

基于数据仓库技术的银行 ACRM 系统设计与实现

王萍萍¹, 王翰虎^{1,2}

(1. 贵州大学 计算机科学与信息学院, 贵州 贵阳 550025;
2. 贵州星辰科技开发有限公司, 贵州 贵阳 550001)

摘要:建立以客户为中心的服务体系,对于银行维持客户资源,争取新客户,挖掘潜力客户,增强市场竞争力有着重要意义。客户关系管理(ACRM)的引入为银行搭建了服务平台,帮助有效整合客户资源。文中提出了一种基于数据仓库、数据挖掘的银行客户关系管理系统解决方案,较好地构建了一个标准的数据仓库系统,详细介绍了该体系的架构与实现,并以客户关系管理为应用主题,确立了相应的数据表及技术指标。以数据仓库为基础,将数据挖掘应用于数据库中的知识发现。

关键词:客户关系管理系统;数据仓库;数据挖掘;支持向量机

中图分类号:TP311.13

文献标识码:A

文章编号:1673-629X(2013)03-0203-05

doi:10.3969/j.issn.1673-629X.2013.03.051

Design and Implementation of ACRM System for Bank Based on Data Warehouse Technology

WANG Ping-ping¹, WANG Han-hu^{1,2}

(1. College of Computer Science and Information, Guizhou University, Guiyang 550025, China;
2. Guizhou Xingchen Science and Technology Development Company Limited, Guiyang 550001, China)

Abstract: Establishing the service system takes the customer as the center has important significance for the bank to maintain customer resources, strive for new customers, explore potential customers and enhance market competitiveness. The introduction of ACRM is to build the service platform for bank, can integrate customer resource efficiently. It brings forward a solution for ACRM based on data warehouse and data mining technologies, build a standard data warehouse system better, discuss the architecture and implementation of the system in detail, establish related tables and technology parameters according to customer relation. On the basis of data warehouse, apply the data mining into knowledge discovery in database.

Key words: ACRM; data warehouse; data mining; support vector machine

0 引言

随着客户资源的重要性在不断增加,ACRM(客户关系管理系统)在银行领域的应用实施也广泛起来,ACRM的核心理念是一对一营销、彻底的个性化营销。对于国内大银行来说,要面对庞大的客户群实施一对一营销是不现实的,银行客户关系管理的核心思想是将银行的客户作为最重要的资源,选择和管理客户,挖掘客户最大的长期价值^[1,2]。

文中建立的 ACRM 系统是以 ODS 为基础平台,

通过建立统一的客户视图,运用一系列先进、完整的银行客户数据模型,对不同层次、不同类型客户进行细分,集客户分析、营销管理、考核管理等诸多分析和应用功能于一体的自动智能化的新型电子化运营支撑系统。

本系统旨在对银行的客户(特别是VIP客户)提供一对一的客户关系管理。通过与客户的不断接触和交流,发现并理解客户的真实需求,为不同的客户提供相对应的优质服务。系统应用对象为总行的个金和公金销售服务部、分行分管业务负责人、各支行网点的业务主管、理财师及客户经理等。系统的四大目标:

- (1)提升客户服务水平;
- (2)增加银行对客户的吸引力;
- (3)增强客户管理及客户经理业绩考核;
- (4)提高客户服务工作效率。

收稿日期:2011-12-26; **修回日期:**2012-04-02

基金项目:贵州省自然科学基金(贵阳市2010年工业科技攻关项目([2010]筑科工合同字第28号));贵州大学2011年研究生创新基金资助项目(校研理工[2011039])

作者简介:王萍萍(1987-),女,硕士研究生,研究方向为数据库技术与应用系统;王翰虎,教授,研究方向为数据库技术与应用系统。

1 ACRM 的架构与实现

ACRM 系统从逻辑整体架构上主要分为三个层次:应用层、数据集成层和外部数据层^[3-5]。系统数据库采用 IBM 的 DB2, ETL 工具采用 Datastage, 应用服务采用 Cognos8, 图 1 是 ACRM 的整体架构图。

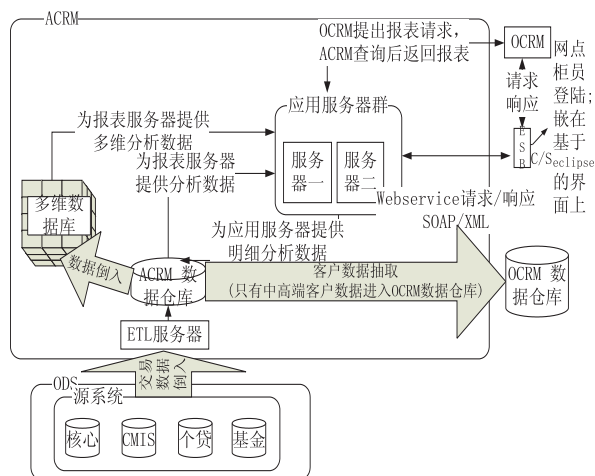


图 1 ACRM 的系统架构

应用层:主要提供数据查询和导出服务。系统根据用户制定的条件,通过固定报表和多维分析的方式展示对应的分析结果。同时系统提供 Webservice 服务,为外围系统查询 ACRM 系统的数据提供了方便的访问接口。

数据集成层:主要是通过标准的 ETL 过程,对源系统数据进行清洗和转换。并把数据加载至数据集市或多维数据库中,为后续的应用提供数据支持。另外系统中提供了多种不同的接口以获得外部数据源,包括:ODBC、JDBC、XML、TEXT FILE 等。

外部数据层:主要是指来源于源系统的数据,是系统中所有分析功能的数据来源基础^[6]。

1.1 外部数据层

外部数据层主要是指 ODS^[7], ODS 位于数据仓库之前,以各业务系统为数据源,经过 ETL 后的数据存放在 ODS 中。建设 ODS 系统的目的是建立全行客户唯一视图,通过整合全行各应用系统的有关客户信息,例如客户的名称、联系地址、联系方式等数据,使银行内的任何层次、任何业务系统通过任何渠道需要数据表时,看到的是一致信息。

在本系统中 ODS 一方面支持完成针对数据库记录级的 OLTP(在线事务处理)查询、报表、监控等,提供了客户经理近实时访问了解客户全面信息的窗口(以后会将整合后的信息发布给各个应用系统,从而完成各个应用系统间的客户信息同步,并向其他各个应用系统提供服务)。另一方面将整合好后的有关客户信息通过 ETL 抽到数据仓库中,来完成对客户贡献

度、忠诚度及客户细分等的各种分析应用。

ODS 会给下游系统提供三种类型的接口:固定接口:将源数据按业务模型整合成固定接口下发给应用,如资产负债管理系统、统计报表系统动态报表等;贴源接口:在保留源数据基本结构的基础上,增加时间戳或拉链控制字段,方便存储和查找。如客户关系管理系统、审计系统等;原貌接口:指将源数据保留原貌转发给下游应用,如历史数据存储系统。ODS 会以文本的形式供数给 Datastage,每天于 6 点前完成数据从 ODS 至数据集市的转换和加载。

1.2 数据集成层

数据集成层是银行 ACRM 系统的核心,指在从 ODS 进行数据抽取、转换和装载,并按照主题进行重新组织,最终确定数据仓库的物理存储结构。同时,它还对数据仓库的元数据进行组织与存储,具体包括数据仓库的数据字典、记录系统定义、数据转换规则、数据加载频率以及业务规则等信息^[8,9]。

该过程的处理流程主要通过两种方式进行:

一种通过 Datastage 从 ODS 抽取数据进行清洗、转换得到需要的数据^[4],该过程的处理流程是:首先,业务分析师会根据业务需求制定 Mapping 文档,然后由 ETL 工程师根据 Mapping 编写 Etljob,通过运行 Etljob,将生成的目标表落地到数据仓库中。Etljob 通过 Shell 脚本每天定时从 ODS 取数,完成日常的数据加载,加载的方式有两种,全量加载和增量加载。

另一种方式是通过存储过程来生成需要的表,系统每个月需要提供相应的月报表供 OCRM 查询,通过存储过程生成固定报表和多维分析报表需要的数据。

1.3 应用层

应用服务一方面主要根据业务需求制定相应的报表,为 OCRM 提供报表查询^[10],这里有两种形式的报表,固定报表和多维分析报表,固定报表是由 Framework 上传数据源到 Cognos8 服务器上,然后利用 Report Studio 和 Query Studio 生成供前端展现的固定报表,多维分析报表是由 Transformer 生成 Cube 然后上传到服务器上,利用 Analysis 生成多维分析报表。

另一方面应用层为最终用户提供 HTTP 服务,负责整个系统的前端展现和流量及权限控制等。同时提供报表工具服务和 JAVA 服务,实现与报表相关的数据库访问,实现查询、报表的解析和数据的生成;实现与数据库的访问操作及相应前端。

2 系统设计

2.1 主题确定

在银行 ACRM 系统中,数据仓库中的数据都是围绕着银行客户和产品来划分,其中以产品为主题进行

模型设计,划分为下面几个主题:存款、贷款、基金、理财、国债、借记卡、贷记卡、票据贴现、保函、国际业务,以客户为主题进行模型设计,划分为下面主题:客户基本信息,产品主题与客户主题通过客户号关联^[11,12]。

通过对客户进行群体划分、资产分析、行为分析等把客户划分为高端客户、关注客户、潜在客户。一方面对其进行分析和跟踪,更好地了解客户的真切需求,制定更适合的营销策略,另一方面通过数据挖掘和多维分析等方法,建立客户行为预测模型,提升营销效果。实现了客户信息服务、客户分析、营销活动/考核三个应用。

2.2 数据表组织

“主题”在数据仓库中是由一系列表来实现的。数据表的定义可以是由于对数据的综合程度不同,也可以是由于数据所属的时间段不同而进行的。本系统采用星形模型组织数据表。围绕客户的事实包括:个人客户基本信息表、个人客户资产信息表、客户系统评级历史表等,图2是围绕客户基本信息表的关系图。

在事实数据的周围,又从多个角度多个层次进行汇总,形成维表。如银行组织结构营业网点维、业务类别维、交易方式维、客户类型维、时间维。汇总的层次就是粒度,它是维划分的单位,如营业网点维(总行、分行、支行)、业务类别维(活期、整存整取、大额定期、存本取息、定活两便等)、交易方式维(柜台、ATM、POS、网银等)、时间维(年、季、月、日)。

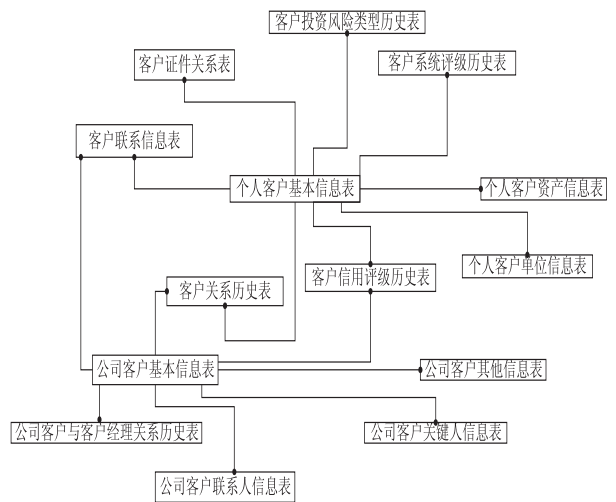


图2 围绕客户基本信息表的关系图

3 系统应用

3.1 客户关系管理

银行系统中,涉及到的技术指标很多,这也是银行数据仓库复杂的重要因素。客户关系管理有很多技术指标,如客户贡献度、客户可信度趋势、客户流失性走向、利率敏感度、组合风险等^[5]。

下面是对客户贡献度的计算:

客户贡献度细分就是充分考虑客户选用不同产品对于银行的贡献,计算一段时间内客户为银行创造的价值,找出真正对银行贡献大的那部分客户。要计算客户的贡献度,首先要统计客户不同产品的持有情况,根据不同产品的贡献度计算客户总的贡献度。

3.1.1 存款贡献度计算

存款产品贡献=内部资金计价收入-(存款利息支出+渠道交易成本),其中:

内部资金计价收入=存款日均余额×(1-准备金率-备付金率)×存款资金中间转移价格×实际天数+存款日均余额×准备金率×准备金日利率×实际天数+存款日均余额×备付金率×备付金日利率×实际天数

交易成本=Σ(交易渠道笔数×单笔交易成本)。交易渠道包括ATM、POS、柜面、电话银行、网上银行等。

不同的储蓄产品及储蓄期限对应不同的中间转移价格,产品的提前支取、到期未支取、正常支取等不同条件存管产品贡献算法也不尽相同。

3.1.2 贷款贡献度计算

贷款产品贡献=贷款利息收入+贷款保险手续费收入-[贷款税费支出+内部资金计价支出+贷款(预期)损失+渠道交易成本],其中:

内部资金计价支出=贷款日均余额×贷款资金中间转移价格×实际天数

贷款利息收入:根据当月贷款借据明细性质判断贷款利息收入。若贷款处于正常状态,计算正常本金利息发生额;若贷款处于违约状态,计算表内欠息发生额、表外欠息发生额、逾期本金利息发生额、表内欠息利息发生额、表外欠息利息发生额。

交易成本=Σ(交易渠道笔数×单笔交易成本)。交易渠道包括ATM、POS、柜面、电话银行、网上银行等。

不同性质、不同期限贷款的利息收入不一,同时内部资金计价也不同。

3.2 数据挖掘

在本系统中进行数据挖掘主要是通过建立预测模型,通过对预测模型的分析来挖掘出知识,图3是模型预测窗口视图^[6]。预测模型是定量管理建设的核心部分。利用统计方法开发预测模型,即从观测到的数据中找出并固化成业务规律: y 和 x 之间的数学方程,包括函数形式 (f)、预测变量 (x):

$$y = f(x)$$
(1)

应用这些函数可以对每一个客户的将来行为评分。利用这些分数,可以对客户进行高级细分,实行最高层次的差别化管理。

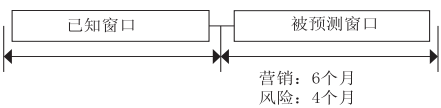


图 3 预测模型窗口视图

3.2.1 模型设计

模型在设计上满足统计要求：

- (1) 观测到的数据量必须足够大,其中包括:客户数、时间长度、变量数。
- (2) 具有业务意义:应用预测模型时,必须产生足够大的经济效益。
- (3) 可操作、可持续、可执行:建设好的预测模型需要置入运营系统,产生运营中应用的智能化预警功能。

系统采用线性 ε -支持向量回归机发展预测模型^[13,14]。该回归技术将最有预测功能的变量(一般十几个)结合在一起,形成如下方程:

$$\min_{\omega, b, \xi^*} 1/2 \parallel \omega \parallel^2 + c \sum_{i=1}^l (\xi_i + \xi_i^*) \tag{2}$$

$$st \ ((\omega \cdot x_i) + b) - y_i \leq \varepsilon + \xi_i, i = 1, \cdots, l \tag{3}$$

$$y_i - ((\omega \cdot x_i) + b) \leq \varepsilon + \xi_i^*, i = 1, \cdots, l \tag{4}$$

$$\xi^{(*)} \geq 0, i = 1, \cdots, l \tag{5}$$

引入松弛变量 $\xi^{(*)} = (\xi_1, \xi_1^*, \cdots, \xi_l, \xi_l^*)^T$ 和惩罚系数 $c = 1$ 求得解 $(\bar{\omega}, \bar{b}, \bar{\xi}^{(*)})$ 后,可得到模型分数 y :

$$y = g(x) = (\bar{\omega} \cdot x) + \bar{b} \tag{6}$$

使用线性 ε -支持向量回归机的原因是此技术最成熟,最通用,安全,可靠;直观,易于理解,符合惯例;有效提高预测模型的稳健性(Robustness),压制数据中的误差,这一点特别重要,因为观测到的数据包含大量的误差和不可控、不可解释因素。

3.2.2 模型的应用

选取客户年龄、客户性别、客户开户时长(月)、最近半年定期存款占总存款比例、最近半年储蓄存款余

额上升趋势等十个指标作为预测变量,求得每个客户在模型中的模型分数^[9]。

1) 利用存款余额提升预测模型预测效果。

利用模型提升分数对 2009 年 6 月半年日均存款 20 万以上的客户排序和分层后,表 1 是存款余额提升预测模型预测效果统计表。

- (1) 提升可能性最高的 3% 客户:其中 7.7% 的客户余额提升 100% 以上,是全部客户平均值(1.0%)的 7.7 倍;
- (2) 提升 80%、60% 和 50% 以上的客户占比也是最高;
- (3) 其它各单元:余额提升 100%、80%、60% 和 50% 以上的客户占比按顺序下降。

2) 利用存款余额提升预测模型验证效果。

利用提升分数对 2009 年 8 月半年日均存款 20 万以上的客户排序和分层后,表 2 是存款余额提升预测模型验证结果统计表。

- (1) 提升可能性最高的 3% 客户:其中 9.1% 的客户余额提升 100% 以上,是全部客户平均值(1.1%)的 8.2 倍;
- (2) 提升 80%、60% 和 50% 以上的客户占比也是最高;
- (3) 其它各单元:余额提升 100%、80%、60% 和 50% 以上的客户占比按顺序下降。

3) 存款余额提升预测模型指标分析。

- (1) 模型稳定性:稳定性系数(PSI)为 0.0266,模型比较稳定;
- (2) 模型预测变量稳定性:稳定性系数(PSI)均小于 0.05,变量比较稳定;
- (3) 模型预测变量共线性:条件指数(Condition Index)为 11.5,不存在严重的共线性;
- (4) 模型预测变量两两相关性:相关系数均小于 0.38,变量间不存在显著的相关性。

表 1 存款余额提升预测模型预测效果统计表

按模型分数细分			总客户数	余额提升客户占比			
等级分类	最低分	最高分		存款提升 100% 以上	存款提升 80% 以上	存款提升 60% 以上	存款提升 50% 以上
0% ~ 3%	976	1,000	1,822	7.7%	11.7%	16.1%	19.3%
4% ~ 10%	909	976	4,249	2.9%	4.1%	6.7%	8.5%
11% ~ 20%	792	909	6,070	1.2%	1.8%	3.2%	4.5%
21% ~ 40%	577	792	12,141	0.8%	1.3%	2.0%	2.7%
41% ~ 100%	0	577	36,422	0.4%	0.6%	1.1%	1.5%
全体	0	1,000	60,704	1.0%	1.5%	2.3%	3.1%

表2 存款余额提升预测模型验证效果统计表

按模型分数细分			总客户数	余额提升客户占比			
等级分类	最低分	最高分		存款提升 100%以上	存款提升 80%以上	存款提升 60%以上	存款提升 50%以上
0% ~ 3%	976	1,000	1,915	9.1%	12.1%	16.3%	20.2%
4% ~ 10%	909	976	4,469	2.6%	4.1%	6.9%	8.7%
11% ~ 20%	792	909	6,383	1.4%	2.0%	3.4%	4.4%
21% ~ 40%	577	792	12,767	0.9%	1.3%	2.0%	2.6%
41% ~ 100%	0	577	38,299	0.6%	0.8%	1.2%	1.6%
全体	0	1,000	63,833	1.1%	1.6%	2.4%	3.1%

4 结束语

本系统较之传统的数据仓库系统加入了 ODS 作为独立的数据库为 ACRM 系统提供数据,做到了系统的负载均衡,系统基于数据仓库的架构,并在此基础上建立挖掘模型,挖掘出知识,这些知识可以为所有的应用所用。客户关系管理系统在我国金融界应用越来越广泛,通过建立银行数据仓库,进行数据分析和挖掘,营销和管理部门可以对已有的资源进行分析,优化客户关系,为银行在这个以人文本的激烈竞争环境中抢占先机。

参考文献:

[1] 冯健文,陈启买,林 璇. 基于 CIF 的银行分析性 CRM 系统的研究与设计[J]. 计算机工程与设计,2006,27(12): 2288-2290.

[2] 周晶平. 基于数据仓库的银行客户关系管理系统设计[J]. 中南民族大学学报,2006,25(12):64-66.

[3] 郑 华. 基于数据挖掘银行客户关系管理系统构建研究[J]. 广西轻工业,2008,24(10):82-83.

[4] 路 川,阎文丽,胡欣杰. 基于数据仓库的航空客户关系管理研究[J]. 微计算机应用,2010,31(4):49-52.

[5] Xie Yaya,Li Xiu,Ying Weiyun. A Process Driven Architecture of Analytical CRM Systems with Implementation in Bank Industry[C]//ISECS International Colloquium on Computing,

Communication,Control and Management. [s.l.]:[s.n.], 2008:57-61.

[6] William H I. Building the Data Warehouse[M]. 4th ed. New York:A Wiley - Qed Publication, John Wiley & Sons Inc, 2005:166-437.

[7] 冯健文,林 璇. 基于 ODS 的数据仓库模型研究[J]. 微计算机应用,2006,27(4):468-470.

[8] Tu Yan,Yang Zijiang,Benslimane Y. Towards an optimal classification model against imbalanced data for customer relationship management[C]//2011 Seventh International Conference on Natural Computation. [s.l.]:[s.n.],2011:2401-2045.

[9] Crone S F, Lessmann S,Stahlbock R. Empirical comparison and evaluation of classifier performance for data mining in customer relationship management neural networks[C]//Proceedings of 2004 IEEE International Joint Conference. [s.l.]:[s.n.],2004.

[10] Alur N,Takahashi C. IBM Infosphere Datastage Data Flow and Job Design[M]//IBM Redbooks. [s.l.]:[s.n.],2008.

[11] 马超群,兰秋军,陈为民. 金融数据挖掘[M]. 北京:科学出版社,2007:150-162.

[12] Han Jiawei,Kamber M. Data Mining:Concepts and Techniques [M]. Singapore:Elsevier(Singapore)Pte Ltd. ,2006.

[13] 王观玉,郭 勇. 支持向量机在电信客户流失预测中的应用研究[J]. 计算机仿真,2011,28(4):115-117.

[14] 邓乃扬,田英杰. 支持向量机-理论、算法与扩展[M]. 北京:科学出版社,2009:77-80.

(上接第 202 页)

source[C]//Waveform Diversity and 17 ~ sign Conference. [s.l.]:[s.n.],2007:204-208.

[7] Skonnard A,Gudgin M. Essential XML Quick Reference;a Programmer's Referee to XML,XSLT,XML Schema,SOA,and more[M]. US:Pearson Education,Inc,2002:304-307.

[8] 杨 悦,许 琪. 浅谈企业办公自动化网络安全[J]. 信息技术,2009(22):251-253.

[9] 赵立军. 基于 SysML 的需求分析研究[J]. 计算机技术与发

展,2011,21(12):139-141.

[10] 唐文忠,李 剑. 基于产品线需求分析的过程改进研究[J]. 计算机技术与发展,2008,18(7):104-108.

[11] 张国生. 基于层次着色 Petri 网的功能需求模型[J]. 计算机技术与发展,2011,21(11):81-83.

[12] 王 莉,吴洁明. 软件项目中的需求变更管理的研究[J]. 计算机技术与发展,2007,17(1):119-122.