信息论的基本概念

自信息：

I(x)=−logP(x)

熵：

H(X)=∑−P(x)logP(x)

计算，不等式证明

相对熵

【定义 3.15】 （相对熵）

两个概率分布为 $p(x)$ 和 $q(x)$ 之间的相对熵定义为

$$
\begin{aligned}
D(p\|q) &= \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)} \\
&= \mathbf{E}_{p(x)}\log\frac{p(X)}{q(X)}.
\end{aligned}
$$

在上述定义中，我们约定 $0\log\frac{0}{0}=0$，$0\log\frac{0}{q}=0$，$p\log\frac{p}{0}=\infty$ (基于连续性)。因此，若存在字符 $x\in\mathcal{X}$ 使得 $p(x)>0$，$q(x)=0$，则有 $D(p\|q)=\infty$ 。

互信息

【定义 3.16】 （互信息）

考虑两个随机变量 X 和 Y，它们的联合概率密度函数为 $p(x,y)$，边际概率密度函数分别是 $p(x)$ 和 $p(y)$。互信息 $I(X;Y)$ 为联合分布 $p(x,y)$ 和乘积分布 $p(x)p(y)$ 之间的相对熵，即:

$$
\begin{aligned}
I(X;Y) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\
&= D(p(x,y)\|p(x)p(y)) \\
&= \mathbf{E}_{p(x,y)}\log\frac{p(X,Y)}{p(X)p(Y)}.
\end{aligned}
$$

证明互信息>=0（利用相对熵>=0）

$$I(X;Y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

$$= D\big(\underset{\substack{\Downarrow \\ p}}{p(x,y)} \,\|\, \underset{\substack{\Downarrow \\ q}}{p(x)p(y)}\big)$$

$$D(p\|q) = \sum_i p_i \log \frac{q_i}{p_i}$$

$$= -E\left(\log \frac{q_i}{p_i}\right)$$ 函数为下凸函数

Jensen不等式 $$\geq -\log\left(E \frac{q_i}{p_i}\right)$$

$$= -\log\left(\sum p_i \times \frac{q_i}{p_i}\right)$$

$$= 0$$

$$\therefore I(X;Y) \geq 0.$$

条件熵:

【定义 3.13】 （条件熵）

对于一对服从联合分布 $p(x,y)$ 的离散随机变量 (X,Y)，其条件熵 $H(Y|X)$ 定义为

$$\begin{aligned}
H(Y|X) &= \sum_{x\in\mathcal{X}} p(x) H(Y|X=x) \\
&= -\sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} p(y|x) \log p(y|x) \\
&= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log p(y|x) \\
&= -\mathbf{E}_{p(x,y)} \log p(Y|X).
\end{aligned}$$

如何证明条件熵不增?

用自信息>=0证明

$$I(X;Y)=\sum_x\sum_y p(x,y)\log\frac{p(x|y)}{p(x)} = \cancel{\sum}$$

$$= -\sum\sum p(x,y)\log p(x) - \left(-\sum_x\sum_y p(x,y)\log p(x|y)\right)$$

$$= -\sum p(x)\log p(x) - \left(-\sum p(x)\log p(x|y)\right)$$

$$= H(X)-H(X|Y) \qquad \because I(X;Y)\geq 0$$

$$= H(Y)-H(Y|X) = I(X;Y) \qquad \therefore H(X)\geq H(X|Y)$$

证明H(x)<log(X)

$$H(X)\leq \log|X|$$

$$H(X)-\log|X| = -\sum p(x)\log p(x) - \sum\frac{1}{|X|}\log\frac{1}{|X|} \qquad \because \frac{1}{|X|}\text{ 是 [illegible]}$$

$$= \sum p(x)\log p(x) - \sum p(x)\log\frac{1}{|X|} = -\sum p(x)\log\frac{p(x)}{\frac{1}{|X|}}$$

令 $q(x)=\frac{1}{|X|}$

$$= -D(p\|q) \quad \cancel{\geq 0}$$

$$\leq 0$$

得证

信源编码算法

kraft不等式

haffman编码

【例题 3.6】 设 X 是 $\mathcal{X}$ 上的随机变量，记 $\mathcal{X}=\{1,2,3,4,5\}$，其对应的概率分布为 $0.25, 0.25, 0.2, 0.15, 0.15$。图示如下：

1. 初始化 5 棵树。

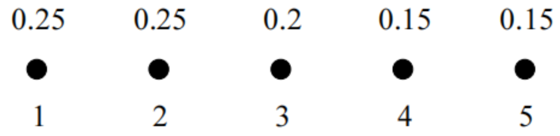
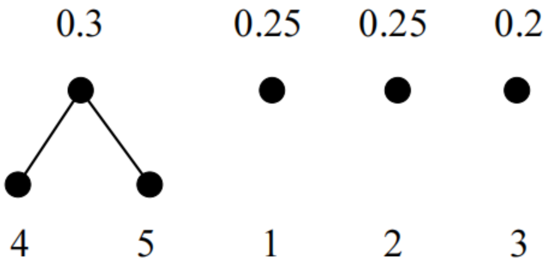


图 3.1

2. 将度量最小的两棵树合并后，得到 4 棵树。



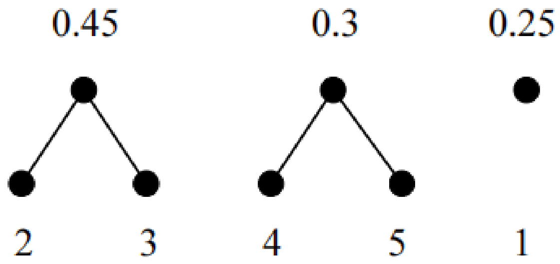
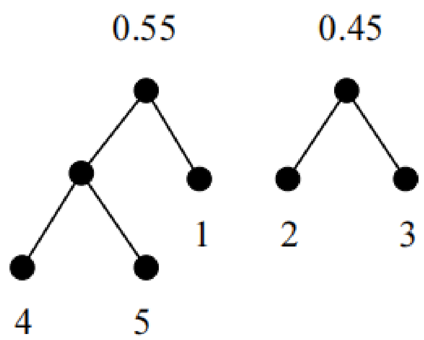


图 3.3

3. 继续上述过程，得到 3 棵树。

4. 继续重复，得到 2 棵树。



5. 当只剩 1 棵树时，退出循环，得到一棵二元树。从根节点开始，左分支标记为 0，右分支标记为 1。

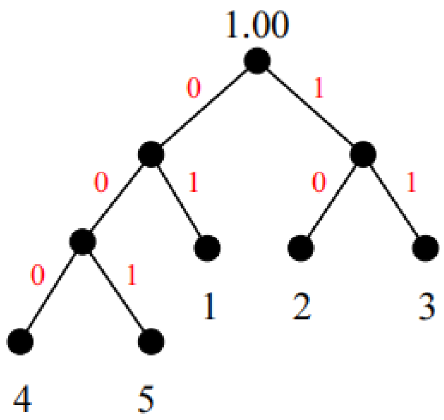


图 3.5

得到码字如下：

4：000

5：001

1：01

2：10

3：11

信道编码定理

信道模型

BSC，AWGN

判决准则MAP(极大后验），MLP（极大似然），

汉明距离

调制方式 2PSK,QPSK等

线性分组码

求生成矩阵，校验矩阵

重复码

奇偶校验码

hamming码

最小汉明距离

汉明码 $\ell\,[n,k]$ 信息位数　　$m=n-k$

分组长度 $n=2^m-1$

$G_{k\times(2^m-1)}$　　$G_{k\times n} \xrightarrow[\text{列置换}]{\text{初等行变换}} [I_k, P_{k\times m}]$

$H_{m\times(2^m-1)}$　　$H_{m\times n} \longrightarrow [-P^T, I_m]$

行宽为$m\times k$

最小汉明距离 $d_{min}=3$，纠 $\left\lfloor \frac{d_{min}-1}{2} \right\rfloor = 1$ 位错误

$\vec{u}G=\vec{v} \xrightarrow[\text{噪声}]{\text{channel}} \vec{r}$　　校验 $H\vec{r}^T=0$?

H的 $\forall\ d_{min}-1$ 列线性无关

且存在某 d_{min} 列线性相关

LDPC码

根据校验矩阵画正规图

=号为变量节点，+号为方程节点

一轮迭代的计算过程

节点输出概率

判决结果

硬判决：

[LDPC 比特翻转算法 - 知乎]

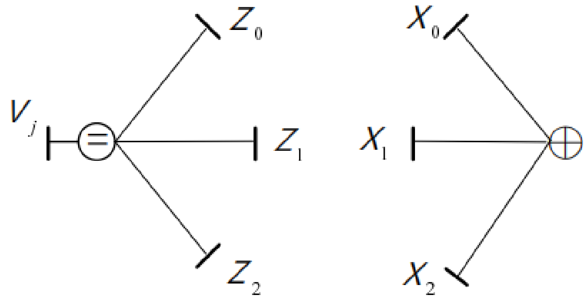
软判决：置信传播

为描述译码算法，我们把一个向量 $(v_0, v_1, \cdots, v_{n-1})$ 看作是随机向量 $(V_0, V_1, \cdots, V_{n-1})$ 的一个实现，每个分量对应一个"半边"。译码的问题就是（准确地或近似地）计算每个分量的概率分布 $P_{V_j}(v), v \in \mathbb{F}_2$。一般地，我们把每条边都看作一个"独立"的随机变量，而把一个节点看作一个约束条件，约束所有与其相连的边所代表的随机变量。在 LDPC 码的正规图表示中，有两类节点：$\ominus$ 与 $\oplus$，其中

$\ominus$：要求所有与其相连的边必须取得相同的值；

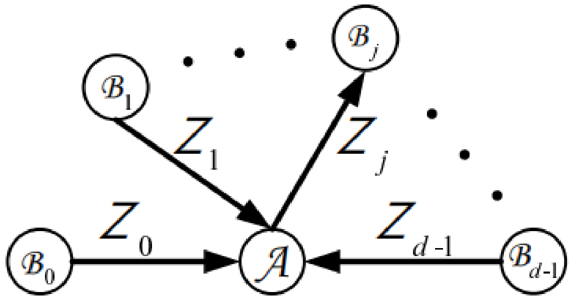
$\oplus$：要求所有与其相连的边必须相加等于 0。

我们看下面的例图，



在上面左图中，我们要求 $V_j = Z_0 = Z_1 = Z_2$，所以，可能的取值仅有两种，$(V_j, Z_0, Z_1, Z_2) = (0, 0, 0, 0)$ 或者 $(V_j, Z_0, Z_1, Z_2) = (1, 1, 1, 1)$。在上面右图中，我们要求 $X_0 + X_1 + X_2 = 0$，所以，可能的取值有四种，即 $(0, 0, 0), (1, 1, 0), (0, 1, 1), (1, 0, 1)$。

为了更清楚地描述译码算法，我们首先介绍在任意类型的节点上进行信息处理的一般规则。假设 $\mathcal{A}$ 是一个任意类型且度为 d 的节点，与它相连的节点为 $\mathcal{B}_0, \mathcal{B}_1, \cdots, \mathcal{B}_{d-1}$，对应的边是 d 个取值空间在 $\mathbb{F}_q$ 上的离散随机变量 $Z_j (0 \le j \le d-1)$，如下图所示。



假设所有输入的信息 $P_{Z_j}^{(\mathcal{B}_j\to\mathcal{A})}(z), z\in\mathbb{F}_q$ 都是可用的，则节点 $\mathcal{A}$ 可看作是一个信息处理器：任意给定变量 Z_j，流出节点 $\mathcal{A}$ 的信息可以通过计算下面的似然函数得到

$$P_{Z_j}^{(\mathcal{A}\to\mathcal{B}_j)}(z) \propto \Pr\{\mathcal{A}\text{ 的约束条件满足 }|Z_j=z\}, z\in\mathbb{F}_q.$$

由于上式的计算与输入信息 $P_{Z_j}^{(\mathcal{B}_j\to\mathcal{A})}(z)$ 是无关的，因此，在这种定义下，$P_{Z_j}^{(\mathcal{A}\to\mathcal{B}_j)}$ 即为所谓的“外信息”。

【例题 6.7】 考虑图 **??** 的两个图，设从其他子系统传向 $\ominus$ 节点的消息已知，例如，

$$P_{V_j}^{|\to\ominus}(0)=0.9,\ P_{V_j}^{|\to\ominus}(1)=0.1$$

$$P_{Z_0}^{|\to\ominus}(0)=0.8,\ P_{Z_0}^{|\to\ominus}(1)=0.2$$

$$P_{Z_1}^{|\to\ominus}(0)=0.4,\ P_{Z_1}^{|\to\ominus}(1)=0.6$$

$$P_{Z_2}^{|\to\ominus}(0)=0.7,\ P_{Z_2}^{|\to\ominus}(1)=0.3$$

则我们可以计算从 $\ominus$ 节点传向其他子系统的消息，比如

$$P_{Z_2}^{\ominus\to|}(0)\propto 0.9\cdot 0.8\cdot 0.4$$

$$P_{Z_2}^{\ominus\to|}(1)\propto 0.1\cdot 0.2\cdot 0.6$$

归一化之后，我们有

$$P_{Z_2}^{\ominus\to|}(0)=\frac{0.9\cdot 0.8\cdot 0.4}{0.9\cdot 0.8\cdot 0.4+0.1\cdot 0.2\cdot 0.6}=\frac{0.288}{0.3}=0.96$$

$$P_{Z_2}^{\ominus\to|}(1)=\frac{0.1\cdot 0.2\cdot 0.6}{0.9\cdot 0.8\cdot 0.4+0.1\cdot 0.2\cdot 0.6}=\frac{0.012}{0.3}=0.04$$

对于 $\oplus$ 节点，若

$$P_{X_0}^{|\to\oplus}(0)=0.9,\ P_{X_0}^{|\to\oplus}(1)=0.1$$

$$P_{X_1}^{|\to\oplus}(0)=0.2,\ P_{X_1}^{|\to\oplus}(1)=0.8$$

$$P_{X_2}^{|\to\oplus}(0)=0.3,\ P_{X_0}^{|\to\oplus}(1)=0.7$$

则我们在 $\oplus$ 节点可以计算，比如

$$P_{X_2}^{\oplus\to|}(0)\propto 0.9\cdot 0.2+0.1\cdot 0.8$$

$$P_{X_2}^{\oplus\to|}(1)\propto 0.9\cdot 0.8+0.2\cdot 0.1$$

归一化之后，有

$$P_{X_2}^{\oplus\to|}(0)=\frac{0.9\cdot 0.2+0.1\cdot 0.8}{0.9\cdot 0.2+0.1\cdot 0.8+0.9\cdot 0.8+0.2\cdot 0.1}=\frac{0.26}{1}=0.26$$

$$P_{X_2}^{\oplus\to|}(1)=\frac{0.9\cdot 0.8+0.2\cdot 0.1}{0.9\cdot 0.2+0.1\cdot 0.8+0.9\cdot 0.8+0.2\cdot 0.1}=\frac{0.74}{1}=0.74$$

给定一个 LDPC 码，假定 $P_{V_j}^{|\to\ominus}(v), 0\le j\le n-1$，已知（比如，可以结合信道规律 $P(y_j|v_j)$，由接收符号 y_j 进行计算），则 LDPC 码的译码算法可以概括如下：

0. 初始化：所有从 $\oplus$ 到 $\ominus$ 的信息初始化为 $P^{\oplus\to\ominus}(0)=\frac{1}{2}$。
1. 在每个 $\ominus$ 节点，计算 $P^{\ominus\to\oplus}$ 的所有外信息。
2. 在每个 $\oplus$ 节点，计算 $P^{\oplus\to\ominus}$ 的所有外信息。
3. 在每个 $\ominus$ 节点，计算 V_j 的总信息，其正比于 $P^{|\to\ominus}\cdot P^{\oplus\to\ominus}$，根据该信息进行判决，得到 $\hat{v}_j$。

若 $\mathbf{H}\hat{\mathbf{v}}^{\mathrm{T}}=\mathbf{0}$，则输出 $\hat{\mathbf{v}}$，译码结束。否则，重复上述 1，2，3 步，直到预设的最大迭代次数。若找不到合法的 $\hat{\mathbf{v}}$，宣告译码失败。

需说明的是，为了推导公式的便利，我们假定在正规图中节点之间传送的消息是概率质量函数。而在工程实际中，可以用任何别的等价量取代。特别地，对于二元变量，我们可以用 $\frac{p_0}{p_1}$ 或者 $\ln\frac{p_0}{p_1}$ 进行表示。消息的表示不同，节点的处理公式也不同。节点的作用就是接收消息，处理并输出消息。

卷积码 网格图与译码算法

画编码电路图

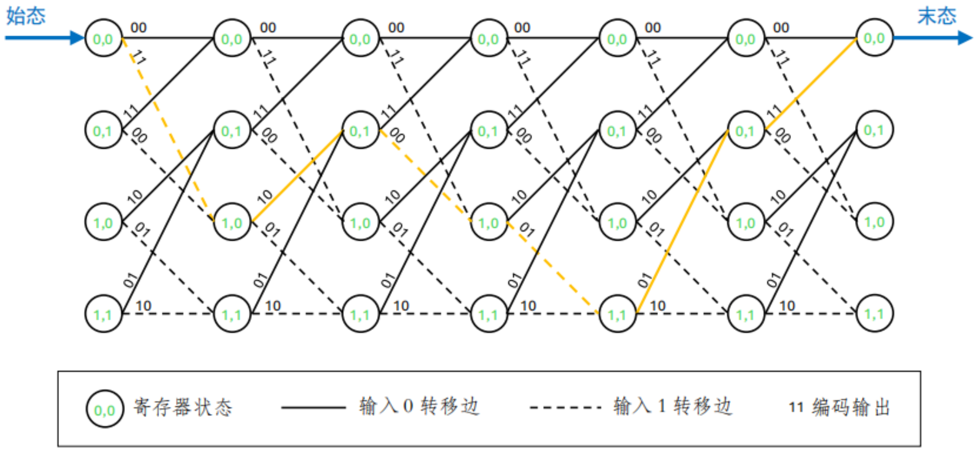
求寄存器状态，输出序列

维特比译码：

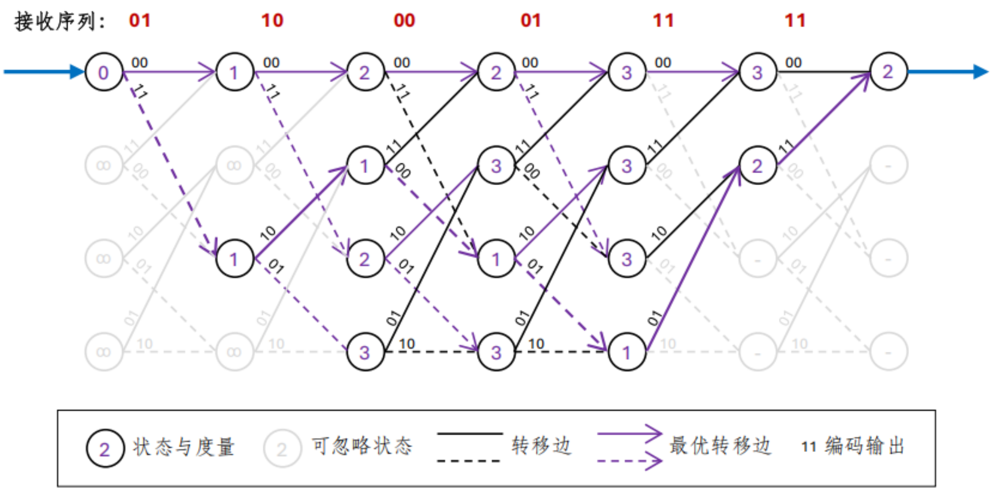
以卷积码为例，展示网格图（Trellis）与 Viterbi 算法。

设卷积码的生成阵为$G(D) = (1 + D + D^2, 1 + D^2)$，网格图与《讲义》图 7.7 一致。约定编码器的始末状态均为0,0，即被编码信息必须以连续 2 个 0 结尾。

假设信息长度为 6 比特，比如$\boldsymbol{u} = (1,0,1,1,0,0)$，对应下图橙色路径，则编码输出为$\boldsymbol{c} = (11,10,00,01,01,11)$，即依次输出边上的标记。



假设某个由上述编码器得到的输出$\boldsymbol{v}$，经过转移概率$p < 1/2$的二元对称信道，接收到了序列(01,10,00,01,11,11)，用 Viterbi 算法试求译码序列。



因为是二元对称信道，转移边的度量等价于[其边标记与接收符号的汉明距离]。最终，译码序列为(11,10,00,01,01,11)，恰好与上述编码输出相同。

Turbo码