基于双面过程的纠错方法

Boris Ryabko

联邦信息与计算技术研究中心 ,
新西伯利亚国立大学，以及
西伯利亚国立电信与信息学大学，新西伯利亚

arXiv:2601.06447v1 [cs.IT] 10 Jan 2026

摘要

本文提出了一种解决通信信道纠错问题的新方法：以一种显著增强符号间相互依赖的方式转换输入序列。随后，序列经信道传输后，利用这一特性进行纠错，从而显著降低剩余错误率。编码和解码的复杂度均为线性。

I. 引言

自克劳德·香农发表著名论文 [1], 以来，如何在噪声信道上高效传输数据一直是信息论的核心课题。每年都有数百篇论文探讨各种纠错方法，我们曾提及其中的几本著作 [2], [3], ，其中可找到许多参考文献。

理论研究和实际应用中的大多数纠错码，都是通过向传输消息添加冗余符号来实现的。这类方法包括分组码[4], [5], 卷积码 [6], 基于哈希函数的码 [7],以及其他一些方法。

本文提出一种新方法：通过特殊变换增强消息符号之间的相互依赖，然后利用这种依赖来纠正错误。

A. 所提方法的主要工具：双面过程

在文章的主体部分，我们将考虑生成字母表 $\{0,1\}$ 中字母的信息源，尤其是伯努利信源 $B(p)$ ，它以概率 $P(0)=p, P(1)=q(=1-p)$ $p \geq 1/2$生成字母。

所提出的码将基于所谓的双面过程，即阶（或记忆）为 s， $s \geq 1$的马尔可夫链。这意味着，对于生成的序列 $...x_{-1}x_0x_1...$，概率 $P(x_{i+1})$取决于前面 s 个符号，也就是说，概率 $P(x_{i+1}=\alpha/...x_{-1}x_0x_1...x_i)$等于 $P(x_{i+1}=\alpha/x_{i-s+1}...x_i)$，其中 $\alpha \in \{0,1\}$。我们将用以下转移概率矩阵来描述这些链：

$$s=1: \quad \begin{array}{|c|c|c|} \hline probability\,of\,zero & p & q \\ \hline after & 0 & 1 \\ \hline \end{array}. \tag{1}$$

这意味着， $Pr\{x_{i+1}=0/x_i=0\}=p$、 $Pr\{x_{i+1}=0/x_i=1\}=q$ ，当然还有 $Pr\{x_{i+1}=1/x_i=0\}=q$、 $Pr\{x_{i+1}=1/x_i=1\}=p.$。显然， $Pr\{x_{i+1}=1/...\}=1-Pr\{x_{i+1}=0/...\}$，而这些概率并未包含在矩阵中。

$$For\ s=2: \quad \begin{array}{|c|c|c|c|c|} \hline probability\,of\,zero & p & q & q & p \\ \hline after & 00 & 01 & 10 & 11 \\ \hline \end{array}. \tag{2}$$

这意味着， $Pr\{x_{i+1}=0/x_{i-1}x_i=00\}=p$，等等。

当 p 较大时（例如0.9），(1)和(2)的“典型”序列可以如下所示：

$$x_1x_2x_3x_4x_5x_6x_7x_8x_9x_{10} = 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ , \tag{3}$$

$$x_1x_2x_3x_4x_5x_6x_7x_8x_9x_{10} = 0\ 0\ 0\ 0\ 1\ 1\ 0\ 1\ 1\ 0, . \tag{4}$$

（在这两种情况下， x_4之后都发生了一次罕见事件，即其概率为 q。）

B. 所提方法背后思想的直观描述

本文将讨论信息论中非常常见的两种二元信道上的传输：擦除信道（EC）和二元对称信道（BSC）。对于

对于第一种信道，来自 $\{0,1\}$ 的任何传输字母都可能丢失，即接收方以一定概率 π收到一个特殊符号 $*$（而不是传输的 0 或 1）。对于二元对称信道，发送方发出的任何字母都可能以一定概率 π被另一个字母替换，参见 [9]。

现在假设一个二进制序列以 π 的概率通过 EC 传输，也就是说，在传输过程中，任意字母以 π 的概率被转换为 $*$。

我们考虑三种情况：i）传输序列由伯努利信源 $B(p)$ $p > 1/2$ 生成。显然，最佳策略是将 $*$ 替换为 0，因为 0 的概率等于 $p > 1/2$。因此，$Pr(error) = q$。今后，对于所有情况，我们都将采用相同的策略，即选择概率较高的字母，因此，$Pr(erroe)$ 将等于较小的概率。

ii）考虑由双面过程生成的序列通过 EC 传输的情况。假设该过程的记忆等于 1，并且传输序列（3）被接收为

$$0\,0\,*\,0\,1\,1\,1\,*\,1\,1. \tag{5}$$

（也就是说，第三个和第八个字母被擦除）。那么，简单计算可得 $P(x_3 = 0/(5)) = P($ $x_3 = 0/x_2 = 0\&(5))) \,\&\, P(x_4 = 0/x_3 = 0\,\&(5))) = p^2/(p^2 + q^2)$。因此，在这种情况下，$Pr(error) = P(x_3 = 1/(5)) = q^2/(p^2 + q^2)$。第八个字母也是如此。

iii) 现在假设双面过程的记忆为 2，并且序列 (4) 通过同一个擦除信道传输后，接收到以下内容：

$$0\,0\,*\,0\,1\,1\,0\,*\,1\,0\,. \tag{6}$$

直接计算概率可得 $P（x_3 = 0/（6） = p^3/（p^3 + q^3）$。因此，在这种情况下，$Pr（error） = q^3/（p^3 + q^3）$。第八个字母也是如此。

因此，从这些简单的例子中我们可以看到，对于伯努利过程，错误概率为 q；对于记忆长度分别为一和二的双边过程，错误概率为$q^2/（p^2 + q^2）$和$q^3/（p^3 + q^3）$。由此，在这个例子中我们可以看到，增加过程的记忆能够降低译码错误（在这些例子中，我们没有考虑信道错误可能出现的所有位置，

但稍后我们将考虑一般方法，并给出一般的错误估计）。因此，从某种意义上说，增加记忆能够纠正大多数信道错误。

所提出的纠错方法的第二个要素，是能够将伯努利过程（以及任何其他随机过程）转换为具有所需记忆的双面过程。

总而言之，所提出的方法是一个两步过程：首先，将待传输的序列转换为双面过程，并通过噪声信道进行传输。然后，根据错误概率估计对接收到的数据进行数据纠错，再将所得序列从双面过程转换回伯努利过程。

事实证明，在某些参数范围内，该方法能够显著降低擦除信道和二元对称信道中的错误率。需要指出的是，该方法的编码和解码复杂度均为线性。

本文其余部分的结构如下。下一节介绍有关双面过程的背景信息。随后一节描述所提出的方法并考察其性质。第三节介绍实验结果，最后一节给出简短的结论。

II. 双面过程与变换

A. 双面过程

定义矩阵 $T_1 = \binom{p\ q}{0\ 1}$、 $T_1 = \binom{q\ p}{0\ 1}$、 $T_2 = T_1 \parallel T_1 = \binom{p\ q\ q\ p}{00\ 01\ 10\ 11}$, $T_2 = \binom{q\ p\ p\ q}{00\ 01\ 10\ 11}$, $T_3 = T_2 \parallel \underline{T_2}$、 ... $T_{n+1} = T_n \parallel \underline{T_n}$ 等。例如，

$$\overline{T}_3 = \binom{q\ \ p\ \ p\ \ q\ \ p\ \ q\ \ q\ \ p}{000\ 001\ 010\ 011\ 100\ 101\ 110\ 111},$$

其中，与之前一样， $Pr\{x_{i+1} = 0/x_{i-2}x_{i-1}x_i = 000\} = q$ 等。

更确切地说，$\overline{T}_k$ 由 T_k 通过改变 T_k 顶行中的 p 和 q 得到，而上行 T_{k+1} 是上行 T_k 和 T_k的连接，下行则是由 $\{0,1\}^{k+1}$中按字典序排列的词构成的序列。

这些过程由 B. Ryabko [8] 提出，并且已证明，对于具有记忆的过程 s，任意 $r \leq s$ 以及长度为 r的词 g，都有 $Pr\{x_i...x_r = g\} = 2^{-r}$，其中每字母香农熵 h_∞ 等于 $p\log_2 p + q\log_2 q$，也就是说， h_∞ 可以非常小（当 p很大时）。（这就是这类过程被称为双面过程的原因。如果我们观察长度为 r， $r \leq s$的词的频率，该过程看起来是均匀分布的（即"完全随机"）；但如果观察长词（长度超过 s的词）的频率，这种分布就与均匀分布相去甚远。

B. 双面变换

首先，我们定义两个映射 $\varphi_{l,\ w}$ ： $\{0,1\}^n \to \{0,1\}^n$ 和 $\psi_{w,\ l}$ ： $\{0,1\}^n \to \{0,1\}^n$ ，其中整数 n， l， $(l \leq n)$ 以及一个词 $w \in \{0,1\}^l$是参数。给定词 w =$w_1...$， w_l, $\in \{0,1\}^l$、 $x_1...x_n \in \{0,1\}^n$

The 变换 $\varphi_{l,\ w}$： $x \to v$，由以下方程定义： $v_j = w_{l+j}$对于 $j = -l+1$，.....， 0，

$$v_i = \begin{cases} x_i & \text{if } \Sigma_{j=i-l}^{i-1} v_j \text{ is even} \\ \overline{x}_i & \text{if } \Sigma_{j=i-l}^{i-1} v_j \text{ is odd}, \end{cases} \tag{7}$$

对于 $i = 1$，.....， n （此处及下文中$\overline{\ } 0 = 1$， $1 = 0$）。对于某个 n位词 v，变换 $\psi_{l,\ w}$: $v \to u$ 定义为

$$u_{i+1} = \begin{cases} v_{i+1} & \text{if } \Sigma_{j=i-l+1}^{i} v_j \text{ is even} \\ \overline{v}_{i+1} & \text{if } \Sigma_{j=i-l+1}^{i} v_j \text{ is odd}, \end{cases} \tag{8}$$

对于 $i = 0$，.....， $n-1$. 例如，对于序列 $x = 0000100000$， $\varphi_{1,\ 0}$（x）等于 0000111111，$\varphi_{2,\ 00}$（x）等于 0000110110， $\psi_{1,\ 0}$（0000111111） $= 0000100000$， $\psi_{2,\ 00}$（0000110110） $= 0000100000$。

注意， $\psi_{l,\ w}$（$\varphi_{l,\ w}$（x）） $= x$。

以下定义在所述码中发挥着重要作用。

定义 1. 设有一个序列 $v = v_1...v_n$、某个整数 l 以及单词 $w \in \{0,1\}^l$。值 $\nu^{i}_{,\tau}{}^{w}(v)$ 表示字母 τ 在单词 $\psi_{l,w}(v)$中出现的次数，其中 $\tau \in \{0,1\}$。

以下来自 [8] 的定理适用于变换$\varphi_{l,w}(x)$：

定理 1. 设 $l > 0$ 为一个整数， w 为由 l 个字母组成的均匀分布的二进制字，并设 $x = x_1...x_n$ 根据参数为$p \in (0, 1)$的伯努利过程随机生成。那么序列 $\varphi_{l,\ u}(x)$ 服从阶数为 l （矩阵为 T_l ）且概率为 p 的双面过程分布。

III. 码

这里我们介绍用于擦除信道和二元对称信道的纠错码。在所有情形下，我们都考虑以下方案。存在一个整数 l 和一个字 $w \in \{0,1\}^l$，它按照均匀分布形成；发送方和接收方都知道这个字 w 和整数 l 。此外，还有一个由伯努利信源 $B(p)$生成的字 $x = x_1...x_n$ ，它必须通过噪声信道传输。当然，接收方不知道这个字。

我们记 $\varphi_{l,\ w}(x) = v_1...v_n$。定义字 $v = w_1...w_l v_1...v_n$，即 v 是一个由$(n + l)$ 个字母组成的字，且 $v_{1-l}...v_0 = w_1...w_l$。字 $v_1...v_n$ 通过噪声信道传输，接收方收到一个受干扰的字 v_1^*，.....，v_n^*。注意，在擦除信道的情形下， $v_i^* \in \{0, 1, *\}$。将 $v_{-l+1}^*...v_0^*$ 定义为 $w_1...w_l$是方便的。（这是可行的，因为接收方知道字 w 。）因此，
$v_{-l+1}^*...v_0^* = v_{-l+1}...v_0$。

解码是一个两步过程：首先，将序列 $v^* = v_1^* v_2^*...v_n^*$ 处理为序列 $v' = v_1' v_2'...v_n'$ ，以期消除一些错误（但同时也可能引入新的错误）。然后应用映射$\psi_{l,\ w}$（v'），所得序列 $x' = x_1'...x_n'$即为解码结果。在此过程中，变换 $v_1^* v_2^*...v_n^* \to v_1' v_2'...v_n'$ 起着关键作用，下面将分别针对擦除信道和二元对称信道进行说明。

另一个重要的量是比特错误率（误码率），其定义为符号的平均频率，其中 $x_i' \neq x_i$，$i = 1$.....，n。我们将其记为 BER（$Code$），其中 $Code$ 是一种特定的编码方法。

A. 擦除信道

假设接收方收到的字为 $v^* = v_1^* v_2^* ... v_n^*$，其中所有字母均属于字母表 $\{0, 1, *\}$。解码器必须将所有符号 $*$ 替换为字母表中的字母$\{0, 1\}$，并得到字 $v' = v_1' v_2' ... v_n'$，力求减少错误的数量，即最小化字母的数量 v_1'，使得 $v_i' \neq v_i$。方法如下：对于 $i = 1..... n$

i) 若字母为 $v_i^* \neq *$，则 $v_i' = v_i^*$。

ii) 若 $v_i^* = *$，且对所有字母 $v_j^* \neq *$，其中 $j = i - l + 2, ..., i - 1, i + 1,..., i + l + 1$, 我们定义

$$\nu' = \nu_0^{l, v_{i-l}^* ... v_{i-1}^*}(\psi_{l, v_{i-l}^* ... v_{i-1}^*}(1\, v_{i+1}^* ... v_{i+l}^*)),$$

$$v_i' = \begin{cases} 1 & \text{if } p^{\nu'}(1-p)^{l+1-\nu'} \geq (1-p)^{\nu'} p^{l+1-\nu'} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

iii) 若 $v_i^* = *$，且至少对一个 j、$v_j' = *$、$j \in \{i - l + 2, ..., i - 1, i + 1, ..., i + l + 1\}$, 或 $i > n - l$ 成立，则

$$v_i' = \begin{cases} 0 & \text{with probability } 1/2 \\ 1 & \text{with probability } 1/2 \end{cases}. \tag{10}$$

我们将此方法记作 $Code_{EC}$。

我们来看一个例子。假设 $x = 0000100000$, $Pr(0) = p > 1/2$, $l = 1$, $w = 0$.那么

$$\varphi_{1,0}(x) = 0000111111 = v_1 ... v_{10},\ v = v_0 v_1 ... v_{10} = 00000111111\,.$$

设传输的码字为 $v_1 ... v_{10} = 0000111111$，接收到的码字为 $v_1^* ... v_{10}^* = 0 * 00011 * 11$。根据定义，$v^* = v_0^* v_1^* ... v_{10}^* = 00 * 00011 * 11$。现在根据 (9) 对 $v_2^* = *$ 和 $v_7^* = *$进行解码，即纠正错误。因此，对于 v_2^*，我们得到

$$\nu' = \nu_0^{1,0}(\psi_{1,0}(10)) = \nu_0^{1,0}(11) = 0\,.$$

显然，$p^0(1-p)^{2-0} < (1-p)^0 p^2$（对于 $p > 1/2$），因此 $v_2' = 0$（见（9））。类似地，对于$v_7^* = *$,

$$\nu' = \nu_0^{1,1}(\psi_{1,1}(11)) = \nu_0^{1,0}(00) = 2$$

并且 $p^2(1-p)^0 > (1-p)^2p^0$，因此，$v'_7 = 1$。这样，两个错误都得到纠正，且 $v'_1...v'_{10}$ =0000111111 $(= v_1...v_{10})$。因此，$\psi_{1,\ 0}(v'_1...v'_{10}) = 0000100000$，也就是说，等于一个初始序列。

第二个例子针对记忆 $l = 2.$。设

$$x = 0000100000,\ w = 00, \varphi_{2,00} = v_1...v_{10} = 0000110110,\ v = v_{-1}v_0v_1...v_{10} = 00\,0000110110\,.$$

假设单词 $v_1...v_{10}$ 通过擦除信道传输，接收到单词 $000 * 110 * 10$。对于 v^*_4，由式 (9) 可得:

$$\nu' = \nu_0^{2,00}(\psi_{2,00}(111)) = \nu_0^{2,00}(101) = 1\,;\ p^1(1-p)^2 < p^2(1-p)^1$$

因此，$v'_4 = 0$。类似地，对于 $v^*_8 = *$。

$$\nu' = \nu_0^{2,10}(\psi_{2,10}(110)) = \nu_0^{2,10}(000) = 3\,;\ p^3 > (1-p)^3$$

因此，$v'_8 = 1$，两个错误都得到了纠正。

下面描述所述码的性质。

定理 2. 设有一个伯努利信源，其取 *0* 的概率为 $p, p > 1/2$，生成字母序列 $x_1...x_n$，该序列通过错误概率为 π 的擦除信道传输。若使用参数为 l 的$Code_{EC}$ 码来纠正错误，则比特错误率可估计如下:

$$BER(Code_{EC}) \le \pi\Bigg(\Big((1-l/n)(1-\pi)^{2l+1}\sum_{i=\lfloor (l+1)/2\rfloor}^{l}\binom{l+1}{i}p^i(1-p)^{l+1-i}\Big) + \Big(((1-(1-\pi)^{2l+1}) + l/n)/2\Big)\Bigg)(l+1). \quad (11)$$

证明。我们估计随机选取的值 $v'_i \ne v_i$ 的概率，然后将该概率乘以 $l+1$，得到对 BER 的估计。(11) 中的两部分分别对应解码描述中的选项 ii) 和 iii)。更准确地说，第一个求和对应于 (9)，下一部分对应于 (10)。（显然，第一个选项 i) 给出 $(1-\pi)0 = 0$。）

□

该定理提供了一种求得使 (11) 最小化的 l 的方法。

考虑当信道错误概率π很小时，估计值 **(11)** 的渐近行为，即 $\pi \to 0$。令 $l = (1/\pi)^{\gamma}$, $\gamma \in (0,\ 1)$，且 $n >> 1/\pi$。我们将使用霍夫丁不等式:

断言 1。（霍夫丁不等式 *[10]*）。设序列 $x_1 ... x_m$ 由伯努利信源 $B(p)$ $p > 1/2$生成。那么

$$\sum_{i=k}^{m} (_i^m) p^i (1-p)^{m-i} \le \exp(-2m(k/m - p)^2),$$

对于 $k > mp$。

对于 $n > 1/\pi$，由该断言及已知不等式 $(1-\varepsilon)^r \le 1 - r\varepsilon$ 和 $l = (1/\pi)^{\gamma}$，根据式 **(11)**，我们得到 $BER(Code_{EC}) \le \pi(\exp(-const\ l)) + (2l+1)\pi$, $const = 2(p - 1/2)^2$。该不等式表明差异 $p - 1/2$ 对误差的影响。此外，若 $l = 1/\pi^{\gamma}$, $\gamma \in (0,1)$，则 $= \pi\ o(1)$。

B. 二元对称信道

现在我们介绍一种用于二元对称信道（$Code_{BSC}$)的码。与之前一样, $v^* = v_1^* v_2^* ... v_n^*$ 由接收方获得，然后处理成序列 $v' = v_1' v_2' ... v_n'$。定义

$$\Delta = \min\{s : (1-\pi) p^s (1-p)^{l+1-s} \ge \pi (1-p)^s p^{l+1-s} \ \&\ 1 \le s \le l+1\}. \tag{12}$$

将词语 $v^* = v_1^* v_2^* ... v_n^*$ 按如下方式处理为序列 $v' = v_1' v_2' ... v_n'$ 。对于任意$i = 1,..., n-l$, 解码器计算 $\nu_i' = \nu^{l}_{0,\ v_{i-l}^* ... v_{i-1}^*}(\psi^l, v_{i-l}^* ... v_{i-1}^*}(v_i^* ... v_{i+l}^*))$ 以及

$$v_i' = \begin{cases} v_i^* & \text{if } \nu_i' \ge \Delta \\ \overline{v_i^*} & \text{otherwise}, \end{cases} \tag{13}$$

以及 $v_i' = v_i^*$对于 $i = n - l + 1$。

定理3。 设伯努利信源中 *0* 的概率为 p, $p > 1/2$, 生成字母序列 $x_1 ... x_n$, 该序列通过误差概率为 π 的二元对称信道传输。如果

使用参数为 l 的 $Code_{BSC}$ 来纠正错误，则比特错误率可估计如下：

$$BER(Code_{BSC}) \leq \left(\left((1-\pi) \sum_{i=0}^{\Delta-1} \binom{l+1}{i} p^i (1-p)^{l+1-i} + \pi \sum_{i=\Delta}^{l} \binom{l+1}{i} (1-p)^i p^{l+1-i} \right) (1-l/n) + \pi\, l/n \right) (l+1)\,. \tag{14}$$

证明。 这里，第一项对应于传输字母未发生变化但被错误解码的情况（即 $v_i^* = v_i$，但 $v_i' \neq v_i^*$）。第二项对应于传输字母被信道改变，但该错误未得到纠正的情况。

□

应用霍夫丁不等式（断言 1），我们可以看到，一般而言，错误的渐近行为与擦除信道的渐近行为相近。特别是，当信道错误概率 π 较低时， $BER(Code_{BSC})$ 的渐近行为，即 $\pi \to 0$，以及 $l = 1/\pi^\gamma, \gamma \in (0,1)$。 $BER(Code_{BSC}) = \pi o(\pi)$。因此，未纠正错误的频率趋于 0。

IV. 结论

我们可以看到，对所提出的码进行编码和解码，每个字母只需常数次操作。从某种意义上说，可以说这些方法“免费”减少了错误数量，因为对传输序列进行编码所需的时间以及对接收序列进行解码所需的时间都与序列长度成正比。

这项研究的进一步发展方向，可以是创建一种能够对解码后的序列进行最大似然估计或简单近似的码。

香农关于噪声信道信息传输的著名定理表明，只有当信源熵小于信道容量时，所述编码才能将错误概率降至零（两种信道的容量估计见 [9]）。如果信源熵超过信道容量，那么一般来说，可以先使用已知的数据压缩解码器将原始序列“拉伸”，再使用所述编码；不过，这种可能性仍需进一步研究，尤其需要评估其复杂度。

参考文献

[1] Shannon CE. 通信的数学理论. 贝尔系统技术期刊. 1948年7月;27(3):379- 423.[2] Moon TK. 纠错编码：数学方法与算法. 约翰·威利父子公司; 2020年12月15日.[3] Etzion T. 完美码及相关结构. 2022年3月14日.[4] Berlekamp ER. 分组码性能. 美国数学学会通报. 2002年1月;49(1):17-22.[5] Moon TK. 纠错编码：数学方法与算法. 约翰·威利父子公司; 2020年12月15日.[6] Johannesson R, Zigangirov KS. 卷积编码基础. 约翰·威利父子公司; 2015年[7] Ryabko B. 线性哈希函数及其在错误检测与纠错中的应用 // 离散数学、算法与应用. - 2024年. - 第16卷. - 第6期. - 文章编号2350070. -[8] Ryabko B. 用于生成k-分布序列和正规序列的低熵随机过程，以及这些过程与随机数生成器之间的关系. 数学. 2019年9月10日;7(9):838.[9] T. M. Cover and J. A. Thomas, 信息论基础. 美国纽约州纽约市：Wiley-Interscience，2006年.[10] Hoeffding, Wassily (1963). “有界随机变量之和的概率不等式”（PDF）. 美国统计学会会刊. 58 (301): 13–30. doi:10.1080/01621459.1963.10500830