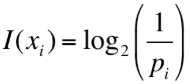
# 课件

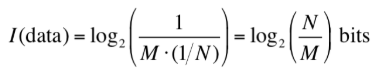
## 什么是信息

**交换、接收的数据。数据表明了特定事实或环境的不确定性。**

## 量化信息



## 数据传达的信息



## 例：信息内容

52张扑克中抽取一张为红心

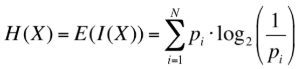
N=52 M=13 信息内容=log2(52/13)=2 bits

## 可能性与信息内容

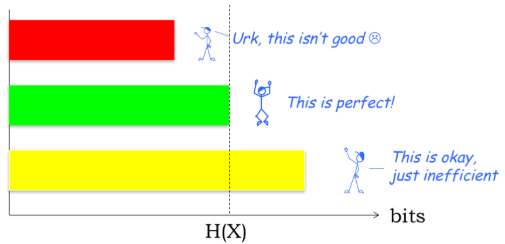


更多可能性，更多内容

## 熵

每组数据表示信息量的平均情况

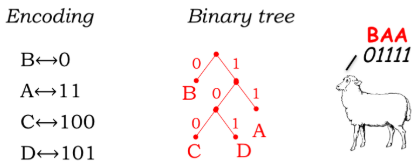
## 熵的意义

传递信息最恰当的位数

## 编码

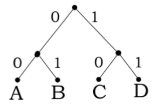
符号与二进制串的对应关系

## 编码如二叉树

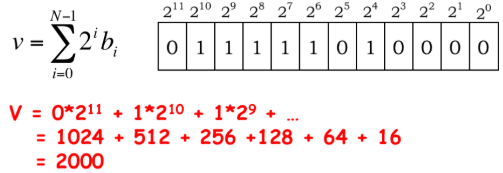


所有叶子节点为符号，所有节点间的树枝有标签0/1，从root到叶子节点的路径标签排列好，即符号对应的二进制串

## 固定长度编码

所有选择等可能，选择固定长度编码合适

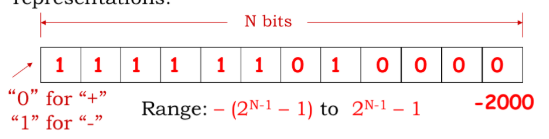
## 正整数编码



## 十六进制概念

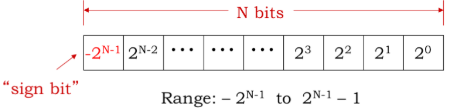
四位二进制合为一个十六进制，a:10，b:11，c:12，d:13，e:14，f:15

## 有符号数编码



10000000表示-0，00000000表示+0，11111111表示最小值-(2N-1-1)，01111111表示最大值2N-1-1

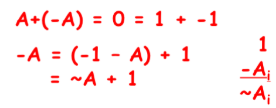
## 补码



00000000表示+0，10000000表示最小值-2N-1，01111111表示最大值2N-1-1

## 补码运算

补码加法即每个补码的二进制位相加

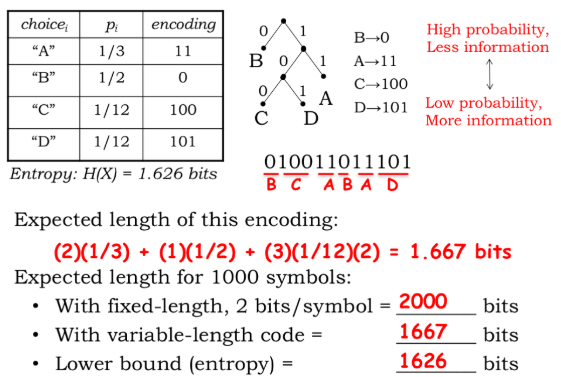


B-A可视作B+(-A)，-A可由~A+1得到

## 可变长度编码

出现次数多的用更少长度，出现次数少的用更多长度

## 例：可变长度编码



对于某一信息，用更少长度二进制串来表示

## 霍夫曼算法

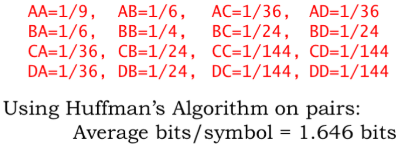
给定一组符号和它们的可能性，构建一个最优的变长编码。

用两个有最小pi的符号组成一个子树

每步选择两个最小pi的符号或子树，构成新子树

自底向上构建出最优树

## 可以做的更好

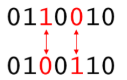


之前是对单符号进行编码，每个字符平均1.667位，可以对符号对进行编码，每个字符平均1.646位

## 错误检测和纠正

bob向alice发送0，传输途中被破坏为1，alice无法区分收到的1是破坏后的，还是想要传的

## 汉明距离



编码长度一致，同一位置处数字不相等的这种位置数量，称为汉明距离，上图为汉明距离为2

## 汉明距离和位错误

有效二进制字、发生单位（single-bit）错误的该字，汉明距离为1。

0、发生单位错误后为1，汉明距离为1

0、1，汉明距离为1

## 单位错误检测

为了信息传输中，发生单位错误后，不会变成另外一个有效字：

每个有效字后面追加一个奇偶校验位，偶数个1追加0，奇数个1追加1。

之前两个有效词之间汉明距离为1，加入奇偶校验位后，汉明距离变为2

## 奇偶校验=检测单位错误

加入奇偶校验位的字，所有1的计数为偶数，以此作为奇偶校验。

若传输中发生单位（奇数个位）错误，所有1的计数变为奇数。

若传输中发生偶数个位错误，无法通过奇偶校验来判断是否发生单位错误。

## 检测多位错误

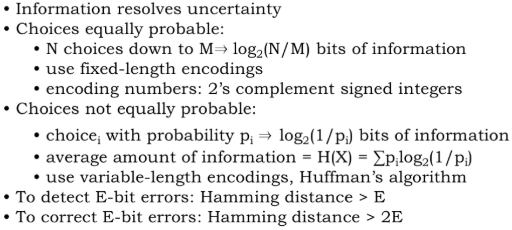
要想检测E位错误，要保证两个有效字之间的汉明距离至少位E+1。

## 单位错误修正

要想修正E个错误，要保证两个有效字之间的汉明距离至少为2E+1。

有效字之间汉明距离足够，保证可以错误检测和修正。

## 总结



# 习题

## 信息内容和熵

A：log2(8/3)

B：log2(3/2)

C：0.6\*log2(5/3)+0.4\*log2(5/2)

D：log25/2

E：log2256/8=5

F：log28/7

G：0.5\*log22+0.125\*log28+0.25\*log24+0.125\*log28

H：log22

I：log226/23

J：log216/7

## 二进制补码

A：101011

B：11001101，CD

C：11010110，-128+86=-42

D：9

E：10110011，01000111，0xFA

F：10110011，10110001，no

G：-77

H：-128

I：0x24，0x70

J：-16，15

K：-14

## 变长编码

A：A：1，B：01，C：000，D：001

B：2

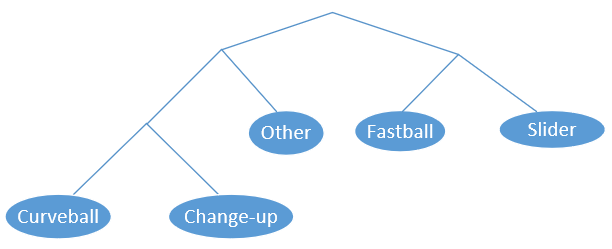
C：3

D：3

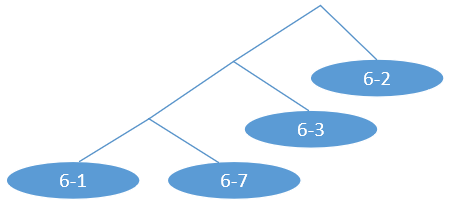
E：1

F：log2(1/0.66)

G：



H：



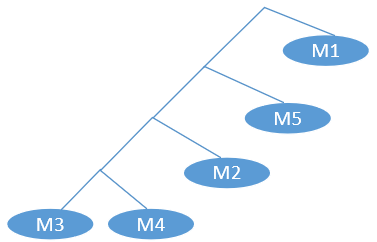
I：5

J：3

K：log2(100/29)

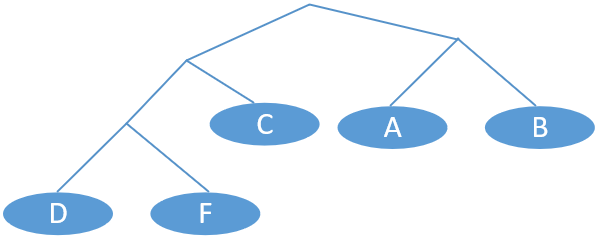
L：2

M：M1:1，M2:001，M3:0000，M4:0001，M5:01



0.6\*1+0.08\*3+0.02\*4+0.01\*4+0.29\*2

N：



O：D

## 错误检测和修正

A：1，0

B：01001，可能是Diamond发生错误导致，可能是Club发生错误导致

C：1，0

D：1

E：可检测不可修正

F：汉明距离为4，可检测3位错误，修正1位错误

G：4

H：无

I：a[3][1]错

J：a[0][0]错

K：a[3][3]错