竞赛题解答提要

下面就竞赛题中涉及的数学原理做一下介绍:

一. 满足方程的随机布尔函数的概率分析

记Ω是所有k元布尔函数组成的样本空间, 事件A为满足Walsh谱点值的布尔函数f的集合, 事件B是满足线性方程组的布尔函数f的集合. 假设事件A, B相互独立, 即有$P(AB) = P(A)P(B)$.

我们首先考虑$P(A)$. 不失一般性, 可以设$\alpha = 0, \beta = e_0, \gamma = e_1, \delta = e_0 + e_1$. 对$y \in \mathbb{F}_2^k$, 按照$y$的最后两比特把$\mathbb{F}_2^k$等分成四个集合$A_i, 0 \le i \le 3$. 记

$$\tilde{p_i} = \frac{|\{y \in A_i | f(y) = 0\}| - |\{y \in A_i | f(y) = 1\}|}{|A_i|}.$$

则给定的Walsh谱点值等价于下面的方程组

$$\begin{cases} \tilde{p}_0 + \tilde{p}_1 + \tilde{p}_2 + \tilde{p}_3 = 4 \cdot \left(\frac{1}{8}\right) \\ \tilde{p}_0 - \tilde{p}_1 + \tilde{p}_2 - \tilde{p}_3 = 4 \cdot \left(\frac{1}{8}\right) \\ \tilde{p}_0 + \tilde{p}_1 - \tilde{p}_2 - \tilde{p}_3 = 4 \cdot \left(-\frac{1}{8}\right) \\ \tilde{p}_0 - \tilde{p}_1 - \tilde{p}_2 + \tilde{p}_3 = 4 \cdot \left(-\frac{1}{8}\right) \end{cases}.$$

易见, 方程的解为$\tilde{p_0} = \tilde{p_1} = \tilde{p_3} = 0, \tilde{p_2} = \frac{1}{2}$. 所以, 满足谱值条件的$f$的个数为$\binom{2^{k-2}}{2^{k-3}}^3 \binom{2^{k-2}}{2^{k-4}}$, 即$P(A) = \frac{\binom{2^{k-2}}{2^{k-3}}^3 \binom{2^{k-2}}{2^{k-4}}}{2^{2^k}}$.

代入$k = 12$, A中f的计数为

$$\binom{1024}{512}^3 \binom{1024}{256} \doteq 2^{3881.65}, \quad P(A) = \frac{2^{3881.65}}{2^{2^{12}}} \doteq 2^{-214.35}.$$

思考一下, 若题目中谱值条件改为$w_f(\alpha) = w_f(\beta) = w_f(\gamma) = w_f(\delta) = \frac{1}{8}$, 问题难度有无变化? 改为$w_f(\alpha) = w_f(\beta) = w_f(\gamma) = \frac{1}{8}, w_f(\delta) = -\frac{1}{8}$呢?

再考虑事件B的概率, 即存在$x \in \mathbb{F}_2^n$, 使得f满足$Ax \oplus f(\mathbb{F}_2^k) = b$的概率. 记$H_{A,b} = \{Ax \oplus b \mid x \in \mathbb{F}_2^n\}$为$\mathbb{F}_2^{2^k}$中的仿射子空间. 由于$A$列满秩, $|H_{A,b}| = 2^n$. $\mathbb{F}_2^{2^k}$中随机向量属于$H_{A,b}$中的概率为$P(B) = \frac{2^n}{2^{2^k}}$. 所以,随机选取的布尔函数$f$满足方程组的概率为

$$P(AB) = P(A)P(B) = \frac{\binom{2^{k-2}}{2^{k-3}}^3 \binom{2^{k-2}}{2^{k-4}}}{2^{2^k}} \cdot \frac{2^n}{2^{2^k}}$$

代入参数$n = 90, k = 12$, 得

$$P(AB) = P(A)P(B) \doteq 2^{-214.35} 2^{-4006} = 2^{-4220.35}.$$

即方程组除了给定真解外, 随机的假解的个数约为

$$P(AB) \cdot |\Omega| \doteq 2^{-124.35} \doteq 0.$$

即方程组存在真解外的随机解基本可以认为不可能事件.

二. LPN问题解的唯一性分析

由上面的分析中, 问题可以转化为含有$N = 1024$个方程的正确率为0.75的LPN问题的求解, 下面分析该问题解的唯一性问题.

(1)信息论角度的解的唯一性

对于伯努利随机变量ξ, 记其分布为$(p_0, p_1), p_0 + p_1 = 1$. ξ的信息熵定义为

$$H(\xi) = H(p_0) = -p_0 \log_2 p_0 - p_1 \log_2 p_1.$$

$H(\xi)$是随机变量ξ的不确定度的度量, 当$p_0 = p_1 = \frac{1}{2}$时, $H(\xi)$达到最大值1. 每个正确率为$\frac{3}{4}$的方程含有的信息为

$$1 - H\left(\frac{3}{4}\right) \doteq 0.1887.$$

1024个方程总共信息量为

$$1024 \cdot 0.1887 \doteq 193.25 > N = 90$$

故, 方程量能保证存在唯一解. 事实上, 从信息论的角度, 能保证唯一解的方程量约为

$$\frac{N}{1 - H\left(\frac{3}{4}\right)} \doteq 476.89,$$

即有477以上的方程量即存在唯一解.

(2)正态分布下解的唯一性

把每个方程看成$\mathbb{F}_2^n$上的仿射函数, 对于随机的$x \in \mathbb{F}_2^n$, 函数在x上的取值是分布为$(0.5, 0.5)$的伯努利均匀随机变量. N个i.i.d的均匀随机变量的和满足二项分布. 当N充分大时, 二项分布近似于正态分布. 故和变量随机到p_0N时, 对应的标准正态函数的T值为

$$T = \frac{(p_0 - 0.5)N}{\sqrt{\frac{N}{4}}} = (2p_0 - 1)\sqrt{N}.$$

代入$p_0 = 0.75, N = 2^{10}$得, $T = 2^4$. 而

$$1 - \Phi(16) \doteq 2^{-190} << 2^{-90}.$$

所以, 对于随机选取的$x \in \mathbb{F}_2^n$, 不可能使得1024个方程的正确率波动到0.75. 这说明了真解的唯一性. 事实上, 2^{-90}对应的T值约为10.9, 此时对应的方程量为

$$(2 \cdot 10.9)^2 \doteq 475.$$

这与上面通过信息论方法估计出的唯一解方程量477差距不大.

上面两种方法都可以用来讨论一般参数$(k, m = 2^k, n)$条件下, 方程解的唯一性问题.

三. LPN问题的求解算法

LPN问题就是编码理论中随机线性码的纠错解码问题, 该问题是NP完全问题. 一个LPN问题实例的困难性由三个参数:变元规模n, 方程量N和正确率p共同决定. 求解LNP问题实例的算法具有很多的开放性. 较常见求解LNP问题的算法有:

1. 穷尽法

穷尽变量x的2^n可能, 当n较大时就不现实. 但可以发展部分穷尽的思想.

2. 快速Walsh-Hadamad变换

3. BKW算法

通过扩张方程量方法做消元. 注意, 增加方程量的同时, 降低了方程的正确率.

4. 基于随机线性码的基内纠错算法

给定$\mathbb{F}_2^n$的一组基对应的方程, 从中纠正少量错误, 解出未知的x, 代入原方程验证是否为真解. 通过计算随机挑选$\mathbb{F}_2$上n阶方阵满秩的概率, 基内错误少于给定值的概率等理论问题, 可以估计出该算法的复杂度.