

8.4.3 K-means 的应用

2023 年 9 月 18 日

K-means 的应用——百货商场会员

在零售行业中，会员价值体现在持续不断地为零售运营商带来稳定的销售额和利润，同时也为零售运营商策略的制定提供数据支持。零售行业会采取各种不同方法来吸引更多的人成为会员，并且尽可能提高会员的忠诚度。

背景

- 当前电商的发展使商场会员不断流失，给零售运营商带来了严重损失。此时，运营商需要有针对性地实施营销策略来加强与会员的良好关系 - 完善会员画像，使会员的形象更具体，帮助商家了解客户 - 加强对现有会员的精细化管理，提供个性化的服务，可与会员建立稳定的关系



客户分析 RFM 模型

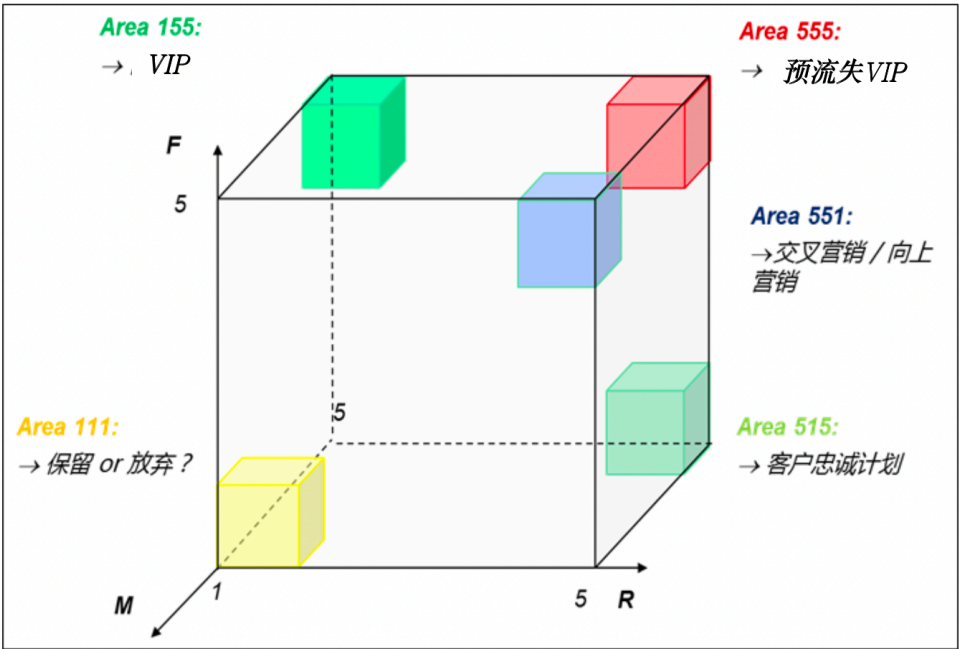
通过客户购买数据识别不同价值的客户，识别客户价值应用最广泛的模型是 RFM 模型。

R (Recency) 指的是最近一次消费时间与截止时间的间隔。通常情况下，最近一次消费时间与截止时间的间隔越短，对即时提供的商品或是服务也最有可能感兴趣。

F (Frequency) 指顾客在某段时间内所消费的次数。可以说消费频率越高的顾客，也是满意度越高的顾客，其忠诚度也就越高，顾客价值也就越大。

M (Monetary) 指顾客在某段时间内所消费的金额。消费金额越大的顾客，他们的消费能力自然也就越大，这就是所谓“20% 的顾客贡献了 80% 的销售额”的二八法则。

RFM 模型包括三个特征，使用三维坐标系进行展示，如图所示。X 轴表示 Recency，Y 轴表示 Frequency，Z 轴表示 Monetary，每个轴一般会分成 5 级表示程度，1 为最小，5 为最大。



数据信息如下：

销售流水表记录的是该商场的销售数据，其中包括会员与非会员的消费数据

数据时间范围是 2015 年 1 月 1 日至 2018 年 1 月 3 日

字段	描述
会员卡号	会员的唯一标志
消费产生的时间	消费产生的时间
商品编码	商品的唯一编码
销售数量	购买某一商品的数量
商品售价	商品的单价
消费金额	消费的金额=单价*数量
商品名称	商品名称
此次消费的会员积分	此次消费累计的会员积分
收银机号	收银机号
单据号	相同的单据号可能不是同一笔消费
柜组编码	分类柜组的编码，如Amani
柜组名称	分类柜组的名称，如阿玛尼

目标：

- 对商场的经营数据和会员信息数据进行处理 - 分析商场的经营特征 - 对商场会员进行用户画像描

绘，方便更了解会员，对会员进行针对性的服务 - 根据会员的消费特征对会员进行精细划分，方便针对不同群体制定对应的营销策略或管理方案，从而提升商场的销售利润

读取数据

```
[1]: import pandas as pd
data = pd.read_csv('数据/sales.csv', encoding='gbk')
data.head()
```

```
[1]:      会员卡号  此次消费的会员积分  积分等级  年龄  年龄段  入会时长  消费次数  消
费频率      消费金额      消费水平  \
0  000186fa      5267.0  积分低等级  41  中年  977      4  低频消费  11880.7  中
等消费水平
1  000234ad      11850.0  积分中等级  43  中年  1097      7  中频消费  12850.0  中
等消费水平
2  000339f1      6141.0  积分低等级  30  青年  1010      8  中频消费  6340.8  低
消费水平
3  0004bad2      8964.0  积分低等级  34  青年   69      1  低频消费  8964.0  低
消费水平
4  000cd735     66423.0  积分中等级  55  中年  1056     40  高频消费  123759.5  高
消费水平
```

	平均每单金额	价值属性	最后一次消费距今时长	柜组名称
0	2970.175000	单均价值一般	101	雅诗兰黛 ESTEE LAUDER
1	1835.714286	单均价值一般	63	雅诗兰黛柜
2	792.600000	单均价值一般	18	Wacoal
3	8964.000000	单均价值高	69	朗姿柜
4	3093.987500	单均价值一般	39	OHUI/后

```
[2]: data.shape
```

```
[2]: (31301, 14)
```

通过数据观察，本案例以消费次数 F，消费总金额 M，最近消费距今时长 R 和入会时长 L 4 个特征作为百货公司识别客户价值的关键特征，记为 LRFM 模型。

```
[14]: X = data[['消费次数', '最后一次消费距今时长', '消费金额', '入会时长']]
X.head()
```

```
[14]:
```

	消费次数	最后一次消费距今时长	消费金额	入会时长
0	4	101	11880.7	977
1	7	63	12850.0	1097
2	8	18	6340.8	1010
3	1	69	8964.0	69
4	40	39	123759.5	1056

数据标准化

```
[19]: from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
feature = sc.fit_transform(X)
feature = pd.DataFrame(feature, columns=X.columns, index=X.index)
```

```
[21]: feature.head()
```

```
[21]:
```

	消费次数	最后一次消费距今时长	消费金额	入会时长
0	-0.162766	-0.852102	-0.123851	0.766248
1	0.238058	-0.953237	-0.105034	1.105873
2	0.371666	-1.073002	-0.231395	0.859645
3	-0.563591	-0.937268	-0.180472	-1.803583
4	4.647126	-1.017111	2.048021	0.989835

K-means 聚类算法

这里我们主观定义聚类后的簇数量为 5，也就是分成 5 个类别

```
[22]: import pandas as pd
from sklearn.cluster import KMeans
model = KMeans(n_clusters=5, random_state=0)
```

训练

```
[23]: model.fit(feature)
```

```
[23]: KMeans(n_clusters=5, random_state=0)
```

训练结果

```
[24]: model.cluster_centers_ # 类中心
```

```
[24]: array([[ 0.35405089, -0.61558198,  0.04441617,  0.67136318],
          [-0.3221002 , -0.58831476, -0.18918965, -1.11544549],
          [-0.41060379,  1.27719539, -0.25501726,  0.71403056],
          [ 2.70992706, -0.94293576,  1.91191234,  0.54151326],
          [ 6.78682128, -1.08705261,  9.7763573 ,  0.71438857]])
```

```
[25]: model.labels_ # 聚类类别
```

```
[25]: array([0, 0, 0, ..., 0, 1, 2], dtype=int32)
```

```
[26]: pd.Series(model.labels_).value_counts() # 数量统计
```

```
[26]: 1    11883
      2    10334
      0     7253
      3     1695
      4      136
      dtype: int64
```

```
[27]: data['客户类别'] = model.labels_ # 给原始数据添加聚类标签
```

绘制雷达图

```
[28]: # -*- coding: utf-8 -*-
import matplotlib.pyplot as plt
import numpy as np

plt.rcParams['font.sans-serif'] = ['SimHei'] # 用来正常显示中文标签
plt.rcParams['axes.unicode_minus'] = False # 用来正常显示负号

# 绘制雷达图，传入参数 1: model_center(聚类中心)，参数 2: label(特征名字)
def radarplot(model_center=None, label=None):
    n = len(label) # 特征个数
    # 对 labels 进行封闭，否则会有因为 matplotlib 版本引起的错误
    label = np.concatenate((label, [label[0]]))
```

间隔采样, 设置雷达图的角度, 用于平分切开一个圆面, *endpoint* 设置为 *False* 表示随机采样不包括 *stop* 的值

```
angles = np.linspace(0, 2 * np.pi, n, endpoint=False)
```

拼接多个数组, 使雷达图一圈封闭起来

```
angles = np.concatenate((angles, [angles[0]]))
```

创建一个空白画布

```
fig = plt.figure(figsize=(8, 8))
```

创建子图, 设置极坐标格式, 绘制圆形

```
ax = fig.add_subplot(1, 1, 1, polar=True)
```

添加每个特征的标签

```
ax.set_thetagrids(angles * 180 / np.pi, label)
```

设置 *y* 轴范围

```
ax.set_ylim(model_center.min(), model_center.max())
```

添加网格线

```
ax.grid(True)
```

设置备选折线颜色和样式, 防止线条重复

```
sam = ['r', 'o', 'g', 'b', 'm', 'y', 'k', 'p', 'c']
```

```
mak = ['4', '8', 'x', '*', 'd', '_', '.', '+', '|']
```

```
labels = []
```

循环添加每个类别的线圈

```
for i in range(len(model_center)):
```

```
    values = np.concatenate((model_center[i], [model_center[i][0]]))
```

```
    print(values)
```

```
    ax.plot(angles, values, c=sam[i], marker=mak[i])
```

```
    plt.yticks(fontsize=15)
```

```
    plt.xticks(fontsize=15)
```

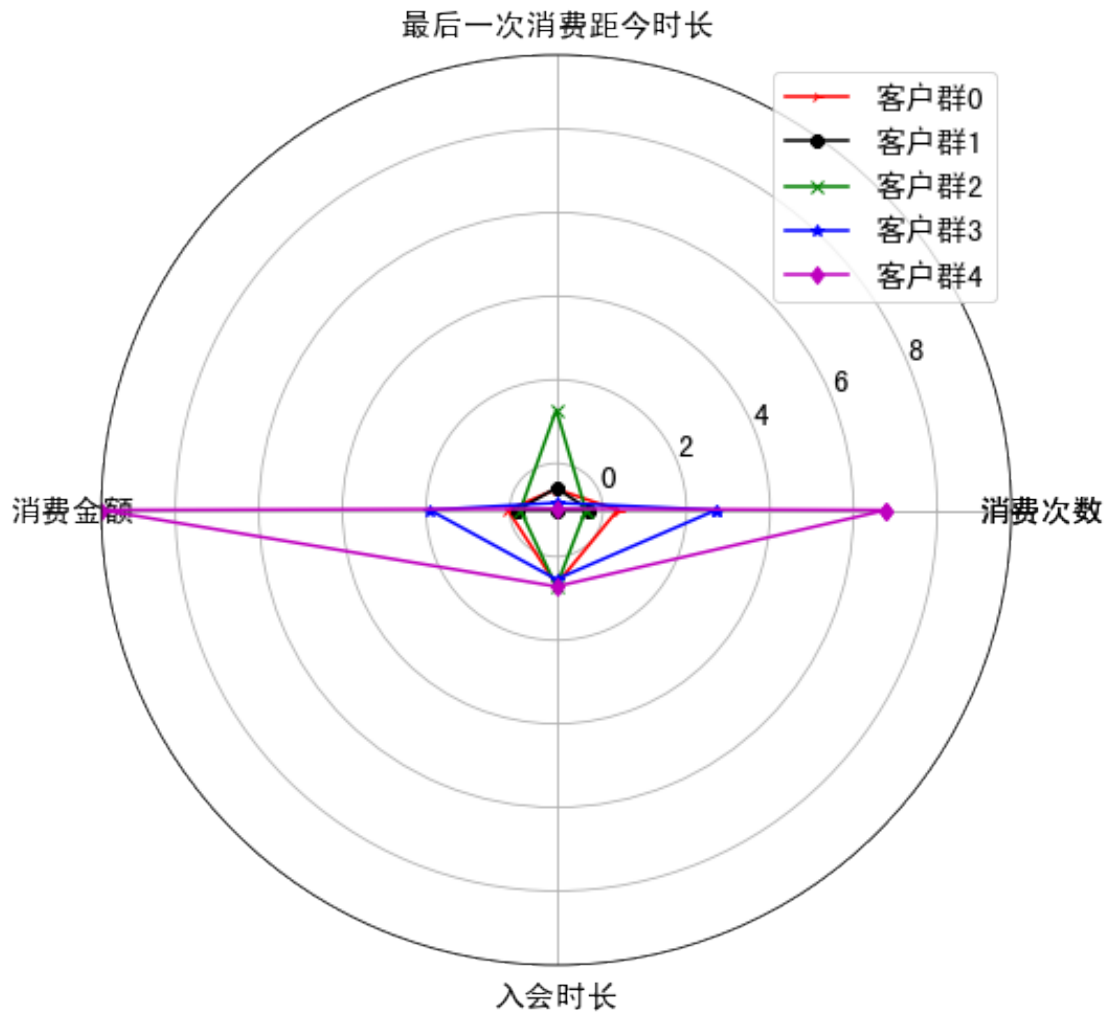
```
    labels.append('客户群' + str(i))
```

```
plt.legend(labels, fontsize=15)
```

聚类的雷达图

```
radarplot(model.cluster_centers_ , feature.columns)
```

```
[ 0.35405089 -0.61558198  0.04441617  0.67136318  0.35405089]
[-0.3221002  -0.58831476 -0.18918965 -1.11544549 -0.3221002 ]
[-0.41060379  1.27719539 -0.25501726  0.71403056 -0.41060379]
[ 2.70992706 -0.94293576  1.91191234  0.54151326  2.70992706]
[ 6.78682128 -1.08705261  9.7763573   0.71438857  6.78682128]
```



根据数据情况，选择使用 Kmeans 算法将客户分成 5 个类别。

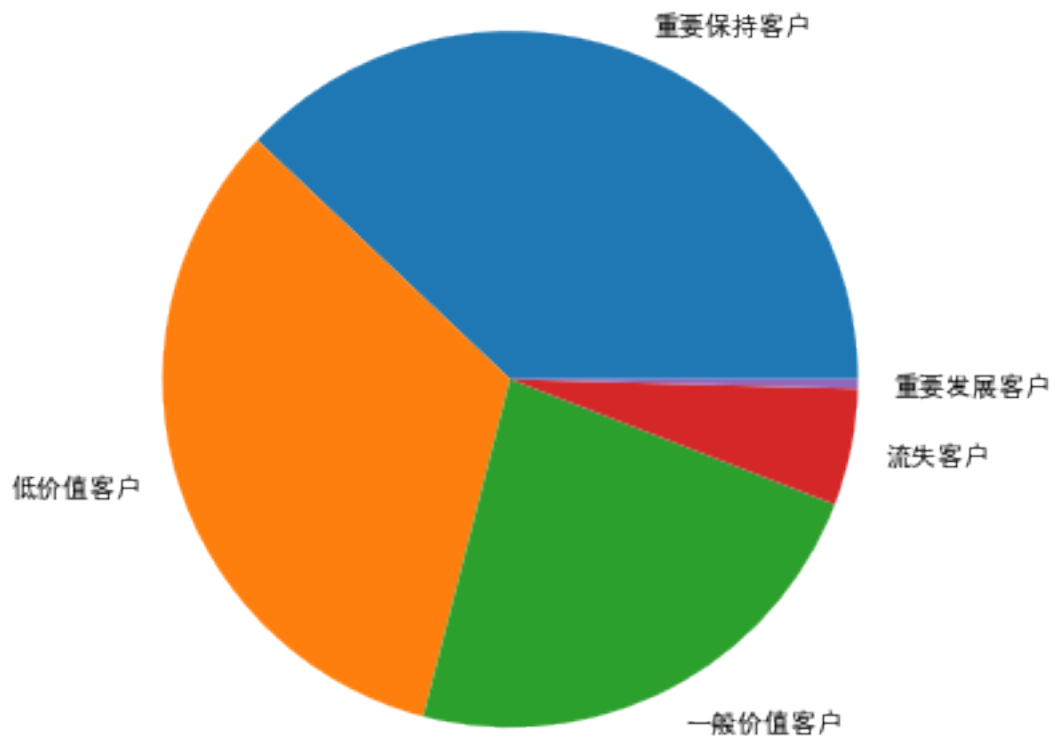
结合业务分析，通过比较各个特征在群间的大小对某一个群的特征进行评价分析，从而总结出每个群的优势和弱势特征。

会员管理方案

```
[50]: s = pd.Series(model.labels_).value_counts()
s = s/s.sum()
s.index = ['重要保持客户', '低价值客户', '一般价值客户', '流失客户', '重要发展客户']
s.name = '聚类类别'
s
```

```
[50]: 重要保持客户    0.379636  
      低价值客户    0.330149  
      一般价值客户    0.231718  
      流失客户      0.054152  
      重要发展客户    0.004345  
      Name: 聚类类别, dtype: float64
```

```
[49]: import matplotlib.pyplot as plt  
fig, ax = plt.subplots(figsize=(6,6))  
ax.pie(s, labels=s.index)  
plt.show()
```



根据对各个客户群进行特征分析，采取下面的一些营销手段和策略，为百货公司的客户管理提供参考。

低价值客户，购买力有限，入会一段时间后不再购买商品，可以通过广告、品牌折扣、类目更新等

勾起客户兴趣，制造品牌效应，增加互动拉新促活为主。

一般价值客户，刚入会的客户，对产品认知度不够，已经有一段时间未进行购买，跟低价值客户一样以增加互动拉新促活为主。

流失客户，此类客户购买力很低，购买频率不高，且数量不多，可以直接丢弃。

重要发展客户，处于新会员阶段，购买力中等，可以提供相似商品的优惠、服务等级提升、免费送货上门等方式，促进消费为主。

重要保持客户，处于新会员阶段，但购买力十足，应按照超级 VIP 的待遇进行管理，按周、按月发送当季新品、商品折扣、服务升级等，以增加用户粘性。