

PEARSON

Beyond 超越大数据

——通过社交主数据管理深入了解客户

(美) 马丁·奥博欧佛
Martin Oberhofer

埃伯哈德·赫克勒
Eberhard Hechler

伊万·米尔曼
Ivan Milman

斯科特·舒马赫
Scott Schumacher

丹·沃尔夫森
Dan Wolfson

著

门群
译



清华大学出版社

Beyond 超越大数据

——通过社交主数据管理深入了解客户

(美) 马丁·奥博欧佛
Martin Oberhofer

埃伯哈德·赫克勒
Eberhard Hechler

伊万·米尔曼
Ivan Milman

斯科特·舒马赫
Scott Schumacher

丹·沃尔夫森
Dan Wolfson

著

门群

译

清华大学出版社
北京

Bigdata

北京市版权局著作权合同登记号 图字:01-2015-6399

Authorized translation from the English language edition, entitled Beyond Big Data: Using Social MDM to Drive Deep Customer Insight, 0-13-350980-X by Martin Oberhofer, published by Pearson Education, Inc, publishing as IBM, copyright © 2014.

All Rights Reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson education, Inc. CHINESE SIMPLIFIED language edition published by TSINGHUA UNIVERSITY PRESS Copyright © 2016.

本书中文简体翻译版由培生教育出版集团授权给清华大学出版社出版发行。未经许可,不得以任何方式复制或抄袭本书的任何部分。

本书封面贴有培生教育出版集团防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

超越大数据:通过社交主数据管理深入了解客户/(美)奥博欧佛(Oberhofer, M.)等著;门群译. —北京:清华大学出版社,2016

书名原文:Beyond Big Data: Using Social MDM to Drive Deep Customer Insight
ISBN 978-7-302-42631-8

I. ①超… II. ①奥… ②门… III. ①企业管理 IV. ①F270

中国版本图书馆 CIP 数据核字(2014)第 008340 号

责任编辑:刘志彬
封面设计:汉风唐韵
责任校对:宋玉莲
责任印制:王静怡

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:北京密云胶印厂

经 销:全国新华书店

开 本:170mm×240mm

印 张:17

字 数:228千字

版 次:2016年2月第1版

印 次:2016年2月第1次印刷

印 数:1~4000

定 价:49.00元

产品编号:055450-01



前 言

内容介绍

社交主数据管理(master data management, MDM)是业务数据处理过程中的全新革命。它把客户和产品中心与大数据衔接在一起,从根本上改进了客户体验和产品策略。传统意义上的主数据管理涉及一整套的流程、做法和技术,其目的是为了创建一个单一视图,供多个业务流程和多个系统(如客户、产品、供应商和位置系统)共享共同的核心业务对象。目前,企业纷纷采用这种数据管理方式,用于改进销售、运营和客户支持的流程。然而传统的主数据管理的核心是结构化数据的管理。目前,与客户和产品相关的有价值的信息被锁定在微博、博客、脸书(Facebook)、电子邮件、呼叫中心谈话记录、呼叫记录中。这些社交数据不仅体量巨大,而且是非结构化的、事务性的数据。随着大数据分析技术 Hadoop 和分析软件 Biginsights 的广泛应用,提取这些信息不再是遥不可及的事情了。但这些技术和软件还没有被绑定到现有的、具备客户和产品信息洞察力的主数据管理系统中,数据的提取往往受限。本书认为,把社交数据、移动数据、位置数据与主数据结合起来,可以实现以下几个目的:

- 与现有客户建立更加密切的关系。
- 采用合适的产品,改进寻找和定位新客户的方法。
- 更加深入地了解客户的想法以及对产品的看法。
- 利用移动技术的即时性,开发吸引客户的新途径。

第1章社交主数据管理概述,介绍主数据和社交主数据管理的基本概念,探讨如何将异构数据有机地关联起来,在主数据管理系统中形成精简、标准的主数据;通过主数据管理的典型案例,探讨如何关注和了解客户(以及管理产品目录);探讨利用社交数据,如何扩展和增强主数据管理系统,从而形成一个功能更加强大的社交主数据管理

系统。

第2章社交主数据管理用例和需求,通过社交主数据管理的相关用例,探讨社交主数据管理如何改进客户体验、提高目标营销、深入了解产品、监管,以及位置数据在社交主数据管理中的作用;介绍通过主数据与社交、移动和位置数据的集成,提取新型洞察力和关系(包括人与人之间、人与产品之间、人与机构之间的关系等);并且讨论这些新型洞察力和关系的需求和功能。

第3章社交主数据管理的功能框架,介绍社交主数据管理运行环境中,有哪些不同的数据域,以及为了实现社交主数据管理的解决方案,信息架构需要哪些核心功能。

第4章社交主数据管理的参考架构,是建立在第3章介绍的功能基础之上的。这些功能可以作为组件部署在参考架构中。因此,参考架构可以体现出各组件之间的交互、层次和关系。

第5章社交主数据管理产品的功能,探讨通过实际的产品和技术,如何将第2章中提到的功能与第3、第4章中提到的架构联系起来。

第6章社交主数据管理和客户关怀,详细探讨通过社交主数据管理,如何设计出引人入胜的用户体验。社交主数据管理和客户关怀的具体参考架构展示的是,通过社交和其他企业数据的离线、在线和实时分析,如何创建以客户为中心的业务。然后把以客户为中心的业务连接到主数据管理系统中,通过各种渠道提供更符合客户需求的体验。

第7章社交主数据管理和营销,介绍社交主数据管理是如何推动传统的广泛市场营销,快速向目标营销转型的。探讨如何深入了解客户和产品,提供令人心动的购物建议;以及基于社交主数据管理,如何策划出高效的(和不同类型的)营销活动,生成适当的购物建议;如何寻找影响者来扩大市场;如何使用内容营销模式,在合适的时间,提供合适的购物建议。

第8章移动主数据管理,带领社交主数据管理朝着新的方向发展。本章阐述主数据管理如何为移动应用程序提供信息,以及如何改进移动应用程序;为了改进客户体验和提高员工的生产力,如何把移动数据融入社交主数据中。本章还介绍了移动数据的特点,探讨如何调整社交主数据管理的架构来满足移动数据和移动渠道的需要。

第9章主数据管理的未来发展趋势,探讨通过大数据平台,如何扩展和加强传统的主数据管理系统的实体解析和匹配功能。介绍主数据管理中,被称为语义主数据管理的新兴技术。语义主数据管理可以通过主数据管理,以新的方式呈现已有的信息,并且通过语义技术,从社交数据中获取新的洞察力和关系,

以便更加深入地了解客户。最后讨论信息的收集、分析以及社交主数据管理系统的使用过程中所涉及的隐私和伦理道德问题,以及在社交主数据管理过程的每一个环节中,必须解决哪些伦理道德问题。

阅读对象

本书可以为许多从业人员和企业员工提供相关的信息和洞察力。对于那些想通过整合社交和主数据来创造新的业务机会,改善现有业务的商业领导者来说,本书提供了一些经典素材,方便了解主数据管理、社交数据、商业价值、社交主数据管理的一些新方法,以及关注隐私的问题。

对于技术领导人员,如企业架构师、信息架构师、系统架构师来说,本书详细探讨了相关的技术和用例,不仅专门介绍了参考架构,而且介绍了开发社交主数据管理最佳范例的具体指南,还讨论了产品映射,为选择哪些技术和产品来解决特定的问题指明了方向。

阅读收获

本书将广泛的探讨社交主数据管理中的商业、技术和伦理道德问题,介绍主数据、主数据管理和社交数据的基本概念。你将了解社交主数据管理是如何处理客户与企业之间关系的,并且更加深入地了解客户(这样才能更好地为他们服务),以便为客户提供更好的服务,创造更大的价值(这样他们才会对产品和服务有更大的购买欲)。

我们所探讨的不同类型的洞察力(文化意识、情绪、详细的客户细分、个人影响力)是在分析社交数据、未充分利用的企业数据、移动数据的过程中获取的。本书还介绍了把洞察力融入社交主数据管理平台的方法。通过具体技术和产品的映射,你将了解社交主数据管理系统的架构和功能。本书深入探讨了通过这些架构和功能,如何改进客户服务、策划先进的营销活动,以及通过深入、全面的客户洞察力,如何向客户提供适当的购物建议和激励措施(规避错误的)。你将了解移动系统带来的新技术,以及移动系统是如何扩展和改变社交主数据管理系统功能的。

你还将了解到,通过新技术,如何扩展传统主数据管理服务的实体解析和关联功能,以及通过语义技术,如何使社交主数据管理在原

有功能的基础上,增加学习和推理功能。最后,你将面临一个挑战,因为即使拥有了所有的数据和洞察力,并不意味着你有权使用它们。隐私保护法和客户的期望将成为有社会责任的主数据管理的核心。

阅读方法

本书在逻辑结构上可以分成以下三部分:

概念、业务价值性能和伦理道德。这部分适合于那些想要了解社交主数据管理,及其与传统主数据和传统分析的差别,深入研究社交主数据管理带来的新的商机,探讨社交主数据管理的功能,并且反思使用社交主数据管理在法律、伦理道德和业务上的影响的商业领袖阅读。

本部分包括以下章节:

- 第1章社交主数据管理概述;
- 第2章社交主数据管理用例和需求;
- 第9章主数据管理的未来发展趋势。

功能、架构和产品映射。这部分致力于向技术领导者介绍社交主数据管理的整体技术,以及底层组件在社交主数据管理架构中的作用。

本部分包括以下章节:

- 第3章社交主数据管理的功能框架;
- 第4章社交主数据管理的参考架构;
- 第5章社交主数据管理产品的功能。

社交主数据管理领域。这部分既向业务领导者也向技术领导者介绍,在以下领域社交主数据管理的具体技术是如何促进业务增长的:

- 第6章社交主数据管理和客户关怀;
- 第7章社交主数据管理和营销;
- 第8章移动主数据管理;
- 第9章主数据管理的未来发展趋势。

使用约定

下面简要介绍本书所使用的主要约定:

●缩写——书中各章节使用了缩写,但在本书第一次使用时,有全称说明。

●引用——本书引用较多。为了方便进一步的研究,在每一章的结束部分一一列出了书中引用的内容。深入研究每一章的主题时,可以在参考书目中找到相关的内容。本书还使用了脚注,把本章的相关语句与相应的参考连接起来。

●脚注——脚注可以提供更多的背景信息,如文中提到的相关产品或工具。

●图表——每一章的图表依次编号。

目录

第 1 章 社交主数据管理概述	1
1.1 社交主数据管理的概念	1
1.2 传统主数据管理	6
1.3 传统主数据管理的业务价值	12
1.4 社交主数据管理	15
1.5 社交主数据管理的业务价值	19
1.6 结论	19
1.7 参考文献	20
1.8 扩展阅读书目	20
第 2 章 社交主数据管理用例和需求	22
2.1 社交主数据管理的业务价值——用例和客户价值	23
2.2 社交主数据管理的基本功能	35
2.3 社交主数据管理中的高级关系	40
2.4 结论	50
2.5 参考文献	50
第 3 章 社交主数据管理的功能框架	54
3.1 简介	54
3.2 数据域	57
3.3 嵌入企业架构的社交主数据管理参考架构	65
3.4 功能框架	66
3.5 结论	86
3.6 参考文献	86

第4章 社交主数据管理的参考架构	91
4.1 简介	91
4.2 架构概述	92
4.3 主数据管理:架构概述	93
4.4 组件模型	97
4.5 组件关系图	101
4.6 面向主题的综合	104
4.7 结论	106
4.8 参考文献	106
第5章 社交主数据管理产品的功能	107
5.1 社交主数据管理	109
5.2 InfoSphere BigInsights 架构	117
5.3 可信任信息的集成	134
5.4 常见的分析功能	140
5.5 参考文献	142
第6章 社交主数据管理和客户关怀	147
6.1 衡量社交媒体数据	148
6.2 以客户为中心	149
6.3 下一个最佳行动	156
6.4 情感分析	162
6.5 社交影响者的确定	166
6.6 社交网络分析	170
6.7 用于客户关怀的社交媒体可信度	174
6.8 参考文献	177
第7章 社交主数据管理和营销	181
7.1 社交媒体营销以及主数据管理的作用	182
7.2 通过社交媒体开展营销活动	185
7.3 兴趣群组	201
7.4 总结	205
7.5 参考文献	207
第8章 移动主数据管理	210
8.1 与消费者互动的演变	210

8.2	移动主数据管理对信息技术的影响	215
8.3	安全	225
8.4	结论	225
8.5	参考文献	226
第9章	主数据管理的未来发展趋势	228
9.1	实体解析和匹配	229
9.2	语义主数据管理	231
9.3	信息中的伦理道德问题	237
9.4	结论	246
9.5	参考文献	247



第 1 章

社交主数据管理概述



本章将讨论企业如何在非传统数据源中,寻找新的洞察力,来了解客户和产品;在提供客户和产品唯一信息源的过程中,传统主数据管理的作用是什么;如何从传统主数据管理系统中寻找到商业价值。最后,介绍把外部源中现有的新洞察力有机地植入主数据管理中,变成社交主数据管理,从而把企业与客户之间的交互提升到一个新的高度。

在社交主数据管理中,企业对客户的了解更加深入。这种了解需要跨越企业的传统界限才能获得。正是因为对客户有了深度的了解,社交主数据管理才成为企业在探索改进客户关系和服务道路上的里程碑。

1.1

社交主数据管理的概念

在采用了主数据管理(master data management, MDM)这种以客户为中心的理念来管理系统和流程之后,企业又开始探索新的技术,

用于开发行之有效的客户洞察力。我们用客户这个术语来代表个人客户和企业客户。为企业客户提供服务的员工,与企业之间不仅存在客户-雇员关系,而且在组织中扮演了一定的角色,所以被视为个人客户。像销售代理这样的内部员工,作为个人,在企业内部担任了一定的职位,同时又与客户建立起了某种关系。传统意义上的洞察力来自于客户交互数据。例如,购买历史记录和网页查看模式,或第三方营销公司提供的客户细分信息。无论这些洞察力来自哪里,只要将这些洞察力汇集在主数据管理中,就可以在操作上支持目标营销、购买建议或客户服务的改进。

大数据技术的出现,使企业开发利用防火墙以外的数据成为可能。换句话说,企业可以充分利用社交媒体信息、博客、推特(Twitter)、脸书(Facebook),以及由客户提供的其他公共数据源。大数据技术具备如基于自然语言处理的情感分析这样的功能,可以大规模地处理数据源。这样的例子包括:在电影中插播特定的广告之前和之后,监测观众的情感变化;或在某事件,如奥斯卡颁奖典礼开始或产品的官方发布会前后,监测情感变化的趋势。

把大数据分析以与客户为中心的理念有机地结合起来,就形成了我们所说的社交主数据管理的核心部分,其目的是要超越大数据分析,充分利用相关信息来改进客户体验。例如,有的公司希望将宏观、或以群体为基础的情感分析转变为微观、或以客户为基础的情感分析;有的公司希望了解,哪个具体的产品受公众欢迎。社交主数据管理通过整合大数据的新技术和主数据管理技术,使社交媒体数据洞察力具备可操作性。

1.1.1 社交数据带来的客户洞察力与机遇

长期以来,面向客户的企业一直在数据中获取客户洞察力——洞察力可以用来改进客户服务,创造追加销售和交叉销售的机会,或开展有针对性的营销活动。具备这种功能的数据有内部数据和公共数据两大类。

(1) 内部数据

例如,客户交互历史记录或客户支持日志,这类数据可以用来提示呼叫中心的接线员,有客户不满意,或用来挖掘补救措施。在植入式广告的决策中,可以利用这类数据,对客户群进行微观细分(例如,“与你类似的客户还看过……”),或创建复杂的预测模型(比如,网飞公司(Netflix)曾用这种模式,建议观众租看下一部电影)。在后面一种方式中,通过分析客户购买的历史信息,创建购买模型,并部署在面向客户的系统中,供客户实时使用。这个例子展示的是客户洞察力运用的一种常见模式,在“分析沙箱”里挖掘历史数据后,在操作系统中部署挖掘出来的洞察力。

企业知道,通过内部数据只能片面地了解客户。因此,许多企业通过购买与个人客户相关的第三方信息来扩充客户信息。这些数据虽然有助于细分客户群,但通常不够精细(例如,通过邮政编码,计算客户的平均收入)。这种方式可以支持广泛市场营销,但不具备对个人客户的洞察力。因此直到最近,人们才知道,能够获取客户详细的、内容丰富的个人档案的数据,只有在内部数据源中才能找到。

(2) 公共数据

社交媒体和消费者愿意为自己创建公共数据,改变了内部数据只能片面了解客户的状况。现在,愿意创建公共数据的人越来越多,主要的生活事件(如生日、周年纪念、毕业典礼)、人口统计信息(如年龄、位置)、兴趣、关系(如配偶、子女)、甚至影响力(如追随者)都出现在了公共的或选择性加入的数据源中。皮尤(Pew)^[1]估计,在18~29岁年龄段中,有27%的互联网用户在使用推特(Twitter)。此外,博客帖子和回复以及在线产品评论也是了解客户对各种问题和产品看法的丰富数据源。

利用社交媒体来提高客户洞察力,在技术上有很大的挑战。例如,如何从非结构化或半结构化的数据中提取相关的信息,以及如何把社交媒体的作者与特定的客户关联起来。部署在大数据平台上的实体提取、情感分析、实体解析功能,可以处理社交界产生的大量数据,使我们具备应对这些挑战的能力。但大数据本身并不能解决全部问题。客户洞察力提升之后,还必须具有可操作性,才能为企业所用。

企业已经采用了主数据管理来存储和分发深度客户洞察力,尽管社交媒体数据带来了新的挑战,但把主数据管理扩展成了社交主数据管理后,就为最终战胜挑战提供了具体的操作步骤。

在应用社交媒体时,还要面对另外两个挑战:法律和伦理道德。例如,在某些国家,客户有权浏览信息系统中所有客户的档案。而事先未经客户明确同意,就利用社交媒体信息来扩展信息系统中的客户信息,这是客户始料未及的。

1.1.2 产品评论带来的产品洞察力与机遇

零售商和制造商可以通过主数据管理,掌握产品信息和客户信息。传统主数据管理系统中的产品数据往往是结构化的数据,而且仅限于产品描述和产品的物理属性(如尺寸、重量等)。而社交媒体、产品评论博客和在线零售网站提供的是消费者自己创建的新的信息源。现在的消费者通过各种方式,研究不同的产品,比较不同制造商的产品,比较产品的价格。通过这一系列的努力,消费者对产品、购买和客户支持经验形成了自己的看法。在线零售商的评论,如亚马逊的评论,就是消费者自己创建产品信息的一个典型例子。在评论里,消费者把产品分成从一到五的五个等级,并通过文本来描述产品体验。下面的段落就是一个典型的产品评论:

这个恒温器太好了。正如预期的那样,不仅与 Nexia 系统兼容,还与其他家庭自动化系统,如 MiCasa Verde 公司无害环保的 Mi Casa Verde VeraLite 家庭控制器、灵活且强大的 Mi Casa Verde Vera3 以及价格实惠的家庭控制器(这两个系统都没有月费)兼容。如果你有一根常见的中性线,安装就会很容易。安装后,样子看起来还不错。通过 Z-wave 发出指令后,会立即响应。这个办法很不错,可以为你省下一大笔电费。

这段评论包含了两种产品属性,对于零售商和制造商来说都很有

用。第一种是产品的功能属性:恒温器与制造商推荐之外的另外几个家庭自动化系统的兼容性良好,安装简便,响应迅捷。这段评论中的功能属性可以使制造商提供的类似属性得到扩展。第二种是零售商感兴趣的非功能属性,还可以提供消费者对产品的喜好信息,如吸引力(“看起来还不错”)和经济实惠(“很不错,可以为你省下一大笔电费”)。

同一产品的第二个评论提供了该产品的其他不同属性。例如:

看过该产品与家庭自动化系统 Nexia 公司的 Z-wave 系统兼容良好的评论后,购买了这种恒温器。后来才意识到,Mi Casa Verde 系统提供了类似的 Z-wave 功能,而且没有月费。购买了三个,安装进展顺利。能够控制家里的电器设备,还能通过 iPhone 的应用程序进行远程控制。

没想到,该设备不允许线上规划。通过 Mi Casa Verde 系统,控制恒温器时,需要遵循多个“规则”,经过复杂的设置,才能创建一个简单的计划。没有方便用户使用的网络接口。

这段评论也表明该恒温器可以与多个家庭自动化系统兼容,安装简单,同时还提到了非功能属性,如设置复杂和用户使用起来不方便。

这两个例子说明,从社交媒体数据中提取产品属性,面临两个技术挑战。第一个挑战与寻找客户洞察力类似,即可供使用的信息是非结构化的文本,需要高级的自然语言处理技术。第二个挑战与客户洞察力不同,从一个评论中提取的产品属性并不可靠。这意味着可用的产品属性的提取需要建立在许多评论和许多评论资源积累的基础上,此外,还需要保留与属性相关的统计数据的属性。例如,80%的评论声称“安装容易”,5%的评论声称“安装困难”,以及 15%的评论中未涉及“安装”。

这些额外的属性可以帮助零售商决定采用哪种广告植入方式,以及使用什么样的广告词(“易于安装”)。制造商也可以从这些属性中,提取重要的客户洞察力,支持产品的开发,改进客户支持和产品使用说明。

把这些属性合并到主数据管理解决方案中,可以方便所有参与方使用。

1.2 传统主数据管理

本节我们介绍(或对一些人来说,再次介绍)主数据的概念——也称为核心业务对象,这些是在企业范围内,被多个业务流程反复使用的数据。接下来,介绍在已有的信息系统中,扩展主数据时,企业需要解决的问题。最后,探讨如何利用主数据管理来解决这些问题。如果想从架构、数据质量和工具选择的角度,全面地研究传统主数据管理,可以参阅“扩展阅读书目”部分提供的参考书目。

1.2.1 主数据的定义

通过主数据的典型例子,了解主数据概念是最简单的方法。这些例子包括(但不仅限于):

- **参与方或成员。**个人或组织。它是指客户、销售商、供应商、雇员、业务合作伙伴、病人、员工、企业服务对象以及与企业打交道的公民。行业不同,参与方的相关细节也不同,但常用数据包括人口统计信息:姓名、地址、电话、电子邮件、隐私、偏好等。更复杂的系统,如客户关系管理(customer relationship management, CRM)系统,可以通过分析系统,不断获取客户信息(例如,该客户是否为黄金客户)。

- **产品。**产品是指企业提供的商品和服务。产品的相关信息包括:产品描述、尺寸、颜色、价格、目录号码以及通用产品代码(universal product code, UPC)。通常以特定的方式,对这些产品进行分类和组合,如软件集(互联网服务+电话+电信公司或有线电视供应商提供的有线电视),或者按照层级结构,进行分类和组合,就像你在在线目录上看到的那样(例如,电子产品→电话→无线设备→智能手机)。

- **位置。**这是核心企业资产的位置,包括办公室、商店、教室、装货码头等。位置主数据也可以是客户的居住地,或者产品的出售地。因此,可以把位置数据和其他主数据捆绑在一起。