基于代价敏感性学习的客户价值细分①

邹 鹏, 李一军, 郝媛媛
(哈尔滨工业大学管理学院, 哈尔滨 150001)

摘要: 在基于价值的客户细分中, 不可避免地产生"拒真纳伪"的两类错误, 由于错误分类代价差异和不同价值客户数量的不平衡分布, 基于总体准确率的数据挖掘方法不能体现由于客户价值不同对分类效果带来的影响. 本研究在代价敏感性学习机制下引入支持向量机作为分类工具, 建立基于客户价值的错分代价函数, 为适应客户价值多类别细分的要求, 将二元分类扩展为多元分类, 建立分类的期望损失函数作为分类效果的评价标准. 实验结果说明, 该方法可以更精确地控制代价敏感性和不同种分类错误的分布, 降低总体的错误分类代价, 使模型能更准确反映分类的代价, 有效识别客户价值.

关键词: 客户细分; 客户价值; 代价敏感学习; 支持向量机
中图分类号: F713 5; TP18 文献标识码: A 文章编号: 1007- 9807(2009) 01- 0048- 09

0 引 言

客户细分 (Customer Segmentation) 是企业在明确的战略、业务模式和特定的市场中, 根据客户的行为、偏好以及价值等因素对客户进行分类, 是一种基础性的分析手段, 对客户管理提供全面的信息支持. 任何高效的客户关系管理都以扎实的客户细分为基础[1]. 客户价值细分 (也有研究称为基于客户价值的细分)是根据客户为企业创造价值的能力对客户进行分类, 提供针对性的产品、服务和营销模式的过程, 能够使企业更合理的分配资源, 有效地降低成本, 同时获得更有利可图的市场渗透.

在数据挖掘领域, 基于客户价值的细分是根据客户为企业创造的价值 (利润)将客户分类, 将客户价值 (利润)作为分类标志, 基本都是采用一种普适性的数据挖掘算法, 将总体准确率作为分类性能的评价标准, 不可避免地产生"拒真纳伪"的两类错误, 忽略了客户价值分类问题中错误分类代价和不同价值客户在数量上的悬殊差异, 没能体现由于客户价值不同对分类效果带来的影响. 本研究尝试在代价敏感性学习机制下引入支持向量机作为分类工具, 解决上述问题.

1 客户价值细分文献综述及问题分析

1 1 客户价值细分文献综述

国内外学者从不同角度采用不同方法对客户价值细分进行了研究, Dwyer将 Jackson的客户流失预测引入到客户价值分类模型中, 提出"永久流失"和"暂时流失"两种客户的价值细分方法[2], 这种方法体现了客户价值的动态性, Berger和 Nasr完善了这一方法, 进一步将两类细化为五类[3], 但还不能实现对单体客户的价值细分. Reichheld和 Sasser在对多个行业的大量实证研究基础上, 根据客户所处在生命周期的不同阶段对客户进行了基于生命周期的价值细分[4,5], 陈明亮等在前人研究的基础上进一步提出客户当前

① 收稿日期: 2007- 06- 26; 修订日期: 2008- 09- 01.
基金项目: 国家自然科学基金资助项目 (70802019); 985二期"技术、政策、管理" 国家哲学社会科学创新基地资助项目.
作者简介: 邹 鹏(1975—), 男, 湖北麻城人, 博士, 讲师. Email zoupeng@ hit edu cn

价值和客户增值潜力是客户细分的两个具体维度[6], 黄亦潇等在建立客户有形和无形价值评价体系的基础上结合客户生命周期理论从客户的生命周期阶段和客户发展潜力两个方面来评价客户价值, 根据客户价值评价结果进行客户分类[7]. 目前为止, 客户价值细分的基本观点得到了共识, 研究者都认为客户生命周期利润是客户价值衡量根本标准, 当前和潜在价值、财务利润和无形价值都可以作为客户价值细分体系中的细分变量, 视情况采用. 而当前客户价值细分面临的主要问题是如何在海量的客户数据中识别客户特征进而预测客户价值, 随着数据量加大, 不仅要考虑到潜在顾客数量增加, 而且每个客户的特征也越来越多, 采用传统的方法已经难以胜任. 就需要引入知识发现和数据挖掘技术, 探寻客户特征与客户价值之间的模式, 据此发动营销活动. 目前客户价值细分采用的数据挖掘方法多数是有监督的学习方法. 比较广泛使用的是关联规则[8,9]、决策树[10]、CHAID算法[11]、神经网络[9]、贝叶斯网络[12]. 基本都是采用一种普适性的数据挖掘算法, 没能体现由于客户价值不同对分类效果带来的影响. 下面本文将在分析客户价值细分问题特点的基础上, 探索解决这一问题的思路.

1.2　客户价值细分问题分析

在统计学上的假设检验存在两类错误, 第一类是当原假设为真时拒绝了它, 第二类是当原假设为假时接受了它. 将这一思想引入客户价值细分问题. 假定根据价值不同把客户分为两类, 一类定义为高价值客户, 另一类定义为低价值客户, 把一个低价值客户分类为高价值是第一类错误, 把高价值客户分为低价值是第二类错误, 一般情况下, 两类错误率存在制约关系, 即在其他条件不变时, 第一类错误的增加会导致第二类错误的减少, 反之第一类错误的减少会导致第二类错误的增加[13], 因此在样本容量一定的条件下, 不能同时减小发生这两类错误的概率.

在某些分类问题中两类错误的代价是一致的, 错误率最小就意味着分类器的性能最佳, 可以使用基于错分率最小的数据挖掘算法, 但在客户价值细分问题中, 两类错误具有不同代价. 很多研究表明, “20/80定律”在客户价值分布中十分显著, 通常 20%的客户创造的利润占据企业利润总额的 80%. 在银行业中, 15%的客户创造了其85%的利润[14], 中小银行甚至有可能是超过 90%的利润来自于 10%的客户[15]. 这说明高价值客户数量远远少于低价值客户, 但为企业创造的利润远远高于低价值客户. 企业要根据客户价值分类的预测结果对不同价值的客户分配资源, 保持和发展客户关系, 以期实现预测的价值. 犯第一类错误至多付出一些成本, 但如果犯第二类错误, 使企业未能充分有效地挖掘高端客户的潜在价值, 不仅会造成大量的收入损失, 还有可能导致高端客户的流失, 而高端客户的流失对企业形象和口碑传播造成的负面影响比普通客户更为严重. 所以对企业而言, 发生两类错误的代价有很大差异.

高价值客户在全部客户中占比例较低, 而普适性的数据挖掘方法都是基于错误率的, 在一个非平衡分布数据集上建立的模型只会对数量上占优的低价值客户有较强的识别能力, 如上文所述, 数据集中高价值客户的比例仅有 10%到 20%, 分类器在该数据集上训练的模型会向低价值客户偏差, 即在正确识别低价值客户的同时会将大部分高价值客户识别为低价值, 这样即使全部高价值客户都被识别为低价值, 模型的准确率也能达到 80%以上. 以上分析说明, 首先在高度不平衡的数据集合上使用基于错误率的分类算法建立的客户价值细分模型并不能有效识别高价值客户, 其次仅仅用准确率来评价客户价值细分模型也是不合理的.

可以从抽样方法解决数据的非平衡性, 例如将训练集中低价值与高价值客户的抽样比例设为 $k:1$(通常 $k=1$), 这样虽然平衡了训练集合的数据分布, 提高了分类器对高价值客户的识别能力, 但损失了很多低价值客户的信息, 而目前还没有 k值选取客观的标准, 需要多次试算.

当错分代价不相等时, 传统的基于准确率的算法不能很好适应客户细分问题, 代价敏感学习考虑了不同类型错分的代价, 并基于最小化总体误分代价的原理来设计分类器, 从而能更好的满足错分代价不同的情况. 代价敏感学习的研究可以追溯到 Granger对文本内容分类的研究[16], Breiman等建立了代价敏感学习的研究框架[17]. Elkan发展了 Granger的研究, 建立错分代价矩阵, 将待预测样本分到期望代价最小的类别

中[18]. Turney和 Kwedlo建立了基于分类准确率和错分代价的拟合函数的遗传算法[19,20], Zadrozny, Langford和 NaokiAbe通过对训练数据赋权建立了基于代价敏感性的决策树分类器[21], Ling提出了决策树代价最小化的剪枝策略[22]. Parag在神经网络中引入错分代价阈值, 将固定的错分代价改进为动态代价函数[23]. 代价敏感学习是数据挖掘领域的前沿[24]. 国内目前的研究还较少, 如郑恩辉研究了代价敏感支持向量机[25], 罗菲菲等研究了基于代价敏感学习的范例推理方法在入侵检测中的应用[26]. 在客户细分方面, 钱苏丽等和夏国恩等用代价敏感学习预测电信客户流失[27], 取得了一些有益的进展, 但只是根据经验和文献主观设定错误分类的代价, 不能根据客户数据的真实情况设定代价函数或取值, 不能精确控制和优化分类预测的效果, 该方法只适合二元的客户分类 (流失和不流失), 而客户价值的细分往往是对多个类别的划分, 所以也不适用于本研究的问题.

近年来支持向量机 (support vector machines, SVM)的研究在广泛开展. 支持向量机是V. Vipnik等根据统计学习理论 (statistical learning theory, SLT)提出的一种新的机器学习方法, 在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势, 有较好的泛化能力, 而且当训练数据中各类样本数量差别较大时, 可以通过参数设置调整偏置[28]. 支持向量机方法是针对二类别的分类而提出的, 而客户价值细分要将二类别分类方法扩展到多类别分类, 这是本研究的关键问题. 本文尝试在代价敏感学习机制下引入支持向量机, 建立以代价最小化为评价标准的客户价值分类方法.

2 基于代价敏感学习机制的 SVM 多分类模型

2.1 基本的 SVM

首先在基本的二元 SVM 分类模型中引入代价敏感学习参数, 然后再将其扩展到多分类的情况. 对二元分类问题, 假定已知观测样本集 $(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_n, y_n)$, $x_i \in R^l$, $y_i \in \{+1, -1\}$, $i = 1, \ldots, n$, (1) 能被超平面 $(w \cdot x) - b = 0$ 分类, 学习问题为最小化目标函数.

$$R(w, \xi) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{n}\xi\right) \qquad (1)$$

$$\text{s.t. } y_i(x_i \cdot w + b) \geqslant 1 - \xi,\ \xi \geqslant 0,\ i = 1, \ldots, n$$

目标函数中各参数意义和优化问题的求解见文献[29,30], 得到最优分类函数为

$$f(x) = sign\{\sum \alpha_i y_i k(x_i \cdot x) + b\} \qquad (2)$$

其中, $k(x_i \cdot x)$ 为核函数, 可以将低维输入空间中的线性不可分问题转化为高维属性空间中的线性可分问题.

2.2 基于客户价值的错分代价函数

式 (1) 是基于准确率的, 没有考虑错分代价, 本研究将基于客户价值的代价敏感性学习引入 SVM. 考虑每个样本都有不同的误分类代价, 样本集可重构为 $(x_i, y_i, cost_i)$ $x_i \in R^l$, $y_i \in \{+1, -1\}$, $cost_i \geqslant 0$ $i = 1, \ldots, n$.

其中 $cost_i$ 为第 i个样本 x_i 的误分类代价, 为正常数, 它依赖于 x_i或 y_i. 本研究假设企业能够从客户获得所识别的全部价值, 建立基于客户价值的代价敏感函数, 即

$$\text{cost}_i = \frac{E(V_{+1}) + C(V_{-1})}{E(V_{-1}) + C(V_{+1})} \qquad (3)$$

其中, $E(V_{+1})$ 和 $E(V_{-1})$ 分别为第一类和第二类客户价值的期望值, $C(V_{+1})$ 和 $C(V_{-1})$ 分别为企业在第一类和第二类客户投入的平均资源消耗, 均可根据企业对客户价值的定义和财务数据计算. (3) 反映了由于错误分类使企业付出的机会成本 (即客户的价值) 和财务成本总和的比较, 实现了错分代价敏感性的精确控制, 也体现了客户价值差异对分类结果的影响.

设样本集 $(x_i, y_i, cost_i)$ 能被超平面 $(w \cdot x) - b = 0$分类, 那么问题为最小化目标函数

$$R(w, \xi) = \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{n}\text{cost}_i\xi\right) \qquad (4)$$

$$\text{s.t } y_i(x_i \cdot w + b) \geqslant 1 - \xi,\quad \xi \geqslant 0,\ i = 1, \ldots, n$$

与式 (1) 不同, 函数 (2) 的经验代价考虑到不同样本的误差具有不同的误分类代价. 为求解优化问题 (2), 使用上述方法得到改进的对偶 Lagrange形式为

$$L_D = \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j (x_i x_j) - \sum_{i=1}^{n}\alpha_i$$

$$s.t \sum_{i=1}^{n} \alpha_i y_i = 0 \qquad 0 \leqslant \alpha_i \leqslant cost_i C,$$
$$i = 1, \cdots, n. \tag{5}$$

此时得到最优分类函数为

$$f'(x) = sign\{\sum_{i,j=1}^{n} \alpha_i y_i k(x_i \cdot x) + b\} \tag{6}$$

2.3　多元分类的 SVM

本研究将充分利用 SVM 二元分类特点, 将多类别分类的各个类别重新组合, 构成 $[\log_2^k]$ 个 SVM 子分类器[31]. 问题描述如下:

假定多类别分类有 K 个类别, $S = \{1, 2, \cdots, K\}$, 训练样本为 $\{(x_i, y_i),\ i = 1, 2, \cdots, l\}$, 其中 $y \in S$. SVM 进行客户价值的多类别划分. 如图 1, 将多类别 SVM 分类问题转化为一组二元 SVM 的分类器集合, 即 (x_i, y_i) 转化为 (x_i, y_i^k), 其中 $k = 1, 2, \cdots, K', K' = [\log_2^K]$

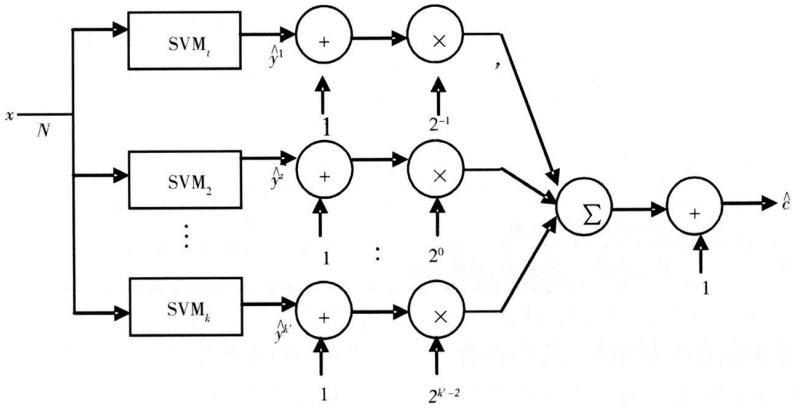

图 1　多分类 SVM 分类器

Fig. 1 multi-classification SVM

此时得到最优分类函数为

$$f(x) = sign\{\sum \alpha_i y_i^k k(x_i \cdot x) + b_k\} \tag{7}$$

分类赋值为 $\hat{c} = 1 + \sum_{k=1}^{K'} (y^k + 1) * 2^{k-2}$ (8)

其中, $y^k = f(x)$

以 4个类别 $\{1, 2, 3, 4\}$ 为例, 构造 $[\log_2^4] = 2$个 SVM 子分类器, 对于第一个 SVM 子分类器, 类别 2和 4所对应的样本数据全标记为正, 类别 1和 3所对应的样本数据全标记为负. 对于第 2个 SVM 子分类器, 类别 2, 3所对应的样本数据标记为正, 类别 1和 4所对应的样本数据标记为负. 由第 1节的方法分别求得这 2个 SVM 子分类器. 测试时, 根据这 2个子分类器的结果可得到测试样本的类别. 例如针对某一测试数据 x, 第 1、2个子分类器子分类器结果分别为 + 1, - 1, 则由第 1个子分类器可知 x 属于类别 2或类别 4, 由第 2个子分类器可知 x 属于类别 1或类别 4, 因此 x 属于类别 4

3　数据试验

3.1　数据描述

本研究以国内某银行信用卡客户数据进行实证研究. 采集了 2006年 1月 1日 — 2006年 12月 31日的 15 260个客户的数据, 银行根据客户的收入和成本, 计算出每位客户在 2006年度在信用卡业务为银行创造的利润作为客户价值在该年度的实际体现. 考虑银行的业务目标和客户价值的分布将客户按照价值从高到底分为 A、B、C 三类, 各类客户比例分别为 9. 7%、31. 6%、58. 7%. 按照 3∶1 的比例划分为训练集和测试集.

3.2　指标体系

根据数据的来源, 用于预测客户价值的数据可分为三类指标. 第一类是客户在申请信用卡时提供的个人基本信息, 第二类是根据客户账户和交易记录衍生出的汇总、比率和评分指标, 第三就是企业内部数据库所缺少的、可以从外部补充的指标数据, 如客户忠诚度数据[32~34].

客户价值与客户忠诚有紧密联系, 由客户忠诚产生的持续购买和增加购买会直接提高客户的现实赢利性, 客户因对企业的信任和归属感而向他人进行推荐也会带来客户的推荐价值, 企业在高度忠诚的客户身上只需花费较少的成本就会获得较高的收益. 客户忠诚度可以反映客户关系的深入和稳定程度, 有助于预测潜在的和未来的客

户赢利性. 这类数据的搜集过程如下.

本研究与该银行合作, 在该银行对客户进行的年度问卷调查中设置有关题项, 具体来说采用利科特 5级量表设计了钱包份额[35]、口碑推荐[36]和品牌认同[37]的题项来测量忠诚度: “使用* * 银行信用卡交易的金额占各种信用卡总交易金额的比例”, “您会推荐* * 银行信用卡吗?”, “您信赖* * 银行信用卡这个品牌吗”. 问卷通过电子邮件和电话访问系统进行调查和回收, 根据客户填选的答案得到该客户在各题项的得分, 这样获得样本客户在忠诚度方面的数据.

这样得到信用卡客户数据的字段见表 1

表 1　信用卡客户数据的字段

Table 1 Credit customer data fields

客户基本属性	性别、年龄、婚姻状况、学历、收入、职业、职位、工作年限、住房状况、居住城市
账户汇总信息 (每个周期)	持卡时间、信用额度、信用额度使用率、最低还款额、未清偿额度、现金提取额度、现金提取次数、消费次数、拖欠额、拖欠时间
模型评分	信用风险评分、破产评分、流失倾向评分等
营销数据	提前批准、利息率、优惠条款、参加促销活动记录
客户价值	2006年度利润贡献
客户忠诚度	钱包份额、口碑推荐、品牌认同

用于分析的数据包含大量属性, 其中含有一些与挖掘任务不相关, 是冗余的. 通过属性相关性分析, 过滤掉统计上不相关或弱相关的属性, 这样就筛选出对挖掘任务最相关的属性组成相关属性集. 本文将信息增益分析技术和基于多维数据分析的方法集成在一起删除信息量较少的属性, 收集信息量较多的属性[10, 38]. 首先计算出初始属性集中每个属性的信息增益度, 再用这一度量将属性排序. 接下来需要用户或领域专家参与进行. 首先根据专家的领域知识和经验确定哪些属性与挖掘任务不相关: 如在为客户利润贡献度的分类中, 客户的具体住址, 邮编这样的属性会被认为与利润贡献度关系不大或基本无关, 这样的属性就可以被排除出相关属性集, 可以被称为标志性可删除属性, 就是说这样的属性的信息增益度的值成为一个选择的标准, 即信息增益度的值比它小的属性也被删除.

可以保守地先将信息增益度的值最小的标志性可删除属性选取为初始阈值, 并将信息增益度的值比它小的属性都删除, 这样形成一个相关属性集, 接下来进行分类产生分类结果, 对其准确度进行检验.

再按照选出的标志性可删除属性的信息增益度的值从小到大的顺序定为测试阈值, 重复以上方法, 得出的几种分类结果比较它们的准确度, 选取最终阈值.

如果无法确定标志性可删除属性, 则可采用 John, G. H 等人提出的 wrapper模型, 该模型采用贪心算法和逐步反向删除 (backward stepwise elimination, BSE) 的策略, 即从包含所有属性的集合开始, 逐个删除属性直到再删除属性就会降低决策树的分类正确性为止. 而这个使准确性开始降低的属性的信息增益值为阈值, 信息增益值比这个阈值小的属性都被删除. 这样保留下来的属性就构成了相关属性集. 该方法详见参考文献 [10]. 最终, 利润贡献指标作为客户价值的分类变量, 即输出变量, 选择了 18个指标模型的相关属性集作为预测所需的输入变量, 如表 2 共 19个指标进入模型.

表 2　信用卡客户相关属性集

Table 2 Credit customer related attributes set

年龄、学历、收入、职业
持卡时间、信用额度、信用额度使用率、现金提取次数、消费次数、拖欠额
信用风险评分、流失倾向评分
利息率、优惠条款、参加促销活动记录
钱包份额、口碑推荐、品牌认同

3.3　模型构造

样本集合 (x_i, y_i), 其中 x 为客户样本, 是多维向量 (表 2的 18项指标), $y \in \{1, 2, 3\}$, 代表高、中、低价值三类客户. 对于核函数和参数选择, 本研究用 MATLAB6.5试验比较了多项式核函数 $K(x_i, x_j) = (\delta x_i^T x_j + r)^d$, sigmoid核函数 $K(x_i, x_j) = \tanh(\alpha x_i^T x_j + \beta)$ 和径向核函数 $K(x_i, x_j) = \exp\{-\lambda \times |x_i - x_j|^2\}$, 当 $\lambda = 0.01$ 时的径向核函数取得较好的效果. 对三个类别的分类, 用 (3) 计算其错分代价矩阵表 3, 关系为 $cost_{A,B}$ (A 错分为 B) = $1/cost_{B,A}$ (B错分为 A); $cost_{B,C}$ (B错分为 C) = $1/cost_{C,B}$ (C 错分为 B); $cost_{A,C}$ (A 错分为 C) = $1/cost_{C,A}$ (C错分为 A).

表 3　分类代价矩阵

Table 3 Cost of classification matrix

	分为 A类	分为 B类	分为 C类
A类客户	0	8.62	21.12
B类客户	0.12	0	2.45
C类客户	0.05	0.41	0

3.4　实验结果分析

用 VC编程实现算法, 进行试验, 因为 A类客户的平均价值远远高于 B和 C类客户, 为了突出显示本研究方法对数量少、价值高的 A类客户的分类性能, 着重比较 A类和 B+C类客户的分类效果, 而 B类和 C类客户数量和平均价值差异不大, 改进算法对这两类客户分类效果的改善不显著. 实验产生了如下的结果, 表 4为分类结果中各种样本的数量.

表 4　分类矩阵

Table 4 Classification matrix

	分为 A类	分为 B类	分为 C类
A类客户	a	b	e
B类客户	c	d	f
C类客户	g	h	l

1) 将 A类客户分为 B或 C类客户的错误比率 $F_{A,B\cup C} = b + e/a + b + e$;

2) 将 B或 C类客户分为 A类客户的错误比率 $F_{B\cup C,A} = c + g/c + d + f + g + h + l$;

3) 将 A类客户分为 A类客户的正确率 $T_{A,B\cup C} = a/a + b + e$;

4) 将 B或 C类客户分为 B或 C类客户的正确率 $T_{B\cup C,A} = d + l/c + d + f + g + h + l$;

5) 整体准确率 $T = a + d + l/a + b + c + d + e + f + g + h + l$

6) 期望损失函数定义为[39]:

$$C = b * cost_{A,B} + c * cost_{B,A} + e * cost_{A,C} + g * cost_{C,A} + f * cost_{B,C} + h * cost_{C-b} \quad (8)$$

7) 平均错分代价

$$AVE(C) = [(b+e)F_{A,B\cup C} * cost_{A,B\cup C} + (c+f+g+h)F_{B\cup C,A} \times cost_{B\cup C,A}] / (a + b + c + d + e + f + g + h + l) \quad (9)$$

图 1是算法试验结果, 可以看到随着训练样本数量的增加, 总体准确率在开始阶段有所提高, 随后保持平稳, 同时总体期望损失明显下降, 而 $F_{A,B\cup C}$ 在错分代价系数的控制下也能够显著降低, 由于 B、C类训练样本数量占明显优势, $F_{B\cup C,A}$ 也能控制在较低水平.

本研究又采用了基本 SVM, 代价敏感 SVM和设计样本比例 SVM 建模进行比较. 设计样本比例 SVM建模实验中当 $k = 3$ 时, 结果如表 5 基本 SVM在训练数据集直接建模, 虽然将 96% 的高价值客户分为中低价值, 但由于高价值客户数量很少, 所以总体准确率仍然较高, 这样不能识别绝大多数高价值客户, 对客户价值细分和高端客户管理是没有意义的. 设计样本比例 SVM 实验过程中, 当 $k = 3$ 时分类效果最好, 高价值客户正确识别率明显提高, 但平均代价下降也不显著. 代价敏感 SVM的实验不仅将 96% 的高价值客户正确识别, 同时能将其他类别客户的正确识别率也控制在较好的范围, 同时显著降低平均代价. 结果证明, 基于代价敏感学习机制下的 SVM比传统的 SVM 能在保证分类准确率的同时, 显著降低模型分类的代价.

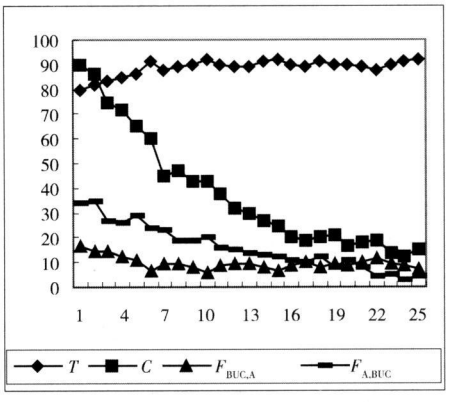

图 2　改进算法试验结果

Fig. 2 Result of improved method test

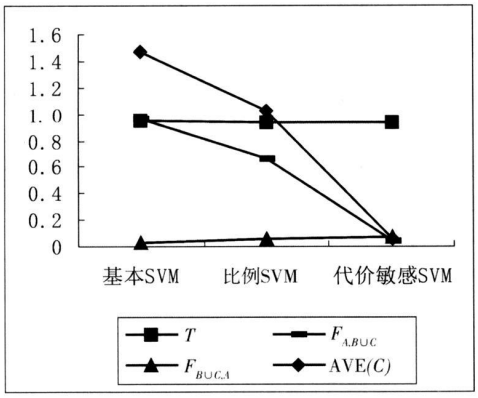

图 3　对比试验结果

Fig 3 Result of comparative test

表 5　对比试验结果

Table 5 Result of comparative test

	T	$F_{A,B\cup C}$	$F_{B\cup C,A}$	$AVE(C)$
基本 SVM	95.7%	96%	3%	1.465
设计样本比例 SVM	93.2%	67%	5.7%	1.021
代价敏感 SVM	93%	4%	7.6%	0.061

4　结论与展望

传统的基于数据挖掘的客户价值分类方法是普适性的学习方法, 本研究在支持向量机的基础上引入代价敏感学习机制, 一定程度上解决了客户价值细分问题中的不平衡数据分布和不同错分代价问题, 通过计算客户平均价值设定错分代价, 改善模型对高价值客户的识别能力, 提高了分类性能和实用价值.

现有基于支持向量机的客户流失预测的研究是对客户的二元分类 (流失与不流失), 不适用于根据客户价值的多元分类, 本研究引入了多分类支持向量机的算法将客户细分问题中的二元类扩展到多类别划分, 完善了支持向量机在客户细分领域的应用.

试验结果说明, 在基于代价敏感学习机制下的分类方法通过设定错分代价, 可以控制不同种分类错误的分布, 降低总体的错误分类代价, 但不一定能降低总体的错分率, 所以仅仅用总体准确率来评价分类模型的性能是不完全合适的, 应考虑实际应用背景, 引入错分代价或分类期望损失来评价模型的性能. 理想的客户价值细分模型尽可能要避免未能正确识别高价值客户造成客户价值挖掘不充分甚至高端客户流失, 同时也要考虑到把低价值客户误认为高价值客户所增加的客户保持和发展成本. 这一方法的思想对客户关系管理实践也有一定启示, 客户关系管理的核心任务不是要保持发展和所有客户的关系, 它应该在保有一定客户规模的基础上关注那些对企业价值最大的高端客户群体, 力求识别他们的特征和需求, 强调客户关系的经济性, 保证货币和其他有限的资源分配给高价值客户, 实现客户关系价值的最大化. 本方法体现和贯彻了客户价值细分的原则, 即价值取向、合理区隔, 差异化定位. 本方法的局限性在于, 仅仅把客户在某段时间的利润贡献作为客户价值, 这也是考虑企业的业务要求和数据的可获得性, 没能全面考察客户较长时期的历史价值和未来的潜在价值, 以及各种无形价值. 今后的研究应在全面评估上述客户价值变量的基础上完善代价函数, 使模型能更准确反映分类的代价, 精确识别各类客户, 更全面、准确和深入地进行客户价值细分.

参 考 文 献:

[1] Darrell Rigby, Frederick F Reichheld. Avoid the four perils of CRM [J]. Harvard Business Review, 2002, (2): 101—109.

[2] Dwyer F Robert. Customer lifetime valuation to support marketing decision making[J]. Journal of Direct Marketing, 1997, 11(Autumn): 6—13.

[3] Berger Paul D, Nada I Nasr. Customer lifetime value: Marketing models and applications[J]. Journal of Interactive marketing, 1998, 12(Winter): 17—30.

[4] Reichheld Frederick F, Earl W Sasser. Zero defections: Quality comes to services[J]. Harvard Business Review, 1990, (September—October): 105—111.

[5] Reichheld Frederick F. The loyalty effect: the relationship between loyalty and profits[J]. European Business Journal, 2000,

12(3): 173—179.

[6]陈明亮. 客户关系管理理论与软件[M]. 杭州: 浙江大学出版社, 2004.
Chen Ming-liang. The Theory and Software of CRM[M]. Hang Zhou: Publishing House of Hang Zhou University, 2004. (in Chinese)

[7]黄亦潇, 邵培基, 李菁菁. 基于客户价值的客户分类方法研究[J]. 预测, 2004, 8(3): 31—35.
Huang Yi-xiao, Shao Pei-ji, Li Jing-jing. The research about customer segmentation method based on customer value[J]. Forecasting, 2004, 8(3): 31—35. (in Chinese)

[8] Duen Ren Liu, Ya Yueh Shih. Integrating AHP and data mining for product recommendation based on customer lifetime value[J]. Information & Management, 2005, 42(1): 387—400.

[9] Nan Chen Hsieh. An integrated data mining and behavioral scoring model for analyzing bank customers[J]. Expert Systems with Applications, 2004, 27: 623—633.

[10]邹　鹏, 李一军, 叶　强. 客户利润贡献度的数据挖掘方法[J]. 管理科学学报, 2004, 7(1): 53—59.
Zou Peng, Li Yi-jun, Ye Qiang. Study on method of evaluating customer profitability based on data mining[J]. Journal of Management Sciences in China, 2004, 7(1): 53—59. (in Chinese)

[11] Jedid Jah Jonkera, Nanda Piersmab, Dirk Van den Poel. Joint optimization of customer segmentation and marketing policy to maximize long-term profitability[J]. Expert Systems with Applications, 2004, 27(5): 159—168.

[12] Bart Baesens a, Geert Verstraeten b, Dirk Van den Poel. Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers[J]. European Journal of Operational Research, 2004, 156(3): 508—523.

[13]蔡越江. 论假设检验中两类错误[J]. 数理统计与管理, 1999, 18(3): 30—35.
Cai Yue-jiang. Two type errors in hypothesis test[J]. Statistic and Management, 1999, 18(3): 30—35. (in Chinese)

[14] Michael Hales. Focusing on the 15% of the Pie[J]. Bank Marketing, 1995, 27(4): 30.

[15] Alan Bloomquist. Small business, big profits[J]. Bank Marketing, 1998, 30(8): 18.

[16] Granger C W J. Prediction with a generalized cost of error function[J]. Operational Research Quarterly, 1969, 20(2), 199—207.

[17] Leo Breiman, Friedman JH, Olshen R A, Stone C J. Classification and Regression Trees[M]. Belmont, California: Wadsworth International Group, 1984.

[18] Elkan C. The Foundations of Cost-sensitive Learning[C]. Proceedings of the 17 International Joint Conference on Artificial Intelligence, 2001.

[19] Turney P D. Cost-sensitive classification: Empirical evaluation of a hybrid genetic decision tree algorithm[J]. Journal of Artificial Intelligence Research, 1995, (2): 369—409.

[20] Kwedlo W, M Kretowski L, De Raedt P. Flach. An Evolutionary Algorithm for Cost-Sensitive Decision Rule Learning[C]. ECML 2001, LNAI 2167, Heidelberg: Springer, Berlin: 288—299.

[21] Zadrozny B, Langford J, Naoki Abe. Cost-Sensitive Learning by Cost-Proportionate Example Weighting[C]. Proceedings of the Third IEEE International Conference on Data Mining, 2003: 435.

[22] Ling C X, Q Yang, J Wang, S Zhang. Decision Trees with Minimal Costs[C]. Proceedings of the 21st International Conference on Machine Learning, 2004.

[23] Parag C. Pendharkar. A threshold varying bisection method for cost sensitive learning in neural networks[J]. Expert Systems with Applications, 2007(online).

[24] Elkan C. The Foundations of Cost-Sensitive Learning[C]. Proceedings of the 17 International Joint Conference on Artificial Intelligence, 2001.

[25]郑恩辉, 李　平, 宋执环. 代价敏感支持向量机[J]. 控制与决策, 2006, 9(4): 473—476.
Zheng En-hui, Li Ping, Song Zhi-huan. Cost sensitive support vector machines[J]. Control and Decision, 2006, 9(4): 473—476. (in Chinese)

[26]罗菲菲, 刘贵全, 安景琦, 张婷慧. 一种基于代价敏感学习的范例推理方法及其应用研究[J]. 计算机应用, 2005, 12(10): 2444—2446.
Luo Fei-fei, Liu Gui-quan, An Jing-qi, Zhang Ting-hui. Research on a case based reasoning method using cost sensitive

learner and its applications[J]. Computer Applications, 2005, 12(10): 2444—2446. (in Chinese)
[27]夏国恩, 金炜东. 客户流失预测中两类错误的平衡控制研究[J]. 营销科学学报, 2006, 1(12): 1—7.
Xia Guo-en, Jin Wei-dong. Tradeoff of errors of two types in customer churn prediction[J]. Journal of Marketing Science, 2006, 1(12): 1—7. (in Chinese)
[28]边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 2002
Bian Zhao-qi, Zhang Xue-gong. Patten Recognization[M]. Beijing: Publishing House of Tingsua University, 2002. (in Chinese)
[29] Clarke D W, Mohtadi C. Generalized predictive control[J]. Part I & part II. Automatic, 1987, 23(2): 137—160.
[30]胡耀华, 贾欣乐. 广义预测控制综述[J]. 信息与控制, 2000, 29(3): 248—256.
Hu Yao-hua, Jia Xin-le. Forecast and control review[J]. Information and Control, 2000, 29(3): 248—256. (in Chinese)
[31] Sebald D J, Buchlew J A. Support Vector Machines and the Multiple Hypothesis Test Problem[C]. IEEE Trans on Signal Processing, 2001, 49: 2865—2872.
[32]Wouter Buckinx, Geert Verstraeten, Dirk Van den Poel. Predicting customer loyalty using the internal transactional database[J]. Expert Systems with Applications, 2007, 1: 125—134.
[33] Chris Baumann, Suzan Burton, Greg Elliott. Determinants of customer loyalty and share of wallet in retail banking[J]. Journal of Financial Services Marketing, 2005, 9(3): 231—248.
[34] Kamakura, Wagner A, Michel Wedel, *et al*. Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction[J]. International Journal of Research in Marketing, 2003, 20(1): 45—65.
[35] Jones T O, Sasser. Determinants of customer loyalty and share of wallet in retail banking[J]. Journal of Financial Services Marketing, 1995, 9(3): 231—248.
[36] Zeithaml V A, Berry L L, Parasuraman A. The behavioral consequences of service quality: An examination of moderator effects in the four-stage loyalty model[J]. Journal of Marketing, 1996, 60(2): 31—46.
[37] Chaudhuri A, Holbrook M B. The chain of effects from brand trust and brand affect to brand performance: The role of brand loyalty[J]. Journal of Marketing, 2001, (4): 81—93.
[38] Jiawei Han, Micheline Kamber. Data Mining Concepts and Techniques[M]. San Mateo: Morgan Kaufmann Publishers Inc, 2001.
[39] Joos P, Vanhoof K, Ooghe H. Credit Classification: A Comparison of Logit Models and Decision[C]. European Conference On Machine Learning, Chemnitz, Germany, 1998.

Customer value segmentation based on cost-sensitive learning

ZOU Peng, LI Yi-jun, HAO Yuan-yuan
School of Management, Harbin Institute of Technology, Harbin 150001, China

Abstract Two type errors of "rejecting the true and accepting the false" are inevitable in customer value segmentation. The traditional data mining method which is based on the total accuracy rate can not reflect the influence caused by the great difference of misclassification costs and unbalanced quantity distribution of customers who have various values. The research proposes the cost-sensitive SVM classifier by presenting a misclassification cost function based on customer value, and expands binary-classification to multi-classification for the feature of customer segmentation, which is evaluated by the function of the exceptional lost. The data test result proves that the method can control the different types of error distribution with various cost of misclassification, reduce the total misclassification cost, and distinguish the customer value effectively.

Key words customer segmentation; customer value; cost-sensitive learning; support vector machines