

信息论 基础教程

李 梅 李亦农 王玉峰◎编著

第 3 版



北京邮电大学出版社
www.buptpress.com

信息论基础教程

(第 3 版)

李 梅 李亦农 王玉皞 编著



北京邮电大学出版社
www.buptpress.com

内 容 简 介

信息论是现代信息通信领域的基础理论,是研究信息传输和信息处理的一般规律的科学。我们在借鉴了国内外众多的信息论优秀教材和参考资料之后编写了《信息论基础教程》。本书为第3版,以香农的三个编码定理为中心,重点讲述了相关的基本概念、基本原理和基本方法。鉴于目前各大专院校都在削减学时,教材只是讲述经典信息论的内容,没有涉及过多的分支。

本书可作为通信及电子信息类相关专业高年级本科生和研究生教材,也可作为相关专业科研人员的参考书。

图书在版编目(CIP)数据

信息论基础教程 / 李梅, 李亦农, 王玉峰编著. --3 版. --北京: 北京邮电大学出版社, 2015. 8
ISBN 978-7-5635-4468-4

I. ①信… II. ①李… ②李… ③王… III. ①信息论—高等学校—教学参考资料 IV. ①G201

中国版本图书馆 CIP 数据核字(2015)第 186179 号

书 名: 信息论基础教程(第3版)

著作责任者: 李 梅 李亦农 王玉峰 编著

责任编辑: 刘春棠

出版发行: 北京邮电大学出版社

社 址: 北京市海淀区西土城路 10 号(邮编:100876)

发行部: 电话: 010-62282185 传真: 010-62283578

E-mail: publish@bupt.edu.cn

经 销: 各地新华书店

印 刷: 北京鑫丰华彩印有限公司

开 本: 787 mm×1 092 mm 1/16

印 张: 16.75

字 数: 419 千字

印 数: 1—3 000 册

版 次: 2005 年 1 月第 1 版 2008 年 10 月第 2 版 2015 年 8 月第 3 版 2015 年 8 月第 1 次印刷

ISBN 978-7-5635-4468-4

定 价: 34.00 元

• 如有印装质量问题,请与北京邮电大学出版社发行部联系 •

前 言

信息论是现代信息通信领域的基础理论,是研究信息传输和信息处理的一般规律的科学,因此目前各高等院校相关专业的本科生、研究生都开设了这门课。

在借鉴了国内外众多的信息论优秀教材和参考资料之后,作者根据多年的教学实践经验编写了《信息论基础教程》。本书以使读者掌握基本概念和方法为目的,力图以读者最易接受的方式介绍信息论的基本内容及应用。本书可作为通信及电子信息类相关专业高年级本科生和研究生的教材,也可作为相关专业科研人员的参考书。

教材以香农的三个编码定理为中心,重点讲述了相关的基本概念、基本原理和基本方法。鉴于目前各大专院校都在削减学时,教材只是讲述经典信息论的内容,没有涉及过多的分支。同时,我们认为信息论是一门理论和实践紧密结合的课程,因此选择了很多与日常生活密切相关,具有一定趣味性的习题,同时增加了上机练习。

本书共分为7章,外加3个附录。第1章主要介绍了香农信息论的研究对象、目的和内容;第2章介绍了关于信息度量的几个重要概念:自信息、互信息、信息熵、平均互信息以及数据处理定理;第3章研究定量度量信源产生信息的能力和信源冗余度的问题;第4章研究定量描述信道传递信息能力的问题,并介绍了信道容量的计算方法;第5章的核心内容是香农的无失真信源编码定理,围绕这个定理介绍了无失真信源编码的基本概念,讲述了几种实用的无失真信源编码方法;第6章讲述香农的有噪信道编码定理以及纠错编码的主要内容,介绍了信道编码的基本概念、基本理论;第7章介绍香农的限失真信源编码定理,引入了信息率失真函数的概念并介绍了信息率失真函数的性质以及计算方法,然后还介绍了几种常用的熵压缩编码算法;附录A是一些学习本课程需要用到的数学知识;附录B是上机作业;附录C是习题解答。

本教材自2005年1月出版以来,承蒙广大读者的厚爱,被许多兄弟院校作为教材,因此一版再版。截至2015年7月,两版共印刷了16次,印数超过5万册。本教材在得到广泛采用的同时,读者也提出了许多宝贵的修改意见。此次的再版在保持原版特色的基础上,修正了一些错误,并调整了部分内容,力求更

简洁、流畅、易懂、无误,更适合教学和自学的需要。

在此感谢王莉、王永峰、任新成、汪小威、赵继军、吴韶波、李红、杨师、徐益民、杨世忠和徐安庭等老师提出的宝贵意见,同时感谢邢开颜、赵永平、刘旭、甄晓丹、明舒晴同学为收集习题做的大量工作。

本书在编写过程中,参阅了国内外很多信息论经典著作(列于本书参考文献中),同时参考了大量国内外知名大学信息论课程的课后习题及解答,在此向有关作者表示感谢!

尽管我们在再版中力求更加符合读者的要求,但仍无法避免错漏和不当,恳切希望广大读者提出宝贵意见。

简历:http://iec.cugb.edu.cn/xgweb/Showteadetail?tea_id=10040049

博客:<http://blog.sciencenet.cn/u/maggielimei>

邮箱:maggieli@cugb.edu.cn

(我校是211/985创新平台高校,欢迎勤奋、主动、有潜力、对电子信息专业有专业热情的同学加入我的研究课题组。)

李梅

2015年6月

目 录

第 1 章 绪论	1
1.1 信息的概念	1
1.2 信息论的研究对象、目的和内容	2
习题 1	5
第 2 章 信息的度量	6
2.1 自信息和互信息	6
2.1.1 自信息	6
2.1.2 互信息	8
2.2 平均自信息	8
2.2.1 平均自信息的概念	8
2.2.2 熵函数的性质	9
2.2.3 联合熵与条件熵	13
2.3 平均互信息	15
2.3.1 平均互信息的概念	15
2.3.2 平均互信息的性质	17
2.3.3 数据处理定理	20
习题 2	21
第 3 章 信源及信源熵	25
3.1 信源的分类及其数学模型	25
3.2 离散单符号信源	26
3.3 离散多符号信源	27
3.3.1 离散平稳无记忆信源	27
3.3.2 离散平稳有记忆信源	28
3.3.3 马尔可夫信源	31
3.3.4 信源的相关性和剩余度	34
* 3.4 连续信源	37
3.4.1 连续信源的微分熵	37
3.4.2 连续信源的最大熵	40

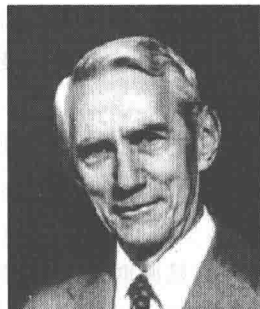
3.4.3 连续信源的熵功率	41
习题 3	42
第 4 章 信道及信道容量	46
4.1 信道的分类	46
4.2 离散单符号信道及其信道容量	47
4.2.1 离散单符号信道的数学模型	47
4.2.2 信道容量的概念	49
4.2.3 几种特殊信道的信道容量	50
4.2.4 离散对称信道的信道容量	52
4.2.5 一般离散信道的信道容量	55
4.2.6 信道容量定理	58
* 4.2.7 信道容量的迭代算法	62
4.3 离散多符号信道及其信道容量	64
4.4 组合信道及其信道容量	67
4.4.1 独立并联信道	67
4.4.2 级联信道	68
* 4.5 连续信道及其信道容量	69
4.5.1 连续随机变量的互信息	69
4.5.2 高斯加性信道的信道容量	70
4.5.3 多维高斯加性信道的信道容量	71
* 4.6 波形信道及其信道容量	72
习题 4	72
第 5 章 无失真信源编码	77
5.1 信源编码的相关概念	77
5.1.1 编码器	77
5.1.2 码的分类	79
5.2 定长码及定长信源编码定理	81
5.3 变长码及变长信源编码定理	84
5.3.1 Kraft 不等式和 McMillan 不等式	85
5.3.2 唯一可译码的判别准则	86
5.3.3 紧致码平均码长界限定理	87
5.3.4 无失真变长信源编码定理(香农第一定理)	89
5.4 变长码的编码方法	92
5.4.1 香农编码	92
5.4.2 香农-费诺-埃利斯编码	94
5.4.3 二元霍夫曼码	94
5.4.4 r 元霍夫曼码	97

5.4.5 费诺码	98
5.5 实用的无失真信源编码方法	99
5.5.1 游程编码	99
5.5.2 算术编码	102
5.5.3 LZW 编码	104
习题 5	107
第 6 章 有噪信道编码	111
6.1 信道编码的相关概念	111
6.1.1 错误概率和译码规则	112
6.1.2 错误概率与编码方法	117
6.2 有噪信道编码定理	123
6.3 纠错编码	125
6.3.1 纠错码分类	126
6.3.2 纠错码的基本概念	127
6.3.3 线性分组码	129
* 6.3.4 卷积码	143
习题 6	146
第 7 章 限失真信源编码	151
7.1 失真测度	152
7.1.1 失真函数	152
7.1.2 平均失真	154
7.2 信息率失真函数	155
7.2.1 D 失真许可信道	155
7.2.2 信息率失真函数的定义	155
7.2.3 信息率失真函数 $R(D)$ 的性质	156
7.3 限失真信源编码定理	161
* 7.4 信息率失真函数的计算	161
7.4.1 应用参量表示式计算 $R(D)$	161
7.4.2 率失真函数的迭代算法	167
7.5 常用的限失真信源编码方法	170
7.5.1 量化编码	170
7.5.2 子带编码	171
7.5.3 预测编码	172
7.5.4 变换编码	173
习题 7	174
附录 A 数学预备知识	176
A.1 概率论与随机过程	176

A.1.1	概率论的基本概念	176
A.1.2	随机变量及其分布	178
A.1.3	多维随机变量及其分布	179
A.1.4	随机变量的数字特征	181
A.1.5	随机过程	181
A.2	凸函数及 Jensen 不等式	185
A.3	信道容量定理引理	187
A.4	渐进等分割性和 ϵ 典型序列	188
附录 B	上机作业	191
B.1	信道容量的迭代算法	191
B.2	唯一可译码判决准则	192
B.3	Shannon 编码	192
B.4	Huffman 编码	193
B.5	Fano 编码	193
B.6	LZW 编码	194
B.7	BSC 模拟器	194
B.8	Hamming (7,4) 编译码器	194
B.9	通信系统仿真	195
附录 C	习题解答	197
C.1	第 1 章习题解答	197
C.2	第 2 章习题解答	197
C.3	第 3 章习题解答	211
C.4	第 4 章习题解答	224
C.5	第 5 章习题解答	237
C.6	第 6 章习题解答	247
C.7	第 7 章习题解答	255
参考文献		260

第1章 绪 论

克劳德·艾尔伍德·香农(Claude Elwood Shannon, 1916—2001 年)美国数学家,信息论的创始人。1948 年香农在《贝尔系统技术杂志》(Bell System Technical Journal)上连载发表了著名的论文《通信的数学原理》。1949 年,香农又在该杂志上发表了另一影响深远的论文《噪声下的通信》。在这两篇论文中,香农阐明了通信的基本问题,给出了通信系统的模型,提出了信息熵的数学表达式,并解决了信道容量、信源统计特性、信源编码、信道编码等一系列基本问题。这两篇论文成为信息论的奠基性著作。



1.1 信息的概念

信息论是通信的数学基础,它是随着通信技术的发展而形成和发展起来的一门新兴的横断学科。

信息论创立的标志是 1948 年香农发表的论文“A Mathematical Theory of Communication”。为了解决在噪声信道中有效传输信息的问题,香农在这篇文章中创造性地采用概率论的方法来研究通信中的问题,并且对信息给予了科学的定量描述,第一次提出了信息熵的概念。

在日常生活中,人们往往对消息和信息不加区别,认为消息就是信息。例如,当人们收到一封电报,或者听了天气预报,人们就说得到了信息。

人们收到消息后,如果消息告诉了我们很多原来不知道的新内容,我们会感到获得了很多信息,而如果消息是我们基本已经知道的内容,我们得到的信息就不多,所以信息应该是可以度量的。那么怎样度量信息呢?人们需要有一个可以用数学模型来表示的信息概念。

1928 年,哈特莱(Hartley)首先提出了对数度量信息的概念,即一个消息所含有的信息量用它的所有可能的取值的个数的对数来表示。比如,抛掷一枚硬币可能有两种结果:正面和反面,所以当我们得知抛掷结果后获得的信息量是 $\log_2 2 = 1 \text{ bit}$ 。而一个十进制数字可以表示 0~9 中的任意一个符号,所以一个十进制数字含有 $\log_2 10 = 3.322 \text{ bit}$ 的信息量。这里对数取以 2 为底,信息量的单位为 bit。

哈特莱的工作给了香农很大的启示,他进一步注意到消息的信息量不仅与它的可能值的个数有关,还与消息本身的不确定性有关。例如,抛掷一枚偏畸硬币,如果正面向上的可

能性是90%,那么当我们得知抛掷结果是反面时得到的信息量会比得知抛掷结果是正面时得到的信息量大。

一个消息之所以会含有信息,正是因为它具有不确定性,一个不具有不确定性的消息是不会含有任何信息的,而通信的目的就是为了消除或部分消除这种不确定性。比如,在得知硬币的抛掷结果前,我们对于结果会出现正面还是反面是不确定的,通过通信,我们得知了硬币的抛掷结果,消除了不确定性,从而获得了信息。因此,信息是对事物运动状态或存在方式的不确定性的描述。这就是香农信息的定义。

用数学的语言来讲,不确定性就是随机性,具有不确定性的事件就是随机事件。因此,可运用研究随机事件的数学工具——概率——来测度不确定性的大小。在信息论中,我们把消息用随机事件表示,而发出这些消息的信源则用随机变量来表示。比如,抛掷一枚硬币的试验可以用一个随机变量来表示,而抛掷结果可以是正面或反面,这个具体的消息则用随机事件表示。

我们把某个消息 x_i 出现的不确定性的大小定义为自信息,用这个消息出现概率的对数的负值来表示:

$$I(x_i) = -\log p(x_i) \quad (1.1)$$

自信息同时表示这个消息所包含的信息量,也就是最大能够给予收信者的信息量。如果消息能够正确传送,收信者就能够获得这么大小的信息量。

信源所含有的信息量定义为信源发出的所有可能消息的平均不确定性,香农把信源所含有的信息量称为信息熵。信息熵定义为自信息的统计平均,即

$$H(X) = -\sum_{i=1}^q p(x_i) \log p(x_i) \quad (1.2)$$

这里的 q 表示信源消息的个数。信息熵表示信源的平均不确定性的大小,同时表示信源输出的消息平均所含的信息量。因此,虽然信源产生的消息可能会含有不同的信息量,比如抛掷一枚偏畸硬币的结果是正面和是反面这两个消息所含的信息量不同,但是可以用它们的平均值来表示这个信源(抛掷一枚偏畸硬币的实验)的平均不确定性。

在收信端,信源的不确定性得到了部分或全部的消除,收信者就得到了信息。信息在数量上等于通信前后“不确定性”的消除量(减少量)。

这种建立在概率模型上的信息概念排除了日常生活中“信息”一词主观上的含义和作用,而只是对消息的统计特性的定量描述,所以信息可以度量,而且与日常生活中信息的概念并不矛盾,因此是一个科学的定义。根据这样的信息定义,同样一个消息对于任何一个收信者来说,所含有的信息量都是一样的。而事实上信息有很强的主观性和实用性,同样一个消息对不同的人常常有不同的主观价值或主观意义。例如,同一则气象预报对在室外工作的人和室内工作的人可能会有不同的意义和价值,因此所提供的信息量也应该不同。所以香农信息的定义在某些情况下也具有一定的局限性。

1.2 信息论的研究对象、目的和内容

信息论从诞生到现在,虽然只有短短的50多年,但它的发展对学术界及人类社会的影

响是相当广泛和深刻的。如今,信息论的研究内容不仅仅包括通信,而且包括所有与信息有关的自然和社会领域,如模式识别、机器翻译、心理学、遗传学、神经生理学、语言学、语义学甚至包括社会学中有关信息的问题。香农信息论迅速发展成为涉及范围极广的广义信息论——信息科学。

信息论的研究对象是广义的通信系统,它把所有的信息流通系统都抽象成一个统一的模型,如图 1.1 所示。

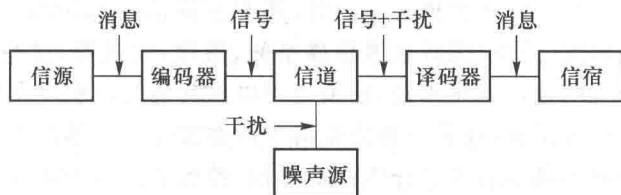


图 1.1 通信系统模型

这个模型不仅包括电话、电报、传真、电视、雷达等狭义的通信系统,还包括生物有机体的遗传系统、神经系统、视觉系统,甚至人类社会的管理系统。信息以消息的形式在这个通信系统中传递,人们通过研究通信系统中消息的传输和处理来得到信息传输和处理的规律,目的是提高通信的可靠性和有效性。

在任何一个信息流通系统中,都有一个发出信息的发送端(信源),有一个接收信息的接收端(信宿),以及信息流通的通道(信道)。在信息传递的过程中不可避免地会有噪声,所以有一个噪声源。为了把信源发出的消息变成适合在信道中传输的信号,还需要加入编码器,在送到信宿之前要进行反变换,所以要加入译码器。

这个通信系统主要分成 5 个部分。

(1) 信源

顾名思义,信源是产生消息和消息序列的源。信源可以是人、生物、机器或其他事物。比如,各种气象状态是信源,能够产生独特的气味吸引蜜蜂来采花蜜的花朵是信源,人脑的思维活动也是一种信源。信源的输出是消息(或消息序列)。

消息有着各种不同的形式,如文字、符号、语言、图片、图像、气味等。消息以能被通信双方所理解的形式,通过通信进行传递和交换。消息携带着信息,是信息的载体。信源输出的消息是随机的、不确定的,但又有一定的规律性,因此用随机变量或随机矢量等数学模型来表示信源。

(2) 编码器

编码就是把消息变成适合在信道传输的物理量,这种物理量称为信号(如电信号、光信号、声信号、生物信号等)。信号携带着消息,它是消息的载体。

编码器可分为信源编码器和信道编码器。信源编码的目的是压缩信源的冗余度(即多余度),提高信息传输的效率,这是为了提高通信系统的有效性。信源编码又可分为无失真信源编码和限失真信源编码。信道编码是为了提高信息传输的可靠性而有目的地对信源编码器输出的代码组添加一些监督码元,使之具有纠、检错能力。比如,老师讲课需要把知识进行加工和提炼,以提高信息传输的有效性,而为了让学生听得明白,有时又需要适当地重复,这是为了提高信息传输的可靠性。

在实际的通信系统中,可靠性和有效性常常是相互矛盾的,提高有效性必须去掉信源符号的冗余部分,但是这会导致可靠性的下降,而提高可靠性就需要增加监督码元,这又降低了有效性。有时为了兼顾有效性,就不一定要求绝对准确地在接收端再现原来的消息,而是可以允许一定的误差或失真,也就是说允许近似地再现原来的消息。

(3) 信道

信道是指通信系统把载荷消息的信号从发送端送到接收端的媒介或通道,是包括收发设备在内的物理设施。信道除了传播信号以外,还有存储信号的作用。在狭义的通信系统中,实际信道有明线、电缆、光缆、无线电波传播空间、磁盘、光盘等,这些都属于传输电磁波能量的信道。对于广义的通信系统来说,信道还可以是其他的传输媒介。

在信道中引入噪声和干扰,这是一种简化的表达方式。为了分析方便起见,把在系统其他部分产生的干扰和噪声都等效地折合成信道干扰,看成是由一个噪声源产生的,它将作用于所传输的信号上。这样,信道输出的已是叠加了干扰的信号。噪声源的统计特性是划分信道的依据,并且是信道传输能力的决定因素。由于干扰或噪声往往具有随机性,所以信道用输入和输出之间的条件概率分布来描述。

(4) 译码器

译码就是把信道输出的已叠加了干扰的编码信号进行反变换,变成信宿能够理解的消息。译码器也可分成信源译码器和信道译码器。译码器需要尽可能准确地再现信源输出的消息。

(5) 信宿

信宿是消息传送的对象,即接受消息的人、机器或其他事物。

以上我们考虑的是收发两端单向通信的情况,它只有一个信源和一个信宿,信息传输也是单向的。在组网通信的情况下(如电话网、计算机网等),可能有很多分开的信源、信道和信宿同时进行信息交换。例如,广播信道是一个输入、多个输出的单向信道,而卫星通信则是多个输入、多个输出的多向传输的通信,这就需把两端单向通信的模型作适当的修正,得出多用户通信系统的模型,把两端单向通信的信息理论发展成为多用户通信信息理论。

信息论研究的是关于这个通信系统的最根本、最本质的问题。举例如下。

① 什么是信息?如何度量信息?

② 怎样确定信源的输出中含有多少信息量?

③ 对于一个信道,它传输信息量的最高极限(信道容量)是多少?

④ 为了能够无失真地传输信源信息,对信源编码时所需的最少的码符号数是多少?这是无失真信源编码,即香农第一定理的内容。

⑤ 在有噪信道中有没有可能以接近信道容量的信息传输率传输信息而错误概率几乎为零?这是有噪信道编码,即香农第二定理的内容。

⑥ 如果对信源编码时允许一定量的失真,所需的最少的码符号数又是多少?这是限失真信源编码,即香农第三定理的内容。

毫无疑问,如果我们对这些问题都有了确定的答案,那么在设计通信系统时就有了目标和指导方向,同时也有了评价通信系统优劣的标准。

在这里,我们举几个成功地应用信息论的概念和方法指导通信系统设计的例子。

(1) 无失真信源编码的应用:计算机文件的压缩

由于数据库的广泛应用,存储计算机文件所需的存储量问题日益突出。在过去的20多

年中,至少已有 20 种不同的对计算机文件的压缩算法问世,其中较好的算法能使文件压缩后所需的存储量只为原文件的 30% 左右。

(2) 有噪信道编码的应用:模拟话路中数据传输速率的提高

最早的调制解调器速率只有 300 bit/s,此后,调制解调器的速率从 4 800 bit/s 到 33.6 kbit/s,已经非常接近于理论极限(56 K 调制解调器由于下行只经过一次模/数转换,所以下行速率更快一些)。

(3) 限失真信源编码的应用:语音信号压缩

按照信息理论的分析,语音信号(也就是话音信号)所需的编码速率可以远远低于按奈奎斯特采样定理和量化噪声理论所确定的编码速率。几十年来,人们在这方面的的工作取得了巨大的进展。CCITT 关于长途电话网的语音编码速率标准已从 1972 年的 64 kbit/s 降低到 1992 年的 16 kbit/s。在移动通信中,1988 年欧洲 GSM 标准中的语音编码速率为 13.2 kbit/s,而 1989 年美国 CTIA 标准中的语音编码速率仅为 7.95 kbit/s。目前,声码器的速率可低于 100 bit/s,已接近信息论指出的极限。

目前,对信息论的研究范围一般有 3 种理解。

(1) 狭义信息论:又称香农信息论。主要通过数学描述与定量分析,研究通信系统从信源到信宿的全过程,包括信息的测度、信道容量以及信源和信道编码理论等问题,强调通过编码和译码使收、发两端联合最优化,并且以定理的形式证明极限的存在。这部分内容是信息论的基础理论。

(2) 一般信息论:也称工程信息论。主要也是研究信息传输和处理问题,除香农信息论的内容外,还包括噪声理论、信号滤波和预测、统计检测和估计、调制理论、信息处理理论以及保密理论等。

(3) 广义信息论:也称信息科学,不仅包括上述两方面内容,而且包括所有与信息有关的自然和社会科学领域,如模式识别、机器翻译、心理学、遗传学、神经生理学、语言学、语义学,甚至包括社会学中有关信息的问题。

本课程主要研究香农信息论的内容。

习 题 1

1.1 下列不属于消息的是()。

A. 文字 B. 信号 C. 图像 D. 语言

1.2 为了提高通信系统的有效性可以采用_____编码,为了提高通信系统的可靠性可以采用_____编码。

1.3 简述信息的特征。

第2章 信息的度量

关于信息的度量有几个重要的概念。

(1) 自信息(量):一个事件(消息)本身所包含的信息量,它是由事件的不确定性决定的,比如抛掷一枚硬币的结果是正面这个消息所包含的信息量。

(2) 互信息(量):一个事件所给出关于另一个事件的信息量,比如今天下雨所给出关于明天下雨的信息量。

(3) 平均自信息(量),或称信息熵:事件集(用随机变量表示)所包含的平均信息量,它表示信源的平均不确定性,比如抛掷一枚硬币的试验所包含的平均信息量。

(4) 平均互信息(量):一个事件集所给出关于另一个事件集的平均信息量,比如今天的天气所给出关于明天的天气的信息量。

我们在最简单的离散随机变量的情况下引入这些概念。

2.1 自信息和互信息

2.1.1 自信息

在绪论中我们讲过,信源发出的消息(事件)具有不确定性,而事件发生的不确定性与事件发生的概率大小有关,概率越小,不确定性越大,事件发生以后所含有的信息量就越大。小概率事件,不确定性大,一旦出现必然使人感到意外,因此产生的信息量就大,特别是几乎不可能出现的事件一旦出现,必然产生极大的信息量;大概率事件,是预料之中的事件,不确定性小,即使发生,也没什么信息量,特别是概率为1的确定事件发生以后,不会给人以任何信息量。因此随机事件的自信息量 $I(x_i)$ 是该事件发生概率 $p(x_i)$ 的函数,并且 $I(x_i)$ 应该满足以下公理化条件。

(1) $I(x_i)$ 是 $p(x_i)$ 的严格递减函数。当 $p(x_1) < p(x_2)$ 时, $I(x_1) > I(x_2)$, 概率越小,事件发生的不确定性越大,事件发生以后所包含的自信息量越大。

(2) 极限情况下,当 $p(x_i) = 0$ 时, $I(x_i) \rightarrow \infty$; 当 $p(x_i) = 1$ 时, $I(x_i) = 0$ 。

(3) 从直观概念上讲,由两个相对独立的不同的消息所提供的信息量应等于它们分别提供的信息量之和,即自信息量满足可加性。

可以证明,满足以上公理化条件的函数形式是对数形式。

定义 2.1 随机事件的自信息量定义为该事件发生概率的对数的负值。设事件 x_i 的

概率为 $p(x_i)$, 则它的自信息量定义为

$$I(x_i) = -\log p(x_i) = \log \frac{1}{p(x_i)} \quad (2.1)$$

从图 2.1 中可以看到上述自信息量的定义正是满足公理性条件的函数形式。在它的定义域 $[0, 1]$ 内, 自信息量是非负的。

$I(x_i)$ 代表两种含义: 在事件 x_i 发生以前, 等于事件 x_i 发生的不确定性的度量; 在事件 x_i 发生以后, 表示事件 x_i 所含有或所能提供的信息量。在无噪信道中, 事件 x_i 发生以后, 能正确无误地传输到受信者, 所以 $I(x_i)$ 就等于受信者接收到 x_i 后所获得的信息量。这是因为消除了 $I(x_i)$ 大小的不确定性, 才获得这么大小的信息量。

自信息量的单位与所用对数的底有关。

(1) 通常取对数的底为 2, 信息量的单位为比特 (bit, binary unit)。当 $p(x_i) = 1/2$ 时, $I(x_i) = 1$ bit, 即概率等于 $1/2$ 的事件具有 1 bit 的自信息量。例如, 一枚均匀硬币的任何一种抛掷结果均含有 1 bit 的信息量。比特是信息论中最常用的信息量单位, 当取对数的底为 2 时, 2 常省略。注意: 计算机术语中 bit 是位的单位 (bit, binary digit), 与信息量单位不同, 但有联系, 1 位的二进制数字最大能提供 1 bit 的信息量。

(2) 若取自然对数 (以 e 为底), 自信息量的单位为奈特 (nat, natural unit)。理论推导中或用于连续信源时用以 e 为底的对数比较方便。

$$1 \text{ nat} = \log_2 e \text{ bit} = 1.443 \text{ bit}$$

(3) 工程上用以 10 为底较方便。若以 10 为对数底, 则自信息量的单位为哈特莱 (Hartley), 用来纪念哈特莱首先提出用对数来度量信息。

$$1 \text{ Hartley} = \log_2 10 \text{ bit} = 3.322 \text{ bit}$$

(4) 如果取以 r 为底的对数 ($r > 1$), 则

$$I(x_i) = -\log_r p(x_i) \quad r \text{ 进制单位}$$

$$1 r \text{ 进制单位} = \log_2 r \text{ bit}$$

【例 2.1】 (1) 英文字母中“a”出现的概率为 0.064, “c”出现的概率为 0.022, 分别计算它们的自信息量。

(2) 假定前后字母出现是互相独立的, 计算“ac”的自信息量。

(3) 假定前后字母出现不是互相独立的, 当“a”出现以后, “c”出现的概率为 0.04, 计算“a”出现以后, “c”出现的自信息量。

解 (1) $I(a) = -\log 0.064 = 3.966 \text{ bit}$

$$I(c) = -\log 0.022 = 5.506 \text{ bit}$$

(2) 由于前后字母出现是互相独立的, “ac”出现的概率为 0.064×0.022 , 所以

$$I(ac) = -\log(0.064 \times 0.022) = -(\log 0.064 + \log 0.022)$$

$$= I(a) + I(c) = 9.472 \text{ bit}$$

即两个相对独立的事件的自信息量满足可加性, 也就是由两个相对独立的事件的积事件所提供的信息量应等于它们分别提供的信息量之和。

(3) “a”出现的条件下, “c”出现的概率变大, 它的不确定性变小。

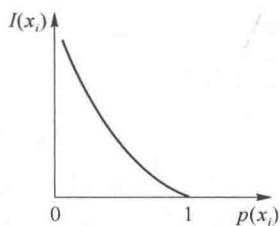


图 2.1 自信息量

$$I(c|a) = -\log 0.04 = 4.644 \text{ bit}$$

2.1.2 互信息

定义 2.2 一个事件 y_j 所给出关于另一个事件 x_i 的信息定义为互信息, 用 $I(x_i; y_j)$ 表示。

$$I(x_i; y_j) = I(x_i) - I(x_i | y_j) = \log \frac{p(x_i | y_j)}{p(x_i)} \quad (2.2)$$

互信息 $I(x_i; y_j)$ 是已知事件 y_j 后所消除的关于事件 x_i 的不确定性, 它等于事件 x_i 本身的不确定性 $I(x_i)$ 减去已知事件 y_j 后对 x_i 仍然存在的不确定性 $I(x_i | y_j)$ 。互信息的引出, 使信息的传递得到了定量的表示。

【例 2.2】 某地二月份天气出现的概率分别为: 晴 1/2, 阴 1/4, 雨 1/8, 雪 1/8。某一天有人告诉你: “今天不是晴天”, 把这句话作为收到的消息 y_1 , 求收到 y_1 后, y_1 与各种天气的互信息量。

解 把各种天气记作 x_1 (晴), x_2 (阴), x_3 (雨), x_4 (雪)。收到消息 y_1 后, 各种天气发生的概率变成了后验概率:

$$\begin{aligned} p(x_1 | y_1) &= \frac{p(x_1 y_1)}{p(y_1)} = 0 \\ p(x_2 | y_1) &= \frac{p(x_2 y_1)}{p(y_1)} = \frac{1/4}{1/4 + 1/8 + 1/8} = \frac{1}{2} \\ p(x_3 | y_1) &= \frac{p(x_3 y_1)}{p(y_1)} = \frac{1/8}{1/4 + 1/8 + 1/8} = \frac{1}{4} \end{aligned}$$

同理

$$p(x_4 | y_1) = \frac{1}{4}$$

根据互信息量的定义, 可计算出 y_1 与各种天气之间的互信息:

$$\begin{aligned} I(x_1; y_1) &= \log \frac{p(x_1 | y_1)}{p(x_1)} = \infty \\ I(x_2; y_1) &= \log \frac{p(x_2 | y_1)}{p(x_2)} = \log \frac{1/2}{1/4} = 1 \text{ bit} \\ I(x_3; y_1) &= \log \frac{p(x_3 | y_1)}{p(x_3)} = \log \frac{1/4}{1/8} = 1 \text{ bit} \\ I(x_4; y_1) &= \log \frac{p(x_4 | y_1)}{p(x_4)} = \log \frac{1/4}{1/8} = 1 \text{ bit} \end{aligned}$$

2.2 平均自信息

2.2.1 平均自信息的概念

自信息量是信源发出某一具体消息所含有的信息量, 发出的消息不同它的自信息量就不同, 所以自信息量本身为随机变量, 不能用来表征整个信源的不确定度。我们用平均自信息量来表征整个信源的不确定度。平均自信息量又称为信息熵、信源熵, 简称熵。