



内容算法

把内容变成价值的
效率系统



闫泽华——著

图书在版编目 (CIP) 数据

内容算法：把内容变成价值的效率系统 / 闫泽华著

北京：中信出版社，2018.4

ISBN 978-7-5086-8527-4

I. ①内… II. ①闫… III. ①网络信息资源-资源利用 IV. ①G250.73

中国版本图书馆CIP数据核字 (2018) 第 007771 号

内容算法——把内容变成价值的效率系统

著 者：闫泽华

出版发行：中信出版集团股份有限公司

(北京市朝阳区惠新东街甲4号富盛大厦2座 邮编 100029)

承 印 者：北京通州皇家印刷厂

开 本：880mm×1230mm 1/32

印 张：8.25 字 数：120千字

版 次：2018年4月第1版

印 次：2018年4月第1次印刷

广告经营许可证：京朝工商广字第8087号

书 号：ISBN 978-7-5086-8527-4

定 价：58.00元

版权所有·侵权必究

如有印刷、装订问题，本公司负责调换。

服务热线：400-600-8099

投稿邮箱：author@citicpub.com

< 目 录 >

推荐序	… VII
自序	… XI
引言	… 001

Part 01

关于内容推荐

走近内容推荐 … 007

初步了解推荐系统架构，有助于我们更好地认知这一过程：信息是如何匹配给用户的，用户的选择又是如何影响后续信息分发过程的。

推荐系统架构初探 … 007

YouTube 和 Netflix 推荐架构参考 … 016

基于推荐架构的优化启示 … 018

推荐的起点：断物识人 … 025

断物识人，是一切推荐行为的起点。断物最简单的方式就是“贴标签”，识人是给目标用户“贴标签”的过程。

内容算法——把内容变成价值的效率系统

断物 … 025

识人 … 032

推荐算法：物以类聚，人以群分 … 039

推荐算法匹配用户和内容这一过程所做的就是“物以类聚，人以群分”。

物以类聚：基于内容属性的相似性推荐 … 039

人以群分：基于用户行为的协同过滤 … 041

从算法到应用 … 047

作为一套提供综合服务的推荐应用，仅有骨架是不够的，还需一副花樣皮囊，通过场景划分和交互界面来实现效能最大化。

场景划分 … 047

推荐系统评估指标 … 053

连接内容与人 … 059

从0到1积累基础数据的过程就是冷启动，其效果的好坏直接关系到用户端、作者端的满意度和留存率。

冷启动 … 059

兴趣探索 … 071

自媒体与平台 … 075

常见的推荐问题 ... 087

内容分发的初心：如你所愿，阅你所悦。

推荐重复 ... 087

推荐密集 ... 091

易反感内容 ... 093

时空限定内容 ... 097

带着偏见看推荐 ... 101

现在的内容推荐产品还远远没有达到理想态，其迭代过程中是存在种种瑕疵和待改进空间的。但发展的进程并不会因为人们的误解而止步。

信息茧房 ... 102

推荐会导致Low? ... 105

编辑、算法与社交，三分天下? ... 113

面对推荐系统 ... 123

从茫然无知到过度自信，再从过度自信到心存敬畏，对于自媒体和内容分发业务，我的感受如是。

人机大战：效率与目标之争 ... 124

数据分析驱动产品迭代 ... 127

个性化的好与好的个性化 ... 135

Part 02 关于自媒体

好内容为什么没人看 ... 139

当越来越多的内容分发平台开始引入机器推荐的时候，自媒体也许需要了解不同平台的分发规则，让有价值的内容获得应得的流量，进而让自己的品牌有更高的价值。

自媒体的数据分析 ... 143

量化一篇内容在不同平台的分发和受众反馈情况，用数据更好地辅助创作者稳定持续地产出 80 分以上的作品，在各个平台获得稳定的流量。

他山之石：BuzzFeed 简介 ... 144

内容阅读分析 ... 149

粉丝增长分析 ... 154

自媒体运营 ... 161

站在市场的角度重新审视内容，不夹杂创作情感的包装和营销，从而实现内容商品属性的最大化。

内容快销：标题党的二三事 ... 161

推荐平台优化：从 SEO 到 REO ... 167

粉丝运营：新时代的新问题 ... 173

全平台运营：从小作坊到 MCN ... 176

自媒体变现 ... 191

一切不以赚钱为目的的商业自媒体，都是耍流氓！

变现入门：平台分成 ... 192

广告变现：品牌的溢价 ... 205

自营电商：隔行如隔山 ... 217

内容付费 ... 227

当越来越多的自媒体高喊让内容不再只是流量，让内容回归其价值本质的时候，内容付费是否迎来了自己的春天，能够真正成为一门好生意呢？

缘何付费，规模几何 ... 235

内容付费平台展望 ... 242

< 推荐序 >

我在今日头条的工作是服务于创作者群体的，其间被问到最多、也最令人头痛的有两个问题。其中，第一个问题是：你们能不能帮我推一推这篇文章？

这时，我通常会解释：很抱歉，做不到，这里的内容会由算法自动推荐。

接下来，对方几乎无一例外地会对“算法”这个词表现出极大的兴趣，并提出第二个问题：那你给我讲讲，内容推荐是怎么回事儿？

大哉问。我的天，这可不是几句话能说清楚的事儿。

为了给出足够可靠的答案，我曾专门请教推荐公司算法的总负责人，他笑眯眯地反问我：“先说说

内容算法——把内容变成价值的效率系统

你的理解。”在听完我幼稚可笑的解释后，他把双手一摊：“你看，从字面意义上，我和你懂的也差不多啊……”

这当然是句玩笑话，真实的情况是：推荐算法实在太过复杂、精密，以至于熟稔此道的专业人士也难以向普通人言传，更别提做到简明扼要、面面俱到、通俗易懂了。

这正是闫泽华和他这本书的至为可贵之处。有句北京话叫“门儿清”，泽华对内容推荐就“门儿清”。他沉稳多思，出身于百度工程师，中途遵从内心做了产品经理，随时沉淀记录自己的判断和见解，后来又在顶级自媒体“凯叔讲故事”和算法分发的首倡者今日头条工作过。多种职业经历形成的多重视角使他能看得透彻，将准确的认知置于立体的坐标系之中，这才是超过一般从业者的地方。我相信，这些经历和特质不仅我自己不可比附，也是很多人无力效仿的。

今天，至少有4亿中国人通过算法获取内容，国内创作者群体半数以上流量来自内容推荐系统。但根据我的亲身体会，不仅多数读者和作者对这方面的情况知之甚少，甚至一些名头响亮的新闻学院教授、互联网评论家聊起这个话题，也十有八九是浮皮潦草，停留在大而化之的浅层次。泽华的这本书，技术出色而且界面友好，正适合对内容推荐有兴趣的各界人士拿来了解其机理。

一个事实是：高阶人群对内容推荐常持批判态度。以前，我和泽华讨论过原因，粗浅而言，一方面是由于算法目前仍明显有其不足和局限，另一方面是由于公众不知其运作逻辑。人这种动物，对不了解

的东西总是心存恐惧。

我们共同达成的结论是：作为一个读者，只要你的感受有代表性，我们就该做出改变，就该做得更好。可是如何实现呢？除了在日常工作中力求卓越，如今他写了这样一本小书，或多或少，也会帮助你洗脱成见。

作为一份迟来的答案，如果以后再次被人问到“内容推荐是怎么回事儿”的问题时，那么我会非常荣幸地推荐这本书给他。希望泽华累积了6年的经验和更久见识的认知能够广为人知。

吴达

今日头条前高级运营总监

< 自序 >

作为一个技术出身的产品经理，我一路在互联网领域兜兜转转，却始终在与内容信息打交道：从信息搜索到内容创作，从内容推荐到知识付费；做过平台也做过自媒体，做过图文和音频，也做过视频，甚至触碰过VR（虚拟现实）。

创作的繁荣、载体的多元、分发的进化，内容信息消费在快速的发展中。我很庆幸自己能有机会亲历这个发展历程，与信息时代共同前行。

一方面，我们同信息的联系变得无比紧密。

作为消费者，我们可以随时随地获取信息，甚至有人断言，人类将不需要太多的记忆，只要依赖外脑就能够帮助我们实现信息的有效处理。

作为生产者，我们可以随时随地地创造信息。从博客到微博，再到公众号，我们不仅获得了平等发声的权利，也获得了以内容获取关注的机会，自媒体真正成为可以带来真金白银的行业。

另一方面，海量的信息也不可避免地带来种种问题。

作为消费者，既要坚守本心，经得住八卦娱乐信息的诱惑，让自己的注意力不在 Kill Time（消磨时间）中虚度，还要有一双火眼金睛，能够从营销信息、标题党和虚假信息里识别真伪，不被虚假信息蒙蔽了双眼。

作为生产者，似乎更难选择。究竟是要追逐 10 万次以上的点击率、蹭热点，还是要坚守自己的本心？就算保证了内容上的坚持，又该如何处理运营和变现的问题？如何面对林林总总的渠道，是赚媒体广告费，还是做自营电商或知识付费，赚直营的钱……

凡有痛点，必有机会。

因为有信息查找的痛点，所以有了搜索引擎，让你可以一键直达，找到所需。

因为有信息发布的痛点，所以有了形形色色的创作工具和大大小小的创作平台，让你可以更酣畅地表达。

因为有信息阅览的痛点，所以有了推荐引擎，让人找信息变成了信息找人，被动消费诉求得到了极大的满足。

因为有信息筛选的痛点，所以有了知识付费，让你花钱买时间，让达人帮你筛选和过滤无效信息。

因为有信息消费的痛点，所以有了音频的繁荣，让你在无暇用眼

查看、用手翻阅的时候，亦能消费信息。

因为有信息表意的痛点，所以有了视频的繁荣，让你能以更丰富的方式消费和理解信息。

凡有机会，必有进化。

从信息搜索到内容推荐，是一种进化。

从免费内容到知识付费，是一种进化。

从图文自媒体到音频自媒体，再到视频自媒体，也是一种进化。

每一次进化，都会引发一群人的自问：跟得上吗？会掉队吗？

处于焦虑中的，不光有创作者，亦有平台。

载体迭代，比快更快。

分发迭代，比快更快。

与其焦虑，不如逐浪。

但是，在快速的迭代过程中，我相信一定有可以让我们慢下来的东西——那是内容的核。敬畏内容的价值，尊重它给受众带来的价值。

我带着这份敬畏，写下此书。希望在内容推荐百家争鸣的喧嚣里，给你带来一份关于内容推荐时代的参考。

它可能是错的，但一定是干的，一定是真的。

感谢你的阅读。

祝好！

闫泽华

< 引言 >

也许，你的生活中曾有过这样的场景：在上下班拥挤的地铁中打开资讯应用查看新闻或逸事，午休时间拿出手机登录淘宝想要剁手买买买，坐在家中在大众点评上搜寻周末的美食聚餐去处……这一个个生活中再普通不过的场景，其实都早已处在推荐系统的掌控中：

- 内容推荐，将用户感兴趣的话题和内容呈现在用户的眼前，你看到的是明星逸事，我看到的是影片点评。

- 购物推荐，根据用户的风格和偏好进行商品推介，你看到的是潮流高跟鞋，我看到的是复古马丁靴。

• 美食推荐，结合用户的口味和住址推荐食肆去处，你看到的是日式寿喜锅，我看到的是重庆九宫格。

我们在同一时间打开了同一个应用，选择了同样的频道，甚至输入了同样的搜索词，但投其所好的结果让我们与自己更感兴趣的信息相遇。

这，就是推荐的力量！

科技赋予的神力让个体价值前所未有地凸显了出来：个性化的推荐让人们不再被热点围绕、不再被排行榜左右，而是享受足够定制化的信息集合——资讯新闻、服饰妆容、饮食偏好，以及一个个让自己成为独一无二的个体的理由。

推荐是什么？

一言以蔽之，推荐是特定场景下人和信息更有效率的连接。

本书试图介绍的内容推荐系统，就是个性化推荐技术在内容消费场景里的一种应用，通过提供个性化的消费型内容来满足人们打发时间的需求。当你身边大量的信息获取途径，如今日头条、微博、网易新闻、手机百度，甚至 Wi-Fi（无线网）万能钥匙都涌上这条内容推荐赛道的时候，我们或许有必要对内容的个性化推荐了解一二。

对用户而言，这些纷繁复杂却又引人入胜的内容，究竟是怎么来到自己眼前的？内容推荐系统缘何如此了解自己的偏好，又是否会如评论家警示的那般让自己深陷信息茧房中呢？

对自媒体而言，与订阅分发、编辑分发不同，推荐分发会给分发模型带来怎样的变数？在这个全平台分发的“战国时代”，自媒体又该如何在多平台间左右逢源，给自己的内容曝光量和品牌影响力带来大的提升？

接下来，就让我们翻开这本写给小白和自媒体的内容推荐分发手册，一起走进内容推荐时代。



Part 01
关于内容推荐