

基于群体智能的客户转移模式挖掘*

金 鹏^{1,2}, 朱云龙¹

(1 中国科学院沈阳自动化研究所 沈阳 110016;

2 中国科学院研究生院 北京 100039)

摘 要: 在复杂多变的市场环境中,企业需要及时准确地掌握客户需求及其消费模式的变化趋势。本文提出了客户转移模式挖掘的概念,分析了客户转移模式挖掘的流程,提出了一种基于群体智能的规则变化挖掘方法,将单一的客户数据作为独立个体,给出了其在规则集中搜索的信息素更新策略及项搜索策略。最后通过对两组电信企业客户数据集的实验验证,表明该方法能够有效地挖掘客户转移模式。

关键词: 数据挖掘; 客户转移模式; 群体智能; 规则变化挖掘

Rule change mining based on swarm intelligence

Jin Peng^{1,2}, Zhu Yunlong¹

(1 Shenyang Institute of Automation of the Chinese Academy of Sciences, Shenyang 110016, China;

2 Graduate School of the Chinese Academy of Sciences, Beijing 100039, China)

Abstract: Understanding and adapting to changes of customer behavior is an important aspect of surviving in a continuously changing market environment for a modern company. The concept of customer change model mining is introduced and its process is analyzed in this paper. A rule change mining method based on swarm intelligence is presented, and the strategies of pheromone updating and items searching are given. Finally, an examination on two customer datasets of a telecom company illuminates that this method can achieve customer change model efficiently.

Key words: data mining; customer change model; swarm intelligence; rule change mining

1 引 言

随着电子商务等新型商业模式的发展,市场环境日益复杂,客户需求也在不断变化。在这种复杂多变的 market 环境中,及时准确地掌握客户需求及其消费模式的变化趋势,以及变化产生的原因,企业才能制定出相应的市场营销策略,从而获得成功^[1]。例如,企业需要了解以下问题:对哪些客户群的销售显著增长?哪些客户群所购买的产品发生了改变?客户消费模式如何变化?变化产生的原因是什么?等等。这些问题可以通过客户转移模式分析给出答案。

群体智能是受到群居昆虫群体和其它动物群体的集体行为的启发而产生的算法和解决方案。在群体智能中组成群体的是结构简单的独立个体,通过个体间及个体和环境间的简单交互,最终体现为群体行为^[4]。单一的一个客户数据与群体智能中的个体类似,本身结构简单,不能提供有效的客户模式,但多个近似的客户数据所体现出来的客户模式,则反映了该客户群的群体特征。而数据挖掘则能通过自动或半自动的方法和过程,在海量数据中发现隐含的、未知的和有潜在用途的知识和规则。因此本文采用群体智能和数据挖掘相结合的方法来进行客户转移模式分析。

由于 IF...THEN 规则在客户模式的表示中应用很广泛,关联、分类、聚类数据挖掘方法所得到的客

* 本课题由国家自然科学基金重点项目(70431003)支持

户模式都可以用规则的形式进行描述,因此对规则变化进行分析和挖掘,具有普遍的意义。目前的研究只是关注规则结构的变化,并未发现规则覆盖样本的变化^[2],这对于客户分析是不够的,客户分析需要知道在变化规则中客户的来源和去向。规则变化分析的难点在于:1)规则结构不同,不能直接比较;2)如何判断发生了何种变化、变化幅度多大以及发生变化的原因。本文将每个客户数据看作一个个体,用基于群体智能思想的客户转移模式挖掘方法在两个规则集中搜索和匹配规则,从而发现规则变化规律,以及相应的客户 cluster 的特征。这种方法不从规则结构的角度来发现规则变化,而是从客户转移的角度来挖掘规则变化,能够发现变化规则所对应的客户转移情况,有利于针对性地做出决策。

本文其余内容组织如下:第二节介绍了客户转移模式的定义及其相关研究情况,并对本文所用到的变量和符号进行了说明;第三节详细阐述了本文所提出的基于群体智能的客户转移模式挖掘方法;第四节对算法性能进行了实验验证;最后,对本文进行了总结和展望。

2 客户转移模式的定义及相关研究

客户模式,即客户消费模式或客户行为模式,每种模式代表了某些客户的消费特征和行为特征。由于市场环境处于不断的变化之中,客户的需求和选择也在不断的改变,因此就产生了客户转移模式的概念。客户转移模式是指客户模式的改变及其变化规律。客户转移模式分析的任务,就是发现客户消费模式的改变,揭示其中的变化规律,得出变化产生的原因,用于辅助企业决策者做出正确的市场营销策略。

客户转移模式分析可以通过规则变化挖掘的方法来实现。现有的对不同数据集或规则集的动态或比较研究主要包括以下几个方面^[1]:

1. 变化环境中的规则准确度维持:其中的规则并不变化,只能维持已有的知识;
2. 浮动模式挖掘:只能发现规则支持度变化,而不能发现规则本身的结构变化;
3. 通过规则主观兴趣度的变化发现意外规则:不能发现哪些方面发生了变化、发生了哪些类型变化以及发生了多少变化;
4. 时序数据挖掘:更多关注的是规律性变化;
5. 分类规则比较:只能发现同结构规则的变化;

6. 不同时刻数据集产生的决策树中的变化挖掘:并不能检测到变化的完全集,只能检测到指定顺序属性的变化,而且不提供任何关于变化程度的信息;

7. 从规则结构变化的角度将规则变化模式分为浮动模式、意外模式、增加模式和消亡模式四种,但不能针对其中的样本进行研究。

本文针对上述研究中存在的问题,采用基于群体智能的客户转移模式分析方法,从客户转移的角度出发,对变化规则中的客户来源和去向、变化的程度以及产生的原因进行研究,从而辅助企业制定正确的市场策略。

本文所用到的变量和符号说明如下:

R^t t 时刻客户模式集;

R^{t+k} $t+k$ 时刻客户模式集;

r_i R^t 中的一个客户模式, $r_i \in R^t$;

r_j^{t+k} R^{t+k} 中的一个客户模式, $r_j^{t+k} \in R^{t+k}$;

$|M_i^t|$ r_i 条件部分的属性个数;

$|M_j^{t+k}|$ r_j^{t+k} 条件部分的属性个数;

$|N_i^t|$ r_i 结论部分的属性个数;

$|N_j^{t+k}|$ r_j^{t+k} 结论部分的属性个数;

A_{ij} 同时出现在 r_i 和 r_j^{t+k} 条件部分的属性集;

$|A_{ij}|$ A_{ij} 中的属性个数;

B_{ij}^c 同时出现在 r_i 和 r_j^{t+k} 结论部分的属性集;

$|B_{ij}^c|$ B_{ij}^c 中的属性个数;

X_{ipq} 二元变量, A_{ij} 中的第 p 个属性值相同时

$X_{ipq}=1$,否则 $X_{ipq}=0, q=1, 2, \dots, |A_{ij}|$;

Y_{jqc} 二元变量, B_{ij}^c 中的第 q 个属性值相同时

$Y_{jqc}=1$,否则 $Y_{jqc}=0, q=1, 2, \dots, |B_{ij}^c|$;

RulePair _{i} 规则 i 和规则 j 所组成的规则对;

RulePairsSet 规则对所构成的候选项集合;

ListofRulePair _{i} 项 RulePair _{i} 覆盖的客户列表;

Customer _{n} 客户 n ;

c 客户数量;

a 规则对的数量;

ρ 信息素挥发系数。

3 基于群体智能的客户转移模式挖掘方法

3.1 客户转移模式挖掘流程

客户转移模式分析的最终目标是要通过发现客户模式改变的规律及原因,既能对市场营销活动的效果进行评估,又能进行预测,从而辅助企业制定正确的市场策略。可以分两种情况来考虑:一种是企业决策者

根据市场的变化情况,制定了相应的市场策略,要预测客户在新的市场策略下的客户转移模式;另一种是根据分析客户在两个不同时期的客户转移模式,发现其中的规律和原因,从而制定新的相应的市场策略。在第一种情况下,要找到与当前策略相类似的以前实施过的策略,例如某产品要进行价格调整,就要找到之前实施过的类似的价格调整策略,分析客户在该策略实施前后的客户转移模式,这与第二种情况是类似的。因此客户转移模式挖掘的关键问题就是对两个(或多个)不同时期的客户数据进行分析,发现客户转移模式及其产生的原因。整个客户转移模式挖掘的流程如图1所示。首先应用分类分析、聚类分析等数据挖掘方法,分别对两个(或多个)不同时期的客户数据进行数据挖掘分析,从结果中提取出客户模式,在此基础上使用基于群体智能的客户转移模式挖掘方法,从中发现各种类型的客户转移模式,最后分析其产生的原因,用于决策支持。其中的关键步骤是规则变化挖掘,也是本文所提出的方法要重点解决的问题。下面对规则变化挖掘方法进行详细阐述。

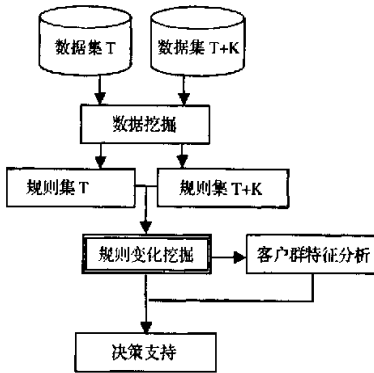


图1 客户转移模式挖掘流程图

3.2 算法的高级语言描述

算法1 基于群体智能的规则变化挖掘算法

RulePairsSet = { (r_i^t, r_i^{t+k}) | r_i^t ∈ R^t, r_i^{t+k} ∈ R^{t+k} }

for (n = 1; n <= c; n++) {

RulePairsSet 初始化;

for (m = 1; m <= a; m++) {

根据信息素量和启发式函数选择 RulePair_{ij};

if (Customer_n ∈ RulePair_{ij}) {

添加 Customer_n 到 ListofRulePair_{ij};

更新信息素量;

break;

}

else if (Customer_n ∈ R_i^t)

在候选项集中只保留包含 R_i^t 的项;

else if (Customer_n ∈ R_j^{t+k})

在候选项集中只保留包含 R_j^{t+k} 的项;

else 将候选项集中包含 R_i^t 或 R_j^{t+k} 的所有项去除;

}

if (Customer_n ∉ ∪ RulePair_{ij})

根据 Customer_n 满足的规则情况将其归类;

}

根据每个项的客户列表中的客户数量及预先定义的客户数量阈值,确定客户转移模式。

3.3 算法的详细阐述

3.3.1 预处理和初始化

对于两个规则集中两条结构完全相同的规则构成的项,一定会有满足该项的客户,因此不能提供有价值的知识。在初始化阶段,根据下列条件判断出两个规则集中结构完全相同的规则,并将其在两个数据集所覆盖的相同样本去除;

$$\begin{cases} |A_{ij}| = |M_i^t| = |M_j^{t+k}| \\ |B_{ij}| = |N_i^t| = |N_j^{t+k}| \\ \sum_p X_{ip} \times \sum_q Y_{jq} = 1 \end{cases} \quad (1)$$

此外预处理和初始化阶段还要进行如下操作:将两个规则集中剩余的规则映射为规则对 (r_i^t, r_i^{t+k}) 的集合,其中每个规则对被称为一个项;对每条规则统计覆盖的样本数,该数据将在算法的后续步骤中被重复用到,通过事先的预计算可以减少计算量,提高算法效率;为每一项构建一个符合该项的客户列表 ListofRulePair_{ij},初始时空;对每一项以相同信息素量 τ(0)=1/a 进行初始化。

3.3.2 信息素更新及项选择策略

本文所提出的信息素更新及项选择策略基于文献[4]提出的蚁群优化算法(ACO)思想,该算法是由蚁群觅食过程中寻找最短路径的方法启发产生的。结合规则变化挖掘的特点,给出如下信息素更新策略。

对有客户符合的项,增加其信息素量,模拟蚂蚁在走过的路径上遗留信息素的过程,更新公式如下:

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \eta_{ij} \cdot \tau_{ij}(t) \quad (2)$$

对没有客户符合的项,减少其信息素量,模拟信息素的挥发,其更新公式如下:

$$\tau_{ij}(t+1) = \tau_{ij}(t) - \rho \cdot \tau_{ij}(t) \quad (3)$$

为了使算法更有效地收敛,采用了基于规则支持度的启发式函数,其公式如下:

$$\eta_{ij} = (s_i + s_j) / 2 \tag{4}$$

其中 s_i 和 s_j 分别表示规则 r_i^k 和 r_j^{k+1} 的支持度, 即满足规则的样本数在总样本数所占的百分比。

在上述计算的基础上, 给出项 RulePair_{ij} 选择概率计算公式如下:

$$p_{ij}(t) = \frac{\tau_{ij}(t) \eta_{ij}}{\sum_{\text{RulePair}_{ij} \in S} \tau_{ij}(t) \eta_{ij}} \tag{5}$$

3.3.3 项搜索过程

为了提高算法的运行效率, 在向搜索过程中加入了判断条件, 避免了遍历所有项。当客户符合当前项, 则将该客户添加到该项的列表中, 根据公式更新信息素, 并结束该客户的搜索过程; 否则, 如果客户符合该项的第一个规则, 则将候选项集中不包含第一个规则的所有项去除; 否则, 如果客户符合该项的第二个规则, 则将候选项集中不包含第二个规则的所有项去除; 如果客户对两个规则都不符合, 则将候选项集中包含第一个规则或第二个规则的所有项去除。通过判断, 可以有效减少候选项集中项的数量, 减少运算量。

3.3.4 客户转移模式的确定

算法的最后是根据算法运行结果来确定客户转移模式。根据事先给定的阈值, 例如支持率(置信度)大于一定比例, 可确定该项为一个客户转移模式。该阈值在应用中由领域专家给出。

4 实验验证

本文选取了某电信企业两组客户数据集作为应用体系的输入, 两组数据的采集时间间隔为三个月。所用到的属性如下表所示:

表 1 数据属性列表

序号	变量名	变量说明
1	Regular_dur	常规期间的通话时间
2	Discount_dur	折扣期间的通话时间
3	Local_dur	本地通话时间
4	Domestic_dur	国内长途通话时间
5	Svc_sms	短消息服务的次数
6	Svc_type	收费服务的种类
7	Svc_time	收费服务的次数
8	Card_type	电话卡种类
9	Disc_type	服务套餐种类
10	Age	客户年龄
11	Gender	客户性别
12	Owed_time	欠费次数
13	ARPU	客户月消费额
14	Churn	客户是否流失

试验平台采用的是自主开发的基于群体智能的数据挖掘软件系统 SIMiner, 计算机 CPU 主频 2.4 GHz, 内存 512 MB。应用本文提出的客户转移模式挖掘的流程, 首先应用文献[3]中的分类方法, 对两组数据集分别进行挖掘, 获得两组规则集, 然后采用上述基于群体智能的客户转移模式分析算法, 对两组规则集进行规则变化挖掘, 输出如图 2 所表示的变化规则集。

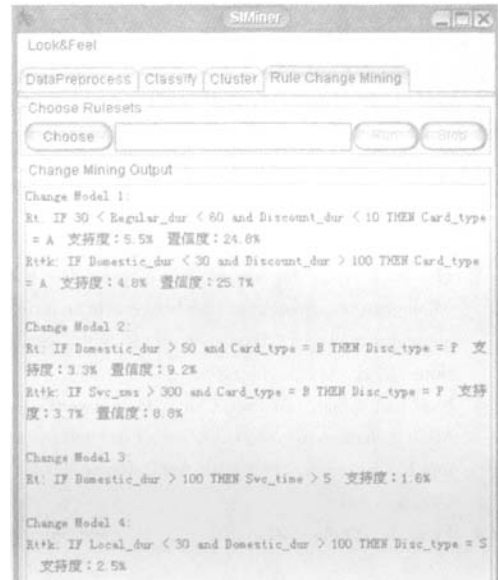


图 2 客户转移模式分析结果

从图中可以看出所得到客户转移模式具有不同类型。其中前两个是意外模式, 即两条规则的结论部分相同, 但条件部分不同。第三个是消亡模式, 即存在于以前所产生的客户模式集但不存在于当前所产生的客户模式集的客户模式。第四个是增加模式, 即新产生的客户模式, 该客户模式是以前所产生的客户模式集里所没有出现过的。

通过对不同类型的客户转移模式的分析, 可以有效地进行决策支持。例如从第一个意外模式中, 我们可以发现 A 电话卡的部分客户, 从常规期间的通话时间较长的消费模式, 向折扣期间的通话时间较长的消费模式转变, 说明 A 电话卡所采取的折扣优惠活动取得了一定的效果。

5 结 论

日益发展的商务模式以及不断变化的客户需求和

客户行为模式,给客户数据的动态分析和客户关系管理带来了新的挑战。本文提出了客户转移模式分析的概念,分析了客户转移模式分析的流程,提出了一种基于群体智能的规则变化挖掘方法,将单一的客户数据作为独立个体,给出了其在规则集中搜索的信息素更新策略及项搜索策略,挖掘出客户转移模式,使企业能够及时准确地掌握客户需求及其消费模式的变化趋势,从而制定出相应的市场营销策略。

在下一步的研究中应该按照浮动模式、增加模式、消亡模式和意外模式等四种客户转移模式类型,提出客户转移模式度量方法,有效区分不同类型的客户转移模式,计算客户转移模式的变化程度,从而更加有针对性地制定出相应的市场决策。

参考文献

- [1] Hee Seok Song, Jae kyeong Kim, Soung Hie Kim. Mining the change of customer behavior in an internet shopping mall[J]. Expert Systems with Applications. 2001, 21(3): 157-168
- [2] Mu-Chen Chen, Ai-Lun Chiu, Hsu-Hwa Chang. Mining changes in customer behavior in retail marketing[J]. Expert Systems with Applications. 2005, 28(4): 773-781
- [3] Peng Jin, Yunlong Zhu, Kunyuan Hu, and Sufen Li. Classification Rule Mining Based on Ant Colony Optimization[A]. ICIC 2006: Intelligent Control and Automation[C]. Lecture Notes in Control and Information Sciences. Springer-Verlag, Berlin Heidelberg New York, 2006, 344: 654-663.
- [4] Eric Bonabeau, Marco Dorigo, Guy Theraulaz. Swarm Intelligence: from Natural to Artificial Intelligence [M]. New York: Oxford University Press, 1999
- [5] Alex Berson, Stephen Smith, Kurt Thearling 著. 贺奇,郑岩,魏黎等译. 构建面向 CRM 的数据挖掘应用[M]. 北京:人民邮电出版社,2001. 8
- [6] Bing Liu, Wynne Hsu, Heng-Siaw Han, and Yiyuan Xia. Mining changes for real-life applications[C]. the 2nd international conference on data warehousing and knowledge discovery, London Greenwich, UK, 2000, 337-346
- [7] Sung Ho Ha, Sung Min Bae and Sang Chan Park. Customer's time-variant purchase behavior and corresponding marketing strategies; an online retailer's case[J]. Computers & Industrial Engineering. 2002, 43(4): 801-820
- [8] 李纯青,徐寅峰,张洋. 基于知识管理的动态客户关系管理研究[J]. 中国管理科学. 2004,12(2): 88-94