

第一章 信息论历史与概率论 (第一部分)

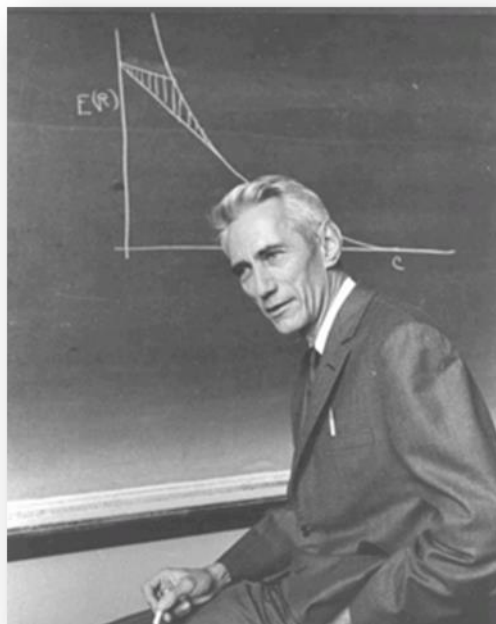
授课教师：樊平毅教授
清华大学电子工程系



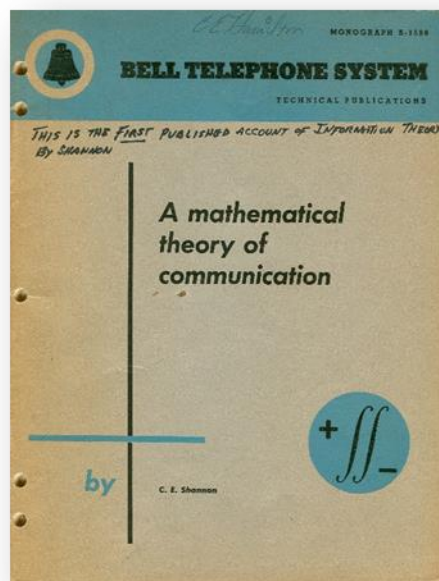
2021年9月20日

1948年

“A mathematical theory of communication,” Bell System Technical Journal. 27 (3): 379-423, 1948.



Claude Elwood Shannon
(1916-2001)



面向的通信技术的基本工程问题

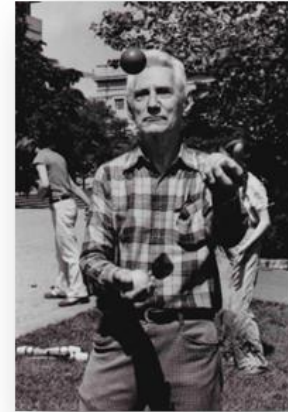
1. 信源编码定理
2. 信道编码定理
3. 联合信道信源编码定理

Shannon 的学术贡献

---开创了数字信息时代

- ✓ Mathematician
 - ✓ Ph.D. in Mathematics from MIT. Worked at **AT&T Bell Labs** and RLE in **MIT**
- ✓ Electrical engineer
 - ✓ Master's Thesis: electrical applications of Boolean algebra could construct any logical, numerical relationship
- ✓ Cryptographer
 - ✓ "A Mathematical Theory of Cryptography," 1949.
- ✓ Friend of Turing
 - ✓ For two months early in 1943, Shannon came into contact with the leading British mathematician Alan Turing. Shannon and Turing met at teatime in the cafeteria. Turing showed Shannon his 1936 paper that defined what is now known as the "Universal Turing machine"

- Claude. E. Shannon (1916-2001)
https://en.wikipedia.org/wiki/Claude_Shannon
- Movie: The bit player
- Blackjack



- 数字逻辑电路的奠基者;
- 保密通信的创始人



通信系统的基本结构

Coding Theorems for a Discrete Source With a Fidelity Criterion*

Claude E. Shannon**

Abstract

Consider a discrete source producing a sequence of message letters from a finite alphabet. A single-letter distortion measure is given by a non-negative matrix (d_{ij}) . The entry d_{ij} measures the "cost" or "distortion" if letter i is reproduced at the receiver as letter j . The average distortion of a communications system (source-coder-noisy channel-decoder) is taken to be $\bar{d} = \sum_{i,j} P_{ij} d_{ij}$ where P_{ij} is the probability of i being reproduced as j . It is shown that there is a function $R(\bar{d})$ that measures the "equivalent rate" of the source for a given level of distortion. For coding purposes where a level \bar{d} of distortion can be tolerated, the source acts like one with information rate $R(\bar{d})$. Methods are given for calculating $R(\bar{d})$, and various properties discussed. Finally, generalizations to ergodic sources, to continuous sources, and to distortion measures involving blocks of letters are developed.

In this paper a study is made of the problem of coding a discrete source of information, given a fidelity criterion or a measure of the distortion of the final recovered message at the receiving point relative to the actual transmitted message. In a particular case there might be a certain tolerable level of distortion as determined by this measure. It is desired to so encode the information that the maximum possible signaling rate is obtained without exceeding the tolerable distortion level. This work is an expansion and detailed elaboration of ideas presented earlier [1], with particular reference to the discrete case.

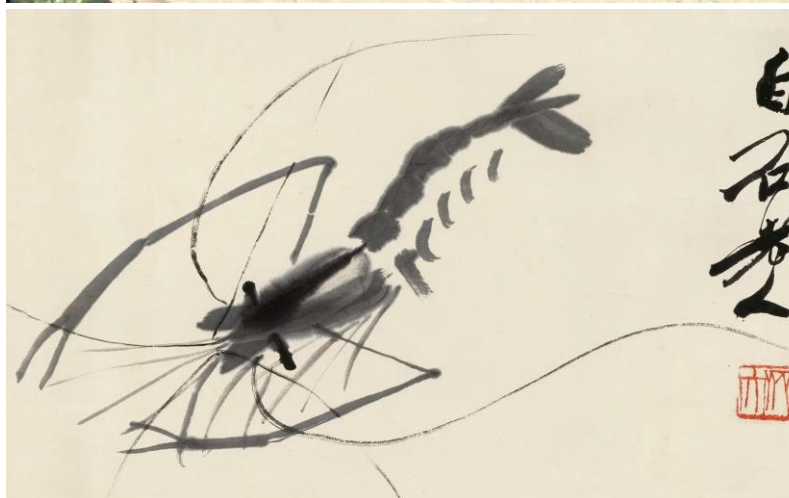
We shall show that for a wide class of distortion measures and discrete sources of information there exists a function $R(\bar{d})$ (depending on the particular distortion measure and source) which measures, in a sense, the equivalent rate R of the source (in bits per letter produced) when \bar{d} is the allowed distortion level. Methods will be given for evaluating $R(\bar{d})$ explicitly in certain simple cases and for evaluating $R(\bar{d})$ by a limiting process in more complex cases. The basic results are roughly that it is impossible to signal at a rate faster than $C / R(\bar{d})$ (source letters per second) over a memoryless channel of capacity C (bits per second) with a distortion measure less than or equal to \bar{d} . On the other hand, by sufficiently long block codes it is possible to approach as closely as desired the rate $C / R(\bar{d})$ with distortion level \bar{d} .

Finally, some particular examples, using error probability per letter of message and other simple distortion measures, are worked out in detail.

The Single-Letter Distortion Measure. Suppose that we have a discrete information source producing a sequence of letters or "word" $m = m_1, m_2, m_3, \dots, m_L$, each chosen from a finite alphabet. These are to be transmitted over a channel and reproduced, at least

* Institute of Radio Engineers, International Convention Record, vol. 7, 1959.

** This work was supported in part by the U.S. Army (Signal Corps), the U.S. Air Force (Office of Scientific Research, Air Reserve and Development Command), and the U.S. Navy (Office of Naval Research).



文字，图片，图像，数字表格
网页，语音，声音，动作，电磁等

源于：声，光，电，文等

波形与数字

听，看，感，思等



电缆



光缆

模拟和数字通信

模拟电话与数字电话

无线：磁场和光场 通信

磁场：短波、长波、微波、毫米波

光场：可见光通信、光光子通信等

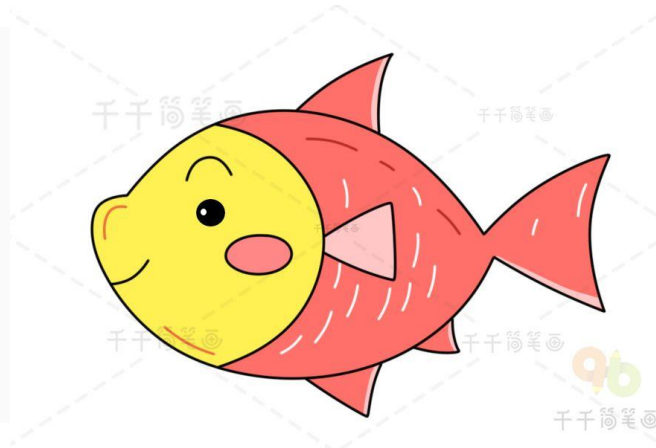
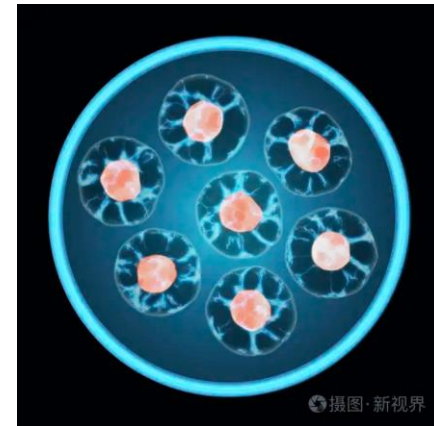
声场：声波通信；

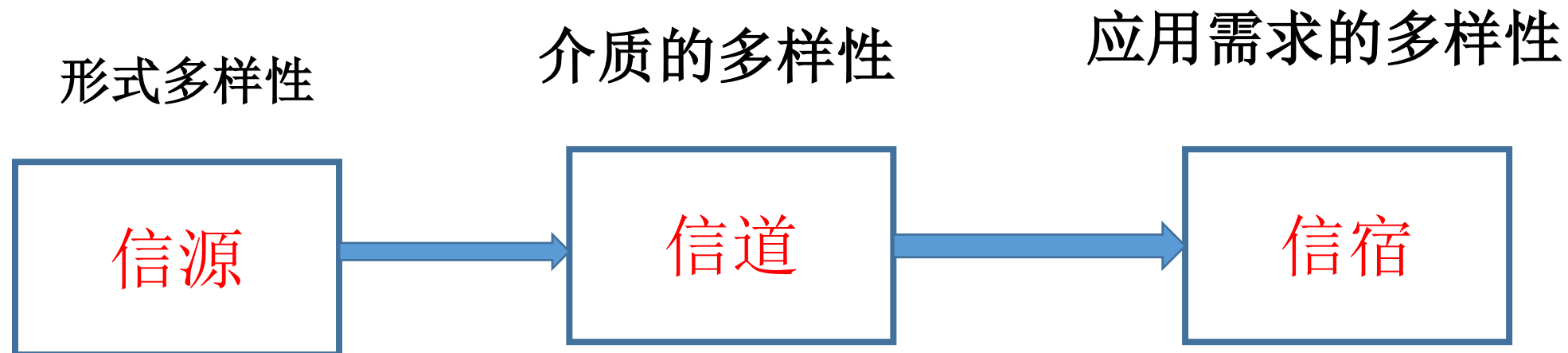
水声通信；

与通信介质有关；

磁存储：读与写 也是一种通信

- 重现原始信号形态的实用者，也成为接收端
- 接收端可以是机器、人、动物和植物等，也可以是一个生物组织等。

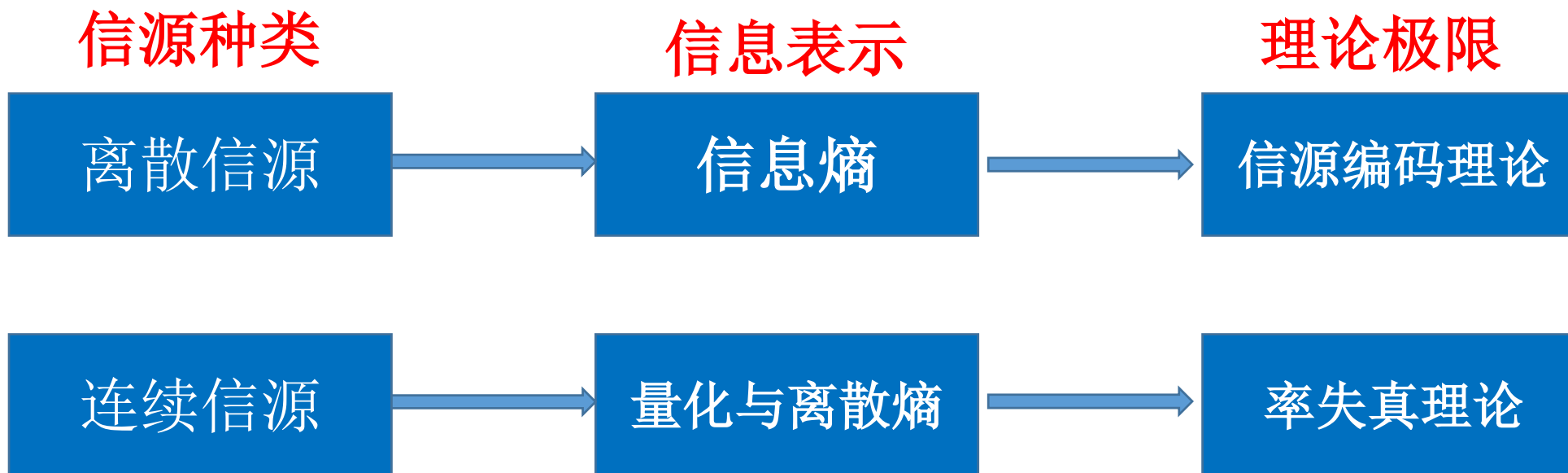




通信系统的基本结构

划时代的技术挑战
? ? ? ! ! !

通信的多样性---导致了工程实现的多样性



问题和挑战:

图像的信息表示缺少核心的理论支撑，无论是模拟图像（光学照片）还是数字图像（像素点态模式）都存在类同的问题；

数字图像的无失真压缩方法和数字图像的信息熵？？？！！！！

传统的信息论观点：

离散信源---信息表示的基本极限是信源的熵（entropy），这就是Shannon 信源编码理论；

连续信源---信息的数字化和可表示理论，这就是Shannon理论中的 Rate-Distortion理论；

主要的信息交流模型：信息从发送端按照信息流的模式（Information flow）传送到接收端。这就是说，**无论信源的原始形态是什么，都需要转化为信息序列。**

问题和挑战：图像的信息表示缺少核心理论支撑，无论是模拟图像（光学照片）还是数字图像（点态模式）都存在类同的问题；

数字图像的无失真压缩方法和数字图像的信息熵？？？！！！！

信道类型

信道刻画工具

无差错理论极限



问题与挑战:

非模型化的信道容量分析与计算，逼近信道容量的信息传输模式；

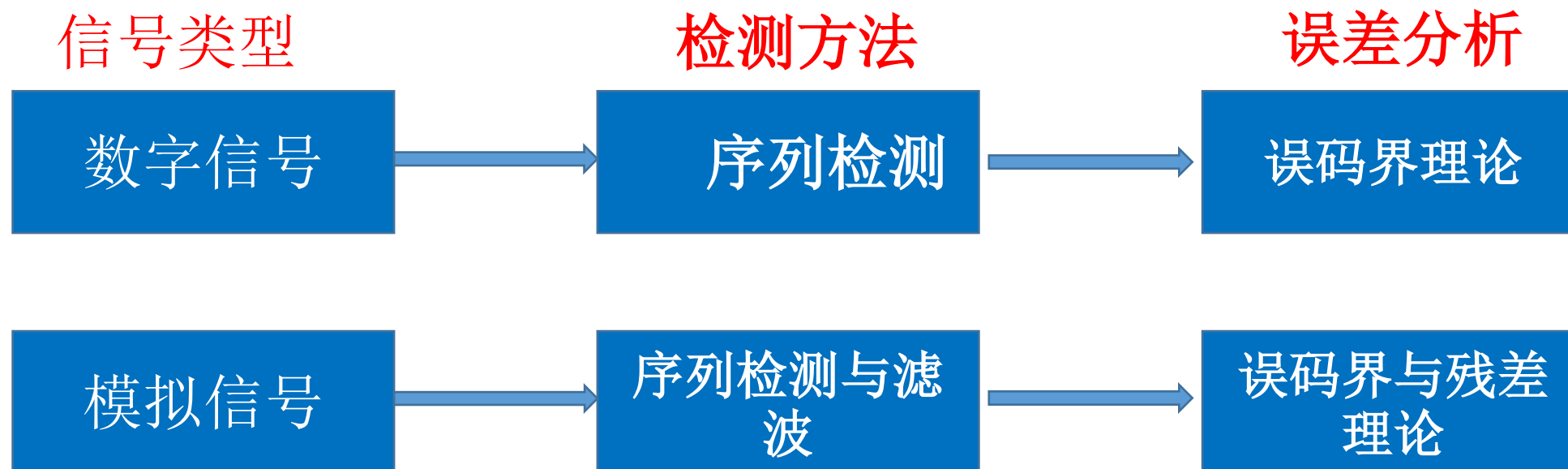
(机器学习带来的新思路和新方法，智能计算与智能信息模式)

- 传统的信息论观点：
- 离散信道---互信息最大化---Shannon信道容量（给定信道，寻找适合信道传输的信源分布）
- 连续信道---信道的离散化+样本点的互信息最大化---- 代表成果
H-Shannon 抽样定理+高斯信道容量公式
(基本策略: Bandlimited Channels + Additive White Gaussian Noise (AWGN) Mode)

离散信道的多样性---从简单（二元对称、二元删除）到复杂（Markov 记忆信道）；
连续信道的多样性---从简单（基带、带通）到复杂（时变、多通道、广播、中继）；
通信链路的网络化---网络信息论---最大流最小割理论与网络编码等；

问题与挑战：

非模型化的信道容量分析与计算，逼近信道容量的信息传输模式；
(机器学习带来的新思路和新方法，智能计算与智能信息模式)



问题与挑战:

对信源和信道具有强依赖特征，信息传输与信息应用是采用独立处理模式；
信息应用被明确标定时，系统的设计如何？ 如何反映用户的主管要求？

---语义信息论， 信息重要性？？？！！ 电磁信息论？?? (联合信源信道处理)

- 传统的信息论观点：

发端信息的恢复---

数字信号---低误码、低畸变的信号检测--- 相干检测、最大似然检测、Bayes 检测（似然比分类）、序列联合检测（软解码、解调算法）；

模拟信号---需要附加光滑滤波等；

问题与挑战：

对信源和信道具有强依赖特征，信息传输与信息应用是采用独立处理模式；
信息应用被明确标定时，系统的设计如何？ 如何反映用户的主管要求？

---语义信息论， 信息重要性？？？！！ 电磁信息论？？？！！

目标迁移理论与技术----（信息论的泛化理论）----机器学习迁移与泛化

基本工具:

信息熵、互信息、条件熵、
交叉熵，典型集理论、
信息不等式、概率单纯形
的凸性理论，
多维数据的流形理论

应用:

博弈论中的收益率；大数定律；大偏差理论；条件极限定理；假设检验；Fisher 信息矩阵和Cramer-Rao 不等式；高维矩阵计算；

问题和挑战:

- 交叉熵、大偏差理论与机器学习;
- 新型信息熵与**GAN**;
- 机器学习预测误差分析;
- 可解释性机器学习方法;
- 概率模型的模式坍塌问题 (机器学习方法)
- 数据添加与更新导致的概率模型迁移问题

- 信息论的目标迁移特征

信息论的原始特点是集中解决通信中的基本理论问题：信息表示和信息传输；发展的数学工具---信息熵、互信息、交叉熵、条件熵、典型集理论、信息不等式、概率单纯形的凸性理论；

数学作为一种基本的逻辑推理工具，许多数学公式本身不具有明确的物理意义，具有繁化性；应用：

博弈论中的收益率；大数定律；大偏差理论；条件极限定理；假设检验；Fisher 信息矩阵和Cramer-Rao 不等式；高维矩阵计算；

问题和挑战：

交叉熵、大偏差理论与机器学习；新型信息熵与GAN；机器学习预测误差分析；可解释性机器学习方法；

- 信源部分

- 数字图像的信息熵理论
- 数字图像的无失真压缩算法

- 信息传输部分

- 基于**MIM**的信息传输（非模型化信道的容量分析）

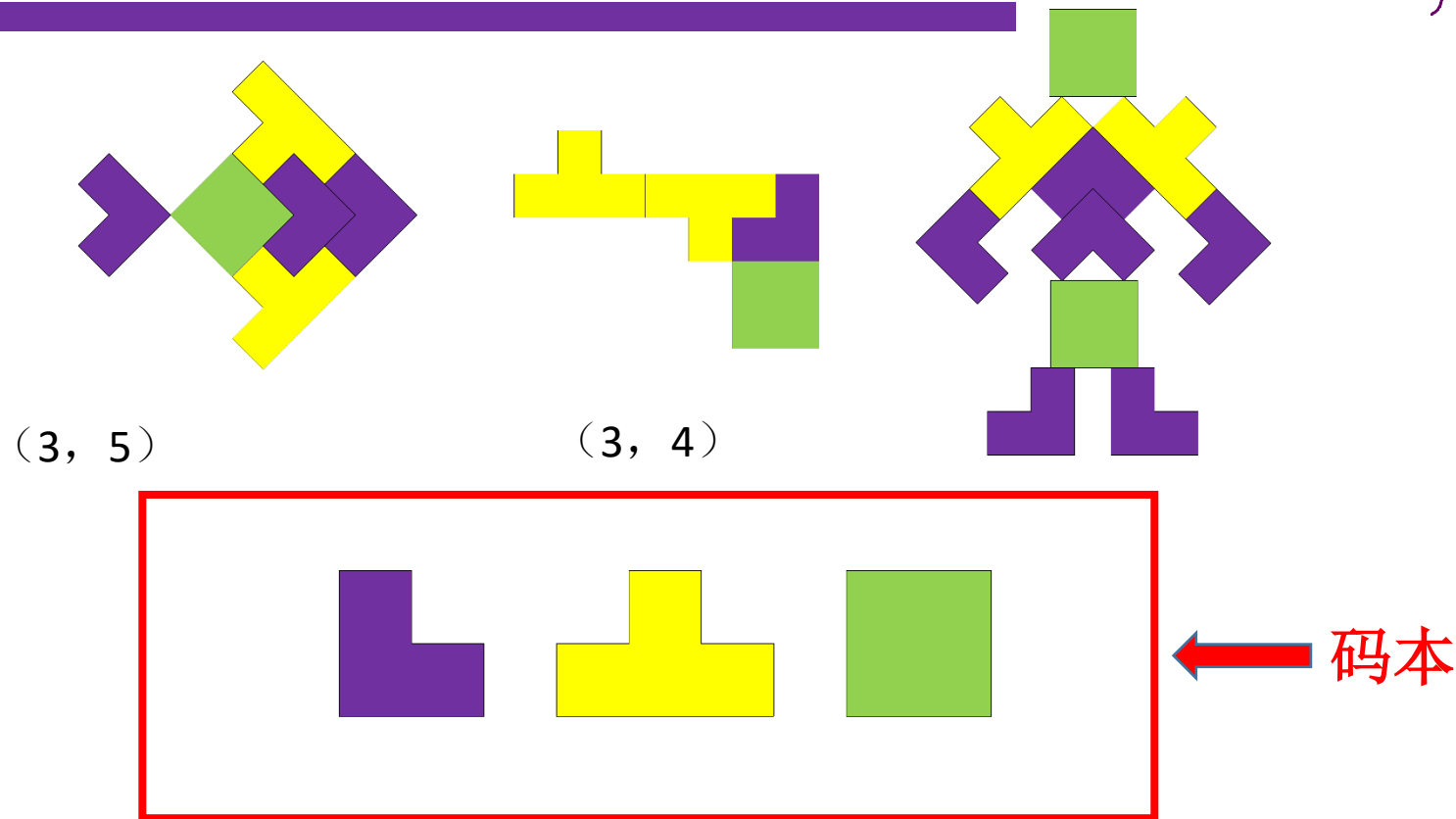
- 信宿部分

- 基于信息重要性的信源编码、信道编码和小概率事件甄别等

- 信息论的拓展应用

- **MiM-GAN**的开发与信道容量的智能计算

信源软压缩原理介绍---积木拼图



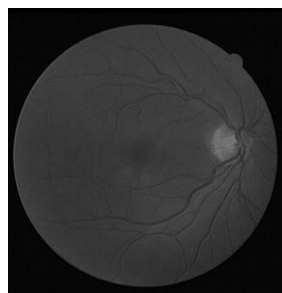
事实：三个基本形状排列组合产生了不同的图案



思考：如果采用上述码本的形状作为表示图像的基本单元会怎样？



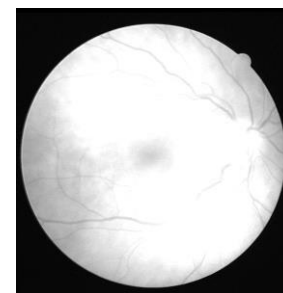
(a)



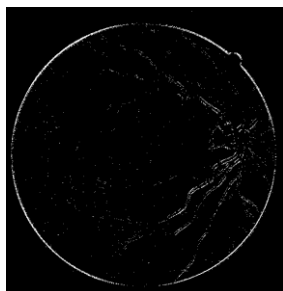
(b)



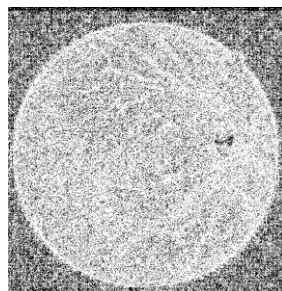
(c)



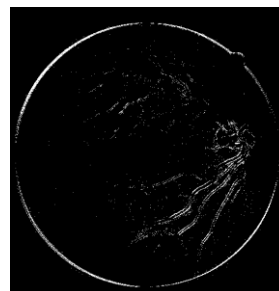
(d)



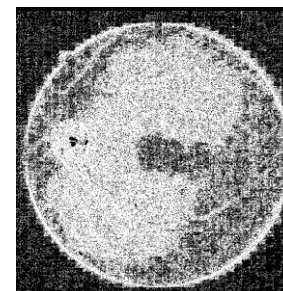
(e)



(f)



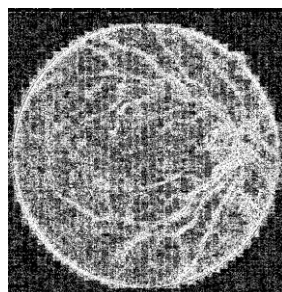
(g)



(h)



(i)



(j)



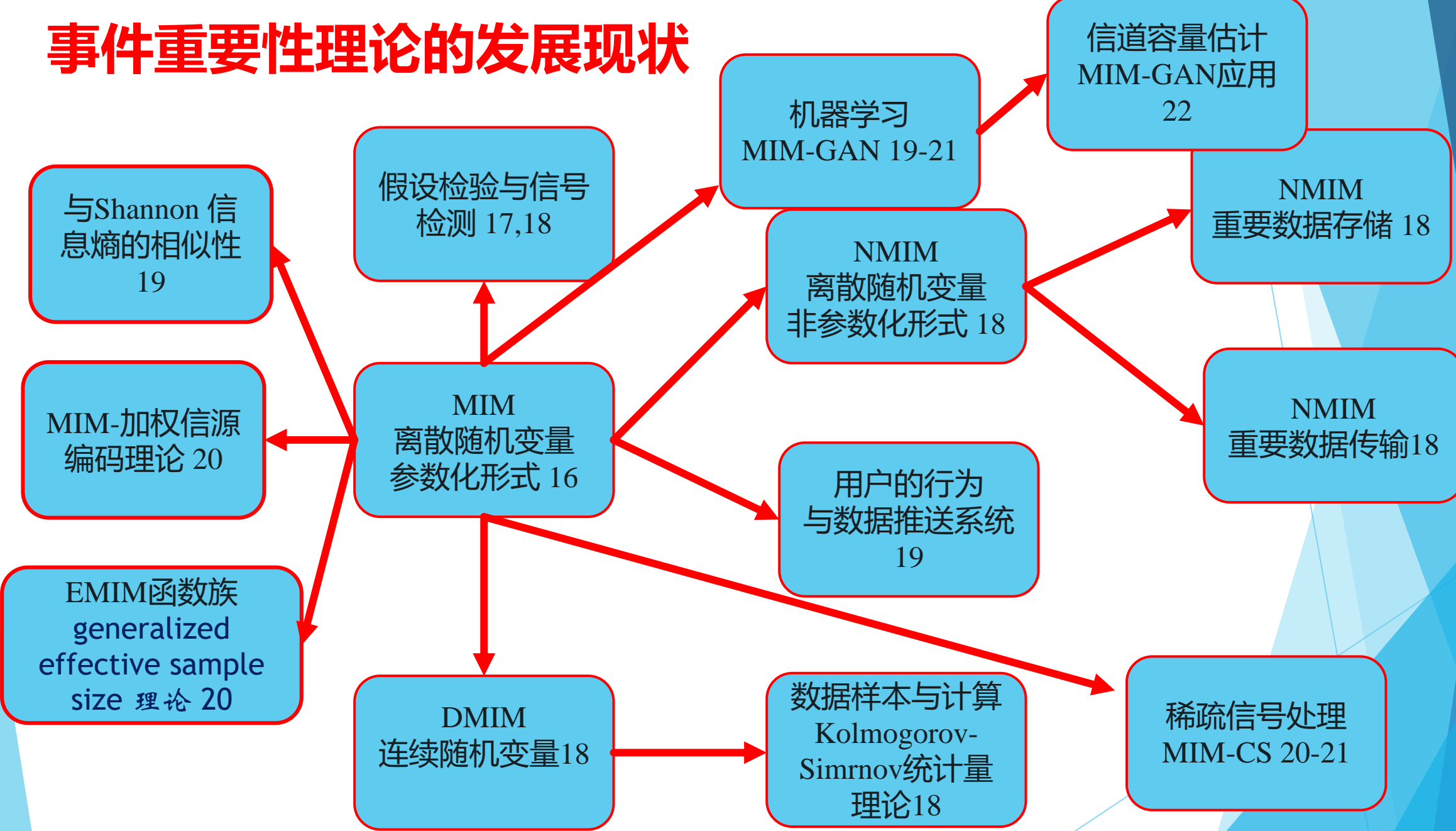
(k)

多成分图像的软
压缩算法举例

多个数据集上，软压
缩算法和主流图像压
缩算法的对比

Dataset	Statistic	Method			
		Soft compression	JPEG (lossless mode)	PNG	JPEG2000 (lossless mode)
DRIVE	Mean	3.16	2.26	2.43	2.97
	Minimum	2.85	2.20	2.33	2.79
	Maximum	4.03	2.43	2.76	3.67
	Variance	0.0555	0.0026	0.0072	0.0333
Malaria	Mean	3.80	2.53	2.83	3.56
	Minimum	2.35	1.77	1.87	2.50
	Maximum	6.88	4.41	7.46	8.92
	Variance	0.8058	0.3401	1.0340	1.2497
BCCD	Mean	4.71	2.49	2.75	4.01
	Minimum	4.25	2.39	2.58	3.72
	Maximum	5.76	2.80	3.13	4.78
	Variance	0.0496	0.0039	0.0066	0.0196
Melanoma	Mean	3.44	1.84	1.93	3.15
	Minimum	1.81	1.29	1.31	2.02
	Maximum	5.17	2.81	3.11	4.70
	Variance	0.4758	0.0596	0.0867	0.2365

事件重要性理论的发展现状



信道特征： 从远场模式迁移到近场模式，球形磁场的分析

发端和接收端： 呈现多输入多输出的多通道通信模式

- ✓ 信源符号的设计
- ✓ 信道特征的适配
- ✓ 收端的多分集接收技术
- ✓ 信道容量的分析
- ✓ 智能学习与环境模拟技术

- 新型物理介质的出现导致新兴技术的创新与拓展（未知领域）；
- 新型研究工具的出现导致原有理论的更新，提升人们对客观世界的认知水平；
- 继承原有信息理论和技术的特征，有助于发现新的研究点；
- 急需学科交叉和综合能力的提升，才能适应新领域、新场景的研究。

- 信息论的历史发展脉络和目标迁移
- 新智能时代的信息论与机器学习的联系
- 机器学习中重要的信息表示、传输和加密等技术与信息论的内在联系

- 课程应体现---“起、承、转、合”
中国哲学思想---传统文化内涵式发展

“起：开头；承：承接上文加以申述；
转：转折；合：结束。泛指文章的做法。”

出处：

元·范德玘《诗格》：“作诗有四法：
起要平直；承要春容；转要变化；合要渊永。”

中国写意画讲座：也需要遵循“起、承、转、合”

图20



中国画：牡丹画法