



《人工智能的数据基础》

——信息编码

主讲教师：冯恺宇





目录

CONTENTS

2.1 二进制

2.2 基本数据表示



2.3 信息编码

2.4 信息论及其应用



“

什么是编码？

编码

用数字、字母等按规定的方法和位数来代表特定的信息。

目的：为了人与计算机之间信息交流和处理。

在计算机中要将数值、文字、图形、图像、声音等各种数据进行二进制编码才能存放到计算机中进行处理，编码的合理性影响到占用的存储空间和使用效率。

多媒体形式

人们将文本、音频、视频、图形、图像、动画的综合体统称为“多媒体”。由此可见，多媒体信息包括以下5种：

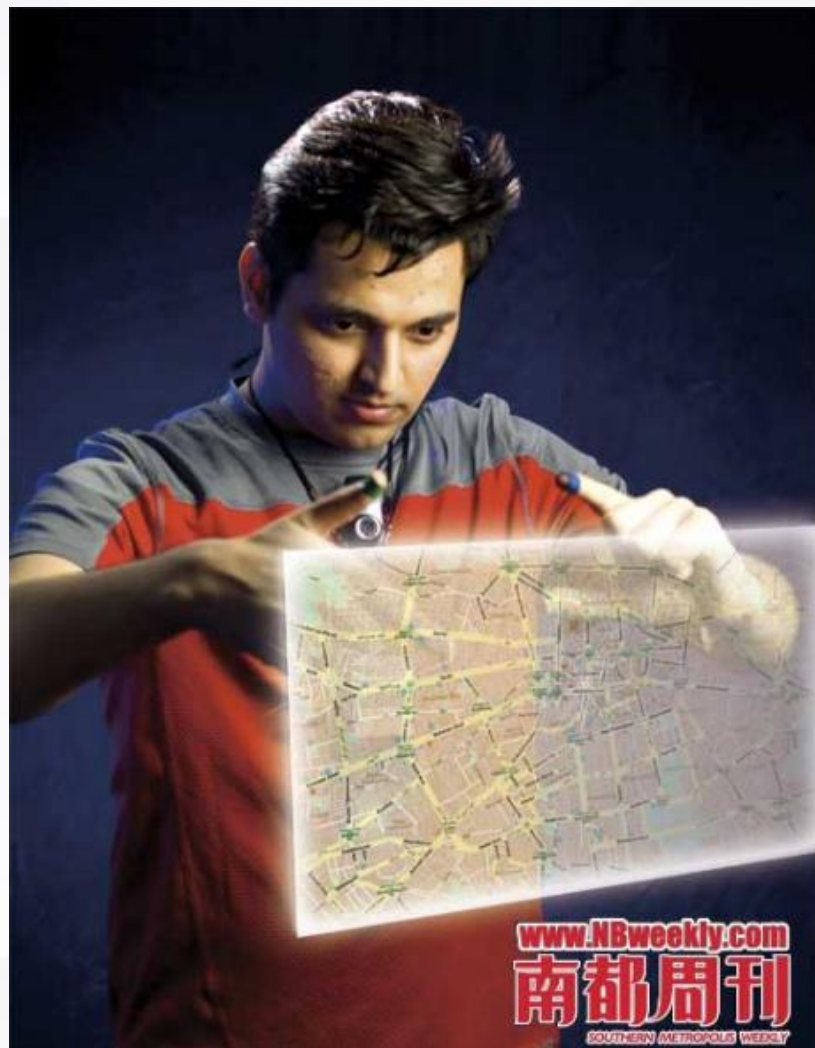
- 文本
- 图形、图像
- 动画
- 声音
- 视频影像



2011最震撼的个人多媒体系统——“第六感”装置



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY



成本: \$350

“第六感装置”是一个可佩戴的姿态交互多媒体系统，它将用户周围的事物，与互联网无缝连接在一起，用户可以通过自然手势与电子设备进行输入输出交互。



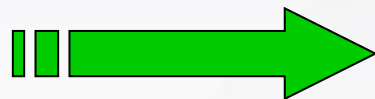


媒体类型与人类感官

- 人类感官的收集的信息比例
视觉：65%;听觉：20%;触觉：10%;味觉：2%
- 心理学研究结果：在进行面谈交流时，词语表达信息占7%，语调占38%，视觉占55%。
- 用鼻子闻 → 用手摸摸 → 听 → 看文字 → 看图片 → 看视频 → 看3D电影 → 看4D电影 → ?



谁获得的信息多？



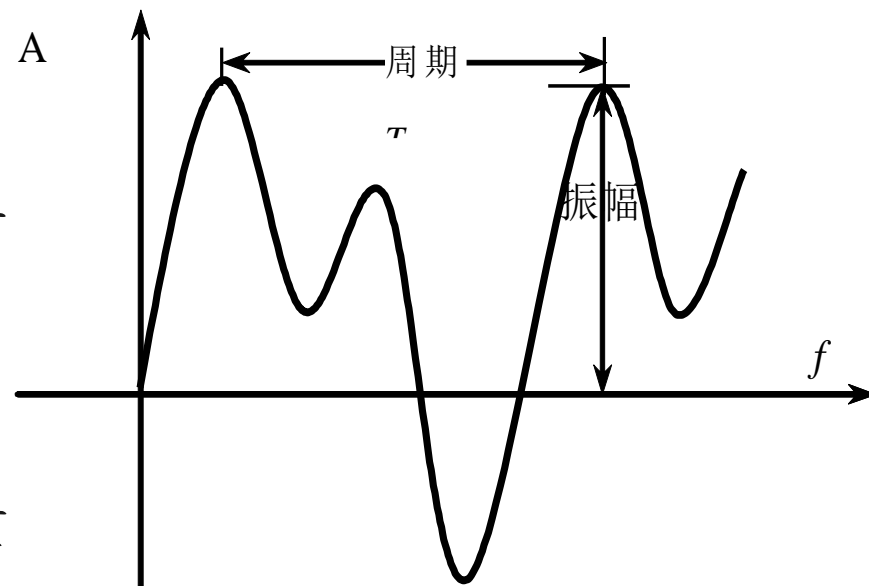
基本概念

声波：声源体发生振动会引起四周空气振荡，振荡方式就是声波。复杂的声波由许许多多具有不同振幅和频率的正弦波组成

周期 T ：重复出现的时间间隔

振幅 A ：波形相对基线的最大位移，表示音量的大小

频率 f ：信号每秒钟变化的次数，即 $1/T$ 以赫兹(Hz)为单位



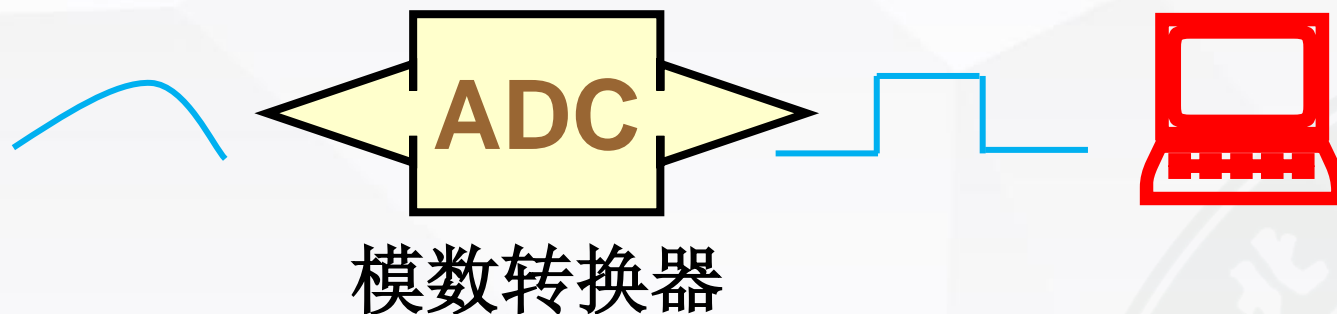
声波示意



模拟声音信号的数字化



为什么要进行转换？

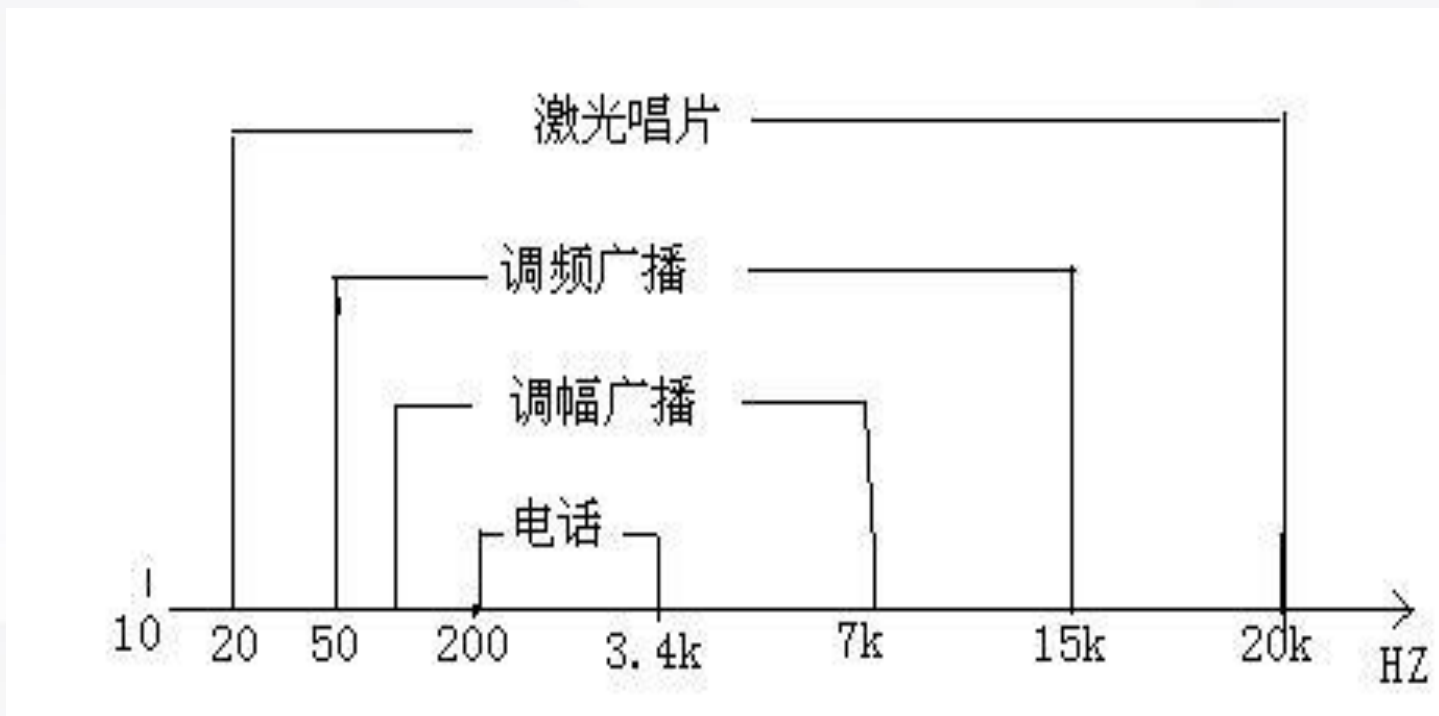


计算机处理音频信号时，必须将模拟信号转换为数字信息，这个过程由模数转换器**ADC (Analog to Digital Converter)**完成



多媒体各种类型表示—按音频带宽分布分类声音

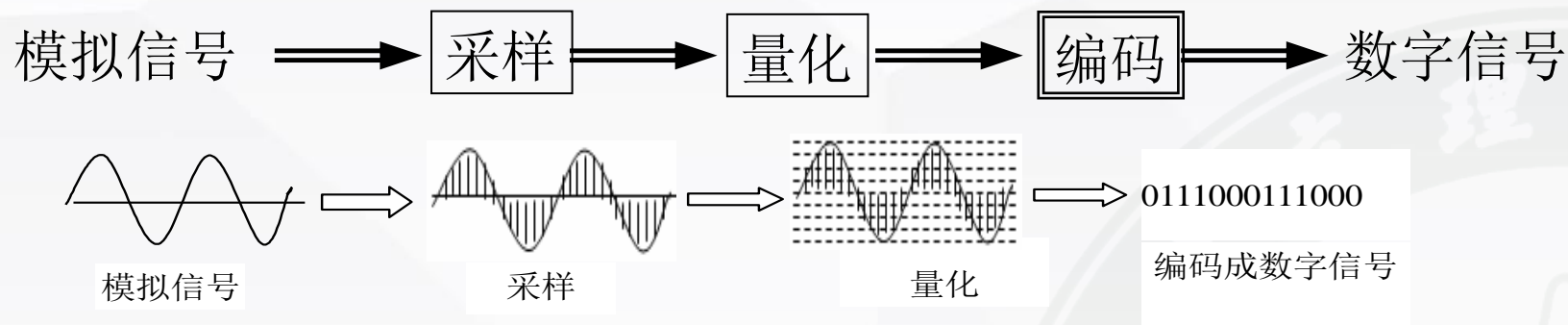
- 人的听觉频率范围为20—20K赫兹，超过这个范围的音频人耳无法听到
- 占用带宽越大，声音越丰富





模拟音频的数字化

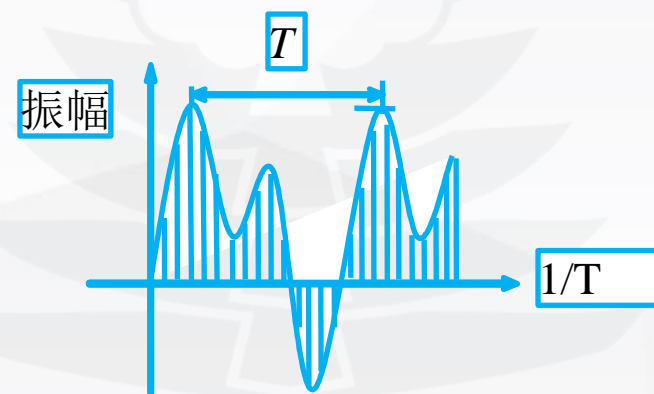
用计算机对音频信息处理，就要将模拟信号（如语音、音乐）转换成数字信号。



采样 每隔一定时间间隔对模拟波形上取一个幅度值

量化 将每个采样点得到的幅度值以数字存储

编码 将采样和量化后的数字数据以一定的格式记录下来





模拟音频的数字化

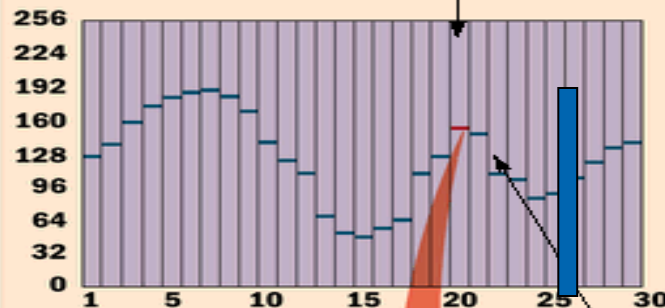


Sampling a sound wave.

连续的声音波形信号



对一秒钟的声音波形采样30次



Sample	Sample Height (Decimal)	Sample Height (Binary)
1	130	10000010
2	140	10001100
3	162	10100010
4	175	10101111
5	185	10111001
6	188	10111100
7	195	10111111
8	200	11001000
9	190	10111110
10	170	10101110
11	150	10011110
12	120	01111100
13	68	01000100
14	50	00110010
15	45	00101101
16	60	00111100
17	66	01000010
18	115	01110011
19	129	10000001
20	155	10011011
21	153	10011001
22	115	01110011

此点高度量化，即转换为二进制数
10011011

要保存此1秒钟声音，保存30个二进制数即可



数字音频的技术指标



采样频率、量化位数和声道数

采样频率： 每秒钟的采样次数

量化位数(采样精度)： 存放采样点振幅值的二进制位数。通常量化位数有8位、16位，分别表示有 2^8 、 2^{16} 个等级。

声道数： 声音通道的个数，立体声为双声道。

每秒钟存储声音容量的公式为：

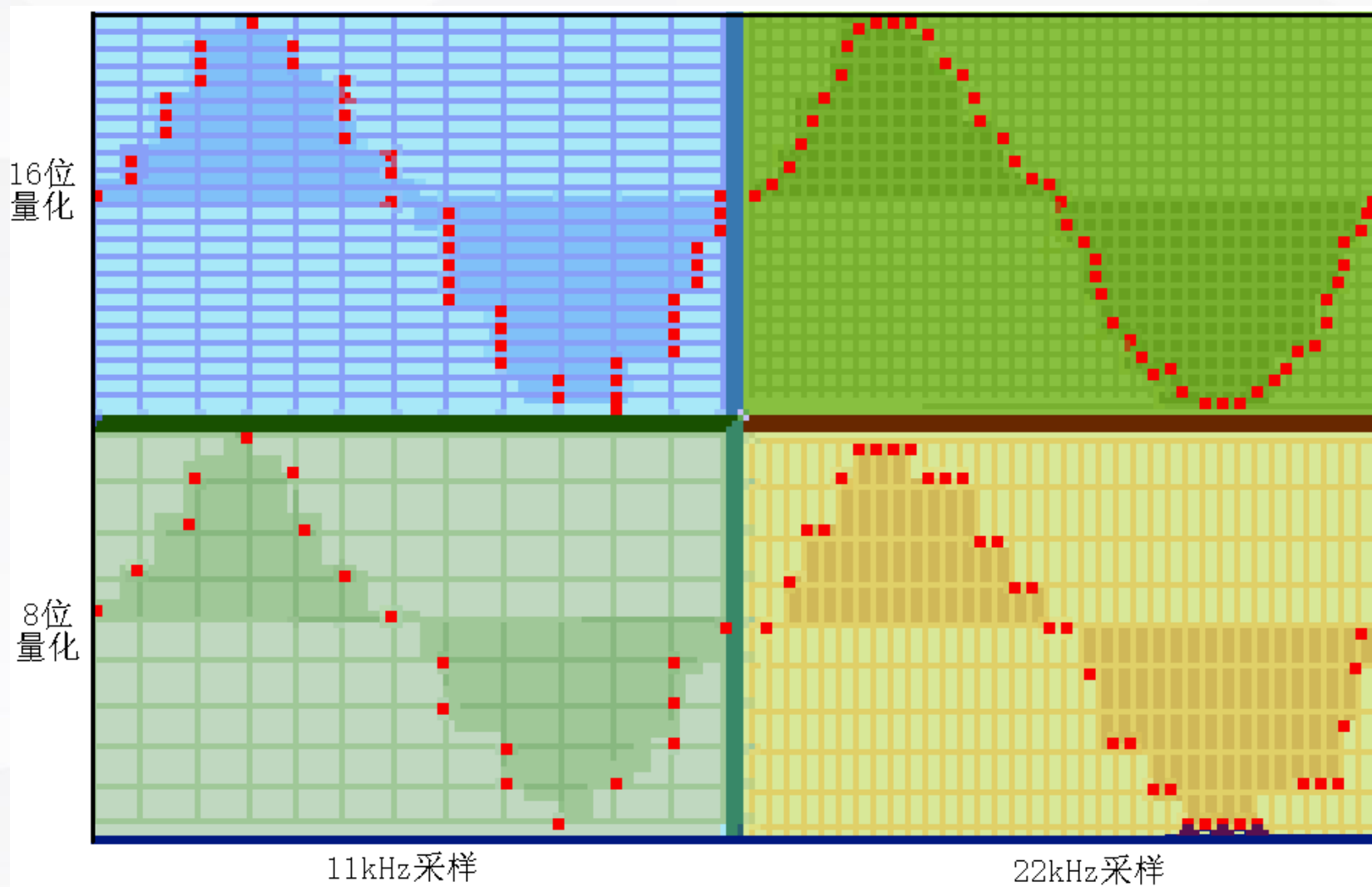
$$\text{采样频率} \times \text{采样精度} \times \text{声道数} / 8 = \text{字节数}$$

例如，用44.10 kHz的采样频率，16位的精度存储，则录制1秒钟的立体声节目，其WAV文件所需的存储量为：

$$44\ 100 \times 16 \times 2 / 8 = 176\ 400 \text{ (字节)}$$



采样频率和量化参数比较



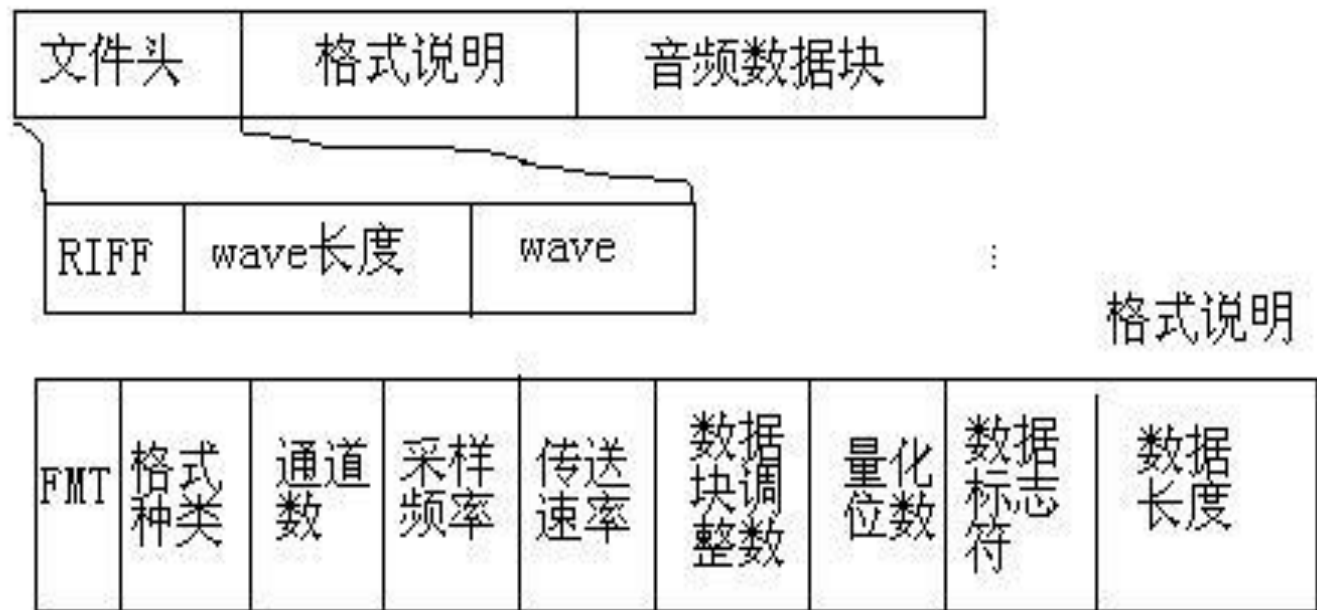


多媒体各种类型表示—常见音频文件格式

WAV音频文件格式

音频数据块保存PCM编码数据，无压缩。

若是单声道，0代表左，1代表右；双声道声音数据交替存放。





常见音频文件格式

MIDI（电子乐器数字接口）是用于在音乐合成器、电子乐器、音序器和计算机之间交换音乐信息的一种标准协议。

MIDI音乐是使用微处理和数字信号处理器代替发声部件，模拟出声音波形数据，然后将这些数据通过数模转换器转换成音频信号并发送到放大器，**合成**出声音或音乐。

由于MIDI只是记录**音乐信息的数字代码如984526**，所以生成的文件比较小，便于传播，也便于编辑修改。

MIDI音乐常作为背景音乐。



layer 3的典型数据
压缩率在1:10至
1:12左右。编码速
率为128kbps的
MP3就**相当于**CD
的音质，按此编码
速率，一分钟CD音
质的MP3文件在
1M左右



MP3文件格式

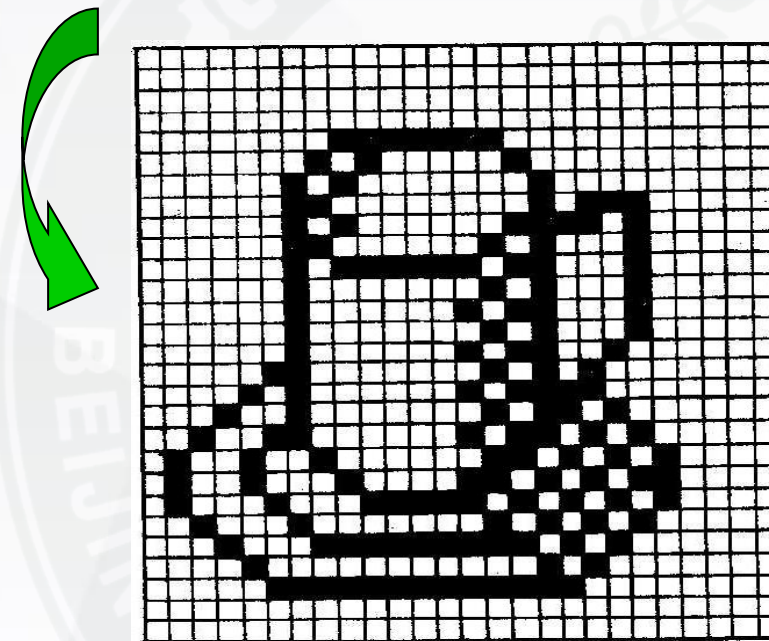
全称MPEG1 layer3，是MPEG-1（VCD）标准中的声音技术



多媒体各种类型表示—图像



- 图像一般是指是由扫描仪、摄像机等**输入设备**捕捉实际的画面产生的数字图像，是由像素点阵构成的位图
- 静止的图像是一个矩阵，由一些排成行列的点组成，这些点称为像素点（pixel），这种图像称为**位图**（bitmap），如文件名“Sample.bmp”。





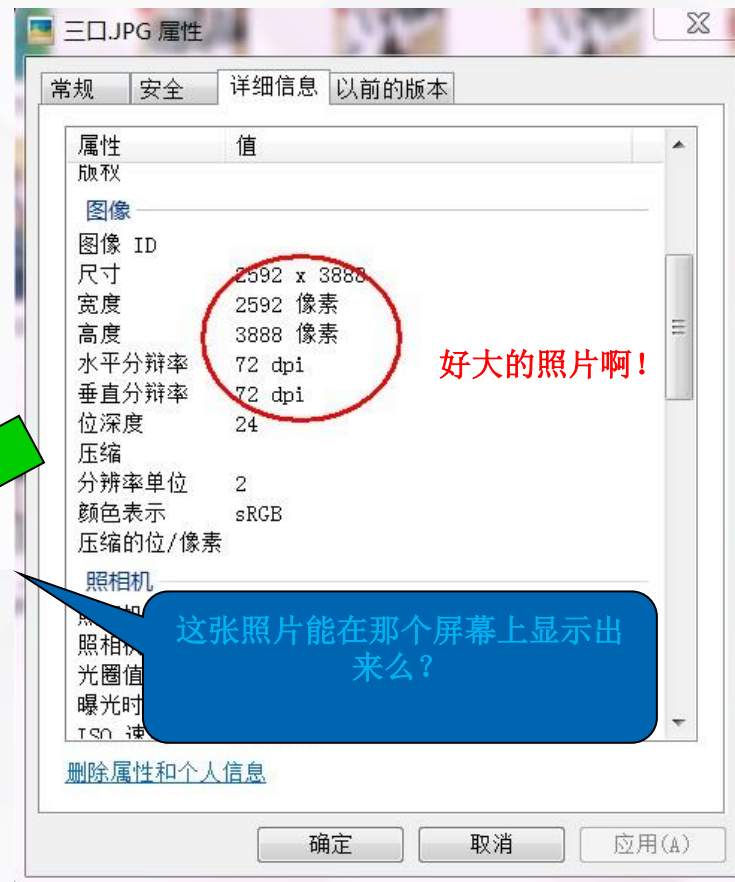
多媒体各种类型表示—图像



- 图像的主要参数—分辨率
- **屏幕分辨率**: 计算机显示器屏幕物理点数
- **图像分辨率**: 数字化图像的大小(尺寸)



这个屏幕的分辨率是多少？





多媒体各种类型表示—图像



图像的颜色

- 二值图像
 - 黑白双色 (0或1)
- 灰度图像
 - $2^8=256$ 种黑白变化色
- 彩色图像
 - 采用RGB彩色空间 (R,G,B)
 - 一般写成 (0~255, 0~255, 0~255)

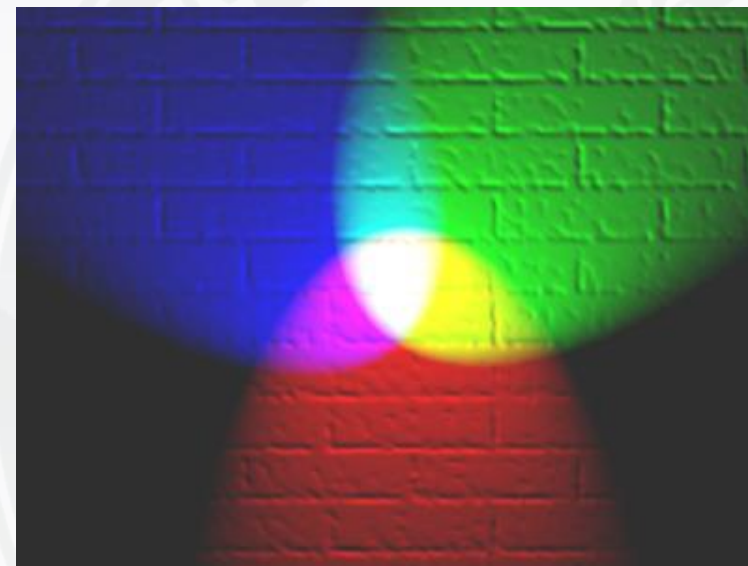


注意：平常所说的“黑白照片”实际是灰度图像，不是二值图像



多媒体各种类型表示—图像的颜色表示RGB

- 颜色深度（位数）
 - 在计算机中用多少位来存放RGB颜色值
 - 可以是：8bit, 16bit, 24bit, 32bit, 48bit, …….
 - 三个8位无符号整数为24bit，可以产生16777216种颜色，而人眼只能识别1000万种颜色
 - 超过24bit的颜色称为真彩色
- 计算图像大小
 - 照片： $2592 \times 3888 \times 24 = ? \text{ b}$
 - 数据太大了，怎么办？



图像的分辨率像素位的颜色深度决定了像文件的大小
计算公式为：

列数×行数×颜色深度÷8=图像
字节数

例8.2

当要表示一个分辨率为640×480的“24位真彩色”图像，则需要：

$$640 \times 480 \times 24 \div 8 \approx 1 \text{ MB}$$

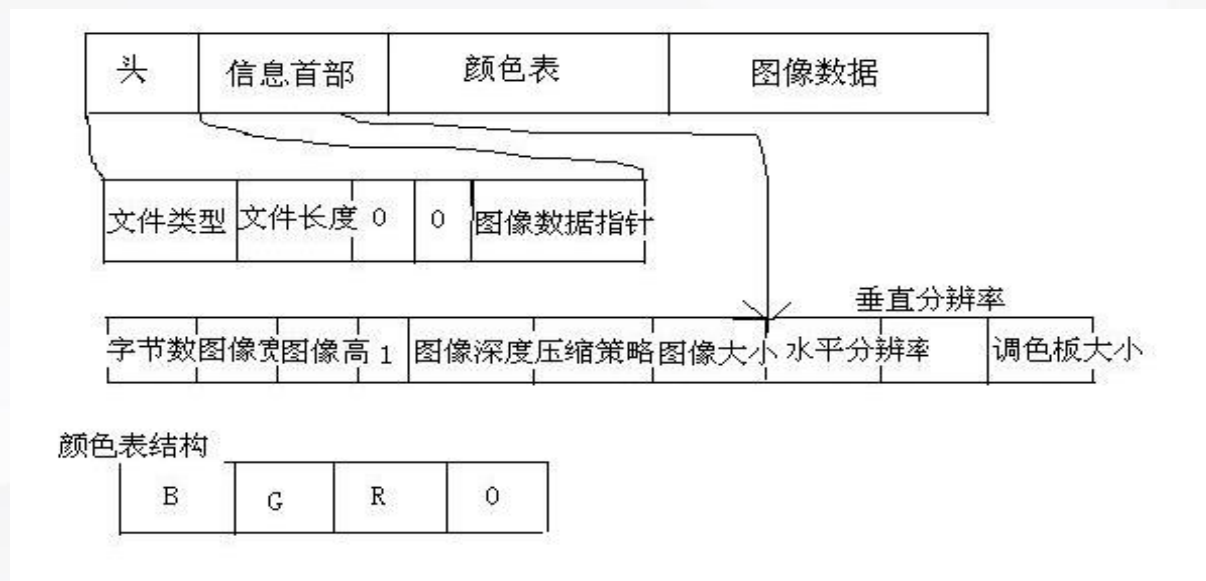
由此可见，数字化后的图像数据量十分巨大，必须采用编码技术来压缩信息。它是图像传输与存储的关键。



多媒体各种类型表示—bmp图像的格式

- 位图bmp是Windows采用的图形文件格式，在Windows环境下的所有图象处理软件都支持BMP文件格式。
- 计算图像文件大小时需要加上文件头和信息头数据。

位图文件头（14字节）	
位图信息	位图信息头（40字节）
	调色板数据（8、16或1024字节） 真彩色图像无调色板
像素数据	



- BMP和DIB格式文件

与设备无关的位图格式文件，Windows环境中经常使用。

- GIF格式文件

Internet上的重要文件格式之一，最大不超过64 KB，256色以内，压缩比较高，与设备无关。

- JPEG格式文件（.JPG）

利用JPEG方法压缩，Internet上重要文件格式之一，适用于处理256色以上、大幅面图像。

- WMF格式文件

位图与矢量图的混合体，Windows中许多剪贴画图像是以该格式存储的。广泛应用于桌面出版印刷领域。

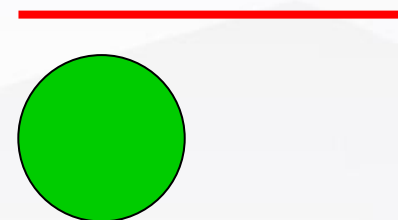


图形

- 一般指用计算机绘制的画面，如直线、圆、圆弧、矩形、任意曲线和图表等。
- 图形的格式是一组描述点、线、面等几何图形的大小、形状及其位置、维数的指令集合。
- 在图形文件中只记录生成图的算法和图上的某些特征点。
- 图形文件也称矢量图。

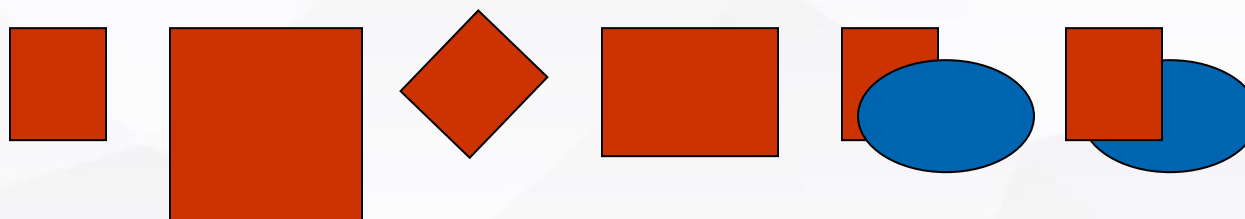
— line (x1,y1,x2,y2,color)

— circle (x,y,r,color)



图形的特征

- 图形是对图像进行抽象的结果（人工或自动）；
- 图形的矢量化使得有可能对图中的各个部分分别进行控制（放大、缩小、旋转、变形、扭曲、移位等）
- 图形的显示需要计算时间（CPU参与计算）



图形与图像的区别

- 图形是矢量概念，基本元素是图元；图像是位图概念，基本元素是像素；
- 图形可以任意缩放无失真；图像缩放有失真；
- 图形以图元为单位修改属性、编辑；图像只能对像素或图块处理；
- 图形文件比较小，图像文件比较大
- 图形适合描述轮廓不很复杂，色彩不是很丰富的对象如几何图形，工程图纸等；图像适合表现含有大量细节（如明暗变化、场景复杂、轮廓色彩丰富）的对象，如：照片、绘图等。

视频

视频源于电视技术，它由连续的画面组成。这些画面以一定的速率连续地投射在屏幕上，使观察者具有图像连续运动的感觉（错觉）

眼睛欺骗了你！

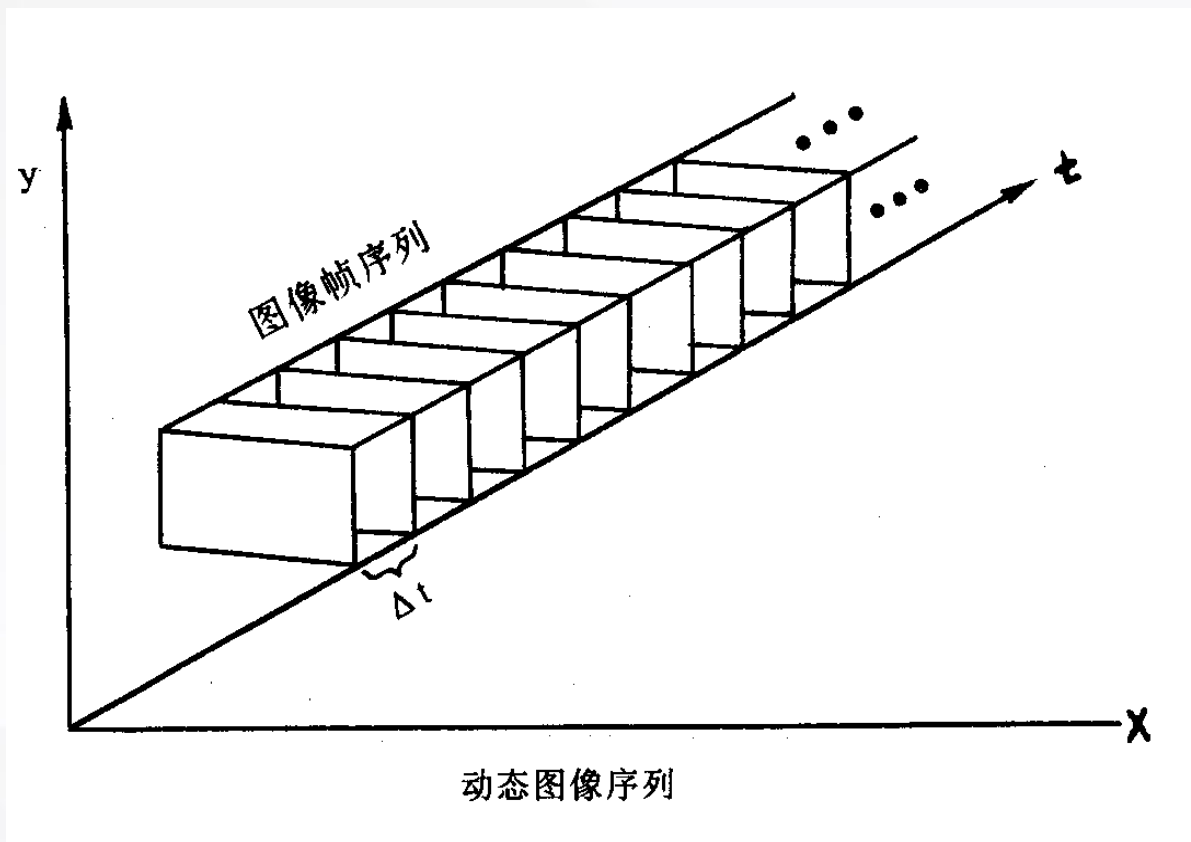


视觉生理与视觉心理规律

- 视觉调节力：视距调节、亮度范围调节。
- 视觉暂留性：视觉惰性维持0.02-0.2秒。视觉暂留性是近代电影、电视、计算机动画的基础。
- 视觉锐度-视力：眼睛分辨景物细节的能力。亮度与对比度影响视力。人眼对彩色的分辨率低于对亮度的分辨率。
- 视觉范围与对比灵敏度：人眼所能感受到的亮度范围。
- 注视点：注视点一般集中在高频部分。
- 适应性：亮适应规律：亮适应一数秒，暗适应一数分钟。颜色变化的眼适应力较差。



动态图像由多个静态图像组成的





严谨的定义和分类



平常所说的“实景动画”，“粘土动画”实际上属于视频

- 动态图像

由多幅连续的图像构成的序列称为动态图像。它利用了人眼的视觉暂留性，沿时间轴每一幅图像保持 $1/20-1/30$ 秒，顺序地在感觉不到的速度下更换另一幅图像，连续不断，就形成了运动图像的感觉。每一幅图象称做一帧。

- 分类

视频：每一帧为实时获取的自然景物时。

动画：每一帧为计算机产生或人工画出的图形。

实际工作中，由计算机生成的特技镜头在制作**单帧**时是图形，但在动态播放时CPU根本无法快速计算和重绘。所以制成影片时都将图形转换成**所需清晰度**的图像。

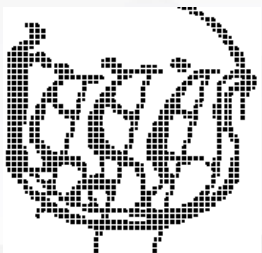
为什么不能有动态图形？



“

动态图像的特点

- 时间连续性：适合于表示过程，具有丰富、自然、生动的自然表现力。
- **帧间相关性**：统计表明，邻帧之间有10%以下的象素有亮度变化，1%以下的象素有色度变化。相关性是动态图像连续动作形成的基础，也是进行压缩等处理的基本条件。
- 帧速：实时性要求高，**16fps** (frame per second) 可以达到满意的动态效果。
- 数据量：每帧数据量 \times 帧速。压缩后数据量变化很大，如MPEG可压缩上百倍。且压缩后的码率是不均匀的。
- 视频质量：压缩比过大会影响回放视频的质量。视频质量的度量使用MOS（主观意见）打分法。



图像的数字化

分辨率（行、列）和颜色深度：

真彩色每像素点占3个字节，
 $2^{24} = 16777216$ 种颜色

计算存储一张图像公式：

列数 \times 行数 \times 像素的颜色深度 / 8 = 字节数

视频的数字化

计算存储一秒图像公式：

列数 \times 行数 \times 像素的颜色深度 / 8 \times 帧 / 秒 = 字节数

例：1280 \times 1024分辨率的“真彩色”电视图像，
按每秒30帧计算，显示1分钟，则需要：

1280 \times 1024 \times 3 \times 30 \times 60 \approx 6.6 GB



特征编码



- 在人工智能的处理过程中，往往需要将离散属性数据转化为可供模型使用的数值表示。
 - 学校（文字）
 - 学号（数字）
 - 成绩类别（ABCD）
- 分类特征
 - 离散
 - 非连续



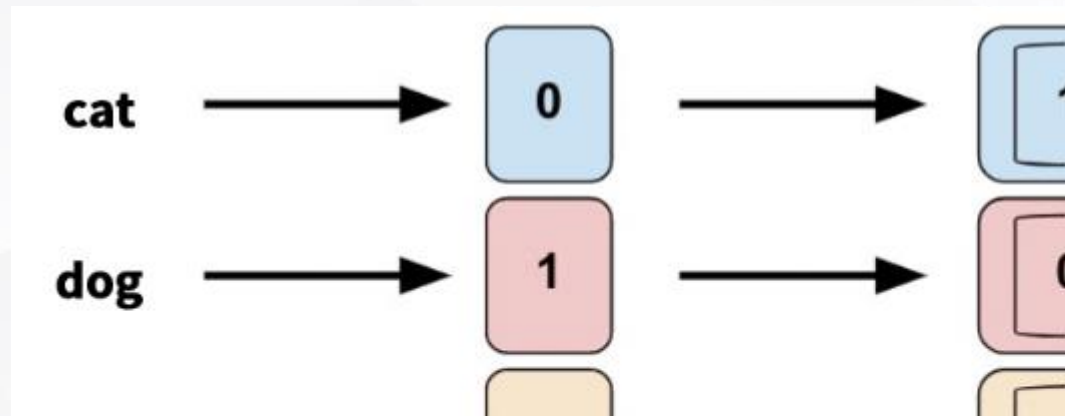
小型分类特征



- 序列编码

适用于类别间本来就有一定的排序关系。

例如：学士-0，硕士-1，博士-2



- 独热编码

处理类别间不具有大小关系的特征

猫表示为 $(1, 0, 0, 0)$ ，狗表示为 $(0, 1, 0, 0)$ ，老鼠表示为 $(0, 0, 1, 0)$ ，猪表示为 $(0, 0, 0, 1)$

- 二进制编码

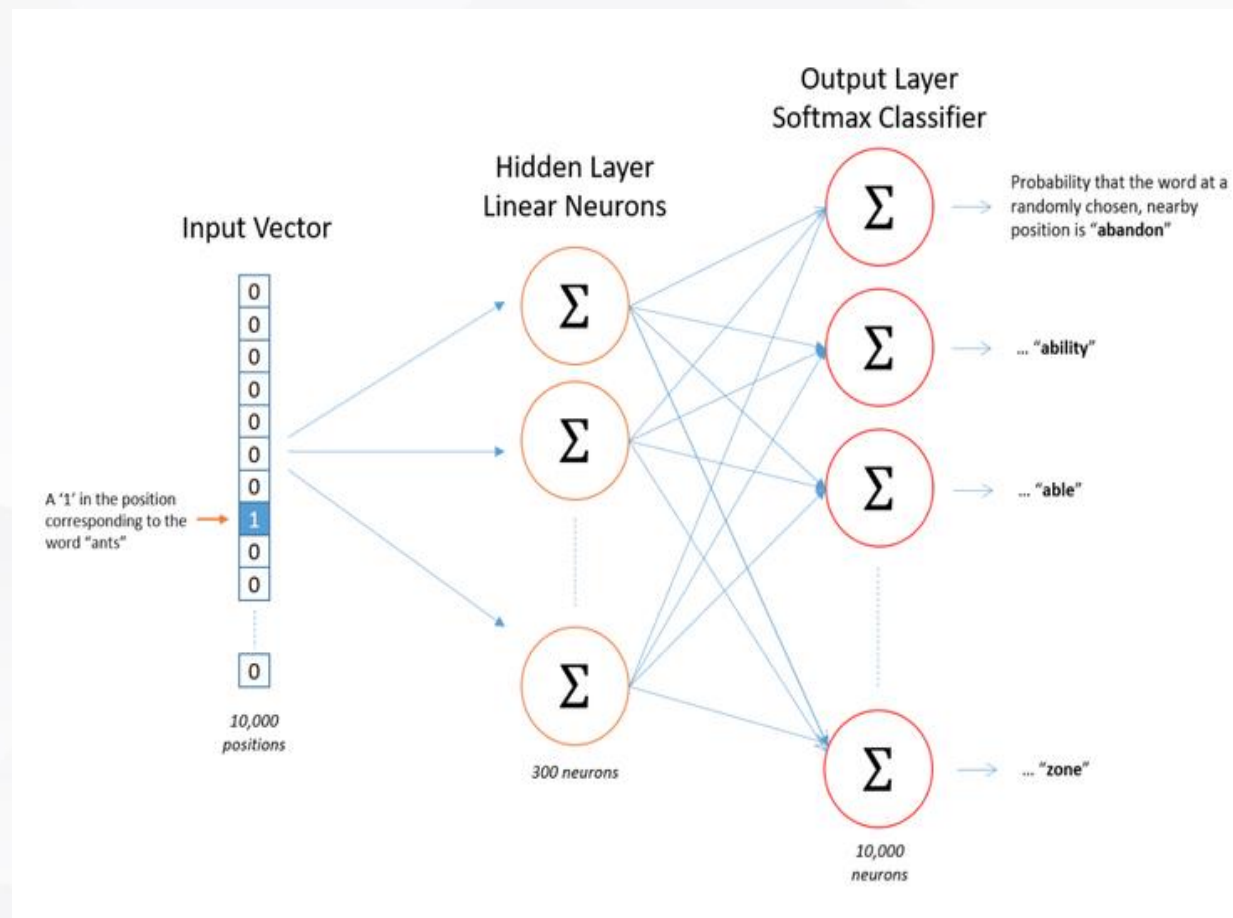


— 独热编码应用

One-hot encoding

	cat	mat	on	sat	the
the =>	0	0	0	0	1
cat =>	1	0	0	0	0
sat =>	0	0	0	1	0
...

知乎 @733TTT





大型分类特征



- 哈希编码
 - 一种将数据转换为固定长度的数字值的方法
 - 将大量的数据压缩为较小的数字值
 - 同时保持数据的唯一性
- 作用
 - 数据验证：防止篡改
 - 提高效率：快速查找