

Foodmart 公司的客户价值评价与客户细分

--基于数据仓库和数据挖掘技术的实现方案

(一) 案例来源:

本案例来源于学术论文（全文见附件）。案例设计的目的是训练大家综合应用数据库技术、统计分析技术和数据仓库技术的能力。提高大家学术研究或商业案例研究的能力。



Expert Systems with Applications 28 (2005) 773–781

Expert Systems
with Applications

www.elsevier.com/locate/eswa

Mining changes in customer behavior in retail marketing

Mu-Chen Chen^{a,*}, Ai-Lun Chiu^b, Hsu-Hwa Chang^c

^aDepartment of Business Management, National Taipei University of Technology, No. 1, Section 3, Chung-Hsiao E. Road, Taipei 106, Taiwan, ROC

^bSAS, Taipei, Taiwan, ROC

^cDepartment of Business Administration, National Taipei College of Business, Taipei, Taiwan, ROC

Abstract

During the past decade, there have been a variety of significant developments in data mining techniques. Some of these developments are implemented in customized service to develop customer relationship. Customized service is actually crucial in retail markets. Marketing managers can develop long-term and pleasant relationships with customers if they can detect and predict changes in customer behavior. In the dynamic retail market, understanding changes in customer behavior can help managers to establish effective promotion campaigns. This study integrates customer behavioral variables, demographic variables, and transaction database to establish a method of mining changes in customer behavior. For mining change patterns, two extended measures of similarity and unexpectedness are designed to analyze the degree of resemblance between patterns at different time periods. The proposed approach for mining changes in customer behavior can assist managers in developing better marketing strategies.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Data mining; Association rules; Retailing; Customer behavior; Change patterns

(二) 研究目标和任务

下面为论文研究流程，我们的任务完成 data preprocessing（数据预处理），customer behavioral 客户行为变量生成和客户细分。Mining customer behavior（挖掘客户行为）是下学期课程的要学习的主要内容之一，感兴趣的同学可继续选修。

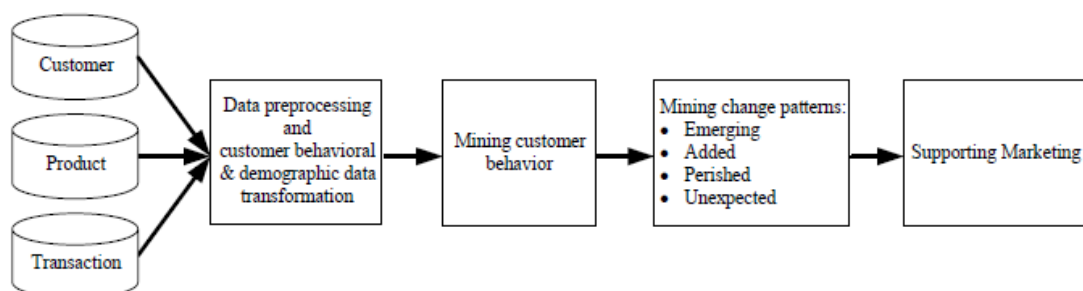


Fig. 1. Flowchart of mining changes for customer behavior.

(三) 相关知识背景

客户价值评价的 RFM 模型

In RFM, recency represents the interval between the most recent transaction time of individual customers and the evaluation time (Stone, 1995). Moreover, frequency represents the number of purchases by individual customers during a specific period. Additionally, monetary represents the

average expenditure of a customer during a specific period. Individual customers' recency, frequency, and monetary are scored to calculate the value of the purchasing behavior of each customer. This study adopts the RFM scoring approach of Miglautsch (2000) to transform the customer behavioral variables.

客户价值评价的 RFM 模型

根据美国数据库营销研究所 Arthur Hughes 的研究，客户数据库中有三个神奇的要素，这三个要素构成了数据分析最好的指标：

- 最近一次消费(Recency)
- 消费频率(Frequency)
- 消费金额(Monetary)

最近一次消费、消费频率、消费金额是测算消费者价值最重要也是最容易的方法，这充分的表现了这三个指标对营销活动的指导意义。而其中，最近一次消费是最有力的预测指标。结合这三个指标，我们就可以把顾客分成 $5 \times 5 \times 5 = 125$ 类，对其进行数据分析，然后制定相应的营销策略。

(四) 实施方案

1. Data source 数据源及处理

- (1) Footmart 公司数据库 (98 年的销售记录表)
- (2) 建立相关表或视图，计算客户行为变量指标 R/F/M
- (3) 对 R/F/M 进行离散化 (5 等分)
- (4) 计算客户价值 (Customer value): $5 \times R$ 离散 + $5 \times F$ 离散 + $5 \times M$ 离散
- (5) 建立 CRM 的 CUBE
- (6) 按“客户价值”进行聚类分析 (SPSS/R)

2. Data transformation 数据转换

Customer behavioral variables (recency, frequency, and monetary, RFM) are the important variables hidden in the database. To reduce the complexity in data explanation, the recency is assessed using five-score in terms of customer value. Customer purchase frequency is calculated based on weighted average and is differentiated using five-score. The monetary quantity of customer purchase indicates average customer expenditure and divides into five behavior scores for market segmentation.

3. Customer valuation and segmentation 客户行为和细分

According to the RFM variables mentioned in Section 3.1, the total **customer value scores** for individual customers are calculated to analyze market segmentation and target marketing. In retailing, recency of store visit is a more important indicator than purchase frequency and average spending per visit; hence, recency is assigned a greater weight (say 5) than frequency (say 3) and monetary amount (say 2). Consequently, the maximum score of individual customer value is 50 (i.e. $5 \times 5 + 3 \times 5 + 2 \times 5$), whereas the minimum score is 10 ($5 \times 1 + 3 \times 1 + 2 \times 1$).

The score of customer value provides a basis for customer clustering intended to demonstrate significant difference in customer behavior scores of individual customer clusters. Customers are divided into four clusters based on differentiated value scores. **客户细分为4类。** Following customers are grouped based on customer values, the growth matrix of Boston Consulting Group (BSG) is employed to differentiate the value of each customer cluster based on purchase frequency and average monetary expenditure and to segment customers into four clusters with different values based on purchase frequency and average monetary expenditure. The clusters are:

best customers (most valuable), frequent buyers, spenders, and uncertain customers (least valuable).

4. Customer segmentation 客户细分

Use K-means approach and ANOVA analysis.

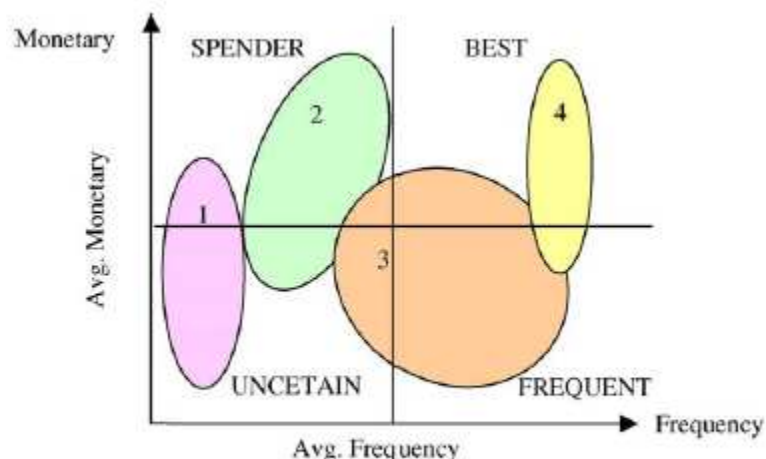


Fig. 3. Customer segmentation via value matrix.

Cluster 4 represents the most valuable customers, 类4代表最有价值客户 while Clusters 1 and 2 represent peripheral customers with lower purchase frequency and monetary expenditure. 类1和类2代表外围客户 When recency is used for cross analysis of the four customer clusters, customers in Clusters 1 and 2 have not made regular purchases recently. Meanwhile, Cluster 4 consists of customers who have recently made regular purchases, and also have higher average purchase size and purchase frequency. Therefore, Cluster 4 is concluded to be the most valuable for the business. This study applies association rules to discover patterns and changes in the behavior of customers in Cluster 4 (Fig. 3).

附：数据预处理SQL脚本及技术细节

--以下脚本为客户关系管理CRM-客户分类准备数据和指标

--(1)通过查询和插入建立“客户行为变量”表

--Recency 最近一次消费

--Frequency 消费频次

--[Money] 消费金额

```
select customer_id,max(datediff(dd,the_date,'1999-01-01')) Recency,count(customer_id)
Frequency,sum(store_sales) [Money]
into Customer_Behavior
from dbo.sales_fact_1998 a join time_by_day b
on a.time_id=b.time_id
group by customer_id
go
```

--(2)创建视图存储C、R、M的最大最小值，为指标的离散提供数据

```
create view RFM_maxmin(maxR,minR,maxF,minF,maxM,minM)
as
SELECT MAX(Recency), MIN(Recency), MAX(Frequency),
       MIN(Frequency), MAX([Money]), MIN([Money])
FROM Customer_Behavior
```

R最少32天光顾，365天前光顾，F最少1次，最多243次，M最大1609，最小1.4

Go

--(3)创建视图计算对R、F、M进行离散化

--注意Recency是越小越好指标，公式同F和M有所不同

```
create view Customer_RFM
as
SELECT customer_id, Recency, Frequency, [Money],
       CASE WHEN (maxR - Recency) <= (maxR - minR) / 5 THEN 1
            WHEN (maxR - Recency) <= 2 * (maxR - minR) / 5 THEN 2
            WHEN (maxR - Recency) <= 3 * (maxR - minR) / 5 THEN 3
            WHEN (maxR - Recency) <= 4 * (maxR - minR) / 5 THEN 4
            WHEN (maxR - Recency) <= 5 * (maxR - minR) / 5 THEN 5
            ELSE NULL
       END AS R,
       CASE
            WHEN (maxF - Frequency) <= (maxF - minF) / 5 THEN 5
            WHEN (maxF - Frequency) <= 2 * (maxF - minF) / 5 THEN 4
            WHEN (maxF - Frequency) <= 3 * (maxF - minF) / 5 THEN 3
            WHEN (maxF - Frequency) <= 4 * (maxF - minF) / 5 THEN 2
            WHEN (maxF - Frequency) <= 5 * (maxF - minF) / 5 THEN 1
            ELSE NULL
       END AS F,
       CASE WHEN (maxM - [Money]) <= (maxM - minM) / 5 THEN 5
            WHEN (maxM - [Money]) <= 2 * (maxM - minM) / 5 THEN 4
            WHEN (maxM - [Money]) <= 3 * (maxM - minM) / 5 THEN 3
            WHEN (maxM - [Money]) <= 4 * (maxM - minM) / 5 THEN 2
            WHEN (maxM - [Money]) <= 5 * (maxM - minM) / 5 THEN 1
            ELSE NULL
       END AS M
FROM Customer_Behavior CROSS JOIN RFM_maxmin
```

--(4)建立客户价值评分表

```
SELECT customer_id, Recency, Frequency, [Money], R, F, M,
       R * 5 + F * 5 + M * 5 as value
into Customer_Value
FROM Customer_RFM
```

(5) 建立CUBE, 进行OLAP分析或制作报表

将细分好的客户资料重新导入数据库, 就可建立CUBE了。

关键表格包括:

Customer_Behavior 客户行为变量表

Customer_RFM 客户RFM表

Customer_Value 客户价值表

Customer_Segment 客户细分表 (customer_id, Recency, Frequency,[Money], R, F, M, value,clusterID)

Segment_category 客户类别对照表 (ClusterID, CluterName) 见EXCEL附件

(6) 基于数据库或数据仓库进行客户细分

采用K-均值快速聚类方法 (根据R/F/M聚类)、描述统计方法、BOX-PLOT等统计方法, 将客户细分为4类, 结果见EXCEL文件。