

基于统计学视角的 数据挖掘在CRM中的应用

基于统计学视角的数据挖掘将统计学的思想方法贯穿于数据挖掘的过程中，对于客户细分、交叉销售及预测等功能都能实现良好的技术支撑作用。

崔冬梅

著



JI YU TONGJIXUE SHIJIAO DE
SHUJU WAJUE ZAI CRM ZHONG DE YINGYONG



西南财经大学出版社
Southwestern University of Finance & Economics Press

基于统计学视角的 数据挖掘在CRM中的应用

崔冬梅 著



西南财经大学出版社
Southwestern University of Finance & Economics Press

中国·成都

图书在版编目(CIP)数据

基于统计学视角的数据挖掘在 CRM 中的应用/崔冬梅著. —成都:
西南财经大学出版社, 2015. 9
ISBN 978 - 7 - 5504 - 2177 - 6

I. ①基… II. ①崔… III. ①统计数据—数据采集—应用—企业管理—销售管理—研究 IV. ①F274

中国版本图书馆 CIP 数据核字(2015)第 227265 号

基于统计学视角的数据挖掘在 CRM 中的应用

崔冬梅 著

责任编辑:李特军

助理编辑:李晓嵩

封面设计:何东琳设计工作室

责任印制:封俊川

出版发行	西南财经大学出版社(四川省成都市光华村街 55 号)
网 址	http://www.bookcj.com
电子邮件	bookcj@foxmail.com
邮政编码	610074
电 话	028 - 87353785 87352368
照 排	四川胜翔数码印务设计有限公司
印 刷	四川五洲彩印有限责任公司
成品尺寸	148mm × 210mm
印 张	3.75
字 数	105 千字
版 次	2015 年 9 月第 1 版
印 次	2015 年 9 月第 1 次印刷
书 号	ISBN 978 - 7 - 5504 - 2177 - 6
定 价	28.00 元

1. 版权所有,翻印必究。
2. 如有印刷、装订等差错,可向本社营销部调换。

前 言

企业离不开市场，激烈的市场竞争、全球化的经济浪潮，导致企业必须以客户为中心，掌握更多、更准确的客户信息，对客户个性化需要做出快速反应，为客户提供更便捷的购买渠道和更好的售后服务以及经常性的客户关怀，这是客户关系管理（Customer Relationship Management, CRM，下同）的核心内容。随着网络和信息技术的发展，企业可获得的信息量日益膨胀，呈几何级数倍增长，CRM 面临着信息爆炸带来的前所未有的严峻挑战。

从庞杂无序的数据中获取真正对企业有用的信息，发现数据背后隐藏着的模式，从而指导企业的生产经营实践，这就是数据挖掘的任务之所在。CRM 的实施需要强大的技术支撑，数据挖掘以其从纷繁、复杂的数据中提取有用知识的强大功能而在 CRM 中有着广泛的应用。在 CRM 中，数据挖掘要发现的是商业活动中各个因素与客户行为之间的不直观的关系。因此，数据挖掘要想在商业活动中真正起到作用，就必须与基本的商业过程关联起来。数据挖掘技术在 CRM 中的应用主要体现在新客户开发、客户效益分类、交叉销售及预测、顾客维系（流失分析）、顾客细分、客户盈利性识别等方面。

本书将统计学的思想、方法和技术贯穿于数据挖掘整个过

程之中，分别从理论和实证两个方面对统计数据挖掘技术在 CRM 中的应用进行研究。本书回顾了前人对 CRM 和数据挖掘的研究成果，从 CRM 和数据挖掘的演变过程、发展现状来论证 CRM 的进一步发展离不开数据挖掘作为技术支撑。本书分为理论研究和实证研究两个部分。理论研究部分定义了 CRM 和统计数据挖掘的内涵，进一步论证了统计数据挖掘技术在 CRM 中的适用性并提出了基于统计学的数据挖掘流程以及数据挖掘模型的建立和实现方法。本书的实证研究部分分为两章，分别为数据挖掘在 CRM 中应用的两个方面，即客户细分和交叉销售及预测。两章内容均以电信企业 CRM 为例，运用统计数据挖掘技术建立了电信客户细分模型和电信增值业务的交叉销售及预测模型，并结合电信企业的实际情况对模型结果进行了分析和解释。研究结果可以为电信企业制定科学有效的营销策略提供依据，从而提高电信企业的 CRM 水平，提升电信企业的竞争能力，对电信企业如何在 CRM 中应用统计数据挖掘技术具有一定的指导意义。

目 录

1 绪论 / 1

1.1 研究问题的提出 / 1

1.2 研究的目的及意义 / 8

1.3 本书的框架结构 / 10

2 数据挖掘与 CRM 概论 / 12

2.1 数据挖掘的内涵、研究和应用现状 / 12

2.1.1 数据挖掘的基本概念 / 12

2.1.2 数据挖掘与其他学科的关系 / 14

2.1.3 数据挖掘的国内外研究和应用现状 / 15

2.2 CRM 的发展、应用概述 / 19

2.2.1 CRM 的内涵、起源和发展过程 / 19

2.2.2 CRM 国内外研究和应用现状 / 24

2.3 数据挖掘在 CRM 中的研究与应用 / 26

3 CRM 中统计数据挖掘技术理论分析 / 29

3.1 统计学与数据挖掘的关系概述 / 29

3.2 数据挖掘技术在 CRM 中的适用性分析 / 31

3.2.1 CRM 的技术支撑 / 31

3.2.2 数据挖掘的几项关键技术 / 32

3.2.3 数据挖掘技术在 CRM 中的几点应用 / 34

3.3 GRM 中的统计数据挖掘模型的建立及实现方法 / 37

3.3.1 数据准备 / 37

3.3.2 分类回归树技术 / 40

3.3.3 购物篮分析——关联规则 / 43

3.3.4 数据挖掘模型的统计评价

——基于统计检验的 49er 搜索算法 / 47

3.3.5 统计框架下的数据挖掘模型 / 49

4 统计数据挖掘在电信客户效益分类中的应用 / 51

4.1 背景分析 / 51

4.2 电信 CRM 中的客户细分 / 55

4.2.1 电信 CRM 的特点及意义 / 55

4.2.2 客户细分 / 57

- 4.3 模型的建立 / 60
 - 4.3.1 问题的定义 / 60
 - 4.3.2 数据准备 / 61
 - 4.3.3 数据挖掘方法选取 / 62
 - 4.3.4 K-means 聚类过程 / 64
- 4.4 模型评价和结果解释 / 67
- 4.5 小结 / 68

5 统计数据挖掘在电信交叉销售及预测中的应用 / 69

- 5.1 商业背景 / 69
- 5.2 交叉销售的理论基础 / 71
 - 5.2.1 交叉销售的概念 / 71
 - 5.2.2 交叉销售的意义及应用 / 72
 - 5.2.3 交叉销售与 CRM 的关系 / 73
- 5.3 电信增值业务交叉销售模型 / 75
 - 5.3.1 选择数据挖掘工具软件 / 75
 - 5.3.2 数据理解与数据准备 / 76
 - 5.3.3 选用的数据挖掘技术及算法介绍 / 78
 - 5.3.4 参数设置及模型建立步骤 / 81
 - 5.3.5 模型评价和结果解释 / 83

5.4 电信增值业务客户预测模型 / 85

5.4.1 数据理解与数据准备 / 85

5.4.2 选用的数据挖掘技术及算法介绍 / 86

5.4.3 参数设置及模型建立步骤 / 87

5.4.4 模型评价和结果解释 / 93

5.5 小结 / 97

6 结论与展望 / 98

6.1 主要结论 / 99

6.2 统计数据挖掘技术应用于 CRM 时应注意的几点问题 / 99

6.3 局限性与研究展望 / 100

参考文献 / 102

附 录 / 105

附录 A CRM 在各行业中的应用 / 105

附录 B 复杂数据挖掘模型结果的解释 (对应分析方法) / 106

1 绪 论

1.1 研究问题的提出

19 世纪 20 年代，随着社会经济的飞速发展，产品日益丰富，市场格局发生了深刻变化，消费者需求层次日趋多元化和复杂化，这就推动了营销观念和营销方式的变革。置身于激烈市场竞争中的企业要想保持领先地位就必须掌握客户更多信息并且更准确地把握客户需求，对客户的个性化需要做出快速反应，为客户提供更便捷的购买渠道和更好的售后服务以及经常性的客户关怀，进行以客户为中心、以客户关系为纽带、以客户关怀为手段、以客户数据为基础的精细化管理，客户关系管理（Customer Relationship Management，CRM）应运而生。

近年来，客户关系管理理论研究备受重视。由于信息时代，互联网和电子商务快速发展，世界范围内，各类企业都在经历深刻的变化。企业面临的商业环境变化巨大，传统的商业模式正日益受到挑战，特别是目前许多行业的产品和服务同质化趋势日益显著，企业要确立竞争优势变得更加困难。这就使得企业之间的竞争情况和经营环境发生了一系列的重大变化，企业之间竞争的范畴逐渐从技术竞争和产品竞争转向以服务为重点

的、以客户为导向的全面客户关系管理策略的竞争，竞争的着眼点逐渐从产品竞争转向客户竞争和服务竞争。客户经济渗透到当今市场上的每一个角落，传统的产品、技术和成本竞争正在发生深刻的变化，企业必须从战略的角度考虑如何建立和维护良好的客户关系，以获得可持续的竞争优势。许多企业，尤其是美国和其他发达国家的企业，把客户资源作为最重要的资源，把客户关系管理作为法宝，通过实施客户关系管理，提高企业的核心竞争力，满足客户的需求，提高客户的满意度和忠诚度，提升客户价值，与客户保持长期良好的合作关系，以获得可持续的竞争优势，实现企业的可持续发展。客户关系管理受到关注有以下几个方面的原因：第一方面的原因是企业之间竞争更加激烈，迫使企业寻找与众不同的和无法模仿的核心竞争力。一般而言，企业的核心竞争力来源于企业经营要素的配置，包括内部人力资源、资金、设备、原材料、技术等；来源于与供应商的业务关系；来源于与客户的关系，包括与中介和最终客户之间的关系。企业只有对这些关系进行科学的管理，才能提高企业的竞争力，客户关系管理也因此提上了日程，被企业高度重视起来。第二方面的原因是管理重点演进。企业的管理重点从内部管理向外部管理转移。在已经过去的数十年中，企业管理的重点经历了三个阶段的转换。第一阶段是 20 世纪 70 年代，产品质量和服务质量受到企业管理的重视，全面质量控制（TQC）在全球各地被普遍接受。第二阶段是 20 世纪 80 年代，制造资源计划（MRP）的概念和技术刮起了信息化浪潮，为提高内部管理效率、降低运营成本发挥了重要作用。第三阶段是 20 世纪 90 年代，企业发现要想获得新的核心竞争优势，必须把研究关注点从企业内部转向企业外部，整合整个供应链的合作伙伴关系。因此，供应链管理（SCM）开始流行。目前，许多企业试图实施的客户关系管理，更要求企业从以产品为中

心的经营理念转向以客户为中心的管理理念。第三方面的原因是管理理论的发展。数据库营销、关系营销、一对一营销等理论对客户关系管理的产生和发展起到了促进作用。研究表明,若客户流失减少5%,企业利润将增长25%~85%。麦肯锡管理咨询公司研究认为,提高客户的忠诚度对增加企业价值是一个非常重要的因素,老客户重复购买的价值相当于新客户购买价值的两倍,客户关系管理可以帮助企业维持与客户更紧密的关系,也因此成为企业创造价值的一个必不可少的环节。第四方面的原因是技术进步。计算机技术、数据库技术、网络技术、通信技术、数据挖掘技术、智能商务技术、知识管理技术和工作流管理技术快速发展,使过去非常困难的客户关系管理变为一个相对简单的过程,互联网为企业与客户建立个性化的和互动的关系提供了低成本的方便工具,信息技术革命使企业可以通过互动媒体与单目标客户在任何时间对话和交流,使得以单向信息传递为特征的大众传播逐渐失去往日的魅力,技术进步使企业与客户建立长期关系的努力成为可能。

虽然企业对客户关系管理日益重视,客户关系管理的研究成果也逐渐丰富起来,但笔者认为客户关系管理理论的发展仍存在许多不足之处。首先,大部分客户关系管理理论的研究角度仍然是从企业角度出发,而没有真正做到“以客户为中心”。例如,现在的研究始终注重研究企业与客户之间的单维关系,这种单维的“客户关系”连接起来,将形成一张“以企业为中心”的关系网络。事实上,从客户的角度看,还存在着一张以客户为中心的客户关系网络,遗憾的是透过客户关系网络对客户关系管理效应进行的研究还非常少见。其次,现有的客户关系管理理论过于倾向于理论研究,许多术语难以被企业正确理解,从而容易导致企业实施客户关系管理时产生实践与理论的偏差。例如,理论研究时习惯将客户关系管理系统分类为操作型、分析型和协

作型,但在企业的实践中,这样的名称分类容易使企业误解为客户关系管理系统是“三中选一”的,企业应选择某一系统类型,而不是建立由这三个系统构成的有机整体。这明显是与客户关系管理系统实施的理论初衷相矛盾的,因此应将其“分类”改为“分层”,这样更易于理解。最后,也是最重要的问题,即客户关系管理的理论研究一直偏重于理论自身的完善,而较少研究将客户关系管理有效应用到实践中去的工具、方法步骤和关键突破点等,因此容易出现客户关系管理理论与客户关系管理实践的严重脱节。笔者经研究认为,客户关系管理理论与应用之间的结合点和突破点在于客户关系管理的效率研究,目前这方面的深入研究在理论界几乎是空白。

世界各国电信业都在轰轰烈烈地进行行业改革、并购与重组。欧盟于1997年发布了《电信、广播与信息技术融合绿皮书》,之后又发起了多次并购。美国继1996年颁布《电信法》之后,也发起了多次并购。这些并购背后的主要目的是放松各种严格管制,使电信业的改革更加深入,进而推动电信市场的全面开放,同时也为电信业的自由竞争和飞速发展奠定了必要的法律基础。世界电信业迎来了跨时代的蓬勃发展。并购不仅大大提高了电信企业自身的竞争能力,同时也使得电信行业的市场竞争越来越激烈。我国电信业的深入改革已经进行了10多年,中国电信运营企业的兼并与重组接踵而来。我国政府对电信的管制正在逐步放松,市场变得越来越开放和透明。

伴随着电信技术的突飞猛进,客户消费需求日益扩大和提高,这为我国电信业的飞速发展创造了千载难逢的机遇。目前我国国内基本形成了中国移动、中国电信、中国联通三家运营企业“三足鼎立”的局面。回顾其近年的发展历程,2008年10月15日,国内电信运营企业又一次完成重组,中国卫通并入中国电信,中国网通并入中国联通,中国铁通并入中国移动,中

国电信全面接收中国联通的码分多址（CDMA，下同）网络。根据重组之后的格局由国家工业和信息化部向各个企业发放第三代移动通信技术（3G，下同）牌照，标志着中国通信网络服务市场正式进入了3G时代。中国联通获得宽代码分多址（WCDMA）牌照，中国移动获得时分同步码分多址（TD-SCDMA）牌照，中国电信获得CDMA2000牌照。自此，三大电信运营企业正式开始了3G通信网络服务的市场竞争：中国移动于2009年1月7日正式启用3G业务“G3”，中国电信于2009年4月16日正式启用3G业务“天翼”，中国联通于2009年9月28日正式启用3G业务“沃”。2011年年底，中国的人口已达13.47亿人，同时手机持有量达到10亿部，也就是说中国的手机普及率差不多达到了75%。截止到2014年，我国手机用户数约为12.4亿户，其数量几乎接近人口总数。在这样一个成熟且近饱和的市场中，开发新用户的难度之大，也就可想而知了。

美国市场营销学会《顾客满意手册》的统计数据显示，就移动通信行业而言，保留现有老客户的利润率是开发新客户的16倍。该数据同时显示，开拓一个新顾客所需要的成本大概相当于保留一个老客户的5倍。一个企业生存和发展的根本途径之一就是保留忠诚的老客户。当市场处于成熟期中，很难通过常规手段发展新客户。那么对于已近饱和状态的通信市场而言，减少客户流失就意味着用较低成本减少利润损失。这样企业不仅可以降低服务成本上的支出，还可能会产生老客户带新客户的正面效应。长期保有忠诚度较高的老客户，不论对企业的营运还是创造收益方面都有非常大的益处。因为忠诚的老客户会带来正面的口碑和宣传效应，可以促使其他人购买企业的产品或服务，进而成为企业的新客户，给企业带来新的收入和利润。

客户对于电信运营商和电信产品具有自主选择权，这使得不同电信运营企业对客户资源进行剧烈的争夺，对高端客户的

争夺尤其愈演愈烈。在世界经济一体化的大环境下,随着我国电信市场的管制不断放松和不断开放,国内电信运营企业的竞争也日益激烈,已近乎白热化。加之通信技术、相关法律法规等的动态变化,客户从一个电信运营商转到另一个运营商也变得更加容易。3G、第四代移动通信技术(4G)时代,客户会比以往更加频繁地更换电信产品和服务。同时,随着转网成本的不断降低,客户在不同运营商之间进行转网的现象也会越发频繁。对于电信行业的客户保持问题的分析区别于传统客户忠诚度的分析。我国手机用户的数量非常庞大,然而用户对电信运营商的依赖程度低、缺乏忠诚度。尤其是中低端用户,他们往往可以为了一些小的利益而频繁地更换不同的电信运营商。电信运营商试图通过非理性的价格竞争手段,比如一再降低服务费用、过度发放赠品或礼品等,并不能起到提升用户忠诚度的作用。相反,这样做的结果是削弱了用户的转网壁垒,同时更降低了用户对企业的忠诚度。在这样的背景下,通过本书的研究,可以采用交叉销售等为用户“量身定做”的服务,进行有效的客户流失管理,从而在日趋激烈的市场竞争中实现“保存量、激增量”的精确营销,使电信运营企业获取最大的价值和利润。

21世纪是以网络为代表的信息技术突飞猛进的时代,随着网络的飞速发展和经济全球化带来的激烈挑战,信息技术已经成为现代企业进行经营管理不可或缺的支撑。近年来,随着数据库技术的不断发展和数据库管理技术的广泛应用,数据库中存储的数据量不断增大。这些大量的数据背后隐藏着重要信息,这些信息将为公司创造很多潜在的利润。随着大数据时代的来临,数据已然成为企业的核心资产。对于大数据的概念,目前尚无权威而统一的定义。大数据的核心思想一般认为大数据分析不是抽样分析、绝对精确和因果分析,而是分析更多的大量甚至是全部的数据,

是基于相关关系分析法的预测。大数据分析追求效率、重视相关关系的分析。大数据不仅将对企业的业务模式产生深远影响,还将进一步影响企业的组织结构和企业文化等诸多方面。消费者与网民的区分日渐削弱,企业疆界的划分日渐模糊。麦肯锡管理咨询公司早在 2001 年 5 月发表了《大数据:下一个创新、竞争和生产率的前沿》的一篇研究报告。此后,大数据开始得到政府、研究机构和大公司的重视。然而在我国,大数据还是一个比较新的事物,在 2013 年之前这方面的研究都还较少,基本以概念介绍、理论分析和应用预测为主。直到 2013 年,大数据开始成为全球关注的热点,各行各业争相加入到轰轰烈烈的大数据研究中,无论是理论研究还是实证研究都取得了阶段性的研究成果。人们对大数据带来的改变进行探讨,对如何更好地应对大数据时代带来的机遇和挑战进行分析。人们通过对相关文献的整理和分析,对大数据的认识和研究从最初定义大数据的概念,探讨大数据的特点、成因、对各行业的影响发展到大数据的应用研究,即将大数据研究与具体行业和具体问题相结合,研究如何在相关领域应用大数据分析,甚至运用大数据技术去开拓全新的学习、工作和生活方式。大数据分析的实践已经在全世界范围内引起了广泛的关注。当前,大数据分析在有些行业和领域已经能够得以实践,其应用结果比传统数据分析方法更为科学和有效,大数据正在改变我们生产、生活、工作、学习的方方面面。各大知名企业,如国际商业机械公司(IBM,下同)、亚马逊公司等开始用大数据技术改进经营和生产流程;各家媒体公司,如谷歌(Google)、脸谱网(Facebook)等将大数据用于客户关系管理、客户细分、广告经营等领域。美国奥巴马政府更是把大数据推高到了国家战略层面,并且制定了一系列推动大数据发展的政策措施。

能否更好地组织、管理和利用数据信息对企业来说越来越重要。企业可获得的数据在急剧增长,对客户数据进行全面收

集、集成、分析与利用是 CRM 实施的基础。如何管理和分析海量的客户信息，并从中找出对 CRM 决策有价值的知识，这需要更先进的技术和工具的支持，即数据挖掘技术。

诺贝尔物理学奖得主彭齐亚斯（Penzias）博士曾评论说：“数据挖掘将变得更加重要，由于数据挖掘如此有价值以至于企业将不再会丢失与其客户有关的事物。如果你不在这方面做些什么，那么你将失去你的生意。”数据挖掘技术的应用正逐渐渗透到现代生活的各个领域。现实世界中堆积了越来越多的各种数据，如何将这些数据转化成有用的信息和知识，已经成为当代人尤其是商业、企业、各种社会机构甚至是政府部门迫切需要解决的问题。

统计数据挖掘技术以其强大的从繁杂数据中发现规律和知识的功能可以为 CRM 的建立和实施提供强有力的技术支撑。

1.2 研究的目的及意义

以客户为中心的经营模式是新一代企业经营的标志，越来越多的企业开始通过实施 CRM 来保持和赢得更多的有价值的客户，进而提高客户的忠诚度、满意度以及降低企业经营成本。因此，要提高企业竞争力、以客户为中心、提高企业利润，实施 CRM 是必然的趋势。

CRM 在国外的应用已经取得了极大的成功，而我国对其认识尤其是应用还处于起步阶段。我国企业面临国内外企业强大的竞争压力，要想在竞争中脱颖而出，就必须加快自身的信息化建设步伐，以自身的基本情况为出发点，参考和借鉴国外先进管理理念、模式和成熟的技术方法，这样才能建设出符合中国特色的 CRM。市场经济浪潮下的企业经济活动的最终目标就