

国家863计划项目，国家自然科学基金项目研究成果

信息共享理论与 网络体系结构

邢 玲 马建国 马卫东 著



科学出版社

内 容 简 介

本书针对信息交换与信息共享两大信息服务方式的博弈推进,研究了两者的关联与本质差别,建立信息共享的理论体系,揭示信息不守恒原理,进一步完善了信息共享领域的理论基础。通过典型信息网络结构的分析,提出了信息共享的播存结构和主动服务模式,从而构建了信息共享的网络体系结构。以语义 Web 为牵引,探讨了信息共享的语义理解和信息安全问题,建立了基于 UCL 的语义计算与内容管理体系,为构建安全聪明的信息网络做了科学探索。以信息共享网络的应用模式作为全书的尾声,提出了双结构信息共享网络的实践和网络融合的战略发展,对于构建我国下一代信息网络具有重要的参考价值和实践意义。

本书可供高等院校通信、计算机、物联网等专业的师生作为教学参考书,也可供信息类相关领域的研究工作者和实践工作者参考。

图书在版编目(CIP)数据

信息共享理论与网络体系结构/邢玲,马建国,马卫东著. —北京:科学出版社, 2011

ISBN 978-7-03-032466-5

I. 信… II. ①邢…②马…③马… III. ①信息-网络-研究 IV. G202

中国版本图书馆 CIP 数据核字(2011)第 200378 号

责任编辑:余 丁 / 责任校对:朱光兰

责任印制:赵 博 / 封面设计:耕 者

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

新 蕾 印 刷 厂印刷

科学出版社发行 各地新华书店经销

*

2011 年 10 月第 一 版 开本: B5(720×1000)

2011 年 10 月第一次印刷 印张: 14 3/4

印数: 1—2 000 字数: 278 000

定价: 58.00 元

(如有印装质量问题,我社负责调换)

前 言

随着人类社会化进程的加快,网络信息服务方式的两极化也越发明显:一类是以信息交换为目的的信息对流;另一类是以信息共享为基础的信息辐射。互联网作为优秀的信息交换结构,应对人类信息共享服务已经显得疲惫不堪,这便激起了作者探索信息共享理论与网络体系结构的积极性。

作者在“十五”和“十一五”期间先后承担了多项与信息共享理论及其关键技术相关的国家级项目。通过国家 863 计划项目“第五媒体物理链路传输协议研究”(2002AA121063)和国家自然科学基金“信息共享的第二类网络”(60272014)的项目研究,明确了现有互联网存在结构性共享信息冲突和语义障碍,从而突出了研究信息共享网络体系结构的重要意义。通过国家 863 计划“具有语义标引的互补结构网络”(2005AA121520)和“多层网络数据语义分类与理解技术研究”(2007AA01Z151)的资助研究,探索了信息共享网络的分层模型、信息内容理解技术。通过国家自然科学基金重点项目“非对称广域覆盖信息共享网络理论与关键技术”(60932005)和“新一代广播网的语义计算与内容管理机制研究”(60902021)以及国家重点基础研究发展计划(973 计划)子课题“融合网络环境下公共信息主动服务理论研究”(2010CB328104-01)的研究,探索了信息共享的理论基础、信息的语义理解与语义计算及其关键技术。通过国家自然科学基金面上项目“互联网冗余流量的测量及其特性分析”(61171109)的前期研究,发现了由 Shannon 忽略语义特性而带来的信息传输有效性和可靠性问题。对信息共享理论与网络体系结构的研究是多方位的、螺旋式推进的,在充满荆棘的研究之路上,我们品尝着成功的喜悦和沮丧的辛酸。

本书以李幼平院士指导的三名博士——邢玲、马建国和马卫东——的博士学位论文以及相关研究成果为内容基础。全书共分为三大部分。第一篇,信息共享理论:介绍了信息共享的基本概念与特征,提出了信息共享系统的一般模型以及栅格模型,分析了信息共享网络的动力学特性,奠定了信息共享网络的理论基础。第二篇,信息共享网络的体系结构:对典型的信息网络结构进行了分析,提出了信息共享的播存结构,解析了信息共享中的语义处理问题,完善了信息共享网络体系结构的系统工程。第三篇,信息共享网络的应用模式:提出了基于播存理论的双结构信息共享网络,实现了网络融合的信息共享重要实践。

本书的撰写汇集了很多人的辛勤劳动。马建国负责全书的内容规划和体系架构搭建,以及撰写过程中的组织和协调工作;邢玲参加了第 1~5、7 章的撰写;

马卫东参加了第 3、5、6 章的撰写;马强参加了第 7、8 章的撰写;黎茂锋参加了第 6 章的撰写;肖宇峰参加了第 7 章的撰写。课题组的研究生贺梅、胡金军、湛煊、刘鹏、杨震泉对本书的排版、参考文献的校对,以及图表的规范等做了大量而细致的工作。全书由邢玲、马建国和马卫东统稿。

本书的撰写得到了国家自然科学基金委员会和国家 863 计划办公室的大力资助。在此,感谢科学出版社对本书的出版给予的关心和支持!同时也感谢在本书撰写前后,很多前辈、同行和学者对于信息共享理论的热情参与、研讨,以及给我们的鼓励。特别要感谢导师李幼平院士带领我们进入极富挑战性的信息共享理论研究领域,并以孜孜不倦的探索精神推动着研究的深入,为我们开拓了极其广阔的思维空间,使我们走在领域前沿!

限于我们的水平,加之时间仓促,书中难免存在不妥之处,恳望读者不吝赐教!

作 者

2011 年 6 月于西南科技大学

目 录

序 前言

第一篇 信息共享理论

第 1 章 网络信息的基本概念与特征	3
1.1 信息的定义	3
1.1.1 狭义定义	3
1.1.2 广义定义	4
1.1.3 基本性质	5
1.1.4 网络信息定义	6
1.2 信息场的数域空间	7
1.3 信息场的空间特性	9
1.3.1 信息通量和散度	9
1.3.2 信息环量和旋度	11
1.4 信息场的统计特性	11
1.4.1 信息热度	11
1.4.2 信息梯度	12
1.4.3 信息阶	13
1.5 信息共享的概念描述	14
1.5.1 信息不守恒原理	14
1.5.2 信息的共享增益	15
1.5.3 信息的共享度	16
1.5.4 信息共享的量纲	17
1.5.5 信息共享的二元属性	18
参考文献	19
第 2 章 信息共享系统模型	21
2.1 信息共享系统的基本特征	21
2.2 信息共享系统的一般模型	22
2.3 信息共享系统的栅格模型	23
2.4 信息传输的能量与效率	25

2.4.1	信息传输的能量代价	25
2.4.2	信息传输的效率问题	28
参考文献		29
第3章	信息共享网络的动力学特性	30
3.1	Internet 与复杂网络	30
3.2	复杂网络演化模型	31
3.2.1	随机网络模型	31
3.2.2	无标度网络模型	32
3.2.3	小世界网络模型	33
3.2.4	Internet 与 WWW 的拓扑特性	35
3.3	Internet 物理特性测量与分析	37
3.3.1	Internet 直径测量	37
3.3.2	Internet 节点度分布	41
3.3.3	互联网络效率测量	42
3.4	Internet 用户需求行为实证研究	44
3.4.1	Internet 用户需求分析	44
3.4.2	用户访问的二分网络模型	45
3.4.3	用户需求实验结果	46
3.4.4	用户需求实验分析	46
3.5	用户需求行为对 Internet 整体动力学特性的影响	50
3.5.1	用户需求幂律特性分析	50
3.5.2	用户需求的虚拟网络演化模型	50
3.5.3	仿真结果	53
3.6	网络资源与用户需求行为幂律特性的启示	57
参考文献		58

第二篇 信息共享网络的体系结构

第4章	典型信息网络结构分析	63
4.1	常见网络体系基础结构	63
4.1.1	OSI	63
4.1.2	TCP/IP	64
4.1.3	CS/BS	66
4.2	WWW 与 Web 服务	67
4.2.1	Web 服务概念与实现架构	68
4.2.2	Web 服务支撑技术及其实现	69

4.3 计算资源共享网络	71
4.3.1 网格技术	71
4.3.2 云计算技术	74
4.4 存储资源共享网络	79
4.4.1 P2P 网络	79
4.4.2 CDN 网络	83
4.4.3 网络 Cache 技术	85
4.5 数据广播网络	87
4.5.1 数字电视广播系统的划分	88
4.5.2 数字电视与数据广播的标准	89
4.5.3 数据广播网络结构存在的问题	102
4.5.4 下一代数据广播结构	103
参考文献	106
第 5 章 信息共享的主动服务与播存结构	108
5.1 信息共享主动服务体系结构	108
5.1.1 主动服务体系构架	108
5.1.2 主动服务网络层次模型	111
5.1.3 主动服务业务流生成与内容解析	112
5.2 信息共享的发布/订阅主动服务	113
5.2.1 Web 信息采集	113
5.2.2 Internet 上的发布/订阅系统	115
5.2.3 信息订阅/发布主动服务系统	118
5.3 信息共享的播存结构	120
5.3.1 播存结构的定义	120
5.3.2 播存结构的特征	123
5.3.3 播存结构的共享意义	124
5.4 基于播存结构的主动服务	125
5.4.1 播存结构与透明计算	127
5.4.2 播存结构的边缘存储机制	129
5.5 播存结构系统成本的最优解	130
参考文献	132
第 6 章 信息共享的语义处理	134
6.1 元数据与语义 Web	134
6.1.1 元数据	134
6.1.2 语义 Web 及其体系结构	137

6.1.3	RDF 与 RDF schema	139
6.1.4	语义网设计语言	142
6.2	语义处理的聚类与分类	143
6.2.1	聚类与分类	143
6.2.2	分类分析与文本分类	145
6.2.3	聚类分析与文本聚类	147
6.2.4	语义网存在的问题	150
6.3	统一内容定位技术	151
6.3.1	UCL 定义	151
6.3.2	UCL 与内容分类	152
6.3.3	UCL 运算	153
6.4	UCL 应用	156
6.4.1	UCL 应用模型	156
6.4.2	UCL 交换模型	156
6.4.3	UCL 应用示例	158
6.5	UCL 设计规范及应用	161
6.5.1	媒体结构与本体结构	161
6.5.2	Web 信息的 UCL 规范设计	164
6.5.3	数字视频广播的 UCL 规范设计	165
	参考文献	168

第三篇 信息共享网络的应用模式

第 7 章	双结构信息共享网络	173
7.1	双结构互联网	173
7.1.1	双结构互联网的模型	173
7.1.2	双结构互联网的结构特点	175
7.1.3	双结构互联网的实现	177
7.2	双结构电视网络	178
7.2.1	数字电视的发展现状	178
7.2.2	播存电视	179
7.2.3	双结构电视的网络架构	180
7.2.4	实验研究	181
7.3	双结构移动通信网络	185
7.3.1	移动通信网络的现状与发展	186
7.3.2	双结构移动通信网络的结构	188

7.3.3 中国移动多媒体广播	189
7.3.4 双结构移动通信网络的展望	190
参考文献	190
第8章 网络融合——信息共享的重要实践	192
8.1 网络的融合发展现状	192
8.1.1 互联网	192
8.1.2 电信网	195
8.1.3 广播网	197
8.1.4 物联网	200
8.2 信息共享技术在网络融合中的应用	202
8.2.1 可管可控的信息共享融合网络	203
8.2.2 语义计算在网络融合中的应用	204
8.2.3 可信计算在网络融合中的应用	206
8.2.4 主动服务在网络融合中的应用	208
8.3 网络融合的发展前景	210
参考文献	211
缩写表	213

第一篇 信息共享理论

第 1 章 网络信息的基本概念与特征

网络,原指用一个巨大的虚拟画面把所有的东西连接起来。用数学语言来描述就是节点和连线构成的图,它表示研究诸对象及其相互关系。除了数学定义外,网络还有具体的物理含义,即网络是从某种相同类型的实际问题中抽象出来的模型,如生物网络、交通网络、社会网络等。我们要研究的是信息网络,它是网络设施及其传输的各类信息所构成的复杂网络。

在对复杂网络进行刻画和分析之前,信息的描述是必不可少的。信息以及由此形成的信息场是信息共享理论存在的基础。本章首先介绍学术界关于信息定义的主要观点,在此基础上,阐明本书作者对于信息,特别是网络信息的理解,给出网络信息的复数定义。同时,对信息场的空间和统计特性进行描述,以便建立信息共享的基本概念,从定性与定量两个方面来揭示信息共享的本质。

1.1 信息的定义

信息,古人称音讯、消息。例如,南唐李中《碧云集·暮春怀故人》诗曰“梦断美人沉信息,目穿长路倚楼台”,明朝李昌祺《剪灯馀话·至正妓人行》云“荡子江湖信息稀,疲兵关塞肌肤裂”,即是此意。信息,英语中常用 information 来表述,它是从原词义“消息、情报、知识”引申而来的。但是,作为一个严谨的科学术语,到目前为止,信息的定义却没有一个统一的观点,这是由它的极端复杂性决定的。信息的表现形式数不胜数:声音、图片、温度、体积、颜色等。信息的分类也不计其数:电子信息、财经信息、天气信息、生物信息等。同时,信息作为知识和智慧的转化基础,它应该从各个不同层面和形态反映事物的本质特征。因此,要对信息作一个严密而又具有普适性的定义,就必须从本质上来把握信息。现在学术界的主要观点可以从狭义和广义范围来概括。

1.1.1 狭义定义

美国数学家,信息论的奠基人 Shannon 在他的著名论文《*A mathematical theory of communication*》中以概率论为工具,探讨了一系列通信工程中的基本理论问题^[1]。虽然文中并没有直接叙述信息的定义,但是他在进行信息定量计算的时候明确地把信息定义为随机不确定性之差,计算信息量 H 的公式为(一个信息由 n 个符号所构成,每个符号出现的概率为 p)

$$H = - \sum_{k=1}^n p_k \log p_k \quad (1-1)$$

这个公式和热力学熵的计算方式一样,故也称为信息熵。从公式可知,当符号出现的概率平均时,不确定性最高,信息熵也最大。所以,信息可以视为不确定性或者选择自由度的度量。

美国数学家,控制论的奠基人 Wiener 在他的《控制论》^[2]一书中认为,信息是“我们在适应外部世界,控制外部世界的过程中同外部世界交换的内容的名称”。在这里,他把人与外部环境交换信息的过程看做是一种通信的过程。正如熵是无组织程度的度量一样,消息集合所包含的信息就是组织程度的度量。事实上,完全可以将消息所包含的信息理解为负熵。

控制论的另一位奠基人,英国生物学家 Ashby 在 *Introduction to Cybernetics*^[3]一书中对信息提出了另一种理解。他认为信息的本性在于事物本身具有变异度。他首先引入了“变异度”的概念。他定义,任何一个集合,它所包含元素的数目的以 2 为底的对数就称为这个集合的变异度。不难证明,变异度实际上是 Shannon 熵的特殊情形:设某集合 X 有 N 个元素,每个元素出现的概率都等于 $1/N$,那么这个集合的 Shannon 熵就等于它的变异度。

意大利学者 Longo 在 *Information Theory: New Trends and Open Problems*^[4]中认为,信息是反映事物的形成、关系和差别的东西,它包含于事物的差异之中,而不在事物本身。

实际上,Shannon、Wiener、Ashby 和 Longo 等人对于信息都有着共同的理解,他们认为信息是负熵,是“用以减少不确定性的东西”,有差异就有信息;这些认识比仅仅把信息看做通信内容要深刻得多。

在熵、变异度这些概念的基础上还延伸出了一些新的说法:“信息是被反映的差异”,“信息是被反映的变异度”等。这些说法在“差异”和“变异度”的概念之上加了限制,其可用范围就更狭窄了。总的说来,虽然信息被赋予了不同的定义,给出了不同的理解,然而人们却不满足狭义的信息定义,还在不停地寻求更合理、更广泛的信息内涵。

1.1.2 广义定义

在漫长的人类历史发展进化和不断积累的科学技术实践过程中,信息一开始都是表现为真伪难辨的消息群——消息是信息的外壳,而信息则是消息的内核;信息也不同于信号——信号是信息的载体,信息则是信号所承载的内容;当然,信息也不同于数据——数据是记录信息的一种形式,同样的信息也可以用文字或图像来表述;信息还不同于情报和知识。虽然可以从侧面讨论“信息不是什么”的问题,但是仍然无法回避“信息是什么”的问题。因此,信息最广泛、最普遍的定义认

为：“信息即事物运动的状态与方式。”这个定义不仅能涵盖所有其他的信息定义，还可以通过引入约束条件转换为所有其他的信息定义。

本体论认为，信息是事物运动的状态和方式，而与是否被其他事物反映无关^[5]。如果对于某个随机信息 X ，它有 N 种可能的运动状态，设其状态空间为 $x_1, x_2, \dots, x_n, \dots, x_N$ ，而这些状态的运动方式是按照某种概率分布规律 $p_1, p_2, \dots, p_n, \dots, p_N$ 进行的，那么，由状态空间与概率分布结合而成的概率空间就能充分地刻画这个信息。

认识论认为，信息是认识主体关于某事物所处状态及其变化方式的表述。若用参量 C 、 T 和 U 来刻画“状态/方式”在形式上的肯定度、逻辑上的真实性和它们对于主体目标而言的效用度，那么，对于一个状态离散且状态数为 N 的事物 X ，它的认识论信息可以用一个相应的矩阵表示^[5]

$$\begin{bmatrix} X & x_1 & x_2 & \cdots & x_N \\ C & c_1 & c_2 & \cdots & c_N \\ T & t_1 & t_2 & \cdots & t_N \\ U & u_1 & u_2 & \cdots & u_N \end{bmatrix}$$

(1-2)

由上可知，Shannon 信息是认识论信息的特殊情形。Shannon 信息论只关注噪声背景下“随机波形”的复制，状态的“形式肯定度”则退化为状态的出现概率，逻辑真实度和效用度也都被忽略。

1.1.3 基本性质

由于信息概念的复杂性，钟义信教授在《信息科学原理》^[5]一书中指出，定义信息的时候应引入约束条件。在不同条件限定下的信息，可以有不同的定义表述。因此，为了得到清晰准确的认识，应当根据不同的条件，区分不同的层次来给出信息的定义。

最高的层次是最普遍的层次，没有约束条件的层次，称为“本体论层次”；在这个层次上定义的信息是最广泛意义的信息，它的适用范围最广。如果在此基础上引入一个约束条件，则最高层次的定义就退变为次高层次的定义；而次高层次信息定义的适用范围就比最高层次定义的范围要窄。所引入的约束条件越多，定义的层次就越低，它所定义的信息的适用范围也就越窄。这样，根据引入条件的不同，就可以给出一系列不同层次和不同适用范围的信息定义，这些不同信息定义的系列就构成了信息定义的体系。这种用区分层次来定义信息的思路，可以用表 1-1 来表示^[5]。

表 1-1 信息的定义体系

约束条件	定义层次	适用范围
无	最高	最广

续表

约束条件	定义层次	适用范围
一个较弱的约束	次高	次广
⋮	⋮	⋮
多个较强的约束	较低	较小
最强的约束	最低	最小

尽管从不同的角度出发对信息有不同的定义,但是对于信息的一些基本性质还是得到了共识:

(1) 普遍性。由于信息以客观事物的状态和运动方式来描述,客观事物的普遍性就决定了信息的普遍性。信息无所不在,充满了整个物质世界。

(2) 无限性。反映客观事物的信息如同客观物质世界一样是无限丰富的、永恒的。在整个宇宙时空中,由于客观事物及其发展变化的无限多样性,所产生的信息自然也是无限多样的。当代社会信息量的暴涨正说明了这一特性^[6]。

(3) 动态性。一切事物都是在不断变化发展的,信息也必然随之运动发展,其内容、形式、容量都会随时间而改变。

(4) 可传递性。信息具有离开产生它们的物质而相对独立存在的能力,它们可以通过各种媒介在人—人、人—物、物—物等之间传递,也可以为适应某种需要而变换内容和形式。

(5) 可识别性。信息代表着客观事物所要表达的内容,在信息传递和解析的过程中,信息具有各方均能识别的内容,可以从具体的信号中获取内容。

(6) 可共享性。信息与物质、能量显著不同的是,信息在传递过程中并不是“此消彼长”。同一信息可以在同一时间被多个主体共有,而且还能够无限地复制、传递。

从以上关于信息的定义和性质的讨论中,我们可以看到:信息既不是物质本身,也不等同于能量。它在不同的约束条件下有着不同层次的定义。在本书中,我们需要讨论的是网络信息,此时的信息有具体的应用环境,有特定的运动状态和运动方式,因而需要有不同的描述方式和定义表达。

1.1.4 网络信息定义

在信息网络中,由于网络环境的特殊性,信息的定义应该适合网络的传输与存储,并具有可识别性和可共享性。因此,在本体论层次上,我们引入一个具有实际意义的约束条件,此时的信息可以表述为:在信息网络中,存储和转发事物的运动状态及其变化方式,包括状态及其变化方式的形式和含义。网络信息运动状态及其变化方式的外在形式称为网络信息的语效信息,用 Q 表示;内在含义称为网

络信息的语义信息,用 R 表示。如果对网络信息进行矢量化的复数表示——实部是网络信息的语效特征、虚部是网络信息的语义特征,将网络信息运动体现的外在效用和其对应的语义内涵关联起来,那么,网络信息 H_w 的定义可以用语效信息和语义信息来表示

$$H_w = Q + R * i \quad (1-3)$$

式中, i 表示虚数单位; Q 和 R 是矢量化的信息分量。语效信息是网络信息一般意义的体现,而语义信息则是沟通信息网络中其他信息的桥梁。通常网络信息之间的物理距离是指信息的实数距离,而逻辑距离则是指信息的语义距离,也就是虚数距离。

信息网络作为信息传输系统来说,应该包含三个基本要素,即信源、信道、信宿。信源是以信号的形式发送信息的主体;信道是指传送信号的通道或媒介。尽管在传输中可能对信号进行编码、调制等各种形式的转换或处理,但到达信宿的仍然是信号而不是信息。只有当信宿具备相应的识别能力时,信号才能转化为信息。因此,信息的语效信息 Q 仅表达了信息的外在形式,语义信息 R 才是真正能转换为有用信息的部分。当信源与信宿都同时对语义信息 R 有所表述的时候,信息才能被理解。例如,两个人想交谈,其中一个是一句中文都不懂的外国人,一位是一句外语都不懂的中国人,双方尽管都听到了对方的语言信号(语效信息 Q),但是由于不具备对方语言的识别能力(语义信息 R),从而无法得到对方传来的信息。再如,一个正常人和一位不能从对方口唇变化识别语言的聋哑人,聋哑人虽为信宿,但依然无法了解对方的谈话内容;如果正常人把表达的内容用手语(语义信息 R)表示,当聋哑人具备手语识别能力时,则会从正常人发出的信号中获取信息^[7]。

以上说明,在信息网络或者信息系统中,只有同时具有语效信息和语义信息的信息才能成为真正意义上的网络信息,而此时信息的定义需要用复数形式来完整地描述。事实上,表示语效特征的信息主体反映了网络信息变化或者各类现象属性标志的集合,它是信息的最基本描述层次;而语义信息是语效信息的高阶量,它对语效信息进行抽象和解析,两者的复数形式在网络信息空间中能确定网络信息之间的距离和相似度等信息间的基本物理量。

1.2 信息场的数域空间

通常人们研究的信息是在“信息路”的概念下描述和表征的,用“路”来分析问题时更简单、实用,但在逻辑严密性和理论完整性上不够完美。因此,我们在网络信息中引入了“信息场”的概念。信息场是信息载体在其中发挥作用的空間,它是在时间和空间上两种分别无限和无边的因素相互结合的虚拟化综合结构。尽管

“场”是看不见摸不着的,但它确实存在。信息路研究的是信息场中某种特定的信息现象,这种特定的现象所遵循的规律必然可以从宏观的场空间中得到,因此信息场的性质更具普适性。

场的概念最初是物理学中研究的问题,往往涉及分布在一定空间区域中的物理量,如引力场、电场、磁场等。从某种意义上讲,引力场、电场、磁场也都是信息场,因为它们都负载着信息。在这里,我们讨论的信息场,是指网络环境中因数据信息流动而存在的空间。

首先将场的概念推广到数域空间,此时我们讨论的数域空间都是实数空间。设有 m 维数域空间 R^m 中的数据对象集 $D = \{X_1, X_2, X_3, \dots, X_n\}$, 其中 $X_i \in R^m$ 且 $i = 1, 2, \dots, n$ 。数据对象集 D 中的每个对象都是信息网络中的一个信息,由于空间中信息数据的流动,就形成了信息场^[8]。若对信息场的数域空间进行定量描述,就有下面的定义。

定义 1-1 设 Q^n 是 m 维欧氏空间 R^m 中的单连通凸域, $\forall a \in Q^n$, 若在 Q^n 中存在唯一的 Shannon 熵值 $H_a(X)$ 与 X 对应, 其中 X 取值于数据对象集 D , 则称 (Q^n, H_a) 为信息场, 有时简称 (Q^n, H_a) 为 H_a 。

在信息场中, 为了将信息的描述定量化, 在空间的任何一点都唯一地对应着一个随机变量 X , 根据 Shannon 信息论, 它含有一定的信息量。相对于其他点, 它既是信源, 同时也是信宿, 我们称之为信息点。也就是说, 每个信息点既向其他点传递信息, 同时, 也接收来自其他信息点的信息^[9]。因此, 在这个过程中存在着信息流, 形成信息场。由此可见, 信息场是双向的, 既有信息往外涌的“源”, 也有信息被吸入的“洞”。源和洞的物理含义可以通过信息的空间特性来体现, 将在 1.3 节中介绍。

信息场的数学刻画证实了它的客观存在, 可是, 其中信息点或者信息点集合之间存在着的信息传递该如何描述呢? 为此, 可以通过定义信息传递指数来揭示信息的传递规律。

定义 1-2 设 U, V 是信息场中的两个信息点集合, 则称

$$D_i(U; V) = I(U; V) / H(V) = 1 - H(V/U) / H(V)$$

为 U 对 V 的有向信息传递指数。同样, $D_i(V; U)$ 为 V 对 U 的信息传递指数。其中, $H(V)$ 、 $H(V/U)$ 分别是 V 的熵和 U 条件下 V 的条件熵, $I(U; V) = H(V) - H(V/U)$ 是 U 与 V 之间的互信息。

一般而言, 有向信息传递指数 D_i 不满足对称性。有时 $H(V/U)$ 被认为是 U 在向 V 传递信息过程中的信息损失。由于 $0 \leq H(V/U) \leq H(V)$, 所以 $0 \leq D_i \leq 1$ 。当 $D_i(U; V) = 0$ 时, U 对 V 不存在信息传递, 即信息全部损失; 当 $D_i(U; V) = 1$ 时, 此时 U 包含了 V 的全部信息, 这时认为没有信息损失^[9]。而在实际的信息网络中, 信息不会主动消失, 而是在传输的过程中发生丢包、延迟等网络现象引发的