

4 信息的增长

Word count: 2049

1. 增长的极限 h1

罗马俱乐部关于人类困境的报告

在罗马组成了一个旨在研究人类当前和未来处境问题的非正式国际性协会——罗马俱乐部。麻省理工学院学者丹尼斯·梅多斯领导的研究小组受罗马俱乐部委托以计算机模型为基础，在1972年提交了第一份报告，即《增长的极限》

正规的世界模型 h2

假如世界人口、工业化、污染、粮食生产与资源消耗按当时的增长趋势继续下去，全球的经济增长将在100年内达到极限

改变这类增长趋势并达到一种生态和经济稳定是可能的，全球经济的均衡状态将会使世界上每个人的基本需要得到满足

越早开始改变成功的可能性越大

指数增长的本质 h2

“当一个量在一个既定的时间周期中，其百分比增长是一个常量时，这个量就显示出指数增长。”

地球增长的极限会在之后11年中发生，较为可能的结果将是人口与工业生产力双方有相当突然的不可控制的衰退——过冲

过冲：状态的突然改变，导致人体难以适应

全球性平衡 h2

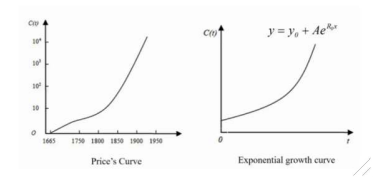
影响人们在各种平衡状态中进行选择的因素是该状态能够存在的时间和社会要想保持长时间的平衡状态，就需要一组最低限度的必要条件：

- 1. 资本设备与人口数量都是稳定的
- 2. 各种输入量和输出量的比例都保持在最低限度「出生、死亡、投资、折旧」
- 3. 资本与人口的水平及两者的比率应该依照社会的各项数值，尤其是技术进步的情况加以调整。

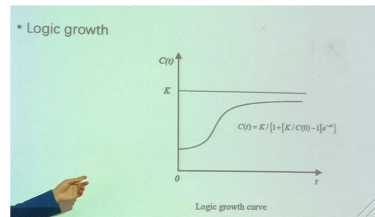
这些条件阐明了一种“动态”的平衡状态，其目的在于为社会创造自由

2. 信息增长的规律：增长 h1

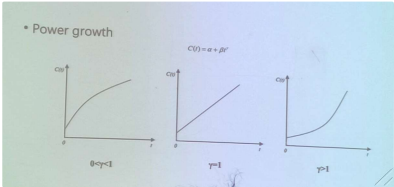
指数增长模型 h2



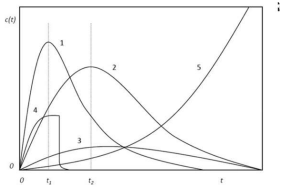
逻辑增长 h2



幂增长 h2



3. 信息增长的规律：老化 h1



Yerkes's framework: (1) initially much printed articles, (2) basic recognized work, (3) severely reflected work, (4) well-received but later extremely qualified work, and (5) genius

- 1. 初期被很多引用的文章
- 2. 基础型文章
- 3. 没什么反响的文章
- 4. 原本挺好突然拉倒的文章：可能证明有错误
- 5. 大天才：起步晚，增长快

Year 50%：一篇文章达到它被引量的50%的年份在文章整个历程中的位置，对同一出版年份的所有文件，将Year50%的函数的分位数25%和75%记录为“P25”和“P75”

- 昙花：“Year 50%”< P25
- 延迟：>P75
- 普通：P25< <P75

文献被引半衰期：达到该期刊当前总被引用次数50%时的年限；被引半衰期越长，证明期刊越长寿 cited half life

引用半衰期：期刊引用的文献中的前一半的文献所处的年代

普赖斯指数：假设现在有1000文献，其中有多少是近五年发表的，近五年的和1000篇的比值就是普赖斯指数；普赖斯指数越小，老化越慢

4. 信息分布规律 h1

4.1 布拉德福定律 h2

- 1. 找核心信息
- 2. 确定核心科学期刊索引：CSCD、ISTIC、SCI、CSSCI
- 3. 将期刊划分区域，每个区域的期刊的发文总量大致相等，每个区域的期刊数量大概为1：n：n²，从而找到核心期刊

文献为什么会分散？

- 1. 在某一学科中，写第一批论文时，人们首先把它寄给最合适的杂志发表，这些杂志随着该学科发展刊载越来越多这样的文章
- 2. 于是许多作者都希望把他们的文章发表在本专业的、以前发表了大量高质量论文的这类杂志上
- 3. 投稿数量大大增加，使杂志选择性增大，因而对文章的质量要求提高，杂志的威信也日益提高，这就产生了一些带“核心”性质的杂志，这种现象就是文献分布中的“堆加”效应——马太效应

4.2 齐普夫定律 h2

- 1. 通常用来预测全文本中的词频：对中频词有效
- 2. 在文本集、图书馆借书记录、指向网页的链接数量等领域都有应用。
- 3. 出现：很少的高频词，一些中频词，大量的低频词
- 4. 齐普夫定律并不是在统计学上精确的，但对于确定搜索引擎自动索引的词语权重具有重要意义
- 5. 齐普夫定律具有广泛的应用
 - 1. 在信息系统建设、词表管理和控制、自动标引、信息存储与检索方面都取得了许多应用成果
 - 2. 通过词频分布来揭示信息的分布规律则是齐普夫定律在理论上的重要开拓

4.3 洛特卡定律 h2

- 1. 描述科学家产出的能力，又被叫做负平方规律
- 2. 在特定领域内，产出x篇论文的作者数量占有所有作者数量比为 C/x^2
- 3. 产出一篇论文的作者数量占有所有作者数量的60%
- 4. 产出n篇文章的作者数是产出一篇文章作者数的 $1/n^2$

4.4 Base模型 h2

一种销售模型，可以预测最大销售量m，应用于引文领域，饱和度=目前的总被引量/m

4.5 延迟认可 h2

一篇重要的论文在发表很长时间后才被人们认识到它的价值

基于引文均值的方法：

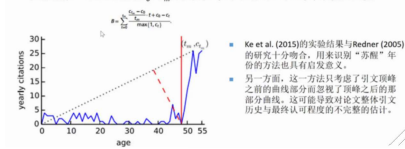
- 1. 设置标准的“沉睡期”「3-5」与“苏醒期”「至少4」
- 2. 当一篇论文在“沉睡期”内被引次数低于某个阈值且在“苏醒期”内被引次数高于某个阈值，就被视为“延迟认可”的论文
- 3. 但是如何划分时期和如何设定阈值是个大问题，还可能遗漏本应该入选的论文

基于百分位值的方法

- 1. Year 50%：论文达到一半引文数量所用的时间
- 2. P25：样本中最先的25%达到一半引文数量的论文所用的时间
- 3. P75，样本中最后的75%达到一半引文数量的论文所用的时间
- 4. Year 50% > P75, 延迟认可，< P25, 昙花一现
- 5. 不足：计算量很大，且标准过于宽松

美丽系数：

- 在引文曲线上的原点与最高点之间连起一条参考线，两点分别对应论文发表年份（记为 t_0 ）与达到最高引文的年份（记为 t_m ）。
- “美丽系数”=通过加和从 t_0 到 t_m 时段内参考线与引文曲线上每年数值的差值



- 这个指标最大的优势在于没有实参，但是可延展性有一定的缺陷，最大值的点可能会发生变化

GS指数

- 1. GS指数与引用延迟（D1）这两个指标并专门非针对量化“延迟认可”现象而被提出，它们考虑到整条引文曲线，适用于任何论文，两个指标具有相关性
- 2. Gs指数，基尼系数的改版指标，用来分析“睡美人”论文在“沉睡期”年度引文分布的不均衡程度
- 3. 如果Gs指数在（0.2，0.6）之间，则有更大可能被“唤醒”并获得认可

