

大嘴巴漫谈 数据挖掘

易向军

第2季
产品篇



大嘴巴带你玩转**数据挖掘**
贯穿**整个产品生命周期**的业务数据挖掘体系
大数据的小科普让你成为“**读心**”魔法师！



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

大嘴巴漫谈 数据挖掘

 易向军 

第2季
产品篇

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

《大嘴巴漫谈数据挖掘》系统而全面地描述了数据挖掘的基本概念、常用算法等。本书是《大嘴巴漫谈数据挖掘》的姊妹篇，在前作的基础上，它以产品为核心，按照产品发展的过程，依次详细分析产品策略研究期、产品概念评估期、产品研发期、产品测试期、产品导入期、产品发展期、产品成熟期、产品衰退期这8个产品发展的必经阶段所必须做的数据挖掘工作。针对产品每一阶段的不同特点，作者分享了数据挖掘的核心技能，并指出了每一阶段数据挖掘需要避免的坑。

本书适合想从事数据挖掘方面工作的初学者、数据分析爱好者、分析师，以及一线的数据挖掘开发人员参考阅读，也适合客户经理针对如何开展针对性的营销活动，避免客户流失而阅读学习，更适合产品经理阅读，因为针对如何预测产品的目标用户，促进用户活跃和业务有效使用，靠经验已经不行了，数据才最有说服力，更适合企业管理者将其作为一本通俗易懂的数据挖掘基础读物阅读学习，对下属的工作方向给予指导，以及适合教师学生数据挖掘课程辅导之用。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

图书在版编目（CIP）数据

大嘴巴漫谈数据挖掘.第2季,产品篇/易向军著.—北京:电子工业出版社,20163

ISBN 978-7-121-28237-9

I.①大… II.①易… III.①产品管理—数据采集IV.①TP274

中国版本图书馆CIP数据核字(2016)第042574号

责任编辑：孙学瑛

印 刷：中国电影出版社印刷厂

装 订：·,三河市华成印务有限公司

出版发行：电子工业出版社

北京市海淀区万寿路173信箱

邮编：100036

开 本：787×980 1/16 印张：14.5 字数：290

版 次：2016年3月第1版

印 次：2016年3月第1次印刷

定 价：66.00元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至zlts@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。

服务热线：（010）88258888。

大咖评论

在大数据连续多年成为流行热词，并不断地被媒体、投资人、政府挂在嘴边时，说实话，大多数人并不真正明白什么是“大数据”。如果你真的希望深入了解什么是大数据，那么大嘴巴漫谈将通过图文并茂的丰富案例讲解，轻松带你揭开大数据的面纱，一窥堂奥。

PreAngel投资人，zixia BBS，叽歪网创始人

李卓桓

世界在变，大数据已经来临。向军在数据行业钻研多年，他服务的客户包括中国电信和中国移动等顶级企业，本书不是学术论文，图文并茂，深入浅出，是一本难得的“大数据”佳作。

复星联合基金董事长，世界粤商大会秘书长

兰晓华

没有人会否认，数据是未来互联网的核心。而数据挖掘则是让数据得以应用的基础，但要用简单的语言说清楚数据挖掘这个专业性很强的事并不容易，向军在这本书中做到了，很了不起，我推荐！

阿里巴巴集团资深总监，yunOS产品总负责人

胡晓东

本书给人以耳目一新的感觉。首先，该书对关键技术阐述准确，可操作性强。其次，该书涵盖了主流的数据挖掘方法，对方法的阐述深入浅出，非常适合初学者作为入门读物。最为关键的是，该书形式生动活泼，使用了大量图表和例子，趣味性强。

北方工业大学信息学院副教授

宋威

大数据近几年在资本市场广受热捧，而相对美国而言，中国在大数据的应用上面还处在初级阶段，还有相当大的发展空间，《大嘴巴漫谈数据挖掘（第2季产品篇）》着重案例方面的提炼，相信会给大数据相关创业企业提供切实的参考价值！

CloseAVentures Founder& CEO

刘立伟

今天，海量涌出的数据刻画着我们生活的方方面面。谁能挖掘出这些数据里隐藏的故事，谁就抓住了未来的发展机遇。但空谈无用，该怎么实现数据挖掘，正是本书令人惊艳之处。易向军既懂专业理论又有深厚的实战经验，对于数据挖掘的剖析见解独到，引人入胜，让我获益良多。

ExceedData联合创始人，CTO

谢宁

商业模式的设计，离不开市场需求的支撑；市场需求的把脉，离不开市场调研的支撑；市场调研的精准，离不开数据挖掘的支撑。大数据时代，创业者一定要学会数据挖掘，《大嘴巴漫谈数据挖掘（第2季产品篇）》是一本不可多得的教材。

继业网创始人

李金宇

一图胜千言，这是我见过的很用心的数据挖掘入门的书，完全轻松的图表表达，而又能将数据挖掘背后复杂的统计知识和原理用图形轻松演绎，相信一个外行也能非常快地读懂本书。

菜鸟网络专家

樊春晖

当我第一次阅读本书的时候，我确实兴奋起来了，因为我很少见到能把数据挖掘介绍得这么通俗易懂，深入浅出而又理论联系实际的书。易向军的这本书做到了，这与他本人科班出身而且在数据挖掘领域有着十多年的丰富商业实践经验有关。

互联网资深人士

傅志华

大量趣味漫画与插图轻松吸引读者，但是仔细研读下去，会发现作者并不是为了“漫谈”而“漫谈”。本书其实涵盖内容广泛，是把数据挖掘所涉及的数学基础、模型和案例轻松、生动地传递给读者，使得晦涩的概念和技术显得通俗易懂，可以说是一本“讲得清、看得懂、记得住”的科普书籍。

软交所总监

李洪艳

一本好书，带给不同层面的读者不同程度的收获，向军的这本书，把多维度的数据分类、聚类、预测和关联一一呈现出来，带我进入一个立体的大数据王国，让枯燥的数据和挖掘，变得亲切易懂，于炎炎七月如沐春风。

上海研发公共服务平台

王丽萍

很早就与大嘴巴漫谈数据挖掘结缘了，经翻阅之后，深感这是一本很好的关于数据挖掘和统计的入门书，它知识点较为全面，图文并茂，深入浅出，很容易理解。我认为此书提纲挈领，提供数据挖掘理念的全局性，以它作为知识体系的梳理非常合适，为后期深入做了极好的铺垫。

读者

melisa

前言

写作，好比买菜烧饭，众口难调，实属不易。

文字，是食材，讲究用料实在，足斤够两，轻重适宜，刀工精细，厚薄均匀；行文，即烹调，力求挥洒自如，尝出滋味，暖了脾胃，唇齿留香，润了心肺；样式，为摆盘，尽享视觉盛宴，巧妙融合，疏密得当，别出心裁，饱含深意。

煎、炒、炖、炸、煮，油、盐、酱、醋、糖，齐活！

做书，宛若数据挖掘，集聚知识，涌现智慧。

选题策划侧重业务理解，主题思想则如指标体系，词语短句即是抽样样本，构思编辑恰似降噪归约，妙语连珠就像算法调优，谋篇布局等同模型构建，段落搭配在于前端展现，读者反馈实为效果评估。

理解、探索、调整、研发、评价，收工！

为什么要写作本书

什么是好的科普图书，我的理解：一，提供有价值的知识，能够满足读者心灵需要的内容；二，重视阅读体验，比如行文的编排与图画的辅助；三，找到最恰当的词语进行组合，体现其中的节奏和韵律，以及特定的情感，如同求一道最优解的数学题。

大嘴巴漫谈，有货，有形，有味！千里之行，始于足下，希望我可以做到。

本书主要内容

本书基本上涵盖了在产品生命周期的各个阶段，各自不同的研究目的和研究任务，结合业务特性和运营特点，针对用户需求，开展满足不同业务目标的研究工作。通过采用具体的研究方法，建立起完整的用户研究体系框架，以及综合考虑业界最常用的数据挖掘与分析方法，制定通用标准的研究流程。

关于本书作者

作者拥有北京科技大学计算机硕士学历，长期工作于互联网和电信领域，目前正在从事大数据、云计算方面的创业。新浪微博：@大嘴巴漫谈。

致谢

本书的顺利出版离不开电子工业出版社博文视点的大力支持，离不开孙学瑛女士，以及汪达文、王乐等老师的辛勤工作，在此表示最诚挚的感谢！

书中难免有不足之处，请读者及专家指正。

易向军

2016年2月

目 录

0 引言 / 3

0 引言：产品运营靠分析 / 4

0 引言：设计商用八时期 / 6

0 引言：定性定量双途径 / 8

0 引言：数据来源主客观 / 9

0 引言：分析方法全周期 / 10

0 引言：途径来源成体系 / 12

0 引言：用户视角来设计 / 13

0 引言：取长补短建优势 / 14

1 产品策略研究期 / 15

1 产品策略研究期：策略研究找趋势 / 16

1 产品策略研究期：目标用户要细分 / 17

1 产品策略研究期：定性研究划小组 / 18

1 产品策略研究期：样本条件配用户 / 20

1 产品策略研究期：座谈讨论辨特征 / 21

1 产品策略研究期：研究用户挖供需 / 23

1 产品策略研究期：街头面访寻偏好 / 24

2 产品概念评估期 / 34

2 产品概念评估期：需求概念生产品 / 35

2 产品概念评估期：开展工作挖需求 / 37

2 产品概念评估期：定性研究定类型 / 38

2 产品概念评估期：定量分析筛需求 / 43

2 产品概念评估期：概念测试主客观 / 48

2 产品概念评估期：产品概念细描述 / 49

2 产品概念评估期：小组座谈评概念 / 51

2 产品概念评估期：电话访问测指标 / 54

3 产品研发期 / 59

3 产品研发期：产品研发原型 / 60

3 产品研发期：参与设计供方案 / 61

3 产品研发期：卡片分类排架构 / 62

3 产品研发期：专家评审找问题 / 74

4 产品测试期 / 76

4 产品测试期：产品商用需测试 / 77

4 产品测试期：可用测试原型 / 78

4 产品测试期：设备搭建助协调 / 79

4 产品测试期：制定计划做准备 / 80

4 产品测试期：测试完成出报告 / 83

4 产品测试期：实例描述全过程 / 84

4 产品测试期：用户招募选样本 / 85
4 产品测试期：用户情况做统计 / 87
4 产品测试期：任务操作有情境 / 88
4 产品测试期：发现问题给建议 / 89
4 产品测试期：结果分析看指标 / 90
4 产品测试期：试用产品商用前 / 96
4 产品测试期：定性研究分用户 / 98
4 产品测试期：焦点小组谈反馈 / 99
4 产品测试期：研究结果论问题 / 102
4 产品测试期：深入评价靠定量 / 105
4 产品测试期：样本甄别有条件 / 107
4 产品测试期：问卷测试评分项 / 109

5 产品导入期 / 119

5 产品导入期：产品运营重精确 / 120
5 产品导入期：精细运营多角色 / 121
5 产品导入期：用户画像成特征 / 122
5 产品导入期：产品导入先认知 / 123
5 产品导入期：研究内容定策略 / 125
5 产品导入期：内外方面有不同 / 126
5 产品导入期：外部调研看态度 / 127
5 产品导入期：主观客观双结合 / 128
5 产品导入期：定性研究先抽样 / 129

5 产品导入期：焦点小组谈问题 / 130
5 产品导入期：小组座谈出成果 / 133
5 产品导入期：定量分析助定性 / 134
5 产品导入期：电话访问依问卷 / 136
5 产品导入期：分析结果有展示 / 137
5 产品导入期：内部数据识用户 / 141
5 产品导入期：维度角度辨特征 / 142
5 产品导入期：环比同比看趋势 / 144
5 产品导入期：定位目标找用户 / 147
5 产品导入期：数据挖掘来辅助 / 148
5 产品导入期：数据理解选变量 / 149
5 产品导入期：预先处理验数据 / 151
5 产品导入期：分类筛选决策树 / 153

6 产品发展期 157

6 产品发展期：产品发展有关联 / 158
6 产品发展期：用户产品藏联系 / 159
6 产品发展期：研究产品抽样本 / 160
6 产品发展期：数据校验剔冗余 / 160
6 产品发展期：关联规则建模型 / 162
6 产品发展期：关键指标评效果 / 164
6 产品发展期：模型输出生规则 / 166
6 产品发展期：结果展示网状图 / 167

- 6 产品发展期：分析用户找关系 / 169
- 6 产品发展期：数据采集建模型 / 170
- 6 产品发展期：借助决策树结构 / 171
- 6 产品发展期：从根到叶是规则 / 172

7 产品成熟期 / 173

- 7 产品成熟期：产品成熟重差异 / 174
- 7 产品成熟期：差异运营聚用户 / 175
- 7 产品成熟期：因子分析打前站 / 176
- 7 产品成熟期：聚类分析后安排 / 181
- 7 产品成熟期：细分群组有特征 / 184
- 7 产品成熟期：用户体验不可缺 / 187
- 7 产品成熟期：监测产品助运营 / 188
- 7 产品成熟期：定性研究建指标 / 190
- 7 产品成熟期：指标体系全涵盖 / 193
- 7 产品成熟期：定量评测体验度 / 197

- 7 产品成熟期：结构方程生模型 / 201
- 7 产品成熟期：输出权重和得分 / 203
- 7 产品成熟期：应用推广助运营 / 204
- 7 产品成熟期：发现短板定优先 / 205

8 产品衰退期 / 208

- 8 产品衰退期：产品衰退用户离 / 209
- 8 产品衰退期：主动被动分流失 / 210
- 8 产品衰退期：定性定量来剖析 / 211
- 8 产品衰退期：焦点研究分小组 / 212
- 8 产品衰退期：座谈聚焦找原因 / 215
- 8 产品衰退期：定量分析建模型 / 216
- 8 产品衰退期：输入输出选指标 / 217
- 8 产品衰退期：抽取数据按条件 / 220
- 8 产品衰退期：逻辑回归估流失 / 221
- 8 产品衰退期：基于业务验效果 / 222

0

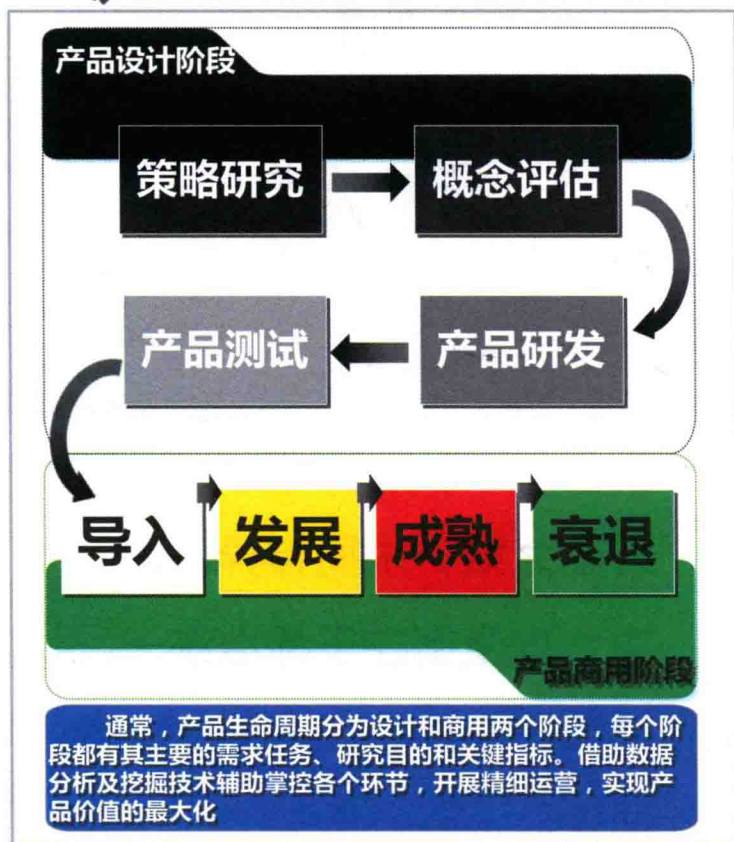
引言

万物之始，求同存异，而其中之度量衡在于标准统一，即如：车同轨，书同文，行同伦。



0

引言：产品运营靠分析



大数据时代，数据的分析及挖掘在企业的经营过程及业务管理中，逐步发挥出越来越显著的作用。无论是在产品的构想、原型设计阶段，还是在测试、上市商用后，用户需求与市场竞争环境都在每时每刻地不断发生变化。在这种情况下，就要秉承以用户为中心的理念，综合技术、市场两种驱动能力，以科学严谨的方法，准确有效地收集并分析用户订购及使用产品的评价、动机及行为等信息，为产品运营、业务支撑和数据管理等相关人员的工作提供系统化的指导，进一步辅助企业管理人员解决经营决策中所面临的问题。

总之，构建一个全面的，贯穿整个产品生命周期的业务数据挖掘体系框架，进行精准的用户研究及市场检验，已成为当前企业业务发展的必然趋势和重要举措。

0

引言：产品运营靠分析



在产品生命周期的各个阶段，有着各自不同的研究目的和研究任务，结合业务特性和运营特点，需要针对用户需求，开展满足不同业务目标的研究工作。通过采用具体的研究方法，建立起完整的用户研究体系框架。

尽管不同企业及研究机构对产品设计和商用阶段的划分和定义有所不同，但总体看来，基本上涵盖了如上所示的过程和方法。针对产品生命周期的各个阶段，综合考虑业界最常用的数据挖掘及分析方法，制定通用标准的研究流程。同时，通过长期研究积累而拓展出的关键指标，更是将用户对产品的感性认识转化为具体的参数形式。



0

引言：设计商用八时期



在产品阶段，主要分为策略研究、概念评估、产品研发和产品测试四个时期。

策略研究时期主要围绕新产品的需求趋势来确定研究方向，通过对市场、技术和竞品等方面的深入分析，宏观把握外部竞争环境和内部用户情况两个方面。

概念评估时期通过挖掘上一时期所确定下来的目标用户群特征，生成满足用户要求的产品概念，并进行有效评估，从中筛选出潜力大、用户关注度高的概念。

产品研发时期基于筛选出的产品概念，按照以用户为中心的设计理念，完成产品原型开发。

产品测试时期利用可用性测试以及邀请用户试用等手段，在产品正式上市商用之前，最大限度地优化改进产品。

0

引言：设计商用八时期



在产品商用阶段，主要分为导入、发展、成熟和衰退四个时期。

导入期，产品认知度较低，应及时监控产品的用户使用行为，掌握用户反馈评价，快速提高产品用户规模，并为产品改进提供重要依据。

发展期，产品订购开始逐渐增多，用户群和销售业绩明显上升，这个时期主要是促进用户活跃和业务的有效使用。

成熟期，随着用户量的增大，市场需求逐渐饱和，用户增速逐渐放缓，开展差异化的运营，改善用户体验是这个时期的主要任务。

衰退期，产品功能不能充分满足市场环境和用户需求的变化，产品订购开始呈现下降趋势，应采取产品优化策略，提高产品质量，并对有可能流失的用户进行挽留。



0

引言：定性定量双途径



定性研究旨在解释用户行为背后的原因，探明用户的意图和见解。通过小规模的用户样本来揭示未知问题的缘由。定量分析则是基于样本数据来证明某种观点的方法，借助定量分析来验证定性研究中发现的假设情况，找出符合统计意义的业务规律和趋势。

主观性研究涉及用户使用产品的目标、体验和感受，以及在产品使用过程中能够实现的自我价值。客观性研究意指基于收集的用户产品使用数据，来了解用户的实际使用行为和特征偏好。一般来说，客观性研究比主观性研究更能发现用户使用产品的真实状况。