

Research for Customer Classification of Clothing E-Business Based on Time Series Transaction Data

Hongping Zhang

Glorious Sun School of Business and Management, Donghua University, Shanghai
Email: zhp0119yy@163.com

Received: Dec. 7th, 2017; accepted: Dec. 22nd, 2017; published: Dec. 29th, 2017

Abstract

Customer classification is the premise of CRM. Based on the time sequence of transaction data and study of customer value in e-business environment, this paper proposed a multi index classification model with time series data and static data. In model application, this paper using feature extraction method for the time series data transform, ReliefF algorithm for index selection, then obtain a multi index weighted classification model. The results obtained from the real clothing business show that the customer groups formed using the method clustering all has statistical significant differences, and with meaningful explanations in terms of marketing strategy. Thus, this study considers useful for discriminative customer relationship management.

Keywords

Customer Classification, Time Series, ReliefF Algorithm, Clustering

基于时间序列交易数据的服装电商客户分类研究

张鸿萍

东华大学旭日工商管理学院, 上海
Email: zhp0119yy@163.com

收稿日期: 2017年12月7日; 录用日期: 2017年12月22日; 发布日期: 2017年12月29日

摘要

客户分类是企业进行客户关系管理的前提。文章在充分利用时间序列交易数据的基础上, 结合电子商务

环境下客户价值研究,建立了时间序列数据与静态数据相结合的多指标分类模型,并提出对时间序列数据进行特征提取处理,应用ReliefF算法对指标选择,最终得服装电商客户多指标加权分类模型。通过某著名服装电商的交易数据的分析,得出了有效客户分类,表明了本方法具有较强的客户分类和客户消费特征的解释能力。

关键词

客户分类, 时间序列, ReliefF算法, 聚类

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着互联网信息技术的发展,网购成为人们生活的重要组成部分。购物方式的改变,使得客户在企业间的选择成本大幅降低,流动性不断增强。为了更好的发展,越来越多的电子商务企业开始重视客户的不同需求。对于企业而言,资源是有限的,客户是不同的。企业如果能有效利用客户信息对客户进行准确的识别,那将会助力企业更好发展。

客户价值是评判客户对于企业重要性的一个标准。经典的评价模型是由 Arthur Hughes 提出来的 RFM 模型,将客户最近消费时间(R)、客户消费频率(F)、客户消费金额(M)作为衡量客户价值的量化模型。Chang 和 Tsay [1]在此基础上增加了客户关系持续时间(L),提出了 LRFM 模型,客户关系持续时间越久,说明客户对于企业的忠诚度越高,相应的客户价值也越高,该模型被广泛应用于零售、银行、运营商等行业。随着电子商务的发展,服装商业消费者也发生了不同的消费行为改变,服装电商企业积累了一定的客户消费数据,有效的利用模型识别服装电商的客户价值,将有利于服装企业集中资源,大力发展高价值客户。

传统的客户分类模型大多基于客户静态数据的分析,本文将立足于服装电商企业都拥有的客户交易数据,考虑客户消费行为的持续性,对服装电商客户进行分类研究。

2. 分类模型

2.1. 时间序列交易数据

传统的客户分类,以客户的静态属性数据作为分类依据,静态数据指的是不随时间变化的属性数据,主要包括人口统计学特征,客户历史消费数据。静态数据在获取处理上比较便捷,但是无法反映更多变化趋势情况,得到的结果往往准确性不够高。

在电子商务时代,企业的运营管理依托于互联网,企业与客户不直接面对面接触,但是企业可以轻松得到客户在电商平台的所有操作行为。客户交易数据是电商数据的重要组成部分,它主要记录了客户在该电商平台所购买的商品信息、购买商品的时间信息以及部分客户的个人信息,例如快递地址、颜色、尺码等。交易数据是客户消费行为的一个重要记录,消费行为是客户价值的一个体现,是客户对该品牌忠诚度的一个体现,大量的数据暗示着客户的价值[2] [3] [4] [5]。客户的交易数据是一串时间序列数据,这些数据包含了客户的消费偏好行为以及客户的未来消费趋势,是十分具有研究价值的,借助这些数据进行客户分类有助于增加客户识别的准确性。

2.2. 电子商务客户价值

本客户价值由客户既成价值和潜在价值组成[6]。前者指的是客户的消费行为为企业带来的直接利润；后者指的是客户可能为企业带来的收益。

客户的消费金额是企业关注的第一要点，直接反映客户对企业的收益的贡献度，是客户的既成价值。除此之外，在电子商务活动中，客户从挑选商品到最终购买商品，可能产生的企业可以获得又能体现客户价值的行为数据主要包括：订单价格，订单商品品类，下单时间、是否晒单、评价好坏、是否评价等。从上述行为数据中，企业可以通过统计汇总得到包括客户关系持续时间、一段时间内的客户消费频率、一段时间内的客户平均消费价格、客户的需求结构、晒单率、好评率等等，这里的大部分数据都可以体现客户的对企业的忠诚度，是衡量客户潜在价值的标准。其中客户所消费的商品品类数，研究认为客户在企业消费的商品品类越多，客户的既成价值和潜在价值都相应增加。此外，促销活动是电商运营的一个重要组成部分，客户对于活动的参与度也是衡量客户价值的一个标准。综上所述，可以得到电子商务客户价值评价指标体系如图 1。

2.3. 多指标分类模型

根据综合时间序列交易数据与电子商务客户价值衡量分析，关系持续时间、最近消费时间、消费频率、消费金额以及活动参与度和消费商品总品类数作为衡量电商客户价值的指标体系。在数据类型选择上，采用时间序列数据与静态数据相结合，消费频率和消费金额采用时间序列数据，更准确地对客户进行分类。具体指标如下。

1) 关系持续时间(L)

客户关系持续时间，一般用客户的最近一次交易与第一次交易的时间间隔，以年、月、日等作为统计单位。关系持续时间越长表示该客户的忠诚度越高，价值越高。

2) 最近消费时间(R)

最近一次消费时间，一般用统计日期前的客户最后一次消费时间到统计日期的时间间隔。

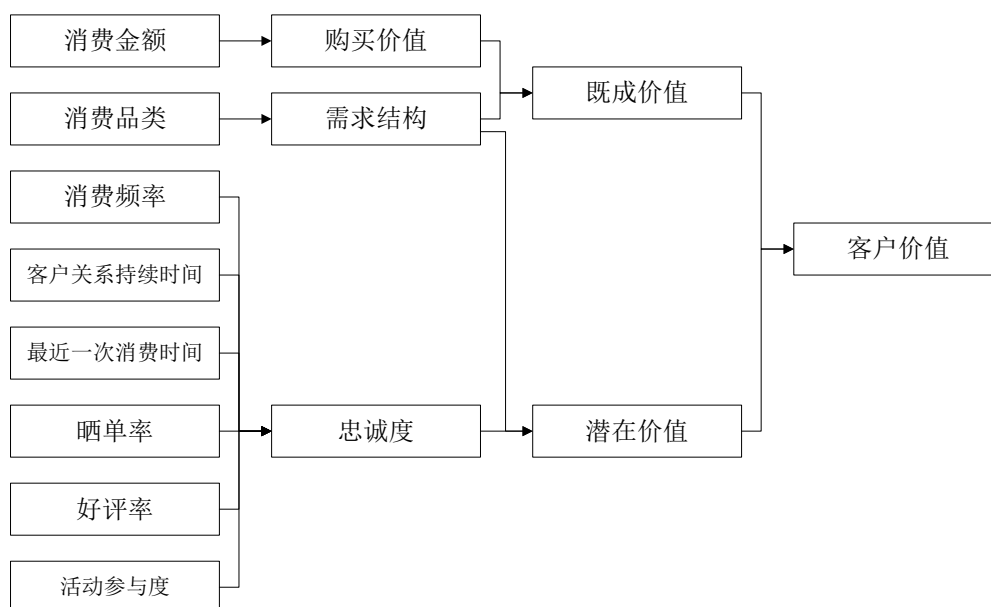


Figure 1. Evaluation index system of e-commerce customer value

图 1. 电子商务客户价值评价指标体系

3) 消费频率序列(F)

消费频率序列, 客户在消费过程中随时间行程的序列数据, 不仅能体现客户历史消费频率, 同时包含未来的发展趋势。

4) 消费金额序列(M)

消费金额序列, 客户在消费过程中随时间产生的金额序列。

5) 活动参与度(P)

活动参与度指的是客户在一次购买过程中对于活动商品的相对购买比例。

6) 消费商品总品类数(Q)

客户消费的总品类数是根据服装行业特点所增加的一个指标, 服装产品是一个有季节性的产品, 在不同的季节, 服装之间差异比较大。本文的消费商品总品类数是指在一定时间内, 客户所消费的商品中涉及的商品品类数量, 商品品类不等于商品, 这里的品类是指衬衫、牛仔裤这样的中类, 衬衫这个品类下可以包含各种不同的衬衫单品。客户购买过的商品品类数量越多, 说明客户对这个品牌的喜爱程度比较高, 会继续在该品牌购买的可能性越大。

本文对服装电商客户进行分类所选用的分类指标见表 1。

2.4. 指标量化

在实际对客户进行分类, 需要对提出的各指标进行量化表示, 具体分类指标的数学表示见表 2。在电商交易数据中, 在 T_1 到 T_n 时间内的单个客户的消费行为数据记录公式(1)、(2):

$$Y = ((X_1, t_1), (X_2, t_2), \dots, (X_n, t_n)) \quad (1)$$

$$X_i = \begin{bmatrix} c_{i1} & q_{i1} & p_{i1} & pro_{i1} \\ c_{i2} & q_{i2} & p_{i2} & pro_{i2} \\ \dots & \dots & \ddots & \dots \\ c_{il_i} & q_{il_i} & p_{il_i} & pro_{il_i} \end{bmatrix} \quad (2)$$

(X_i, t_i) 表示该客户第 i 次消费的消费时间是 t_i , 所消费内容为 X_i ; l_i 表示该客户在第 i 次消费活动购买的商品总数; c_{ij} 表示该客户在第 i 次消费的第 j 件商品的所属类别; q_{ij} 表示该客户在第 i 次消费的第 j 种商品的数量; p_{ij} 表示该客户在第 i 次消费的第 j 种商品的单价; pro_{ij} 表示该客户在第 i 次消费的第 j 件商品的在当时是否参与促销活动;

$$pro_{ij} = \begin{cases} 1, & \text{参与促销活动} \\ 0, & \text{未参与促销活动} \end{cases} \quad (3)$$

T_1 表示统计时间段的开始时间点, T_n 表示统计的截止时间点。

1) 关系持续时间(L)

客户关系持续时间用在统计区间内客户的第一次消费时间与最后一次消费时间的时间间隔表示, 选择以月作为时间单位:

$$L = t_n - t_1 (\text{月}) \quad (4)$$

2) 最近消费时间(R)

最近消费时间表示在统计区间内客户的最后一次消费时间与统计截止时间之间的间隔表示, 同样以月作为时间单位:

$$R = T_n - t_n (\text{月}) \quad (5)$$

Table 1. Customer classification index of clothing e-commerce
表 1. 服装电商客户分类指标

分类指标	说明
关系持续时间(L)	客户最后一次消费与第一次消费的时间间隔
最近消费时间(R)	客户最后一次消费与统计时间的时间间隔
消费频率序列(F)	客户在等时间段内的消费次数序列
消费金额序列(M)	客户每次消费的订单总金额序列
活动参与度(P)	一段时间客户所购买过的商品的总体活动占比
消费商品总品类数(Q)	一段时间内客户所购买过的商品品类的总数

Table 2. Classification index and quantitative description
表 2. 分类指标与量化说明

指标	量化表示
关系持续时间	$L = t_n - t_1$
最近消费时间	$R = T_n - t_n$
消费频率序列	$F = (f_1, f_2, \dots, f_i)$
消费金额序列	$M = (m_1, m_2, \dots, m_n)$
活动参与度	$P = \frac{\sum_{i=1}^n \sum_{j=1}^{l_i} pro_{ij}}{\sum_{i=1}^n l_i}$
消费商品总品类数	$Q = \text{不相等的 } c_{ij} \text{ 个数}$

3) 消费频率序列(F)

消费频率指的是一定周期内客户消费的次数，首先将统计时间开始截点 T_1 和结束截点 T_n 划分成等长的 h 个时间段，每个时间段相隔 Δt ，再统计不同客户每个时间段内的消费次数，一个客户形成一个时间序列 F ：

$$F = (f_1, f_2, \dots, f_h) \quad (6)$$

f_i 与 f_{i-1} 的时间间隔为 Δt ； f_i ：在该 Δt 时间段内，该客户的消费次数。如果该客户某一时间段 i 内未在店铺进行消费， $f_i = 0$

4) 消费金额序列(M)

消费金额序列指的是客户每次购买的订单总价形成的序列：

$$M = (m_1, m_2, \dots, m_n) \quad (7)$$

m_i ：该客户第 i 次消费的商品总价

$$m_i = \sum_{j=1}^m (q_{ij} * p_{ij}) \quad (8)$$

5) 活动参与度(P)

客户每进行消费活动，形成一个消费订单，其中包含一件或多件商品，一件商品可能是促销活动商品也可能不是，客户活动参与度用客户在统计时间段内购买的参与活动的商品数量于购买总商品数量的比值表示。

$$P = \frac{\sum_{i=1}^n \sum_{j=1}^{l_i} pro_{ij}}{\sum_{i=1}^n l_i} \quad (9)$$

6) 消费商品总品类数(Q)

一般来说, 一个服装的品类包含多个服装商品, 消费商品品类总数计算中, 客户购买同属一个品类的多个商品。

$$Q = \text{不相等的 } c_{ij} \text{ 个数} \quad (10)$$

2.5. 特征提取

由于客户的消费金额序列不是等频率的, 因此客户时间序列数据具有维数高、维数不确定、数据间隔不等的特点。而且考虑到特征中还存在静态数据, 本文采用首先对序列数据提取特征, 再与静态数据一起使用聚类方法进行分类。特征的提取选取时间序列的趋势、均值、方差、偏度、峰度等来描述客户时间序列, 在降低了数据的维数同时, 保证描述时序数据的基本统计特征, 反映了客户时序数据的变化特点。假定第 i 个客户的时间序列为 $Y_i = y(1), \dots, y(n)$, 则特征的具体选择方法如下:

1) 趋势

趋势特征用于反映客户消费序列的长期变化趋势, 采用最小二乘法拟合时间序列, 以拟合直线的斜率作为时间序列的变化趋势。

$$T = \frac{\sum_{t=1}^n x_t y(t)}{\sum_{t=1}^n x_t^2} \quad (11)$$

2) 平均值

客户时间序列的平均值反映了客户消费序列的平均水平。

$$u = \frac{\sum_{t=1}^n y(t)}{n} \quad (12)$$

3) 方差

方差反映了客户购买序列的波动程度。

$$\sigma = \sqrt{\frac{\sum_{t=1}^n (y(t) - \mu)^2}{n}} \quad (13)$$

4) 偏度

偏度是用来衡量数据的分布相对与中心点是否看起来一致, 用于度量时间序列的值相对于平均值的对称程度。

$$S = \frac{\sum_{t=1}^n (y(t) - \mu)^3}{n\sigma^3} \quad (14)$$

5) 峰度

峰度是数据分布集中趋势高峰的形状, 用于描述时序数据的分布相对与正态分布来讲是平坦的还是

具有尖峰的。峰度对于标准的正态分布是 3，公式计算：

$$K = \frac{\sum_{t=1}^n (y(t) - \mu)^4}{n\sigma^4} - 3 \quad (15)$$

2.6. 指标选择与赋权

本文的客户分类模型主要涉及 L 、 R 、 P 、 Q 以及 F 、 M 的 5 个统计特征提取值，共 14 个分类指标。在实际客户价值衡量中，不同的特征对于价值的贡献程度是不一致的。因此，为了得到更合理的客户分类结果，需要对不同的指标进行重要性赋值。**ReliefF** 是一种特征权重算法，可以处理多分类问题，通过计算可以得到稳定的特征权值大小，根据权值大小在对不重要的特征进行删除，最终得到相应的关键特征以及其权值大小。

基于 **ReliefF** 算法的权重计算步骤如下，其中 $M_j(C_i)$ 表示第 C_i 类的第 j 个样本， $j=1,2,\dots,k$ ； $P(C_i)$ 表示第 C_i 类目标的概率； $\text{diff}(A, I_1, I_2)$ 表示样本 I_1 和 I_2 在特征 A 上的相似性，相似性用距离表示， $\text{diff}(A, I_1, I_2)$ 计算公式如下：

$$\text{diff}(A, I_1, I_2) = \frac{|I_{1A} - I_{2A}|}{\max(A) - \min(A)} \quad (16)$$

$\max(A)$ ， $\min(A)$ 分别表示特征 A 的最大值和最小值。

输入：训练数据集 S ，迭代次数 m ，最近邻样本个数 k

输出：预测的特征权值向量 W

- 1) 初始化权值向量 $W(A) = 0.0, A = 1, 2, 3, 4$;
- 2) 从 S 中随机选择一个样本 s_i ;
- 3) 找到与样本 s_i 同类别的 k 个最近邻 h_j ;
- 4) 对于不同于 s_i 所属类别 C_i 的每个其他类别，都找出 k 个最近邻 $M_j(C_i)$
- 5) 更新各个特征的权值

$$W(A) = W(A) - \sum_{j=1}^k \text{diff}(A, s_i, h_j) / (m \cdot k) + \sum_{C \neq C_i} \left[\frac{P(C \neq C_i)}{1 - P(C_i)} \sum_{j=1}^k \text{diff}(A, s_i, M_j(C)) \right] / (m \cdot k) \quad (17)$$

- 6) 重复步骤(5) $p = 4$ 次
- 7) 重复步骤(2)~(5) m 次

采用上述算法对现有的 14 个指标进行指标选择。首先，需要准备 **ReliefF** 算法数据集，根据客户价值的分类依据，将客户分类三大类，第一是再次消费可能性较小的客户；第二类是再次消费可能性较高但金额较小的客户；第三类是再次消费可能性较高且金额较大的客户。

经过数据统计分析发现，在交易数据中仅有 8.97% 的客户会在时隔两年之后重新在该企业消费。因此，在数据集准备上，从 2011 年~2015 年 5 年的电商客户交易数据中选择消费总次数大于 5 次且在 2011 年-2013 年有过消费记录的 300 位客户以及这 300 位客户的消费记录。将这 300 位客户分成三大类，按照上述分类标准，分为三大类。

- 1) 再次消费可能性较小的客户

所选客户中最后一次消费时间在 2014 年 1 月 1 日之前的客户，记为属于类别 C_1

- 2) 再次消费可能性较大但金额较小的客户

除去无再次消费行为的客户，其余客户均认为是再次消费客户，统计所有客户每次消费的平均金额 \bar{x} ， \bar{x} 表示如下式：

$$\bar{x} = \frac{\sum_{i=1}^m \sum_{j=1}^n x_{ij}}{\sum_{i=1}^m n_i} \quad (18)$$

在余下客户中，客户最后一次消费金额小于 \bar{x} 的客户，成为再次消费且金额较小的客户记为类别 C_2 客户。

3) 再次消费可能性较大且金额较大的客户

剩余客户未再次消费且金额较大的客户，记为类别 C_3 客户。

结合 14 个特征以及三个分类集，得到客户数据集 S ，记 $S = (s_1, s_1, \dots, s_m)$ ， s_i 表示第 i 个客户的特征及所属类别，具体标号表示见表 3。

3. 聚类算法

3.1. k-means 算法

k-means 算法是使用最普遍的聚类算法之一[7]，适合于处理大数据集，简单、快速，且算法具有可伸缩性和高效性。一般步骤如下：

输入：聚类个数 k 。

- 1) 从 n 个数据对象任意选择个 k 对象作为初始聚类中心；
- 2) 循环(3)到(4)直到每个聚类不再发生变化为止；
- 3) 根据每个聚类对象的均值中心对象，计算每个对象与这些中对象的距离；并根据最小距离重新对相应对象进行划分；
- 4) 重新计算每个有变化聚类的均值中心对象。

3.2. 聚类评价指标

对于 k 值的确定本文利用 DB 指标作为评判标准[8]，是基于样本的类内散度与各聚类中心间距的测度。它是进行类数估计时其最小值对应的类数作为最佳类数。表示如下：

$$DB(k) = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{W_i + W_j}{C_{ij}} \right) \quad (19)$$

Table 3. Data of indicator description

表 3. 指标说明

指标名称	指标编号	指标名称	指标编号
关系持续时间(L)	1	M 趋势	9
最近消费时间(R)	2	M 方差	10
F 均值	3	M 偏度	11
F 趋势	4	M 峰度	12
F 方差	5	活动参与度(P)	13
F 偏度	6	消费商品总品类数(Q)	14
F 峰度	7	分类	15
M 均值	8		

其中 C_{ij} : 聚类中心 C_i 和聚类中心 C_j 之间的距离, W_i 表示聚类中心 C_i 的所有样本到其聚类中心 C_i 的平均距离。

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p \omega_k (x_{ik} - x_{jk})^2} \quad (20)$$

4. 实例应用

4.1. 数据获取与预处理

本文选择在 2011 年至 2015 年在该店铺有 5 次及以上消费行为的客户作为分类的客户样本集总体, 从中随机抽样 5000 名客户作为实例分析样本集。

首先将数据样本预处理得到分类所需的 14 个特征值。其次, 对数据进行标准化处理。本文采用 [0,1] 标准化, 计算公式如式(21):

$$X'_{ij} = \frac{X_{ij} - \min_{1 \leq i \leq n} \{X_{ij}\}}{\max_{1 \leq i \leq n} \{X_{ij}\} - \min_{1 \leq i \leq n} \{X_{ij}\}} \quad (21)$$

4.2. 指标选择与赋权

根据本文提出的客户分类模型, 首先利用 ReliefF 算法对 14 个指标进行权重赋值, 得到权重结果如图 2。

计算各个指标对应的权值的平均值, 见表 4。

由于权值接近 0 或者负值的指标对结果的影响程度很小, 可以抛弃。从结果我们可以看出, L 、 R 、 F 均值、 F 趋势、 M 平均、 M 趋势、 P 、 Q 这 6 个指标的权值均大于 0.1, 其他指标的值均偏低。最终选择这 6 个指标作为客户分类模型, 经过归一化后, 得到这 6 个指标的权值见表 5。权重表达式

$$W = (0.145, 0.158, 0.182, 0.082, 0.187, 0.084, 0.094, 0.067).$$

4.2. 客户分类结果

结合企业分类需求, 在多指标加权分类模型基础上, 本文选取 $k = 3, 4, 5, 6$ 作为聚类的数目, 利用 Python 编程得到聚类数目及其 DB 指标值见表 6, 其中 $k = 4$ 时, DB 指标值相对较小, 所以选择聚类数目 4 作为分类数目。最终分类结果见表 7。

表 8 方差分析结果, 显著性水平均小于 0.05, 说明各个类间有显著差异, 分类结果是合理的。从上述聚类结果可以看出, 服装电商客户大致可以分为 4 大类。

第一簇, 该类客户称为核心客户, 该类客户的关系持续时间长久, 一般都在 2 年以上, 对于该品牌所有品类几乎都购买过, 忠诚度较高, 而且在该品牌的交易次数明显高于其他三类客户, 在消费金额上也高于其余各类。除此之外, 这类客户还在近期有过消费。这类客户的消费频率趋势是明显上升的, 表明该类客户会继续在该品牌消费, 但是在消费金额上, 趋势上比较平缓。核心客户是一个企业最重要的客户资源, 他们往往数量不大, 但是忠诚度比较高, 可以为企业带来稳定又较高的销售额。对于这类客户, 企业应该着重提高这类客户的满意度, 关注该类客户的特殊需求, 继续保持这类客户的高频率消费, 并且加强与这类客户的沟通, 了解这类客户消费金额没有继续上升趋势的原因, 保持这类客户的高频率高金额消费趋势。

第二簇, 该类客户称为潜力客户。这类客户的关系持续时间较短, 但是在短期内有消费行为, 所以这类客户是新客户, 消费频次低于平均水平, 但是呈增长趋势, 消费金额高于平均水平, 也呈现增长趋

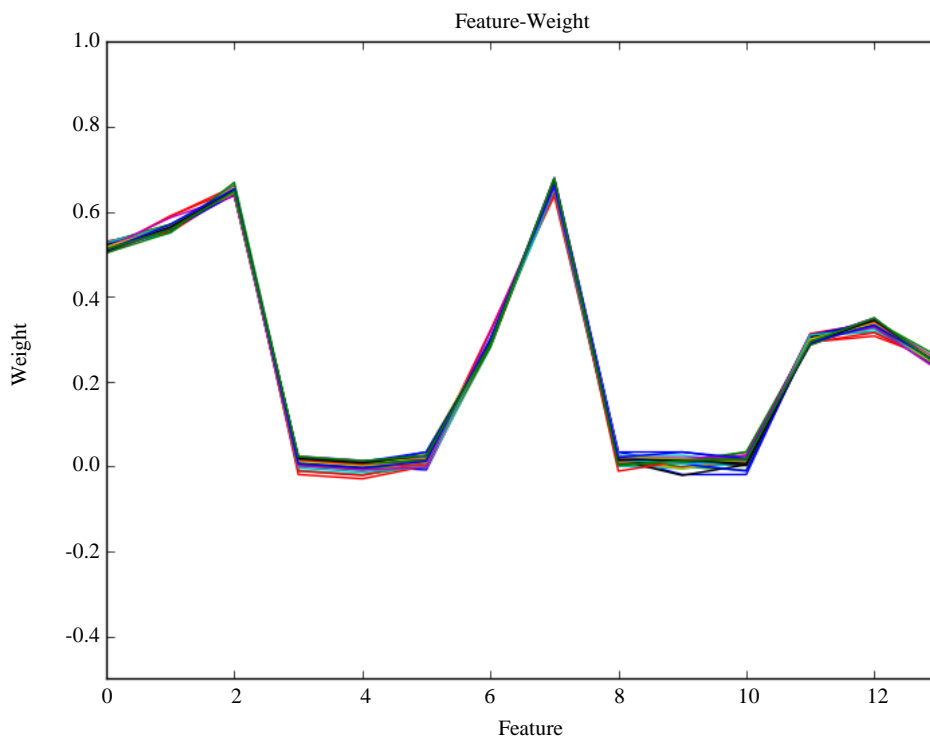


Figure 2. Curve: weight of index
图 2. 指标权重曲线

Table 4. System resulting data of index weight
表 4. 指标权重结果数据

指标	权值	指标	权值
<i>L</i>	0.524	<i>M</i> 均值	0.665
<i>R</i>	0.561	<i>M</i> 方差	0.015
<i>F</i> 均值	0.653	<i>M</i> 峰度	0.010
<i>F</i> 方差	0.009	<i>M</i> 偏度	0.013
<i>F</i> 峰度	-0.001	<i>M</i> 趋势	0.299
<i>F</i> 偏度	0.013	<i>P</i>	0.335
<i>F</i> 趋势	0.293	<i>Q</i>	0.240

Table 5. System weight data of index
表 5. 指标权值表

指标	<i>L</i>	<i>R</i>	<i>F</i> 均值	<i>F</i> 趋势	<i>M</i> 均值	<i>M</i> 趋势	<i>P</i>	<i>Q</i>
权值	0.145	0.158	0.182	0.082	0.187	0.084	0.084	0.067

Table 6. Cluster number and DB index
表 6. 聚类数目与 DB 指标表

聚类数目	3	4	5	6
DB	1.434	1.032	1.543	2.01

Table 7. Data of Final cluster center**表 7.** 最终聚类中心表

簇类别	1	2	3	4	5	6	7	8	占比
1	37.71	3.00	6.17	1.46	161.61	0.75	7.71	0.31	9%
2	14.40	7.34	1.41	0.89	149.19	3.67	4.29	0.40	36%
3	13.56	31.81	1.27	-0.43	118.18	0.14	3.92	0.06	26%
4	37.06	7.96	1.57	0.20	128.75	-2.21	4.57	0.22	29%

Table 8. Data of F test of clustering results**表 8.** 聚类结果 F 检验

指标	<i>L</i>	<i>R</i>	<i>F</i> 均值	<i>F</i> 趋势	<i>M</i> 均值	<i>M</i> 趋势	<i>P</i>	<i>Q</i>
P 值	0	0	0	0	0.025	0.022	0.000	0.000

势，并且趋势比较明显。这类客户对促销活动的参与度比较高，目前消费的商品品类数量较低。潜力客户是成长型的客户，在一个电商企业中，这类客户的占比较大，他们往往会因为网店推出新的款式，参与某些促销活动而成为该品牌的客户。这类客户目前的忠诚度较低，但是是发展成为核心客户的主力军，对于这类客户企业应该关注这类客户购买之后的评论等反馈意见，主动为这类客户推送促销活动信息，提供更好的服务水准，使之升级成为企业的核心客户。

第三簇，该类客户是企业的临时客户。这类客户与企业的交易持续时间低于平均水平，而且已经很长时间没有再消费，消费频率、消费金额都低于平均水平，对品牌活动参与度较低，消费商品类别数较少，对企业的贡献价值也不高。对于这类客户，企业可以减少关注度。

第四簇，该类客户是企业的长期客户。该类客户与企业的交易关系持续时间比较长，但是消费频率一般，在一段时间内有过消费行为，并且消费频次保持的比较稳定，消费金额低于平均水平，并且呈现下降趋势，趋势比较明显，活动参与度一般，消费商品。这类客户属于比较稳定的客户，对于这类客户，企业应该引起重视，重点关注，了解这类客户消费金额明显下降的原因，从产品推荐、客服服务质量等各方面提高这类客户的满意度，从而提高客户的消费金额和消费频次，将这类客户发展成为核心客户。

5. 小结

对客户进行合理分类是现代企业客户关系管理的一个重要组合部分。本文在充分利用时间序列交易数据的基础上，结合电子商务环境下客户价值研究，建立了时间序列数据与静态数据相结合的多指标分类模型，使企业更有效利用数据，提高客户识别准确性。在模型处理上，提出对时间序列数据进行特征提取处理，并应用 ReliefF 算法对指标选择，最终得服装电商客户多指标加权分类模型。通过实例应用，将服装电商客户的多次消费客户分为四大类，并根据每类客户特点，提出相应客户管理策略。大数据时代，科学有效的利用数据资源，能为企业运营管理提供有效支持。

参考文献 (References)

- [1] Chang, H.H. and Tsay, S.F. (2004) Integrating of SOM and K-Mean in Data Mining Clustering: An Empirical Study of CRM and Profitability Evaluation. *Journal of Information Management*, **11**, 161-203.
- [2] 权明富, 齐佳音, 舒华英. 客户价值评价指标体系设计[J]. 南开管理评论, 2004, 7(3): 17-23.
- [3] 夏维力, 王青松. 基于客户价值的客户细分及保持策略研究[J]. 管理科学, 2006, 19(4): 35-38.
- [4] 杨兰, 卢润德. 基于客户价值的客户分类方法研究[J]. 现代管理科学, 2007(11): 95-96.

-
- [5] Chmitt, P., Skiera, B. and Christophe, V.D.B. (2011) Referral Programs and Customer Value. *Journal of Marketing*, **75**, 46-59. <https://doi.org/10.1509/jmkg.75.1.46>
- [6] Verhoef, P.C. and Donkers, B. (2001) Predicting Customer Potential Value: An Application in the Insurance Industry. Erasmus Research Institute of Management (ERIM), ERIM Is the Joint Research Institute of the Rotterdam School of Management, Erasmus University and the Erasmus School of Economics (ESE) at Erasmus University Rotterdam, 189-199. [https://doi.org/10.1016/S0167-9236\(01\)00110-5](https://doi.org/10.1016/S0167-9236(01)00110-5)
- [7] Zhang, H., Zhao, T.F., Wang, Y.Z., *et al.* (2011) The Application of K-means Algorithm Based on Density in Recognizing Gas-Bearing Core Samples from Water-bearing Ones. *Science Technology & Engineering*,
- [8] 赵娟英, 周颖. 一种新聚类评价指标[J]. 陕西师范大学学报(自科版), 2015(6): 1-8.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2160-7311, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: mm@hanspub.org