

营销科学学报

Journal of Marketing Science

JMS

清华大学经济管理学院

北京大学光华管理学院

- 客户流失预测中两类错误的平衡控制研究 / 夏国恩 金炜东 1
- 消费者个人特征及民族中心主义对品牌原创地识别准确度的影响——基于中国城市消费者的实证研究 / 朱 凌 陆雄文 褚荣伟 8
- 具备重复购买机制的新产品扩散模型：理论模型与非线性最小一乘估计 / 常 莹 李 季 王汉生 涂 平 22
- 在线服务满意的决定要素及其满意对后续行为的影响研究 / 胡左浩 杨志林 32
- 电子集市的网络外部性与定价 / 郭 强 42
- 消费者非伦理行为形成过程：理论构建与机理探讨 / 曾伏娥 51
- 旅游虚拟社区服务质量与社区成员忠诚感研究 / 谢礼珊 王 鼎 72
- 感知风险对信息搜寻的影响：消费者知识的调节作用 / 邹德强 王 高 赵 平 83
- 中国市场关系投资对顾客忠诚影响的实证研究——来自四川、云南和浙江三省酒店业顾客赢回管理的调查 / 唐小飞 周庭锐 陈淑青 96



清华大学出版社

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13501256678 13801310933

图书在版编目(CIP)数据

营销科学学报.第2卷第4辑/清华大学经济管理学院,北京大学光华管理学院编.一北京:清华大学出版社,2007.2

ISBN 978-7-302-14523-3

I. 营… II. ①清… ②北… III. 市场营销学—丛刊 IV. F713.50-55

中国版本图书馆 CIP 数据核字(2007)第 003821 号

责任编辑:徐学军 王巧珍

责任校对:宋玉莲

责任印制:孟凡玉

出版发行:清华大学出版社 地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn> 邮 编:100084

c-service@tup.tsinghua.edu.cn

社总机:010-62770175 邮购热线:010-62786544

投稿咨询:010-62772015 客户服务:010-62776969

印刷者:清华大学印刷厂

装订者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:205×282 印 张:7 字 数:182 千字

版 次:2007 年 2 月第 1 版 印 次:2007 年 2 月第 1 次印刷

印 数:1~2000

定 价:30.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770177 转 3103 产品编号:024744-01

《营销科学学报》编委会

主 编

赵 平 (清华大学)

副主编

符国群 (北京大学)

主编助理

陈 荣 (清华大学)

彭泗清 (北京大学)

编委会委员 (按汉语拼音排序)

白长虹 (南开大学)

晁钢令 (上海财经大学)

丁 敏 (美国宾州州立大学)

董大海 (大连理工大学)

范秀成 (复旦大学)

郭国庆 (中国人民大学)

胡左浩 (清华大学)

黄 静 (武汉大学)

黄 沛 (上海交通大学)

黄升民 (中国传媒大学)

贾建民 (香港中文大学)

蒋炯文 (长江商学院)

江明华 (北京大学)

蒋青云 (复旦大学)

景奉杰 (华中科技大学)

刘 益 (西安交通大学)

卢泰宏 (中山大学)

陆 娟 (中国农业大学)

陆雄文 (复旦大学)

吕一林 (中国人民大学)

苏晨汀 (香港城市大学)

田志龙 (华中科技大学)

涂 平 (北京大学)

汪 波 (天津大学)

汪 涛 (武汉大学)

王 高 (清华大学)

王方华 (上海交通大学)

王海忠 (中山大学)

王兴元 (山东大学)

王永贵 (南京大学)

武建安 (美国新奥尔良图伦大学)

夏春玉 (东北财经大学)

谢贵枝 (香港大学)

许敬文 (香港中文大学)

于洪彦 (吉林大学)

张 实 (美国加州大学洛杉矶分校)

郑毓煌 (美国福德汉姆大学)

周 南 (香港城市大学)

周庭锐 (西南交通大学)

周 政 (香港大学)

庄贵军 (西安交通大学)

邹绍明 (美国密苏里大学)

Editorial Board of Journal of Marketing Science

Editor

Zhao Ping, Tsinghua University

Associate Editor

Fu Guoqun, Peking University

Assistant to the Editor

Chen Rong, Tsinghua University

Peng Siqing, Peking University

Members of Editorial Board

Bai Changhong, Nankai University

Chao Gangling, Shanghai University of Finance & Economics

Ding Min, Pennsylvania State University, USA

Dong Dahai, Dalian University of Technology

Fan Xiucheng, Fudan University

Guo Guoqing, Renmin University of China

Hu Zuohao, Tsinghua University

Huang Jing, Wuhan University

Huang Pei, Shanghai Jiao Tong University

Huang Shengmin, Communication University of China

Jia Jianmin, The Chinese University of Hong Kong

Chiang Jeongwen, Cheung Kong Graduate School of Business

Jiang Minghua, Peking University

Jiang Qingyun, Fudan University

Jing Fengjie, Huazhong University of Science & Technology

Liu Yi, Xi'an Jiaotong University

Lu Taihong, Sun Yat-Sen University

Lu Juan, China Agricultural University

Lu Xiongwen, Fudan University

Lu Yilin, Renmin University of China

Su Chenting, City University of Hong Kong

Tian Zhilong, Huazhong University of Science & Technology

Tu Ping, Peking University

Wang Bo, Tianjin University

Wang Tao, Wuhan University

Wang Gao, Tsinghua University

Wang Fanghua, Shanghai Jiao Tong University

Wang Haizhong, Sun Yat-Sen University

Wang Xingyuan, Shandong University

Wang Yonggui, Nanjing University

Wu Jian'an, Tulane University, USA

Xia Chunyu, Dongbei University of Finance & Economics

TSE, David K., The University of Hong Kong

Hui, Michael K., The Chinese University of Hong Kong

Yu Hongyan, Jilin University

Zhang Shi, University of California, Los Angeles, USA

Zheng Yuhuang, Fordham University, USA

Zhou Nan, City University of Hong Kong

Zhou Tingrui, Southwest Jiaotong University

Zhou Zheng, The University of Hong Kong

Zhuang Guijun, Xi'an Jiaotong University

Zou Shaoming, University of Missouri, USA

大 事 记

1. 清华大学经济管理学院学术委员会和北京大学光华管理学院学术委员会分别于 2006 年 9 月和 12 月将《营销科学学报》(JMS) 认定为各自单位的核心学术期刊。

2. “2006JMS 营销科学学术年会”于 2006 年 10 月 21 日~22 日在武汉大学成功举行,会议主题为“中国市场营销:全球化与本土化”。会议共收到论文 242 篇,评选出优秀论文 10 篇。本次会议还邀请了世界著名营销学者、美国杜克大学教授、国际顶级营销学报 Journal of Marketing Research 主编 Joel Huber 等做了主题演讲。

3. “2006JMS 营销科学博士研究生学术论坛”于 2006 年 10 月 20 日在武汉大学成功举行,会议共收到论文 63 篇。经过专家组多轮匿名评审,由西安交通大学管理学院博士生张志勇提交的论文《控制机制使用、态度性承诺与知识转移间的关系:假设与实证检验》获得优秀论文一等奖和 1 万元奖金;由北京大学光华管理学院博士研究生华婷提交的论文《核心与非核心产品考虑域之间相互影响的不对称性》,以及由武汉大学经济与管理学院博士研究生郭锐提交的论文《中国企业品牌战略决策研究:基于环境不确定性和企业家导向》获得优秀论文二等奖和 5 000 元奖金;此外还有 5 篇论文获得优秀奖。

4. “JMS 博士生研究基金”公布了 2006 年到 2007 年度获得资助的博士研究生名单,他们是南开大学的杜建刚、武汉大学的望海军、清华大学的刘佳和北京大学的丁嘉莉,分别获得 5 000 元的研究资助。本年度基金由香港张健儒先生出资。

5. 《营销科学学报》于 2006 年 10 月 21 日召开了编委会和理事会会议,编委会表决通过了新增黄静、丁敏、郑毓煌、蒋炯文、邹绍明和白长虹 6 位编委的决议;理事会表决通过了新增香港城市大学商学院、上海财经大学国际工商管理学院和中南财经政法大学工商管理学院 3 个理事单位的决议。

主编感谢下列评审人与编委们一起贡献他们宝贵的时间和专业知识帮助学报评审从 2005 年 1 月 1 日起至 2006 年 12 月 31 日止的所有来稿。

陈信康

上海财经大学

崔新健

中央财经大学

戴 鑫

华中科技大学

董春莲

University of Alaska Fairbanks

甘碧群

武汉大学

高蕾蕾

Stanford University

郭国庆

中国人民大学

郭贤达

北京大学

何佳讯

华东师范大学

黄劲松

北京航空航天大学

黄京华

中国传媒大学

黄敏学

武汉大学

贾生华

浙江大学

江 岚

University of British Columbia

姜旭平

清华大学

姜宇威

Hong Kong University of Science and Technology

金立印

Paichai University, Korea

雷 静

University of Ontario Institute of Technology

李宝库

清华大学

李纯青

西安工业大学

李海洋

Rice University

李怀斌

东北财经大学

李 娟

City University of Hong Kong

李 蔚

四川大学

李秀平

National University of Singapore

刘丁己

澳门科技大学

陆奇斌

对外经济贸易大学

鲁 直

复旦大学

马振峰

McGill University

毛慧芳

Indiana University

孟露芳

University of Minnesota-Twin Cities

邵兵家

重庆大学

施铁冰

Queen's University

宋学宝

清华大学

宋亦平

复旦大学

苏 萌

北京大学

唐以明

Macquarie University

涂荣庭

北京大学

万 方

University of Manitoba

万 雯

Northwestern University

王长征

武汉大学

王剑锋

City University of Hong Kong

王 霞

中国人民大学

汪旭辉

清华大学

王 毅

清华大学

王正元

University of Technology, Sydney

魏立原

City University of Hong Kong

吴 思

武汉大学

吴泗宗

同济大学

吴小丁

吉林大学

熊元斌

武汉大学

许正良

吉林大学

薛求知

复旦大学

杨晓京

Indiana University Bloomington

杨志林

City University of Hong Kong

于春玲

清华大学

张红霞

北京大学

张 萌

The Chinese University of Hong Kong

张绪彬

Hong Kong University of Science and Technology

张 岩

University of Chicago

张 影

University of Chicago

赵 冰

首都经贸大学

赵广志

University of Kansas

赵 敏

University of North Carolina

赵 昕

University of Hawaii at Manoa

曾伏娥

武汉大学

《营销科学学报》征稿简则

1. 征稿范围

本刊倡导学术研究的科学精神和规范方法,鼓励对市场营销及消费行为现象进行创新性理论探究和求证。来稿要求理论创新显著,研究方法规范,对营销实践具有深刻管理内涵。

2. 投稿要求

- 1) 封面页格式:中文标题→中文作者姓名、学位→工作单位、职称、通讯地址、电话、电子邮箱、传真号码、致谢→中文摘要(不超过150字)→中文关键词→英文标题→英文作者姓名、学位、职称、工作单位、通讯地址、电子邮箱→英文摘要(不超过100个单词)→英文关键词;多名作者署名的论文,请标明通讯作者。
- 2) 论文正文格式:中文标题→中文摘要→中文关键词→英文标题→英文摘要→英文关键词→正文(正文中大的章节标题居中,小的子标题左对齐)→附录。论文正文中不要出现关于作者的任何信息。
- 3) 文章中的数学公式按出现的先后顺序统一编号。数学公式的标号放入圆括号内(例如(1)),并标注在公式行的右边。表或图按在文中出现的先后顺序统一编号(例如图1、图2……)。表的标号和标题放在相应表格上方,图的标号和标题放在相应图的下方。另外,尽可能不要使用脚注或尾注。
- 4) 参考文献在文中出现标注在圆括号中,如(Wensley 1981),如果作者名需在正文中出现,标注成Wensley (1981);多个参考文献出现的顺序按照第一作者的英文字母顺序排列,如(Donnelly and George, 1981; Iacobucci, Grayson 和 Ostrom, 1994; Nevin 和 Smith, 1981);四个作者以上的参考文献只需写出第一个作者,后加“等”,如(Rust 等,1999)。中文参考文献在文中的标注原则相同,形式如(胡晓云,2003;王永贵和董大海,2004;王海忠,于春玲和赵平,2006)。
- 5) 文后所附参考文献先列中文文献,然后列英文文献,接着其他文参考文献。无论是何种语言的参考文献,都遵循相同格式。具体格式如下:

中文参考文献的著录按汉语拼音形式排序(即A、B、…、Z)。格式如下:

作者.专著书名.编者名.出版地:出版社,出版年:起止页码。

作者.期刊文章名.刊名,出版年,卷(期),起止页码。

作者.报纸文章名.报纸名,版次,(出版日期)。

作者.电子文献文章名.文章出处或可获得地址,发表或更改日期/引用日期(可选)。

英文书籍、期刊、书籍中的某篇文章的著录格式按英文拼音排序。示例如下:

[1] CRAWFORD W, GORMAN M. Future libraries: dreams, madness, & reality [M]. Chicago: American Library Association, 1995.

[2] DES MARAIS D J, STRAUSS H, SUMMONS R E, et al. Carbon isotope evidence for the stepwise oxidation of the Proterozoic environment [J]. Nature, 1992, 359:605-609.

[3] TURCOTTE D L. Fractals and chaos in geology and geophysics [M/OL]. New York: Cambridge University Press, 1992[1998 09 23]. <http://www.seg.org/reviews/mccorm30.html>.

请在参考文献部分中每条索引前面加序号;每条索引的各行与第一行对齐。对于任何未列出的格式要求,请参照 Journal of Marketing 或联系本编辑部。

3. 稿件处理

来稿录用与否,本刊都会通知作者。若来稿未被录用,本刊限于人力不予退稿,请作者自留底稿。

4. 通讯地址:北京市海淀区清华大学经济管理学院舜德楼南 302 室《营销科学学报》编辑部

联系电话:62772941 传真:62772941 邮政编码:100084

电子邮件:postmaster@jms.org.cn 网站:<http://www.jms.org.cn>

Journal of Marketing Science

Volume 2, Number 4

December 2006

Tradeoff of Errors of Two Types in Customer Churn Prediction	7	Xia Guo-en, Jin Weidong
The Influences of Consumers' Personal Characteristics and Influence of Consumer Characteristics and Ethnocentric Tendency on Brand Origin Recognition Accuracy: An Empirical Study on Urban Chinese Consumers	21	Zhu Ling, Lu Xiongwen, Chu Rongwei
Diffusion of New Product with Repeat Purchase Potential: Theoretical Model and Nonlinear LAD Estimation	31	Chang Ying, Li Ji, Wang Hansheng, Tu Ping
Determinants of Customer Satisfaction with Online Service and its Impacts on Behavioral Intentions	41	Hu Zuohao, Yang Zhilin
Network Externalities and Pricing in E-marketplace	50	Guo Qiang
Formation of Unethical Consumer Behavior: Theoretical Concepts and Key Mechanism	71	Zeng Fue
Service Quality of Virtual Tourist Community and Community Member Loyalty: An Empirical Study	82	Xie Lishan, Wang Ding
How Perceived Risk Affects Information Search: The Moderating Role of Consumer Knowledge	95	Zou Deqiang, Wang Gao, Zhao Ping
Effects of <i>Guanxi</i> Investment on Customer Loyalty in the Chinese Market: An Empirical Study—Evidence from the Hotel Industry in SiChuan, YunNan and ZheJiang	107	Tang Xiaofei, Chou TingJui, Chen Shuching

夏国恩^①, 金炜东^②

摘要 在客户流失预测过程中不可避免地会出现“拒真纳伪”两类错误。本文通过对“拒真纳伪”两类错误在客户流失预测中不同影响的分析比较,采用支持向量机(SVM)作为预测模型,并利用某电信公司实际数据对两类错误的平衡控制进行了研究。实验结果表明,选取一个适当的损失比例系数,预测模型能在控制两类错误的前提下有效地减少期望损失函数值,这在实际应用中具有反映问题本质的现实意义。

关键词 两类错误, 客户流失, 支持向量机

客户流失预测中两类错误的平衡控制研究

0 引言

客户流失预测是客户关系管理中一项重要的基础性工作。为了有效地预测未来潜在的流失客户,学者们主要提出了以下两类预测模型:第一类是传统预测模型,如决策树(Chih 和 Chiu, 2002)、Logistic 回归(Logistic Regression)(Kim 和 Yoon, 2004; Rosset 和 Neumann, 2003)、贝叶斯分类器(Naive Bayesian Classifiers)(Nath, 2003)、聚类分析(Clustering)(Yi Ming 等, 2004);第二类是人工智能预测模型,如人工神经网络(Artificial Neural Network, ANN)(Yan 等, 2001)、自组织映射(Self Organizing Maps, SOM)(Ultch, 2002)和进化学习(Evolutionary Learning, EL)算法(Au 等, 2003)。在使用模型进行预测的过程中,不可避免地会犯“拒真纳伪”两类错误。这两类错误对于客户保持的影响有很大的差别。然而,无论是传统预测模型还是人

工智能预测模型,都无法直接控制两类错误的分布并且其模型评价标准也没有考虑两类错误的关系,从而影响了它们在实际应用中的效果。

针对上述问题,笔者以支持向量机(Support Vector Machine, SVM)为预测工具,讨论了客户流失预测中的两类错误的平衡控制问题。文中引进了一个损失比例系数 Y 用其来调整第一类错误与第二类错误间惩罚系数的比率,在此基础上定义了评价预测模型的客户流失期望损失函数。通过对美国某电信公司未来潜在的流失客户进行预测,发现在当前和未来时间条件下, SVM 可以在不同的错误类别上采用不同的惩罚系数,从而有效控制“拒真纳伪”两类错误的分布,并且以客户流失期望损失函数值为预测模型的评价标准比模型总错误率更具有实际意义。

1 客户流失预测中的两类错误

1.1 假设检验中的“拒真纳伪”两类错误

在统计学的假设检验中由于样本的随机性,在进行判断时可能犯两类错误。一是当假设 H_0 为真时,拒绝了它,称为犯第一类错误发生的概率或称拒真概率,记作

$$\alpha = P\{X \in W \mid H_0 \text{ 真}\} \quad (1)$$

^① 夏国恩,西南交通大学经济管理学院博士研究生, E-mail: gandlf007711@163.com

^② 金炜东,西南交通大学经济管理学院教授、博士生导师, E-mail: WDJin@vip.sina.com

其中, W 是一个检验的拒绝域。二是当假设 H_0 为非真时, 而接受了它, 称为犯第二类错误发生的概率或称纳伪概率, 记作

$$\beta = P\{X \notin W \mid H_0 \text{ 非真}\} \quad (2)$$

犯这两类错误所造成的影响常常很不一样。以电信业根据客户的可能流失情况判断是否给对客户进行客户保持为例, 原假设为 H_0 (该客户可能流失)。此时, 犯第二类错误会使企业多花费用于客户保持的费用, 但犯第一类错误可能导致客户流失。企业希望根据历史样本的检验结果做出的预测使犯两类错误的概率都尽可能小, 但实际上是不可能的。由于两类错误之间存在制约关系(蔡越江, 1999): 当其他条件不变时 α 减小必导致 β 增大; 反之, β 减小则 α 增大。因此, 在样本容量一定的情况下, 不能同时控制犯两类错误的概率大小。在统计检验中, 一般采取限制第一类错误的概率, 即选一个正数作为 α 的上限, 这个正数通常称为检验水平或显著水平。其他常用的解决方法是: 增大样本容量, 尽量采用单边检验等。在实际应用时, 必须根据客观事物的背景, 恰当选取合适的 α 或合适的 β 。

1.2 客户流失预测中分析两类错误的重要性

构建一个适用的客户流失预测模型是企业进行客户关系管理的有力保障。这里将需要评判的客户分为: 流失客户和非流失客户。影响模型性能的主要因素之一是误判率, 即: 将流失客户评判为非流失客户和将非流失客户评判为流失客户这两类错误所引起的误判比率。根据“尽量使后果严重的错误成为第一类错误”的原则, 将流失客户评判为非流失客户称为第一类错误, 将非流失客户评判为流失客户称为第二类错误。

客户流失预测中出现的两类错误是统计学中两类错误在具体应用中的一种表现, 因此具有上述两类错误的基本性质: 由于样本的随机性及样本容量的有限性, 无法同时控制犯两类错误的概率 α 和 β 都很小, 大多数客户流失研究一味地强调模型整体的准确率, 却忽视了两类错误对企业客户保持的不同影响, 以至于实际应用效果并不

理想, 因此有必要对这两类错误进行深入的分析 and 探讨。就本文所讨论的问题而言, 企业犯第二类错误至多增加一笔进行客户保持的费用, 而犯第一类错误则面临着客户流失的巨大风险, 因此第一类错误的危害性要远比第二类错误严重。Bhattacharya 等(1998)的研究指出: 在客户流失预测中, 第一类错误造成的损失为第二类错误损失的 5 倍~6 倍。两类错误之间的制约关系及对实际问题的不同影响, 要求企业在进行客户流失分析时, 除了尽可能地规避风险较大的第一类错误, 还要减少由两类错误所带来的期望损失。

2 基于改进支持向量机的客户流失预测模型

根据预测问题的不同, 支持向量机(SVM)可分为支持向量分类机(Support Vector Classifier, SVC)和支持向量回归机(Support Vector Regression, SVR)。在标准的 C-SVC (Cortes 和 Vapnik, 1995)中, $C > 0$ 是一个常数, 它控制对错分样本惩罚的程度, 并且对于错分样本的惩罚相同。因此, 没有对第一类错分和第二类错分的样本分别进行统计。为了刻画客户流失预测中两类错误存在的差异, 考虑在标准 C-SVC 目标函数中, 对它们分别采用不同的惩罚系数 C_1 和 C_2 ($C_1, C_2 > 0$), 并通过对 C_1, C_2 的调整来控制两类错误的分布。

设样本集为 $(x_i, y_i), i = 1, 2, \dots, n, x_i \in R^n, y_i \in \{1, -1\}$ 是类别标号。为了对第一类错分和第二类错分的样本类别采用不同的惩罚参数 C , 令 C_1 是第一类错误的惩罚系数, 即对正类点集的惩罚系数, C_2 是第二类错误的惩罚系数, 即对负类点集的惩罚系数。此时 C-SVC 的原始问题形式变为

$$\begin{aligned} \min \phi(w) &= \frac{1}{2} \|w\|^2 + \sum_{y_i=1} C_1 \xi_i + \sum_{y_i=-1} C_2 \xi_i \quad (3) \\ \text{s.t. } y_i[(w \cdot x_i) + b] &\geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, \dots, n \end{aligned} \quad (4)$$

引入 Lagrange 系数 $a_i \geq 0, \beta_i \geq 0$, 构造 Lagrange 函数

$$L = \frac{1}{2} \|w\|^2 + \sum_{y_i=1} C_1 \xi_i + \sum_{y_i=-1} C_2 \xi_i - \sum_{i=1}^n [a_i(y_i(w \cdot x_i + b) - 1 + \xi_i) + \beta_i \xi_i] \quad (5)$$

计算偏微分,并令偏微分等于零,得

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n a_i y_i x_i = 0 \quad (6)$$

$$\frac{\partial L}{\partial b} = - \sum_{i=1}^n a_i y_i = 0 \quad (7)$$

$$\frac{\partial L}{\partial \xi_i} = \begin{cases} C_1 - a_i - \beta_i = 0, y_i = 1 \\ C_2 - a_i - \beta_i = 0, y_i = -1 \end{cases} \quad (8)$$

将式(6)、式(7)、式(8)代入式(5)得到对偶表达式

$$\min Q(a) = \frac{1}{2} \sum_{i,j=1}^n a_i a_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^n a_i \quad (9)$$

$$\text{s. t.} \quad \sum_{i=1}^n y_i a_i = 0; \quad 0 \leq a_i \leq C_i, y_i = 1; \quad 0 \leq a_i \leq C_2, y_i = -1 \quad (10)$$

一般情况下,该优化问题解的特点是大部分 a_i 将为零,其中不为零的 a_i 所对应的样本为支持向量。根据 KKT(Cortes 和 Vapnik, 1995) 条件,在鞍点有

$$a_i [y_i(w \cdot x_i + b) - 1 + \xi_i] = 0, i = 1, \dots, n \quad (11)$$

$$(C_1 - a_i) \xi_i = 0, y_i = 1 \quad (12)$$

$$(C_2 - a_i) \xi_i = 0, y_i = -1 \quad (13)$$

于是可得 b 的计算式

$$y_i \left(\sum_{j=1}^l a_j y_j (x_j \cdot x_i) + b \right) - 1 = 0 \quad (14)$$

因此,可以通过任意一个支持向量求出 b 的值。最后得到最优分类函数为

$$f(x) = \text{sign} \left\{ \sum_{i=1}^n a_i y_i (x_i \cdot x) + b \right\} \quad (15)$$

对于非线性问题, Vapnik 引入了核空间理论:将低维的输入空间数据通过非线性映射函数映射到高维属性空间,将分类问题转化到属性空间进行。可以证明,如果选用适当的映射函数,输入空间线性不可分问题在属性空间将转化为线性可分问题。这种非线性映射函数被称之为

核函数 (Vapnik, 2004)。从理论上讲,满足 Mercer 条件的对称函数 $K(x, x')$ 都可以作为核函数。引入核函数后,最优分类函数为 $f(x) = \text{sign} \left\{ \sum_{i=1}^n a_i y_i K(x_i \cdot x) + b \right\}$ 。

3 客户流失预测中两类错误的平衡控制实证研究

3.1 指标体系的建立

本文以电信业客户流失预测为例,来对两类错误的平衡控制进行研究。目前在电信业客户流失中采用的指标可分为:客户基本特征、客户通话行为、客户接触信息、客户签约信息、产品特征等。根据 Wei Yu 等的 Delta 客户流失管理战略模型 (Wei 和 Jutla, 2005),并综合考虑客户流失的各影响因素和数据的可获得性,选取产品特征、客户方案和客户信息三类指标共 52 个(见表 1)。

3.2 样本数据处理

本文数据来源于美国 DUKE 大学 TERADATA 客户关系管理中心,该数据包括 6 个月的客户数据,因此本研究将前 4 个月的客户数据作为输入指标,第 6 个月的客户状态为输出指标,中间预留 1 个月作为时间延迟。首先对有 30% 以上缺失值的指标进行删除;然后通过指标均值对缺失项进行处理;最后进行稳健性处理,选用两倍、三倍标准差检验进行异常数据剔除。通过抽样,最终获得 5 617 个训练样本数据(其中 2 849 个为非流失客户,2 768 个为流失客户),3 107 个与训练样本相同时间下的测试样本数据(其中 2 705 个为非流失客户,402 个为流失客户),2 484 个比上述测试样本数据晚 2 个月~3 个月的未来时间下的未来客户样本数据(其中 2 191 个为非流失客户,293 个为流失客户)。通过利用 SPSS11.5 对训练样本的数据进行因子分析,并用方差最大正交旋转法 (VARIMAX),在特征值大于 1 的情况下,当前样本数据条件下的解释因子数为 10,其累计方差贡献为 86.16%,从而指

标被分成标准使用因子、生命周期使用因子、付费意愿因子、额外开销因子、使用时间长度因子、客户关怀因子、无线语音使用因子、产品价值因子、产品质量因子、使用变化因子(见表2)。

表1 客户流失预测指标

产品特征	手机价格、日前手机使用的时间、网络掉线语音通话次数、网络阻塞语音通话次数、网络掉线、阻塞语音呼叫总次数
客户方案	无应答语音通话次数、语音通话主叫次数、接听语音通话次数、成功的语音通话次数、客户关怀通话次数、客户关怀通话使用的取整时间长度、客户关怀通话使用的非取整时间长度、少于一分钟的被叫通话次数、成功的语音通话使用的非取整时间长度、接听语音通话使用的非取整时间长度、国际无线语音通话次数、国际无线语音通话使用的非取整时间长度、国内无线语音通话次数、国内无线语音通话使用的非取整时间长度、峰值语音通话次数、峰值语音通话使用的非取整时间长度、非峰值语音通话次数、非峰值语音通话使用的非取整时间长度、通话主叫次数、成功的通话次数、通话等待次数、总的服务月份数、客户生命周期内总的通话次数、客户生命周期内总的通话时间长度、客户生命周期内调整的总的通话时间长度、客户生命周期内调整的总的通话次数、客户生命周期内月均通话时间长度、客户生命周期内月均通话次数、过去3个月月均通话时间长度、过去3个月月均通话次数、月均通话费用、月均通话时间长度、月均连续消费费用、查号呼叫次数、月均超时使用时间长度、过去3个月内月均通话时间长度改变的百分数、过去6个月月均通话时间长度、过去6个月月均通话次数
客户信息	第一家庭成员的年龄、总的通话收益、客户生命周期内调整的总的通话收益、客户生命周期内月均通话收益、过去3个月月均通话收益、月均超额收益、月均语音通话超额收益、过去3个月内月均收益改变的百分数、过去6个月月均通话收益

表2 因子分析结果

因子	指标	因子载荷	累积方差贡献率(%)
标准使用因子	无应答语音通话次数	0.77	25.75
	语音通话主叫次数	0.87	
	接听语音通话次数	0.82	
	成功的语音通话次数	0.86	
	少于一分钟的被叫通话次数	0.82	
	接听语音通话使用的非取整时间长度	0.61	
	峰值语音通话次数	0.81	
	峰值语音通话使用的非取整时间长度	0.58	
	非峰值语音通话次数	0.81	
	通话主叫次数	0.87	
	成功的通话次数	0.86	
	通话等待次数	0.59	
	客户生命周期内月均通话次数	0.61	
	过去3个月月均通话次数	0.76	
	过去6个月月均通话次数	0.70	
生命周期使用因子	总的服务月份数	0.59	38.62
	客户生命周期内总的通话次数	0.86	
	客户生命周期内总的通话时间长度	0.82	
	客户生命周期内总的通话收益	0.79	
	客户生命周期内调整的总的通话收益	0.79	
	客户生命周期内调整的总的通话时间长度	0.82	
	客户生命周期内调整的总的通话次数	0.85	

续表

因子	指标	因子载荷	累积方差贡献率(%)
付费意愿因子	月均连续消费费用	0.73	47.21
	客户生命周期内月均通话收益	0.85	
	月均通话收益	0.72	
	过去3个月月均通话收益	0.73	
	过去6个月月均通话收益	0.76	
额外开销因子	月均超时使用时间长度	0.94	55.11
	月均超额收益	0.94	
	月均语音通话超额收益	0.94	
使用时间长度因子	客户生命周期内月均通话时间长度	0.61	62.50
	成功的语音通话使用的非取整时间长度	0.66	
	非峰值语音通话使用的非取整时间长度	0.69	
	月均通话时间长度	0.57	
	过去3个月月均通话时间长度	0.61	
	过去6个月月均通话时间长度	0.61	
客户关怀因子	客户关怀通话次数	0.82	68.37
	客户关怀通话使用的取整时间长度	0.96	
	客户关怀通话使用的非取整时间长度	0.93	
无线语音使用因子	国际无线语音通话次数	0.57	74.13
	国际无线语音通话使用的非取整时间长度	0.65	
	国内无线语音通话次数	0.78	
	国内无线语音通话使用的非取整时间长度	0.84	
产品价值因子	手机价格	0.64	79.22
	目前手机使用的时间	-0.74	
产品质量因子	网络阻塞语音通话次数	0.92	82.72
	网络掉线和阻塞语音呼叫总次数	0.72	
使用变化因子	过去3个月内月均通话时间长度改变的百分数	0.87	86.16
	过去3个月内月均收益改变的百分数	0.87	

3.3 支持向量机模型的构造

根据上述分析,构造样本集 (x,y) 。其中, x 的维数为10; y 是样本的类别属性,对于流失客户, $y=1$,对于非流失客户, $y=-1$ 。对于核函数和参数的选择,通过在MATLAB6.5上进行实验分析并与线性核函数、多项式核函数进行对比发现,采用参数 $\sigma=1.11$ 的径向基核函数 $K(x_i,x)=\exp\left\{-\frac{|x-x_i|^2}{2\sigma^2}\right\}$ 来构造SVC模型,在客户流失率第一个10分位上能取得较好效果(这里利用支持向量回归机对客户流失概率进行预测,模型也采用径向基核函数,其参数为 $\sigma=0.24,C=$

$12,\epsilon=0.02$)。对于SVC惩罚参数,选取适当的 C_1,C_2 来控制两类错误的分布,对惩罚系数相同($C_1=C_2$)及不同($C_1=YC_2,Y=0.2\sim 2$)的情况进行了建模,其中 $Y(Y>0)$ 称为损失比例系数。

3.4 实证结果分析

表4对两类错误平衡控制下的测试集内客户流失率在第一个10分位上预测结果进行了对比。模型评价标准由表3可得:第一类错误率= $B/(A+B)$;第二类错误率= $C/(C+D)$;总的错误率= $(B+C)/(A+B+C+D)$ 。另外,期望损失函数定义为(Joos等,1998)

表3 分类矩阵

单位:个

样本中客户状态	预测流失	预测非流失
实际流失	A	B
实际非流失	C	D

表4 两类错误平衡控制下的预测结果对比

损失比例系数(Y)	当前时间				未来时间			
	第一类 错误率	第二类 错误率	总错误率	期望损失 函数值	第一类 错误率	第二类 错误率	总错误率	期望损失 函数值
0.2	0.972	0.000	0.339	143.759	0.974	0.010	0.157	144.359
0.4	0.528	0.040	0.210	79.308	0.921	0.067	0.198	138.254
0.6	0.454	0.198	0.287	73.170	0.790	0.252	0.335	124.507
0.8	0.514	0.429	0.494	89.071	0.447	0.471	0.468	80.439
1	0.065	0.703	0.481	30.999	0.342	0.667	0.617	70.872
1.2	0.028	0.832	0.552	29.451	0.184	0.795	0.702	51.398
1.4	0.009	0.886	0.581	28.283	0.105	0.885	0.766	42.451
1.6	0.009	0.931	0.610	29.652	0.053	0.914	0.782	35.643
1.8	0.009	0.965	0.632	30.686	0.026	0.938	0.798	32.379
2	0.009	0.980	0.642	31.143	0.026	0.962	0.819	33.109

$EC = P_1 * L_1 * T_1 + P_2 * L_2 * T_2$ (16)

其中,EC为损失的期望, P_1 为训练集中“1”类所占比率, P_2 为训练集中“-1”类所占比率, L_1 为将一个“1”类错分为“-1”类所造成的损失(文中为300美元), L_2 为将一个“-1”类错分为“1”类所造成的损失(文中为50美元)(Athanasopoulos,2000), T_1 为第一类错误率, T_2 为第二类错误率。从表4中可以看出,在预测当前时间下的潜在流失客户时,随着损失比例系数的增加,第一类错误率逐渐下降,第二类错误率逐渐上升,总的错误率有上升的趋势,期望损失函数值起初逐渐减少,当减少到一定值时(当前时间为28.283, $Y=1.4$;未来时间为32.379, $Y=1.8$),到达最小,随后又逐渐增加。上述现象说明,在SVM预测模型中,通过调整两类错分的惩罚系数,可以控制两类错误的分布率,并且使总错误率达到最小的损失比例系数,却不能使期望损失函数值达到最小,即采用总错误率来评价预测模型是不妥的。因此,考虑实际应用背景,较理想的客户流失预测模型应尽可能地避免第一类错误引起客户流失带来的巨大风险,同时也要考虑到第二类错误引起客户保持费用的增加。

4 结束语

“拒真纳伪”两类错误是许多实际应用领域研究的重要问题,如何控制两类错误要视具体问题具体对待。就本文讨论的客户流失预测问题,引入了可以对两类错误进行控制的C-SVC,并定量地研究它们之间的关系。实证结果表明:在SVM预测模型中,通过调整两类错分的惩罚系数,可以控制两类错误的分布率;采用文中定义的期望损失函数来评价预测模型更有实际意义。在上述研究中,其预测模型的范围仅限国外电信业,在后续工作中,可以考虑将其推广到国内电信业和其他客户信息数据的分类预测,这样,两类错误的平衡控制研究就会更有意义。

参考文献

[1] 蔡越江.论假设检验中的两类错误[J].数理统计与管理,1999,18(3):30-35.
[2] VAPNIK V N.统计学习理论[M].许建华,张学工,译.北京:电子工业出版社,2004.

- [3] ATHANASSOPOULOS A D. Customer satisfaction cues to support market segmentation and explain switching behavior[J]. *Journal of Business Research*, 2000, 47 (3):191-207.
- [4] AU W H, CHAN K C C, YAO X. A novel evolutionary data mining algorithm with applications to churn prediction [J]. *Evolutionary Computation, IEEE Transactions*, 2003, 7(6):532-545.
- [5] BHATTACHARYA C B. When customers are members; customer retention in paid membership contexts [J]. *Journal of the Academy of Marketing Science*, 1998, 26 (1):31-44.
- [6] CHIH P W, CHIU I T. Turning telecommunications call details to churn prediction; a data mining approach [J]. *Expert Systems with Applications*, 2002, 23(2):103-112.
- [7] CORTES C, VAPNIK V. Support vector networks [J]. *Machine Learning*, 1995, 20(3):273-297.
- [8] JOOS P, VANHOOF K, OOGHE H, et al. Credit classification; a comparison of logit models and decision trees; ECML 1998; Application of machine learning and data mining in finance; European Conference on Machine Learning, 1998 [C]. Chemnitz (Germany):1998.
- [9] KIM H S, YOON C H. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market [J]. *Telecommunications Policy*, 2004, 28(9):751-765.
- [10] NATH S V. Data warehousing and mining; customer churn analysis in the wireless industry [D]. Boca Raton, Florida; Florida Atlantic University, 2003.
- [11] ROSSET S, NEUMANN E. Integrating Customer Value Considerations into Predictive Modeling; ICDM 2003; Third IEEE International Conference on Data Mining, 2003[C]. Florida (USA):2003.
- [12] ULTCH A. Emergent self-organizing feature maps used for prediction and prevention of churn in mobile phone markets [J]. *Journal of Targeting*, 2002, 4(10):401-425.
- [13] WEI Y, JUTLA D N, SIVAKUMAR S C. A churn-strategy alignment model for managers in mobile telecom; CNSR 2005; Networks and Services Research Conference, Proceedings of the 3rd Annual 16-18 May 2005 [C]. Nova Scotia (Canada):2005.
- [14] YAN L, MILLER D J, MOZER M C. et al. Improving prediction of customer behavior in nonstationary environments; IJCNN 2001. Proc. of the International Joint Conference on Neural Networks, July 15-19, 2001 [C]. Washington, DC:2001.
- [15] YI M, Wan H, LI L, et al. Multi-dimensional model-based clustering for user-behavior mining in telecommunications industry; ICLMC 2004; proceeding of the third international conference on machine learning and cybernetics, August 26-29, 2004[C]. Shanghai:2004.

Tradeoff of Errors of Two Types in Customer Churn Prediction

Xia Guo-en, Jin Weidong

(School of Economics & Management, Southwest Jiaotong University)

Abstract Two types of errors about “rejecting true and accepting false” are inevitable in customer churn prediction. In this paper, a model based on SVM was used to predict customer churn. Through the analysis of different effects by two types of errors in customer churn prediction, issues related to the tradeoff between the errors in an American telecommunication carrier is studied. The results show that by adjusting the “loss-ratio-coefficient”, the model is efficient in reducing expectation loss function value on the condition that errors of the two types can be controlled. This reflects the real essence of the problem and can be a powerful decision tool in customer churn.

Key Words Two Types of Errors, Customer Churn, Support Vector Machine

朱凌^①, 陆雄文^②, 褚荣伟^③

摘 要 在有关来源地的研究历史中,许多学者都默认了这样一个假设:消费者知道产品或品牌的来源地,而且在选择产品时使用了该信息。作者对此提出质疑,于是在中国首次进行了品牌原创地识别准确度的测量,考察了消费者个人特征以及民族中心主义对识别准确度的影响,得出结论:中国城市消费者对品牌原创地有一定的了解,他们对品牌原创地的识别准确度与其受教育程度呈正相关;男性消费者比女性消费者更了解品牌原创地;消费者对外国品牌原创地的了解受到其民族中心主义观念的显著影响;品牌原创地识别准确度会因产品种类的不同而产生差异。

关键词 来源地,品牌原创地识别准确度(BORA),消费者民族中心主义,消费者个性特征

消费者个人特征及民族中心主义对品牌 原创地识别准确度的影响

——基于中国城市消费者的实证研究

1 研究背景和目的

自从中国改革开放以来,中国市场上各种各样消费产品日新月异、丰富多彩,国内外品牌也层出不穷,品牌竞争已经成为中国业界的焦点。各企业纷纷尝试各种手段提升其品牌影响力。其中,品牌原创地(Country of Origin of Brand或者 Brand Origin,也有人译为品牌原产地)的概念及其应用越来越受到营销实践界和理论界的关注。

品牌作为一种符号,成为企业与竞争者区分

开来的重要标志。但从另一个角度来看,品牌却能够遮掩它们所代表的产品或服务的来源地(Country of Origin)(Samiee, Shimp 和 Sharma, 2005)。在对中国市场上各种产品的品牌名称进行观察之后,我们不难发现国内外品牌在命名时存在一种有趣的现象:不少国外品牌在中国市场的名称起得比较中国化,其中很多品牌使用的中国名称能够表达出产品的理念、体现品牌特性(林升栋和邹起寿,2005),如洗发水的飘柔、饮料的百事可乐、超市的家乐福等;而不少本土品牌的名称趋于外国化,比如佰草集、柏丽丝、TCL等。这些例子都说明很多国内外品牌都很注重品牌命名策略,或本土化,或国际化,以掩藏其真实的来源地信息。从而我们可以发现,如果消费者对于品牌原创地识别准确度很高的话,某些外国化或本土化的取名策略可能需要进一步的考量。

品牌原创地概念是由来源地研究发展而来,属于后者的一个分支,是国际营销领域最热门的研究议题之一(王海忠和赵平,2004)。国外学者对来源地的研究最早可以追溯到1965年

① 朱凌,复旦大学管理学院企业管理专业在读博士,华东理工大学商学院讲师, E-mail: zhuling_pearl@163.com

② 陆雄文,复旦大学管理学院教授, E-mail: xwlu@fudan.edu.cn

③ 褚荣伟,复旦大学管理学院企业管理专业在读博士, E-mail: 012025011@fudan.edu.cn

(Schooler, 1965)。Roth 和 Romeo(1992)曾这样定义来源地:“消费者基于他们过去对某国或某地区在制造和营销方面的优劣的了解,形成了对该国或该地区产品的总体认知。”随着国际贸易的增加,消费者有大量来自不同国家的产品可以选择(Bilkey, 1993),而且每个国家的企业在进行国际营销时会遇到来自不同国家的很多竞争对手,所以关于来源地的研究越来越引起学者的重视,至今已有 900 多篇相关文献。国际贸易实践的进一步发展使得来源地概念也不断细化。多国合作制造产品的出现为来源地的研究打开了一个新的领域,于是研究者提出了品牌原创地的概念,它是指品牌始创公司总部所在的国家或地区(Ahmed 和 d'Astous, 1995; Johansson, Douglas 和 Nonaka, 1985; Kim 和 Chung, 1997; Thakor 和 Kohi, 1996; Thakor 和 Lavack, 2003)。我们将品牌的原创地归于母公司所在地,尽管许多产品都是在子公司所在地生产的,但它们都表现出统一的外形和色彩。另外, Papadopoulos 和 Heslop(1993)指出,国际化公司会将品牌定义为源于它们母国的。无论索尼和丰田在哪里生产,它们都会被认定为是日本品牌。同样,尽管耐克公司不在美国生产一双鞋子,它还是被毫无疑问地定义为美国品牌。这些都被实证研究所检验(Kim 等, 1997; Ratliff, 1989)。

在过去几十年中,有关来源地的研究主要关注来源地对于消费者决策的影响。如果某项产品特征能够帮助消费者对产品或品牌进行评价和决策,那么我们认为该产品特征具有判断功能(Attribute Diagnosticity)(Lynch, Marmorstein 和 Weigold, 1988)。显然品牌的来源地作为一种消费者可以获得的产品属性之一,它可能能够帮助消费者进行判断,这就是来源地能够影响消费者决策的内在机理。

以往关于来源地的研究都隐含了这样的一个假设:消费者在购买产品时是知道产品或品牌来源地的,而且他们在选择产品时使用了来源地信息(如 Liefeld, 2003)。然而,也不断有学者对这个假设提出质疑(如 Samiee, Shimp 和 Sharma, 2005)。Bickart(1993)的研究也指出,属性判断功

能随着时间和个人的不同而不同,因此来源地对于消费者决策的影响机制也可能将随着环境的变化而变化。因此当人们在一定条件下对失去判断功能的产品特征信息不加留意时,消费者知道并在决策中使用来源地的假设就值得质疑了。

本文针对这个质疑在中国进行了一次有关品牌原创地识别准确度的实证研究。我们首先总结前人的相关研究结果,然后借鉴 Samiee, Shimp 和 Sharma(2005)的研究成果进行中国本土化研究,以期帮助中外企业了解本土城市消费者的心理,制定出适合中国市场的品牌发展战略。本文第二部分是文献回顾,第三部分提出模型和假设,第四部分介绍研究方法和设计,接下来报告数据收集和分析结果,最后一部分是针对研究结果的讨论及不足的提示。

2 文献回顾

2.1 国外学者对来源地的研究情况

国外学者对来源地的研究已有 40 年的历史,以往的绝大部分研究都暗含了这样一个假设:消费者在购买产品时知道产品的来源地(Liefeld, 2003; White 和 Cundiff, 1978)。然而,早在 20 世纪 80 年代,就有学者在对这方面进行研究之后,得出了截然相反的结论(Hester 和 Yuen, 1987; Hugstad 和 Durr, 1986)。进入 21 世纪之后,有学者开始展开有关消费者是否能够清楚地知道产品来源地的研究,并且对来源地研究中的这个暗含假设提出了质疑(Liefeld, 2003; Samiee, Shimp 和 Sharma, 2005)。Samiee, Shimp 和 Sharma(2005)提出了品牌原创地识别准确度的概念(Brand Origin Recognition Accuracy, 以下简称为 BORA),并且建立了一个结构模型,用来验证社会经济地位、过去国际旅游经历、外语技能、年龄、性别等因素与 BORA 之间的相关性。他们发现美国消费者对于品牌原创地的了解是有限的,而且品牌原创地并不像以往文献中所说的那样对消费者非常重要。这是来源地研究中第一次对消费者的品牌原创地