有错必究 *汉明码（Hamming Code）的原理及其应用*

万精油

上期的题目是帽子的颜色问题。为方便解答，我们把上期题目再列一遍。

帽子的颜色问题：三个人头上都被戴上一顶帽子。帽子的颜色是蓝色或红色，完全独立随机。每个人可以看见别人的帽子，但看不见自己的帽子。每个人可以有两种选择：猜自己帽子的颜色，或者放弃（就是不猜）。每个人把自己的决定写在一张纸上。如果最后的结果是至少一人猜对而且没人猜错，那么他们可以得到一笔巨额奖金。我们的问题是，他们用什么策略才能最大地提高得奖的概率。

这个问题二十年前曾经在美国数学界、计算机界轰动一时。不光因为它是一道趣味题目，而且因为这题目背后蕴藏着计算机编码理论中的一个重要思想。

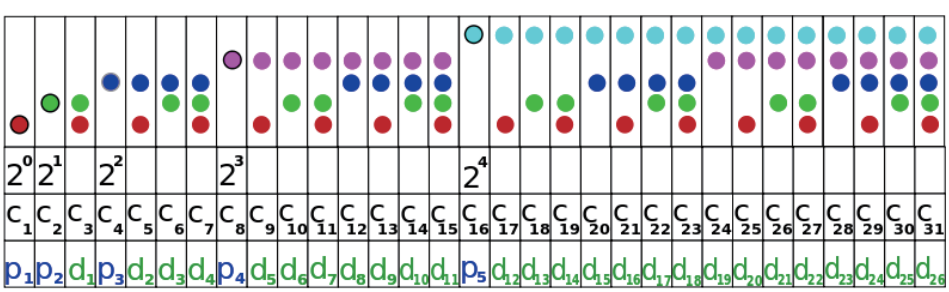
与别的问题不同，这个问题最困难的地方是只要有一个人错则全错。所以不能像别的题那样用数量来搞概率。

如果每个人都随机猜，那么三个人都猜对的可能性是八分之一。除此之外，好像没有什么别的出路。因为帽子都是随机选的，你头上的帽子颜色与别人的帽子颜色独立，似乎没有任何根据让你决定选什么颜色或放弃。其实不然，正因为帽子是随机选的（每个帽子都有二分之一的机会是红色，二分之一的机会蓝色），所以总体帽子的颜色满足一种分布。有些情况多一些，有些情况少一些。我们可以在这上面做文章。

先看三人的情况：三个人的帽子颜色一共有八种情况，红红红，红红蓝，红蓝红，红蓝蓝，蓝红红，蓝红蓝，蓝蓝红，蓝蓝蓝。如果大家商定，当某人看见两个同色的帽子时，他就猜另一种颜色，否则放弃。那么，根据上面的八种分布，我们很容易看出，有六种情况他们都能通过。只有两种情况他们会失败，即全红或全蓝的时候。再仔细数一数，他们答错和答对的时候一样多，都是六次。唯一的区别是，答错的时候大家都一起答错。而答对的时候都只有一人答对，别的人都放弃。

这个题目可以推广到更多人的情况。人数多的时候就不能靠一个情况一个情况地数，必须要有系统方法。这就需要介绍一种叫做汉明码的东西。

现在我们的生活都离不开网络，随时随地都在浏览从网上传来的东西。但是，网上的传递不能保证 100% 都对，经常会出现错误。计算机怎么发现传递有错误？发现了错误以



后又怎样纠正？汉明码就是用来干这个的。

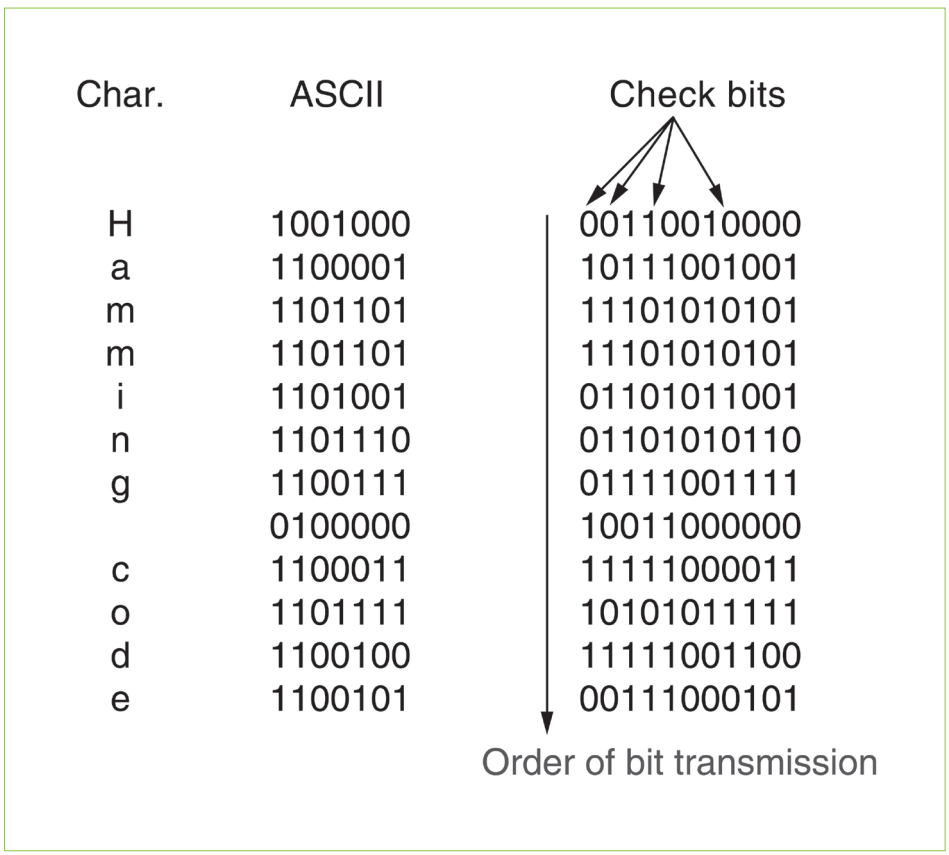
在介绍汉明码以前，先简单介绍一下如何用奇偶性来检查传递的信息是否出错。

如果我们有 8 个比特可以用。那么我们可以用其中的 7 个比特来传递信息，用一个比特来作验证码。如果那 7 个比特传递的信息有奇数个 1，验证码就是 1，否则就是 0。 这样一来，如果信息传递中有一个码出现错误，该是 1 的地方变成了 0，或者该是 0 的地方变成了 1，与这个验证码不符，我们就知道传递有错。这个方法的缺点是它虽然能发现错误，但不能知道错误出在哪里，也不能纠正，只能要求重新传递。汉明码是在用奇偶性来检查传递的信息是否出错的基础上发展出来的更高级的方法。它不但能发现错误，而且能知道错误出现在哪里，从而进行自我纠正。

要知道错误出现在哪里，一个验证码是不够的，汉明码需要用到多个验证码，具体个数根据能够传递的信息长度。假设有 $2^N - 1$ 个比特可以传递。那么我们用其中的 N 个比特来做验证码，剩下的 $2^N - 1 - N$ 个比特来传递信息。要用这 N 个验证码来发现错误并确定其位置，这 N 个码的设置就很有讲究。具体的方法我们用 $N = 4$ 的情况作一个说明。

$N = 4$ 时，我们有 15 个比特可以用。用其中 11 个来传递信息，4 个来做奇偶性验证码。我们假设这 4 个比特的位置是 1, 2, 4, 8。其余的 11 个比特就是真正要传递的信息。如果我们把这 11 个位置都用二进制表示，每个位置就有 4 个比特，我们把它们叫作位置比特，如“3”的位置比特为“0011”。

第一个验证码验证的是所有位置比特第一比特（自右数）是 1 的位置（其实就是所有



单数位置）的奇偶性。第二个验证码验证的是所有位置比特第二比特是 1 的位置（2, 3, 6, 7, 10, 11, 14, 15）的奇偶性。第三个验证码验证的是所有位置比特第三比特是 1 的位置（4, 5, 6, 7, 12, 13, 14, 15）的奇偶性。第四个验证码验证的是所有位置比特第四比特是 1 的位置（8 到 15）的奇偶性。比如，我们要传递的信息是 01100101110。把这个信息放到 15 个比特里，我们就有 _ _0_110_0101110。注意到在 1, 2, 4, 8 这四个位置上没有数，所以下面看位置比特时也不包括这些位置。所有单数位置的比特是 0100010，有偶数个 1，所以第一验证码是 0。位置比特第二比特是 1 的位置的比特是 0101010，有奇数个 1，所以第二验证码是 1。位置比特第三比特是 1 的位置的比特是 1101110，有奇数个 1，所以第三验证码是 1。位置比特第四比特是 1 的位置的比特是 0101110，有偶数个 1，所以第四验证码是 0。把这些验证码放进原信息，我们就得到全部 15 个传递比特，010111000101110。当然，我们的这种安排法主要是为了便于讲解。真正传递的信息没有必要把原信息打散重新组合。而是把原信息放在前 11 位，验证码放在后 4 位。

这样传递过去的码如果有验证码不符，比如传过去变成了 010111000111110，那么 1, 2, 4 验证码不符，我们就知道第 11 个码出了问题，可以把它纠正过来。

仔细想一想，我们意识到，这个方法只能发现有一个错码的时候。在极少数情况下，如果有两个比特同时出错，这个方法就没有办法发现了。于是人们又设计出再加一个验证码验证总体奇偶性，就可以发现有两个错码的时候。

现在再回到我们帽子颜色的题目中来。当总位数是 $2^N - 1$ 时，汉明码的一个特性是所有对错码覆盖了所有 $2^N - 1$ 位数。戴帽子的人可以利用这个特性来设计出一套猜帽子颜色的策略。还是拿 $N = 4$ 来举例。他们自己从 1 到 15 排一个号。当看见别人的帽子颜色以后，把这些颜色放在相应的比特位置上（红色为 1，蓝色为 0）。再把自己的两个颜色带进去得到两个码。如果两个码都是错码，则放弃。如果一对一错，则猜错的那个颜色。如果实际上所有的帽子颜色形成一个对码（1/16 的可能性），则所有人都猜错。如果实际上所有的帽子颜色形成一个错码（15/16 的可能性），则只有错码位置上那个人会猜，其余的人都放弃。所以，这个方法给猜帽子颜色的人 15/16 通过的概率。

这个问题可以推广到任意 N。N 越大，通过的概率越大，$(2^N - 1) / 2^N$。对于帽子数不是 $2^N - 1$ 的时候，不同的数有不同的解法，还没有通解。

本文只是对汉明码做个简单介绍，对汉明码有兴趣的读者可以找相关的书读一读。

下期题目：监狱里有 $2k$ 个犯人。监狱长把犯人找来，说：“你们的名字完全随机地放在这 $2k$ 个盒子里，每盒一个。明天你们轮流到这里来，每个人打开一个盒子，看看是不是自己，不是再开下一个，最多可以开 k 个，看到自己的名字就算通过。如果所有人都通过，就释放你们。现在你们可以讨论一个策略，完了之后不准再有任何形式的交流。”

每个人看到自己名字的概率是 1/2。如果没有策略，释放的概率是 $(1/2)^{2k}$，当 $k = 5$ 时，成功率已经低于千分之一，k 更大时就几乎成为不可能事件。能不能设计一个策略，使得全体被释放的概率有一个与 k 无关的正下界？

注：$2k$ 个盒子从左到右一字排开。每次被打开后马上关上，不能挪动。不能做任何记号。事先商量好对策后犯人之间不能有任何交流。