3]的方法高 30%左右。且当达到较高正确率 90%时，本文方法的信噪比增益将近 2 dB。

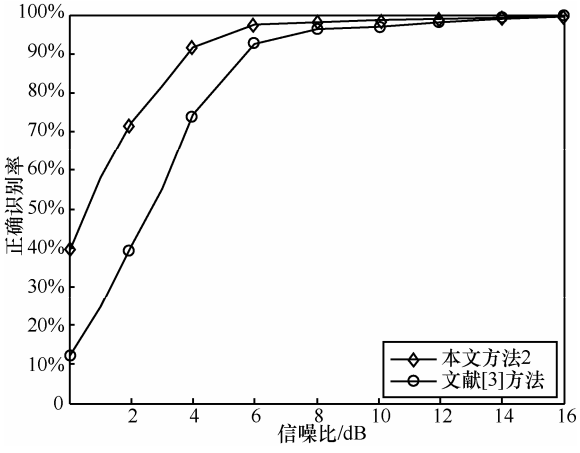


图 16　本文方法 2 和文献[3]方法同步扰码信噪比与正确识别率的关系

从复杂度上分析，对于一个特定的扰码多项式 $P(x)$，本文方法和文献[3]方法的计算量都为 $O((N-L)(d-1))$，但本文方法以乘法操作为主，文

献[3]方法以比特操作为主，也就是说本文方法的复杂度要略高于文献[3]方法，但在性能方面本文方法更优。

6 结束语

本文针对自同步扰码和同步扰码，提出了 2 种利用软信息进行识别的方法。方法 1 建立了扰码系数的代价函数，利用实数域的优化理论进行求解，由于该方法是一种开集识别算法，不再需要对闭集进行遍历，只需进行一次优化求解，利用该方法可大大缩短其识别的时间。方法 2 利用软信息建立了符合度的概念，由于软信息中不仅含有发送比特的符号信息，还包含了发送比特的可靠度信息，实验证明，相对于硬判决识别算法，该算法具有更优的性能。总而言之，本文所提方法识别性能良好，具有较强的抗噪性，更加适用于实际环境。

参考文献：

[1] CLUZEAU M. Reconstruction of a linear scrambler[J]. IEEE Transactions on Computers, 2007, 56(9):1283-1291.

[2] LIU X B, SOO N K, WU X W, et al. Reconstructing a linear scrambler with improved detection capability and in the presence of noise[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(1): 208-218.

[3] 罗向阳, 沈利, 陆佩忠, 等. 高容错伪随机扰码的快速盲恢复[J]. 信号处理, 2004, 20(6): 553-558.
LUO X Y, SHEN L, LU P Z, et al. Fast blind restore of LFSR sequences with high error tolerance[J]. Signal Processing, 2004, 20(6): 553-558.

[4] 廖红舒, 袁叶, 甘露. 自同步扰码的盲识别方法[J]. 通信学报, 2013, 34(1): 137-143.
LIAO H S, YUAN Y, GAN L. Novel blind recognition method for self-synchronized scrambler[J]. Journal on Communications, 2013, 34(1): 137-143.

[5] 于沛东, 李静, 彭华. 一种利用软判决的信道编码识别新算法[J]. 电子学报, 2013, 41(2): 301-306.
YU P D, LI J, PENG H. A novel algorithm for channel coding recognition using soft-decision[J]. Acta Electronica Sinica, 2013, 41(2): 301-306.

[6] 刘骏, 李静, 于沛东. 一种 Turbo 码随机交织器的迭代估计方法[J]. 通信学报, 2015, 36(6): 2015140.
LIU J, LI J, YU P D. Iterative estimation method for random interleaver of Turbo codes[J]. Journal on Communications, 2015, 36(6): 2015140.

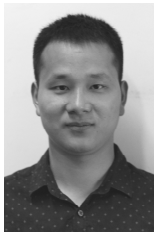
[7] YU P D, LI J, PENG H. Gibbs sampling based parameter estimation for RSC sub-codes of Turbo codes[C]//Wireless Communications and Signal Processing (WCSP), 2014 Sixth International Conference. 2014: 1-5.

[8] 张永光, 楼才义. 信道编码及其识别分析[M]. 北京: 电子工业出版社, 2010.
ZHANG Y G, LOU C Y. Channel coding recognition and analysis[M]. Beijing: Publishing House of Electronics Industry, 2010.

[9] HAGENAUER J, OFFER E, PAPKE L. Iterative decoding of binary block and convolutional codes[J]. IEEE Transactions on Information Theory, 1996, 42(2): 429-445.

[10] YU P D, LI J, PENG H. A least square method for parameter estimation of sub-codes of Turbo codes[J]. IEEE Communication Letters, 2014, 18(4): 644-647.

作者简介：



陈泽亮（1992-），男，湖南岳阳人，解放军信息工程大学硕士生，主要研究方向为信道编码识别分析。

彭华（1973-），男，江西萍乡人，解放军信息工程大学教授、博士生导师，主要研究方向为软件无线电、通信信号处理等。

巩克现（1976-），男，山东泰安人，解放军信息工程大学副教授、硕士生导师，主要研究方向为软件无线电、差错控制编码等。

于沛东（1989-），男，湖南慈利人，解放军信息工程大学博士生，主要研究方向为信道编码及其识别分析。