

基于K-Means的客户群聚类分析研究

陈新, 邓海生

西京学院计算机学院, 陕西 西安

收稿日期: 2025年4月9日; 录用日期: 2025年5月12日; 发布日期: 2025年5月21日

摘要

这篇论文研究了如何利用K-means聚类算法对客户进行细分, 以帮助营销团队进行精准营销。作者使用了百度飞桨的客户群聚类数据集, 其中包含性别、婚姻状况、年龄、教育、收入职业等客户信息。预处理数据后, 通过轮廓系数和SSE确定最佳K值, 最终将客户划分为高质量、中等质量和低质量三类。该研究提供了具体的客户细分结果, 对企业营销策略制定具有一定参考价值。

关键词

客户群, K-Means聚类算法, 数据挖掘, 数据分析

Customer Group Cluster Analysis Based on K-Means

Xin Chen, Haisheng Deng

School of Computing, Xijing University, Xi'an Shaanxi

Received: Apr. 9th, 2025; accepted: May 12th, 2025; published: May 21st, 2025

Abstract

This paper examines how the K-means clustering algorithm can be used to segment customers to help marketing teams with precision marketing. The authors used a customer cluster dataset from Baidu Feizuo, which includes customer information such as gender, marital status, age, education, income and occupation. After preprocessing the data, the optimal K value is determined by contour coefficient and SSE, and the customer is finally divided into high quality, medium quality and low quality. The research provides specific customer segmentation results, which has certain reference value for enterprises to formulate marketing strategies.

Keywords

Customer Group, K-Means, DM (Data Mining), Data Analysis



1. 引言

随着大数据时代的到来,企业亟需通过数据挖掘技术精准识别客户特征,以优化营销策略。K-means 作为经典的划分聚类算法,简洁、高效、快速的特点使其在图像分割、推荐系统、数据挖掘等领域有着广泛的应用[1]。本文利用 K-means 算法对客户数据进行聚类分析,根据客户属性自动将客户划分为若干群组,揭示隐藏的群体差异。通过计算欧氏距离和迭代优化聚类中心,实现客户群的自动分类。该方法不仅能够帮助企业识别高价值客户、潜在流失客户等关键群体,还能为个性化服务与精准营销提供数据支撑。通过轮廓系数或肘部法则评估聚类效果,最终输出具有明确特征的客户分群结果,助力企业提升资源配置效率与市场竞争力。

2. 相关研究

美国市场学家温德尔·史密斯于 1956 年提出“细分”概念,客户细分能够有效降低成本,营销策略是基于建立的客户价值细分模型。谢佳从应用需求出发,构建客户细分框架,并运用聚类分析、决策树、逻辑回归等数据挖掘方法对客户进行细分建模,为电力客户营销服务策略的制定提供依据。李艳君通过对八种不同聚类方法的对比研究,对辽宁省某商业银行的 CRM 结算系统客户细分,得到了五种分类,该细分模型表明银行 80% 的收益是由 20% 比较重要的客户带来的,数据支撑精准营销策略[2]。

对此,目前有很多关于使用 K-means 聚类算法进行客户群体的划分,例如中国计量大学经济与管理学院的靖立峥和吴增源采用基于改进 K-means 算法的电子商务客户细分研究[3],湖南师范大学计算机教学部和中南大学商学院的向坚持、刘相滨、资武成发表的基于密度的 K-Means 算法及在客户细分中的应用研究[4]。都是基于 K-means 进行客户群体的划分。

3. 相关技术

3.1. K-Means 聚类算法描述

K-means 聚类算法是一种经典的无监督学习算法,用于将数据集划分为 K 个互相不重叠的簇。

一般用欧式距离计算衡量两个数据之间的距离,作为相似度。距离越大,相似度越小。算法的核心是确定最佳 K 值作为划分的簇的数目,根据数据点和聚类中心即质心的距离(欧式距离),不断迭代,直到质心位置不再发生变化,从而达到最大迭代次数,即簇内差异小,簇间差异大。

K-means 算法步骤:

1. 初选 k 个凝聚点。
2. 计算每个数据点到各个凝聚点的距离。
3. 进行初始分类。
4. 计算每类的重心,并以此重心作为新的凝聚点。
5. 判断新分类的结果与前一步分类结果是否相同,若不同则按照步骤 4、步骤 5 进一步调整分类。

上述迭代必然收敛,即每个数据点到自身所归属类的重心的距离平方和(SSE)是收敛的。

K-means 聚类算法的优点是对大数据集合聚类效果明显,聚类快速且易于实现[5]。

K-means 聚类算法的缺点是需要预先制定 K 值,对初始质心敏感(可能收敛点局部最优),对噪声和

离群点敏感。

3.2. Pearson 相关系数

Pearson 相关系数(Pearson Correlation Coefficient)是统计学中衡量两个连续变量之间线性关系强度和方向的指标, 是两个变量之间线性关系强度的度量, 它的取值范围在-1 到 1 之间[6]。

$r = 1$: 完全正线性相关(变量同增同减)。

$r = -1$: 完全负线性相关(一个增另一个减)。

$r = 0$: 线性相关性(但可能有其他非线性关系)。

对于两组数据计算公式如下:

$X = \{x_1, x_2, \dots, x_{n1}\}$ 和 $Y = \{y_1, y_2, \dots, y_2\}$, Pearson 相关系数的计算公式为:

$$r = \frac{\text{协方差}(X, Y)}{\text{标准差}(X) * \text{标准差}(Y)} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} * \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

其中:

\bar{x} 和 \bar{y} 分别是 X 和 Y 的均值。

3.3. 轮廓系数

轮廓系数, 即 SC 指标, 表示聚类后各类样本间的紧密程度和各类之间的离散程度。同一类中样本间的距离越小, 异类间样本距离越大, 则 SC(i) 的值越大, 聚类效果越好[7]。是衡量簇内紧密度和簇间分离度。

1986 年, Rousseeuw 提出轮廓系数(silhouette coefficient, SC)的概念, 利用轮廓系数可以评价数据划分结果的好坏, 也可以评价各种算法对同一组数据划分的优劣效果, 轮廓系数主要是利用数据间的结合聚合程度和分离程度两种因素来进行评价计算。

计算样本 i 轮廓系数的计算公式:

$$SC(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

其中:

$a(i)$ 表示样本 i 到同簇内所有其他样本的平均距离, 反映该样本与同簇其他点的相似程度。

$b(i)$ 表示样本 i 到其他所有簇的平均距离的最小值, 反映该样本与其他簇的分离程度。

计算整个数据集的轮廓系数公式:

$$\text{SilhouetteScore} = \frac{1}{N} \sum_{i=1}^N s(i)$$

轮廓系数 SC(i) 的值在[-1, 1] 区间内变化, 当轮廓系数的值接近 1 的时候, 表明数据间的结合聚合程度和分离程度都比较好, 此时的划分效果最好。

3.4. SSE

SSE (Sum of Squared Errors, 平方误差和), 也称为簇内平方和, 是 K-means 聚类中最核心的评估指标, 用于衡量聚类结果的紧密程度。

SSE 计算的是所有数据点与其所属簇中心(质心)的欧氏距离的平方和, 公式如下:

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} \|x - \mu_i\|^2$$

其中:

K 是簇的数量。

C_i 是第 i 个簇的所有数据点。

μ_i 是第 i 个簇的质心。

$\|x - \mu_i\|$ 是数据点 x 到质心的欧氏距离。

SSE 结果:

SSE 越小, 说明簇内数据点越紧密, 聚类效果越好。

SSE 越大, 说明数据点离簇中心较远, 聚类效果较差。

SSE 优点:

计算简单, 适用于 K-Means 等基于距离的聚类算法。

直观, SSE 越小, 聚类效果越好。

适用于大数据集(计算效率较高)。

SSE 缺点:

必须预先设定 K , 无法自动确定最佳簇数。

对异常值敏感(离群点会显著增加 SSE)。

假设簇是凸形, 对非球形簇效果不佳(如流形数据)。

SSE 会随 K 增大而单调递减(极端情况 $K = N$ 时 $SSE = 0$, 但无意义)。

3.5. 手肘法

手肘法(Elbow Method)是确定 K-means 聚类算法中最优簇数(K 值)的常用方法, 它通过分析不同 K 值对应的 SSE (Sum of Squared Errors, 平方误差和)变化趋势来寻找最佳聚类数量。在实际选择不同的 K 值时, SSE 会随 K 增大而单调递减, 样本会随着聚类 K 的增大, 每个簇的聚合度会逐渐提高, SSE 的值也会随着下降。但是, 随着 K 值的逐渐增大, 在没有达到最佳 K 值之前, SSE 的下降率比较明显, 直到随着 K 值的逐渐增大, 达到最佳 K 值之后, SSE 的下降率会平缓, 形成一个手肘状, 这里的手肘点就是最佳 K 值。

优点:

简单直观, 易于实现、计算效率高、适用于初步分析。

缺点:

肘部点有时不明显(“平缓肘部”问题)、对非球形簇效果不佳、依赖主观判断。

4. 数据处理与建模

4.1. 数据信息

本数据来源于百度飞桨数据集 - 客户群聚类数据集, 该数据集包含有关客户的信息, 如性别、婚姻状况、年龄、教育、收入职业等。客户群字段信息如表 1 所示。

Table 1. Customer group field information

表 1. 客户群字段信息

字段名	字段说明
ID	客户 id
Sex	性别 0: male 1: female

续表

Marital status	婚姻状态: 0: single 1: non-single (divorced / separated / married / widowed)
Age	年龄 0: other / unknown
Education	1: high school 2: university 3: graduate school
Income	收入 0: unemployed / unskilled
Occupation	1: skilled employee / official 2: management / self-employed / highly qualified employee / officer
Settlement size	0: small city 1: mid-sized city 2: big city

4.2. 计算变量之间相关性

利用 Pearson 相关系数方法计算性别、年龄、学历、职业、收入、定居城市之间的相关性。方便后面模型建立, 以及剔除不必要的变量。字段相关性信息如图表 2 所示。

Table 2. Field correlation information

表 2. 字段相关性信息

	Sex	Marital status	Age	Education	Income	Occupation	Settlement size
Sex	1	0.566510673	-0.182884799	0.24483756	-0.195146423	-0.202491179	-0.300803219
Marital status	0.566510673	1	-0.21317835	0.374016539	-0.073527728	-0.029489581	-0.097040603
Age	-0.182884799	-0.21317835	1	0.654604987	0.340610164	0.108388053	0.119751291
Education	0.24483756	0.374016539	0.654604987	1	0.233458647	0.064523756	0.034732304
Income	-0.195146423	-0.073527728	0.340610164	0.233458647	1	0.680357396	0.490880929
Occupation	-0.202491179	-0.029489581	0.108388053	0.064523756	0.680357396	1	0.571794844
Settlement size	-0.300803219	-0.097040603	0.119751291	0.034732304	0.490880929	0.571794844	1

由 Pearson 相关系数, 我们得出结论, 性别和婚姻状态具有高度相关性, 收入和年龄、教育水平、职位、居住城市具有较强的相关性。为此, 我们在初步建模后, 依次排除了性别和婚姻状态下的客户群聚类分析。

4.3. 数据预处理

客户群数据存在离散数据例如性别、年龄、学历、职业、定居城市, 以及缺失值。

对于缺失数据, 进行删除处理。

对于性别男女进行数值转换, 即将男设置为 0, 女设置为 1。

将学历进行数值转换, 即 0:other/unknown 1:high school 2:university 3: graduate school

将职业进行数值转换, 即 0:unemployed/sunskilled 1:skilled employee/official 2:management/self-employed/high qualified employee

将居住城市大小进行数值转换, 即 0:small city 1:mid-sized city 2:big city

将年龄进行区间量化, 即划分为(1): 小于 20 岁; (2): 20~25 岁; (3): 25~30 岁; (4): 30~35 岁; (5):

35~40 岁; (6): 40~45 岁; (7): 45~50 岁; (8): 50~55 岁; (9): 55~60 岁; (10): 大于 60 岁。具体数据如下图所示。具体数据如图 1 客户年龄段数量所示。

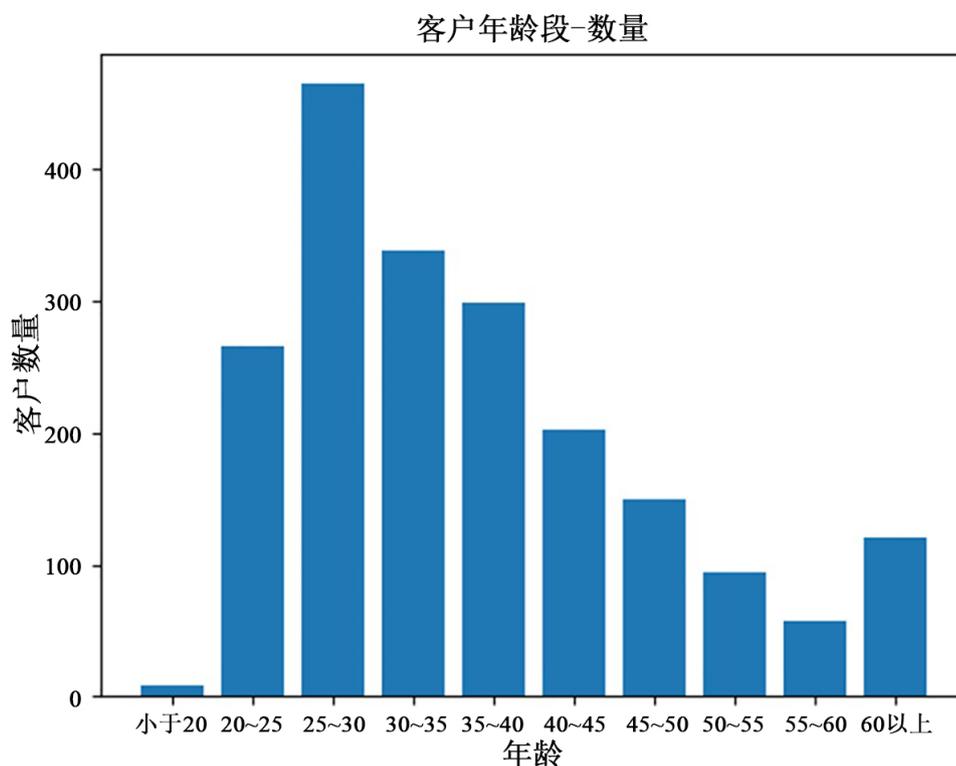


Figure 1. Number of customer segments divided by year
图 1. 客户年龄段数量

4.4. 归一化处理

由于收入和其他字段特征的量纲差距过大, 通过归一化处理, 使不同特征尺度的数据转换到相同的尺度范围内, 解决特征尺度差异问题、加速模型收敛、提高模型精度。

Min-Max (最小 - 最大标准化) 是一种线性变换的方法, 将原始数据线性映射到指定范围内[0-1]。具有线性变换不改变数据间的相对大小和分布形状, 仅压缩或拉伸到目标范围内的优势, 所以, 本次客户群聚类分析采用 Min-Max (最小 - 最大标准化) 进行归一化处理。公式如下。

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

X : 原始数据

X_{min} : 该特征的最小值

X_{max} : 该特征的最大值

X_{norm} : 归一化后的数据 (默认[0, 1]区间)

4.5. 最佳 K 值确定

这里通过迭代在不同的 K 值的情况下, 计算 SSE 和轮廓系数在不同 K 值情况下的结果, 并绘制折线图, 通过手肘法确认最佳 K 值。

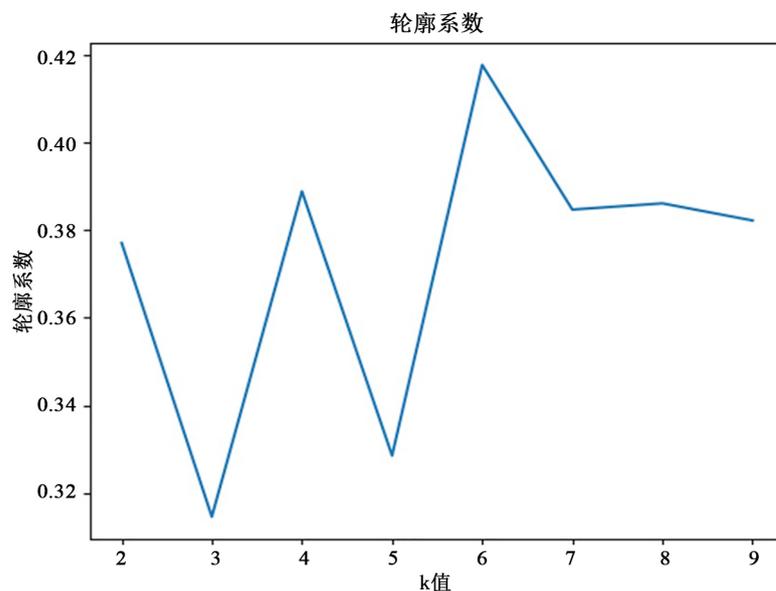


Figure 2. Contour coefficient line diagram

图 2. 轮廓系数折线图

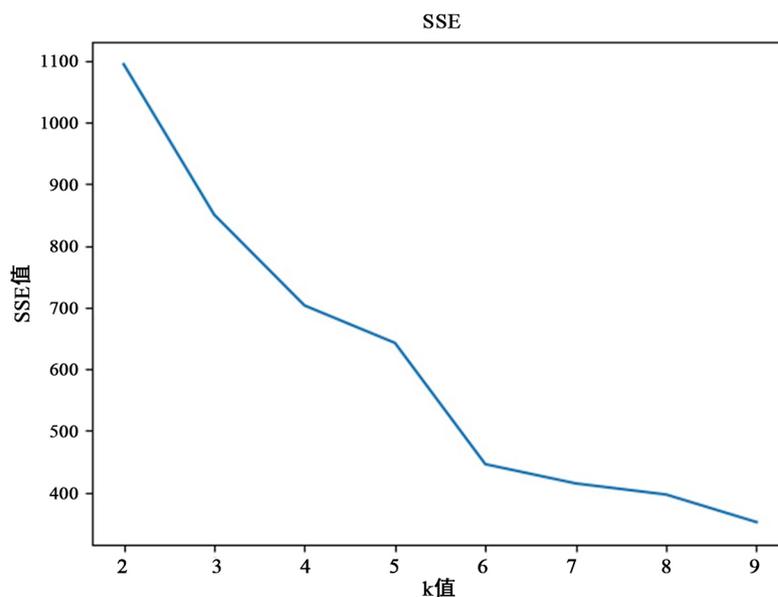


Figure 3. SSE- in-cluster error square sum line graph

图 3. SSE-簇内误差平方和折线图

由图 2 轮廓系数折线图和图 3 SSE-簇内误差平方和折线图可确定, 在 $K = 6$ 时候, 轮廓系数达到最大, SSE 下降率逐渐走向平缓, 因此, 我们可以找到最佳 K 值是 6。

4.6. 建立 K-Means 聚类模型

设置最佳 $K = 6$, 采用 sklearn 库中默认值: 初始化方法(init)为 K-means++, 质心初始化值(n_init)为 10, 最大迭代数(max_iter)为 300, 随机从训练数据中选取初始质心(random_state)为 0。通过训练得出聚类模型, 并以模型为基础计算出预测值。

求的最佳 $SSE = 445.5410560361939$ 。

5. 结果分析

5.1. 初步结果分析

统计各个簇内数据, 各个簇内数据百分比分布如下图 4 所示:

cluster	sex		Marital status		Age (平均)	Education				Income (平均)	Occupation			Settlement size		
	male	female	single	non-single		other/ unknown	high school	university	graduate school		unemployed unskilled	skilled employee official	management self-employed highly qualified employee/officer	small city	mid-sized city	big city
0 簇		100%		100%	34.63		70.97%	23.87%	5.16%	136594.08	1.94%	79.35%	18.71%	1.61%	61.94%	36.45%
1 簇	100%		100%		37.57	33.23%	59.74%	7.03%		102782.32	64.22%	34.50%	1.28%	93.93%	6.07%	
2 簇		100%	100%		35.12	17.98%	71.91%	8.99%	1.12%	97997.24	64.05%	34.83%	1.12%	92.70%	7.30%	
3 簇	100%			100%	33.97		75.88%	21.40%	2.72%	122976.72	23.35%	59.92%	16.73%	38.52%	31.13%	30.35%
4 簇	100%		100%		40.01	29.26%	59.30%	11.44%		145325.6	1.94%	69.57%	28.49%	0.19%	46.32%	53.49%
5 簇		100%		100%	32.15		82.16%	15.26%	2.58%	101777.62	56.81%	43.19%		99.77%	0.23%	

Figure 4. Percentage of data within each cluster

图 4. 各个簇内数据百分比

由上述图表可知:

将客户群体按照收入、居住城市、学历、职位可划分为高质量、中等质量、低质量的三类客户群体。

0 簇内客户群主要特征为女性、已婚、学历为高中为主, 大学为次、工作是大部分为职员和管理者, 居住在大型城市和中型城市, 是属于高质量客户群体。

1 簇内客户群主要特征为男性、未婚、低学历、工作中下层职位, 居住在小型城市, 是属于低质量客户群体。

2 簇内客户群主要特征为女性、已婚、低学历, 工作中下层职位, 居住在小型城市, 是属于低质量客户群体。

3 簇内客户群主要特征为男性、已婚、高中学历为主, 大学学历为次、工作主要为中层的, 部分为高层和下层职位的, 平均居住在大中小型城市, 是属于中等质量的客户群体。

4 簇内客户群主要特征为男性、未婚、学历为高中为主, 工作中上层的职位, 主要居住在大型城市和中型城市, 具有较高收入, 属于高质量客户群体。

5 簇内客户群体主要特征为女性、已婚、高中学历为主、大学学历为次, 工作主要为中下层职位, 基本居住在小型城市, 具有可观收入, 属于中等质量客户群体。

5.2. 不区分性别划分客户群

由上述初步结果分析, 可知, 性别分布在低质量客群、中等质量客群、高质量客群中具有对称性, 因此, 我们在上一步结果的基础上, 进行排查变量年龄进行划分簇。

通过迭代遍历不同 K 值下的轮廓系数和 SSE (簇内平方和), 由轮廓系数和 SSE (簇内平方和) 的走势。

由图 5 不区分性别下的轮廓系数和 SSE 簇内误差平方和折线图, 我们确认此时的最佳 K 值是 4, 在相同条件下, 设置最佳 $K = 4$, 初始化方法(imit)为 K-means++, 质心初始化值(n__init)为 10, 最大迭代数(max__iter)为 300, 随机从训练数据中选取初始质心(random__state)为 0。通过训练得出聚类模型, 并以模型为基础计算出预测值。

求的最佳 $SSE = 401.34229607585337$, 相比初步结果, SSE 小幅度减少, 分类更加精准。

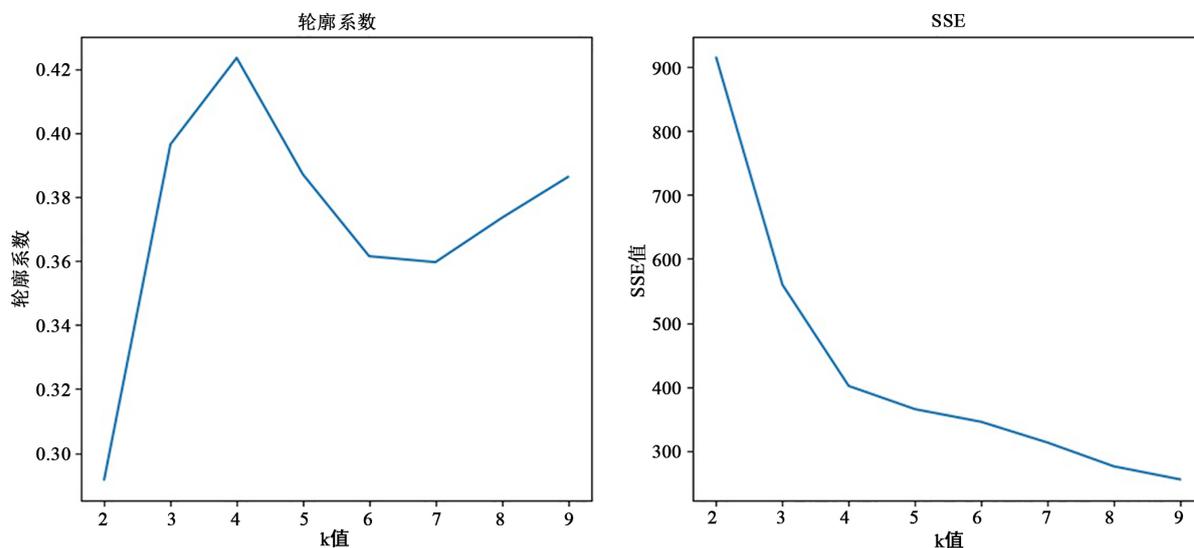


Figure 5. Square and line plots of contour coefficients and SSE cluster errors without distinction of gender

图 5. 不区分性别下的轮廓系数和 SSE 簇内误差平方和折线图

统计各个簇内数据, 各个簇内数据百分比分布如图 6 不区分性别各个簇内数据百分比所示。

cluster	Marital status		Age (平均)	Education			Income (平均)	Occupation			Settlement size		
	single	non-single		other/ unknown	high school	university		graduate school	unemployed unskilled	skilled employee official	management self-employed highly qualified employee officer	small city	mid-sized city
0 簇	100%		39.85	28.44%	60.26%	11.30%	144850.25	1.88%	69.68%	28.44%	0.56%	47.46%	51.98%
1 簇		100%	34.85		70.41%	24.62%	136506.53	2.81%	76.89%	20.30%	0.65%	58.10%	41.25%
2 簇	100%		36.76	28.57%	63.24%	7.77%	100182.57	66.18%	33.40%	0.42%	96.01%	3.99%	
3 簇		100%	32.12		82.83%	15.09%	102082.88	55.66%	43.02%	1.32%	99.25%	0.75%	

Figure 6. The percentage of data within each cluster is not differentiated by sex

图 6. 不区分性别各个簇内数据百分比

由上图表不区分性别各个簇内数据百分比数据可知, 将客户群体按照收入、居住城市、学历、职位可划分为高质量、低质量的二类客户群体。

0 簇内客户群体主要特征为未婚、学历主要以高中为主、工作为中上层职位, 居住咋中型城市 and 大型城市, 具有最高的收入水平, 属于高质量客户群体。

1 簇内客户群体主要特征为已婚、学历主要以高中为主, 大学为次、工作为中上层职位, 居住在中型城市 and 大型城市, 具有相对较高的收入水平, 属于高质量客户群体。

2 簇内客户群体主要特征为未婚、学历主要以高中为主, 高中以下为次, 工作中下层, 主要居住在小型城市, 具有较低的收入水平, 属于低质量客户群体。

3 簇内客户群体主要特征为已婚、学历主要以高中为主, 大学为次、工作中下层, 主要居住在小型城市, 具有较低的收入水平, 属于低质量客户群体。

由上述数据可知, 我们认为排除掉婚姻状态, 依旧可以划分客户群体

5.3. 不区分性别和婚姻状态划分客户群

在排除性别和婚姻状态的变量后, 通过迭代遍历不同 K 值下的轮廓系数和 SSE (簇内平方和), 由轮廓系数和 SSE (簇内平方和) 的走势。

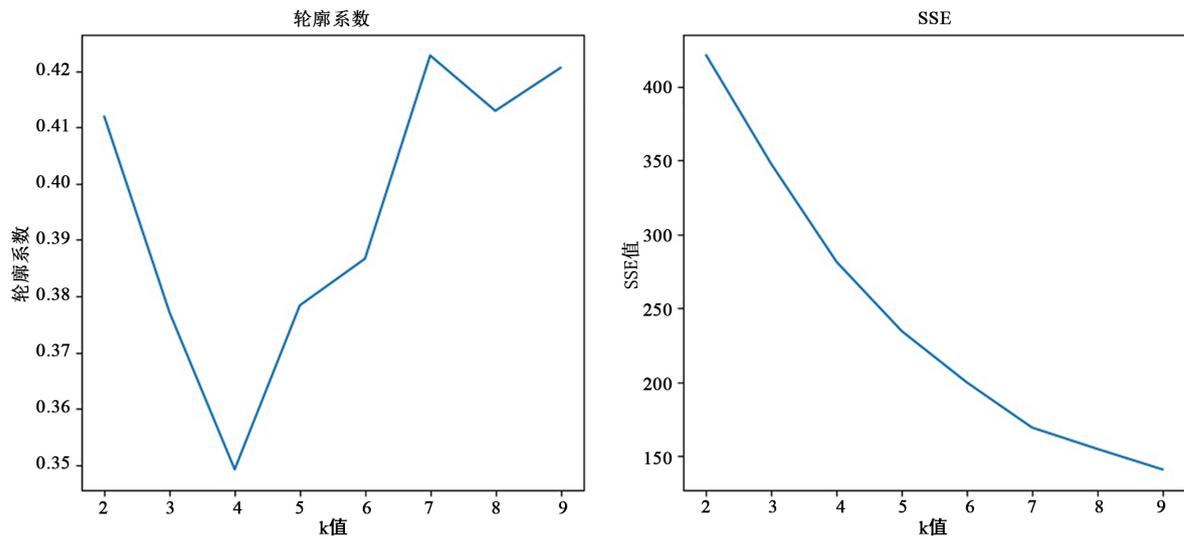


Figure 7. Square and line plots of contour coefficients and SSE cluster errors without distinction of gender
图 7. 不区分性别下的轮廓系数和 SSE 簇内误差平方和折线图

由图 7 不区分性别下的轮廓系数和 SSE 簇内误差平方和折线图, 我们确认此时的最佳 K 值是 7, 在相同条件下, 设置最佳 K = 4, 初始化方法(init)为 K-means++, 质心初始化值(n_init)为 10, 最大迭代数(max_iter)为 300, 随机从训练数据中选取初始质心(random_state)为 0。通过训练得出聚类模型, 并以模型为基础计算出预测值。

求的最佳 SSE = 169.3950010420585, 相比初步结果, SSE 大幅度减少, 分类更加精准。

统计各个簇内数据, 各个簇内数据百分比分布如图 8 不区分性别和婚姻状态各个簇内数据百分比所示。

cluster	age (平均)	Education				Income (平均)	Occupation				Settlement size		
		other /unknown	high school	university	graduate school		unemployed unskilled	skilled employee official	management highly qualified employee officer	self-employed	small city	mid-sized city	big city
0 簇	31.33	12.74%	85.64%	1.62%		117571.56		97.02%	2.98%		100.00%		
1 簇	56.91		14.29%	77.25%	8.46%	161786.49	2.12%	71.43%	26.46%		2.12%	61.38%	36.50%
2 簇	30.29	19.43%	80.57%			84396.25	100.00%				100.00%		
3 簇	32.74	22.86%	74.29%	2.14%	0.71%	126680.58	5.36%	94.64%					100.00%
4 簇	31.17	17.73%	82.27%			123974.69	5.23%	94.77%				100.00%	
5 簇	52.52		23.35%	68.86%	7.79%	110043.97	82.63%	17.37%			94.61%	5.39%	
6 簇	36.1	13.47%	74.61%	9.33%	2.59%	169940.96			100.00%			38.86%	61.14%

Figure 8. The percentage of data within each cluster does not distinguish between sex and marital status
图 8. 不区分性别和婚姻状态各个簇内数据百分比

由上图表不区分性别和婚姻状态各个簇内数据百分比数据可知, 将客户群体按照收入、居住城市、学历、职位可划分高质量、中等质量、低质量的三类客户群体。

0 簇客户群体主要特征为学历以高中为主、工作中层、居住在小型城市、具有中等收入的群体, 属于中等质量客户群体。

1 簇客户群体主要特征为学历以大学及以上学历为主, 工作在高层、居住在大型城市和中型城市、具有最高收入的群体, 属于高质量群体。

2 簇客户群体主要特征为高中学历及以下, 工作在底层、居住在小型城市、具有最低收入, 属于低质量客户群体。

3 簇客户群体主要特征为高中学历为主, 工作中层、居住在大型城市、具有中等收入的群体, 属于中等质量客户群体。

4 簇客户群体主要特征为高中学历为主, 工作中层、居住在中型城市、具有中等收入的群体, 属于中等质量客户群体。

5 簇客户群体主要特征为高中学历为主, 工作中层、居住在小型城市、具有中等收入的群体, 属于中等质量客户群体。

6 簇客户群体主要特征为高中及以上学历, 全部工作在高层, 居住在大型和中型城市、具有较高收入水平的客户群体, 属于高质量客户群体。

6. 总结

本文通过逐步去除变量进行聚类分析, 从而大致了解到客户群体在不同性别、婚姻状态下客户群体的特征信息, 充分分析和挖掘了客户群体信息, 研究结果可以帮助营销部门进行客户精准推送, 扩展市场, 提供了有力的支撑和帮助。

数据源链接地址

本数据来源与百度飞桨: <https://aistudio.baidu.com/datasetdetail/107018>。

参考文献

- [1] 王义武, 杨余旺. 空间投影在 K-means 算法中的研究与应用[J]. 计算机工程与应用, 2020, 56(7): 200-204.
- [2] 刘坤达, 宋红文, 张卫东, 等. 品牌优先的零售客户细分研究[J]. 现代商贸工业, 2021, 42(32): 45-46.
- [3] 靖立峥, 吴增源. 基于改进 K-means 算法的电子商务客户细分研究[J]. 中国计量大学学报, 2020, 31(4): 482-489.
- [4] 向坚持, 刘相滨, 资武成. 基于密度的 K-Means 算法及在客户细分中的应用研究[J]. 计算机工程与应用, 2008, 44(35): 246-248.
- [5] 王勇, 唐靖, 饶勤菲, 等. 高效率的 K-means 最佳聚类数确定算法[J]. 计算机应用, 2014, 34(5): 1331-1335.
- [6] 李彩虹, 刘聪, 宋勇, 等. 基于 Lü 系统的移动机器人完成特殊情况下全覆盖路径规划的参数值选择策略(英文) [J]. 信息与电子工程前沿(英文版), 2023, 24(2): 231-245.
- [7] 马鑫, 段刚龙, 王建仁, 等. 基于改进轮廓系数法的航空公司客户分群研究[J]. 运筹与管理, 2021, 30(1): 140-146.