

基于模糊聚类分析的客户分类算法研究*

郭蕴华^{1, 2}, 陈定方^{1, 3}

(1. 武汉理工大学 智能制造与控制研究所, 湖北 武汉 430063; 2. 西北机电工程研究所, 陕西 咸阳 712099; 3. 中国科学院 计算技术研究所 智能信息处理实验室, 北京 100080)

摘要: 用可定制的评估指标树来描述客户分类的评价体系, 让用户自主确定每个评价指标。在此基础上, 提出了一种基于模糊聚类分析的客户分类算法, 对客户进行分类管理, 并给出了一个计算实例, 取得了正确的计算结果。该计算实例表明, 这个算法可以用于关于客户关系管理的决策支持系统中。

关键词: 模糊聚类分析; 客户分类管理; 评价指标体系; CRM

中图法分类号: TP301.6 文献标识码: A 文章编号: 1001-3695(2005)04-0052-02

Research on Algorithm of Customer-classified Based on Fuzzy Clustering Analysis

GUO Yun-hua^{1, 2}, CHEN Ding-fang^{1, 3}

(1. Institute of Intelligent Manufacturing & Control, Wuhan University of Technology, Wuhan Hubei 430063, China; 2. Northwest Institute of Mechanical & Electrical Engineering, Xiayang Shanxi 712099, China; 3. Key Laboratory of IIP, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China)

Abstract: The customer-classified evaluation system is described by the customization-supporting tree of evaluation indexes, and user can determine any evaluation index independently. Based on which, an algorithm of customer-classified based on fuzzy clustering analysis is put forward to implement customer-classified management. A calculation example is presented, and the correct result is gained. The calculation example indicates that the algorithm can be used in the decision support system of CRM.

Key words: Fuzzy Clustering Analysis; Customer-classified Management; Evaluation Index System; CRM

科技的进步改变了人们的生活方式, 而人们对高品质生活的追求也进一步刺激了科技的进步。如今, 生产的目的已经不是为了满足单一的消费模式, 而是为了满足客户的个性化、多样化的消费需求^[1]。毋庸置疑, 这些变化给企业带来了新的机遇和挑战, 同时也给企业带来了前所未有的压力。尽管个性化的消费需求迫使企业尽最大的努力满足所有客户的各种要求, 但实际上没有任何一家企业能够做到这一点。也就是说, 企业为了生存和发展, 实际上只能优先满足能够给予自己带来重大盈利的客户群体的消费需求。研究表明, 拓展一个新客户的成本, 是保持一个老客户的成本的五倍^[2]。由此可见, 对客户群体进行有效的分类管理, 选择正确的客户策略, 实现以客户为中心的业务流程, 对于企业的发展具有重大的战略意义。

1 影响客户分类管理的评价指标

在实际的客户管理中, 不同的企业对客户分类的评价体系是不同的。为了适用于不同类型的企业, 本文用可定制的评估指标树来描述客户分类的评价体系, 如图 1 所示。

图 1 中所有的评估指标都来自于实际的业务流程, 可以由用户在数据字典中动态地指定。此外, 所有的评估指标都与时

间相关, 所以用户在定制评估指标树时, 同时应该指定当前评估指标树的计算时段。当然, 评估指标的确定和计算时段的选择, 本身就是值得深入研究的课题。本文为了简化模型, 将这两个任务留给客户完成, 同时也为系统留有可扩展的空间。

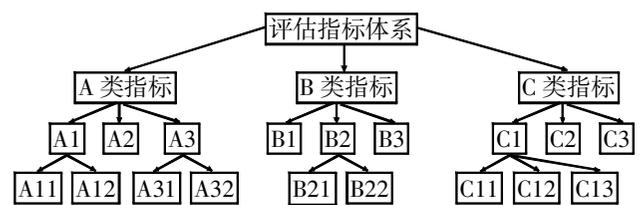


图 1 可定制的客户分类评估指标树

此外, 还应该考虑的是, 不同的指标, 在整个评估指标树中的权重应该是不同的, 但同级的指标的权重比之和应该为 1。例如, A 类指标的权重为 0.35, B 类指标的权重为 0.38, 则 C 类指标的权重一定为 0.27。这样, 根据以上的描述, 可以对评估指标树进行如表 1 所示的数据库设计。

表 1 客户分类评估指标树的数据库设计

字段名称	字段含义	字段类型	是否主码
EvaluateNo	评估指标编号	Varchar (20)	是
EvaluateParentNo	评估指标父节点编号	Varchar (20)	否
EvaluateName	评估指标名称	Varchar (20)	否
EvaluateValue	评估指标值	Float	否
EvaluateRatio	评估指标在父节点中的权重	Float	否

以上各字段中, 除了 EvaluateValue 由业务流程计算或者由专家打分获得, 其余均为建立指标树时由客户指定。

收稿日期: 2004-05-15

基金项目: 国家“863”计划资助项目(2003AA414011); 武汉市重点科技攻关项目(20023005133-04)

2 客户分类算法的计算流程

(1) 将用户指定的计算时段继续细分, 划分为 N 个小计算时段。 N 由客户指定, 同时由客户指定不同小计算时段的影响因子 $C_k, k=1, 2, \dots, N$ 。 计算程序中, 缺省时 $N=5$, 缺省的时间影响因子呈线性分布, 取 $C_k = k/N$, 即 $C = (0.2, 0.4, 0.6, 0.8, 1)$ 。 考虑影响因子是表示距离当前时间最近的指标更为重要。 当然, 如果根据实际的需要, 用户也可以任意指定影响因子的大小。

(2) 形成评估指标矩阵。 设客户集合为 $U, U = \{u_i, i=1, \dots, N_u\}$ (N_u 为客户的个数), 分时段计算每个客户的叶节点指标值, 并求得总计算时段内各客户的叶节点指标综合值 $p, p = \prod_{k=1}^N p_k \times C_k$ 。 然后, 将所有级别的叶节点指标按从左到右的顺序重新编号, 如 p_{A11} 编为 p_1, p_{A12} 编为 p_2, p_{A2} 编为 p_3 , 形成集合 $P = \{p_j, j=1, \dots, N_p\}$, 其中 N_p 为叶节点的个数。 同时, 考虑客户维, 即形成评估指标矩阵 $P = \{P_{ij}, i=1, \dots, N_u, j=1, \dots, N_p\}$ 。 注意, 此时的评估指标矩阵只考虑了时间影响因子, 未考虑各指标的权重。

(3) 求取权重集合 V 。 权重集合 $V = \{v_j, j=1, \dots, N_p\}$, 其中, v_j 的计算方法为每个叶节点指标在不同级别上权重的累乘。 例如, 对于图 1 所描述的指标树, $v_1 = v_{A11} \times v_{A1} \times v_A$ 。 同时, 根据乘法分配率不难证明 $\prod_{j=1}^{N_p} v_j = 1$ 。

(4) 评估指标矩阵的归一化处理。 将各因素的专业评价指标进行归一化处理, 将其转换为 0 ~ 1 区间内的无量纲值, 得到无量纲评估指标矩阵 X , 其中,

$$x_{ij} = P_{ij} / \sum_{i=1}^{N_u} P_{ij} \cdot v_i \quad (1)$$

x_{ij} 的意义在于, 如果将 $\sum_{i=1}^{N_u} P_{ij}$ 视为所有客户在 p_j 指标上的总贡献值, 则 x_{ij} 为各客户在 p_j 指标上考虑了权重的贡献率。

(5) 求取模糊相似矩阵 R 。 采用欧氏距离法求相似矩阵 R , 令 $r_{ij} = 1 - C \cdot \sqrt{\sum_{k=1}^{N_p} (x_{ik} - x_{jk})^2}$ 。 注意, 式 (2) 中的下标 k 在式 (1) 中为下标 j 。 此外, C 的选取原则为保证 $0 < r_{ij} < 1$ 。

(6) 聚类。 用平方法求传递闭包 $t(R)$, 得模糊等价矩阵 $t(R) = R^{2^n} = R^{* [3]}$, 然后选取合适的 λ , 将客户划分为黄金客户、白银客户、具备发展潜力客户和待放弃客户。 由于对于不同的 $\lambda \in [0, 1]$, 可以形成动态聚类图, 所以, 有可能无法完全划分为四类客户。

3 实例分析

从某些制造企业的实际出发, 并参考了文献 [4], 建立了如下的客户分类评价指标集, 如表 2 所示 (用户也可以根据自己的需求, 建立与表 2 完全不同的评价指标集)。 在表 2 中, 赢利指标 A 表现客户累计交易的获利情况, 综合体现了产品的价格、国家退税、采购成本和其他成本等因素; 忠诚度指标 B 反映客户与公司关系密切程度, 由累计交易额 B1、重复交易次数 B2 和客户保持时间 B3 组成; 诚信度指标 C 反映客户的诚信情况, 由付款拖欠天数 C1 和付款拖欠次数 C2 组成; 联络次数反映了客户与企业的交往频繁程度, 可以通过累计呼叫中

心、电子传真和电子邮件中联络次数求得; 客户实力和业务难度都是由专家打分评估的。

表 2 实例中的客户分类评价指标集

指标编号、名称及权重			求取方法
A: 赢利指标	0.45		赢利额 = 销售价格 + 国家退税 - 采购成本 - 其他成本
B: 忠诚度	0.20	B1: 累计交易额 M	$M = \sum_{i=1}^n Q_i$ 为第 i 次交易额
		B2: 重复交易次数	简单累计
		B3: 客户保持时间	简单累计
C: 诚信度	0.15	C1: 付款拖欠天数	简单累计
		C2: 付款拖欠次数	简单累计
D: 联络次数	0.05		可以通过累计呼叫中心、电子传真和电子邮件中联络次数求得
E: 客户实力	0.05		由专家评估, 百分制
F: 业务难度	0.10		由专家评估, 百分制, 业务难度越小则得分越高

设用户将对一年以来的所有九个客户进行分类, 取 $N=4$, 即以季度为小计算时段; 取时间影响因子为线性分布, 即 $C = [0.25, 0.50, 0.75, 1.00]$ 。 这样, 需要从数据仓库中以季度为时间维的单位长度读取以上各指标的统计值 (额), 如表 3 所示。

表 3 某制造企业的客户综合评估指标 (按季度划分)

指标 客户	A (元)	B1 (元)	B2 (次)	B3 (天)	C1 (天)	C2 (次)	D (次)	E	F
X ₁	100 000	200 000	3	90	0	0	28	97	51
	270 000	350 000	4	90	0	0	15	97	62
	300 000	400 000	6	90	0	0	23	98	71
	280 000	560 000	3	90	0	0	26	99	83
X ₂	12 000	20 080	3	90	0	0	4	99	80
	70 000	120 000	2	90	0	0	8	99	80
	50 785	80 000	5	90	2	1	6	99	80
	0	0	0	0	0	0	2	99	0
X ₃	150 000	230 000	2	90	4	1	11	93	98
	230 000	300 000	5	90	13	2	3	88	98
	390 000	560 000	1	90	2	1	5	89	71
	500 000	760 000	2	90	3	2	3	91	92
X ₄	10 000	19 999	3	90	0	0	3	80	59
	270 000	450 000	4	90	0	0	6	85	75
	45 000	70 000	6	90	0	0	2	85	72
	50 000	100 000	3	90	0	0	1	86	86
X ₅	355 555	510 000	4	90	0	0	5	77	55
	40 000	75 000	5	90	0	0	13	77	56
	257 892	400 000	3	90	1	1	17	78	67
	453 452	700 000	2	90	0	0	21	78	81
X ₆	157 811	600 000	1	90	1	1	23	77	15
	10 000	100 000	2	90	0	0	15	77	12
	1 035	150 000	3	90	3	3	22	77	15
	9 000	160 000	4	90	0	0	19	77	12
X ₇	78 056	90 000	5	90	5	1	26	91	99
	50 034	65 000	3	90	0	0	17	94	99
	124 562	140 000	2	90	0	0	25	96	99
	70 000	75 000	2	90	0	0	19	92	99
X ₈	20 567	80 000	3	90	15	3	2	95	43
	21 564	80 000	4	90	0	0	5	95	44
	23 000	80 000	6	90	0	0	7	96	47
	25 892	80 000	3	90	2	1	12	96	49
X ₉	50 000	100 000	4	90	0	0	4	73	91
	51 291	100 000	2	90	0	0	5	76	92
	53 236	100 000	2	90	0	0	7	75	93
	57 987	100 000	2	90	0	0	2	75	95

根据表 3, 考虑时间影响因子, 求取评估指标矩阵 P , 并求得指标权重集合, $V = [0.45, 0.1, 0.05, 0.05, 0.075, 0.075, 0.05, 0.05, 0.1]$ 。 根据式 (1), 可以求得无量纲评估指标矩阵 X 。 (下转第 57 页)

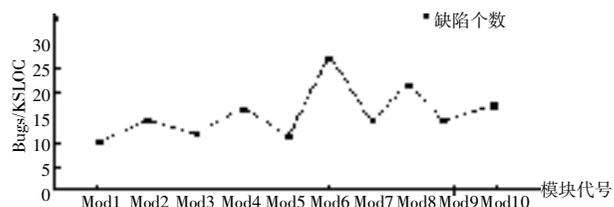


图 3 代码走查阶段的缺陷稠密度

通过跟踪设计和代码走查阶段的缺陷数,我们可以尽可能地获得有关产品质量的信息,以控制图的形式监控缺陷排除过程中项目的稳定性。当超出预定的阈值范围时,我们必须及时展开调查,并且提出可行的改进方案。总之,项目进展是否正常,主要看该阶段开发人员的效率、测试人员的效率与历史数据的偏差是否在预期的阈值区间内。同样如果缺陷的加权走势趋向开放,而项目已经接近交付期限,说明项目开发的某些环节存在严重问题,需要及时调整相关的人员或者检查前期的设计和开发是否存在较大的缺陷。

1.4 度量过程的体会

纵观整个缺陷度量过程,我们可以看到在度量实施过程中,度量的构造和对度量结果的分析占有重要的地位。因此在度量的前期设计时切勿求快求全,成功的度量开始的时候只是针对一些关键的度量源进行设计,随着度量的深入逐渐加入其他的元素。同时实施度量的人员必须对度量的过程和自己所担负的角色有充分的理解,在度量的过程中充分和其他组或角色进行交流,因为不同的人对项目进展的评估分析可能不同,我们在量化的同时需要考虑到与某一现象相关的各种主客观因素,这也就是在后期的分析度量结果时需要注意的问题。

2 结束语

本文中我们从项目管理者的角度出发,通过实际的度量实施过程,将软件度量方法引入缺陷管理过程中,为项目管理者考查相关人员的效率和项目的进展情况提供辅助信息,对传统的缺陷管理模式进行改进。当然后期的数据分析部分还有待进一步挖掘,我们还可以提出更多与项目有关的问题,并找到它们的答案,这也是以后工作的重点。

参考文献:

- [1] umphrey, Watts S. Managing the Software Process [M] . Pittsburgh, PA: Addison-Wesley Publishing Company, SEI Series in Software Engineering, 1989. 50-51
- [2] James A Rozum. Defining and Understanding Software Measurement Data[R] . Pittsburgh: Software Engineering Institute Carnegie Mellon University, 2002. 2-3
- [3] Norman E Fenton, Martin Neil. Software Metrics: Roadmap [R] . Computer Science Department, Queen Mary and Westfield College, 2000. 3-5.
- [4] E Obara, T Kawasaki, Y Ookawa, et al. Metrics and Analyses in the Test Phase of Large-scale Software [R] . 1997. 4-4.
- [5] Robert B Grady, et al. Practical Software Metrics for Project Management and Process Improvement [M] . Prentice Hall PTR, 1992. 192-220.
- [6] Boris Beizer. Software Testing Techniques [M] . New York, NY, USA: John Wiley & Sons, Inc., 1990. 156-160.
- [7] John McGarry, Dabvid Card, Cheryl Jones, et al. Practical Software Measurement [M] . Boston: Addison-Wesley Publishing Company, 2001. 45-60.

作者简介:

张瑞,男,西安人,硕士研究生,主要研究领域为软件工程与理论;郝克刚,教授,博士生导师,主要研究领域为软件工程与相关理论。

(上接第 53 页)

0.0027 0.0012 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
 0.0012 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

根据式(2),取 $C=2.5$,可以求得模糊相似矩阵 R 。

0.00	0.775	0.831	0.831	0.831	0.831	0.831	0.831	0.831	0.831
0.775	1.000	0.682	0.682	0.751	0.666	0.646	0.584	0.574	0.574
0.831	0.682	1.000	0.715	0.708	0.675	0.675	0.677	0.504	0.504
0.831	0.682	0.715	1.000	0.809	0.835	0.975	0.977	0.566	0.566
0.831	0.751	0.809	0.809	1.000	0.945	0.945	0.945	0.515	0.515
0.831	0.666	0.675	0.675	0.945	1.000	0.933	0.975	0.531	0.531
0.831	0.646	0.675	0.675	0.945	0.933	1.000	0.911	0.567	0.567
0.831	0.584	0.677	0.677	0.945	0.975	0.911	1.000	0.564	0.564
0.831	0.574	0.504	0.504	0.515	0.531	0.567	0.564	1.000	1.000
0.831	0.574	0.504	0.504	0.515	0.531	0.567	0.564	1.000	1.000

用平方法求传递闭包 $t(R)$, 求得 $t(R) = R^8$, 即

1.000	0.831	0.928	0.831	0.969	0.831	0.831	0.831	0.831	0.831
0.831	1.000	0.682	0.682	0.831	0.666	0.667	0.584	0.574	0.574
0.928	0.682	1.000	0.821	0.928	0.831	0.821	0.831	0.831	0.831
0.831	0.666	0.831	1.000	0.831	0.967	0.875	0.967	0.967	0.967
0.969	0.831	0.928	0.831	1.000	0.831	0.831	0.831	0.831	0.831
0.831	0.667	0.831	0.967	0.831	1.000	0.967	0.975	0.974	0.974
0.831	0.667	0.831	0.972	0.831	0.967	1.000	0.967	0.967	0.967
0.831	0.584	0.831	0.831	0.831	0.976	0.967	1.000	0.971	0.971
0.831	0.574	0.831	0.831	0.831	0.974	0.967	0.974	1.000	1.000

然后,取 $\alpha=0.928$,可以将客户进行分类(黄金客户、白银客户和具备发展潜力客户),即 $\{X_1, X_5\}, \{X_3\}, \{X_2, X_4, X_6, X_7, X_8, X_9\}$ 。

4 结束语

本文用可定制的评估指标树来描述客户分类的评价体系,提出了基于模糊聚类分析的客户分类算法。该算法的计算实例表明,它可以用于有关客户关系管理的决策支持系统中。希望本文的研究工作对今后的客户分类管理研究能有所帮助。

参考文献:

- [1] 杨叔子,吴波,胡春华,等. 网络化制造与企业集成 [J] . 中国机械工程, 2000, 11(1): 45-48.
- [2] 李宝东,宋瀚涛. 数据挖掘在客户关系管理中的应用 [J] . 计算机应用研究, 2002, 19(10): 71-74.
- [3] 谢季坚,刘承平. 模糊数学方法及其应用 [M] . 武汉:华中科技大学出版社, 2000.
- [4] 张永大. 数据挖掘在客户关系管理系统中的应用 [J] . 福建电脑, 2003, (6): 21-23.

作者简介:

郭蕴华(1975-),男,四川成都人,讲师,博士后研究人员,博士,主要研究方向为电子商务、网络及数据库应用、分布式系统;陈定方(1946-),男,湖北武汉人,教授,博士生导师,主要研究方向为虚拟设计/制造、计算机仿真。