



## 基于改进 SSA-Kmeans 算法的网站用户分群研究

雷雨霄

(中移物联网有限公司, 重庆 401121)

**摘要:** 在数字化时代, 精准把握用户需求是网站实现精准营销与个性化服务的关键。针对该场景, 提出了一种改进的奇异谱分析  $K$  均值 (singular spectrum analysis  $K$ -means, SSA- $K$ means) 算法, 有效地解决了传统 SSA 中窗口长度及特征分量须手动选取的低效问题。该算法先利用 SSA 提取用户访问数据的核心特征分量, 再结合  $K$ -means 进行聚类分析。实验结果表明, 该算法显著地提升了聚类效果, 戴维斯堡丁指数较直接聚类和小波变换去噪后聚类分别降低了 0.407 1 和 0.067 2, 簇划分更精准。基于优化后的聚类结果, 进一步制定了差异化运营策略, 针对不同用户群体提供定制化服务。这一方法为网站精准营销和用户留存提供了高效的解决方案, 具有重要的实践应用价值。

**关键词:** 用户分群; 奇异谱分析;  $K$  均值聚类算法; 戴维斯堡丁指数

**中图分类号:** TP391; TN-9

**文献标志码:** A

**doi:** 10.11959/j.issn.1000-0801.2026023

## Research on website user clustering based on improved SSA- $K$ means algorithm

Lei Yuxiao

China Mobile Internet of Things Co., Ltd., Chongqing 401121, China

**Abstract:** In the digital age, accurately grasping user needs is the key to achieve accurate marketing and personalized services. Aiming at this scenario, an improved singular spectrum analysis  $K$ -means (SSA- $K$ means) algorithm was proposed, which effectively solved the inefficient problem of manual selection of window length and feature components in the traditional SSA. Firstly, SSA was used to extract the core feature components of user access data in this algorithm. Then,  $K$ -means was combined for cluster analysis. The experimental results show that the clustering effect is significantly improved. The Davies-Bouldin index is 0.407 1 and 0.067 2 lower than that of direct clustering and wavelet transform denoising clustering, respectively. Moreover, the cluster division is more accurate. Based on the optimized clustering results, a differentiated operation strategy was further developed to provide customized services for different user groups. This method provides an efficient solution for website precision marketing and user retention, and has important practical application value.

**Key words:** user clustering, SSA,  $K$ -means clustering algorithm, Davies-Bouldin index

收稿日期: 2025-06-18; 修回日期: 2025-10-04

通信作者: 雷雨霄, leiyuxiao@cmiot.chinamobile.com

## 0 引言

近年来,电子商务与互联网信息服务产业蓬勃兴起,宛如强劲引擎,推动网络服务全方位融入大众生活。从日常购物到信息检索,这些产业为大众带来了前所未有的便捷,极大地提升了用户生活效率与品质。随着用户规模的急剧扩张,不同用户群体间的需求差异愈发显著。如何精准捕捉这些差异,为用户提供个性化服务,成为当下亟待解决的关键问题。用户行为模式与兴趣偏好紧密相连,这种内在关联通常通过具体行为得以展现。为深入洞察用户访问电商及互联网信息服务平台的潜在需求与兴趣倾向,业界普遍运用数据挖掘技术,从海量数据中提炼用户访问特征。基于这些特征实施个性化运营策略,能更契合用户需求,进而显著地提升服务效能,满足用户需求<sup>[1]</sup>,增强用户黏性。由此可见,在电商与互联网产业高速发展的当下,用户分群是提升用户体验和销售效率的关键<sup>[2]</sup>。

当前,针对用户分群的研究已成为众多学者和从业者关注的热点领域。苏舒菲和蔺聪<sup>[3]</sup>采用 $K$ 均值( $K$ -means)聚类算法来识别在线评论中的产品需求偏好及客户群组模型,并依据群组特点精准地定位产品,以满足消费者需求。葛青龙<sup>[4]</sup>采用改进 $K$ -means聚类对电商平台的用户行为数据进行分析,以此来评估用户的满意度情况。杨泽丞等<sup>[5]</sup>采用聚类方法在用户侧电流数据中发现用电事件,挖掘用户的用电行为。戴圆圆<sup>[6]</sup>通过对用户需求的分类研究,剖析了不同用户群体的需求特点,进而精准地实施差异化营销策略。徐良军等<sup>[7]</sup>结合 $K$ -means聚类和模糊 $C$ 均值聚类算法,深入分析用户的用电行为模式,助力电力公司更精准地把握用户用电状况。上述研究表明,聚类算法能够有效地对用户群体进行划分,为本文的研究提供了坚实的理论支撑。然而,在网站用户访问数据中,网络波动、异常用

户行为等因素,会不可避免地产生噪声数据。如何有效地抑制这些噪声数据,成为亟待解决的关键问题。

## 1 网站用户分群算法概述

### 1.1 奇异谱分析算法引入

奇异谱分析(singular spectrum analysis, SSA)<sup>[8-10]</sup>常用于处理一维时间序列,能够有效地将时间序列分解为趋势分量、振荡分量和噪声分量,进而可以实现原始时间序列中不同成分信号的提取。由于网站的用户访问数据通常包含噪声,这些噪声可能来自网络波动、异常用户行为等。通过SSA分析可以将含有噪声的用户访问数据分解成一系列特征分量,得到的若干特征分量中有一些主要包含用户访问行为特征的有用信息,有一些主要包含噪声的无用信息,通过提取含有用户访问行为特征较多的分量,实现噪声的抑制,进而更准确地反映用户的真实访问模式和趋势。

奇异谱分析中,各分量的信息权重可通过对应的奇异值大小进行量化表征。具体而言,较大的奇异值通常反映原始用户访问序列中显著的确定性模式,而较小的奇异值通常与噪声扰动相关联。基于该特性,可通过截断较小奇异值对应的分量来实现数据降噪,但在实际应用中,奇异值谱呈现单调递减趋势并逐渐收敛于零,除前几个显著较大的奇异值外,后续分量是否具有保留价值仍存在选择困难。

### 1.2 $K$ -means聚类算法引入

考虑到网站用户访问数据通常具有复杂的分布特性,可能不符合某些特定的分布假设, $K$ -means算法通过计算数据到簇中心的距离来完成数据的划分聚类,能够较好地适应这种复杂的数据分布情况。它不需要依赖历史数据,便能挖掘出数据中的内在结构联系和特征区别<sup>[11]</sup>,并通过迭代优化的方式找到合适的簇划分,从而在一定



程度上保证了聚类结果的准确性,是一种广泛用于大规模数据聚类的简单高效算法<sup>[12]</sup>。而K-means算法中K的选取会影响聚类类别的数量<sup>[13]</sup>,可能会使算法陷入局部最优解,进而影响聚类效果,所以需要选择合适的K值保证较好的聚类效果。

### 1.3 改进SSA-Kmeans算法

由第1.1节和第1.2节分析可知,在使用奇异谱分析对网站用户访问数据的时间序列进行分解时,分解前选择合适的窗口长度 $L$ 和分解后选择合适特征分量是实现最佳分解效果的关键,而分解效果也决定了后续聚类效果的好坏。如果选择的窗口长度过大,则会增加计算复杂度;如果选择的窗口长度过小,则难以有效地提取主成分。对于奇异谱分解后的特征分量选取,若选取过多,可能会引入大量的无用信息;若选取过少,则可能会遗漏部分有价值的信息。在K-means聚类中,选取合适的K值至关重要。若K值过大,会导致原本属于同一簇的用户被拆分为多个簇;若K值过小,则会使簇的分类不够充分。为了解决上述问题,本文提出了一种改进的SSA-Kmeans算法。首先,在SSA阶段,通过设置极大特征值相对变化率阈值的方法来选取合适的窗口长度 $L$ ,为后续奇异谱分解奠定基础。确定合适的窗口长度后,进行奇异谱分解。分解完成后,本文提出进一步采用斯皮尔曼相关系数变化曲线与特征值变化曲线相结合的方法来选取合适的特征分量,为了使选取的特征分量中不同特征在相同的尺度上均衡,在输入K-means聚类算法前进行Z-score标准化处理,对处理后的数据再进行聚类。本文采用手肘法<sup>[14]</sup>解决K-means算法中K值的选取问题,最终实现对用户访问数据的分群。基于SSA-Kmeans算法进行用户分群的流程如图1所示。

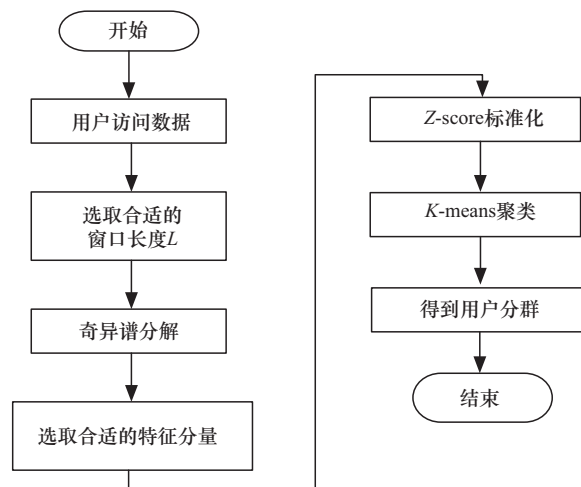


图1 基于SSA-Kmeans算法进行用户分群的流程

## 2 实证分析

### 2.1 实验数据获取

本文使用某电商平台公开的用户访问量数据,选取了该数据集中2014年11月17日至2014年12月17日31天内的用户访问量,该数据集记录了31天中每小时的用户访问量,因此该数据集共有744个数据点。

### 2.2 基于奇异谱分析的噪声抑制

经过第1.3节的分析,通过奇异谱分析进行噪声抑制,其核心步骤为:先确定合适的窗口长度,再据此进行SSA分解,进而获取若干特征分量,挑选含有有用信息较多、噪声较少的特征分量输入至K-means聚类算法,最终得出用户群体的聚类结果。针对窗口长度的确定,本文提出以极大特征值相对变化率阈值为依据进行选取;对于特征分量的筛选,本文提出了结合特征值变化曲线与斯皮尔曼相关系数变化曲线的综合评判方法。

#### 2.2.1 窗口长度的选择

随着窗口长度的增加,对该电商的用户访问数据进行奇异谱分解后的极大特征值虽会增大,但其增长幅度会逐渐减小。为了尽可能保证奇异谱分解的彻底性,以便更好地观察极大特征值相对变化率,本文优先选取一个较大的窗口长度

$L=10$ 。由于 $L=1$ 分解后只有一个特征值，无法计算极大特征值相对变化率，所以，本文计算了窗口长度 $L$ 从2增大至10的过程中极大特征值的相对变化率，极大特征值相对变化率变化情况如图2所示。本文采用设定相对变化率阈值的方法来确定合适的窗口长度，极大特征值相对变化率的计算如式(1)所示。

$$\alpha = \frac{|\lambda_i - \lambda_{i-1}|}{\lambda_{i-1}} \quad (1)$$

其中， $\lambda_i$ 为当前极大特征值， $\lambda_{i-1}$ 为前一个极大特征值。

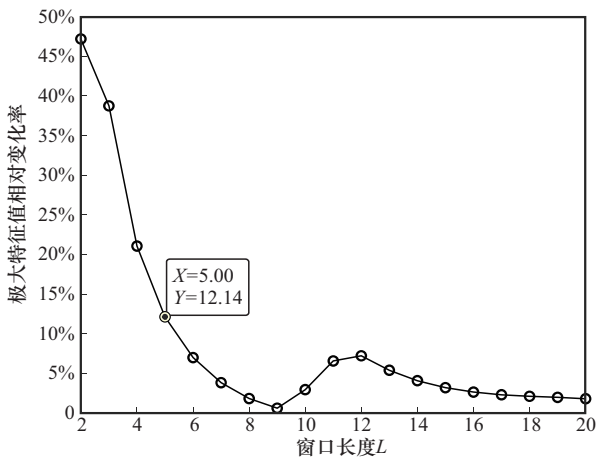


图2 极大特征值相对变化率变化情况

由图2中的极大特征值相对变化率曲线趋势可看出，当变化率处于0~10%范围时，其波动幅度显著收敛并趋于平稳状态，意味着有用信息的增量趋于饱和，此时若继续增加窗口长度，不仅无法显著提升有用信息的获取量，反而会提升计算复杂度。因此，为了尽可能地提取有用信息并且减少计算复杂度，选取极大特征值相对变化率曲线趋于稳定时的数值作为阈值，设定极大特征值的相对变化率阈值为10%。当极大特征值的相对变化率首次降至10%以下时，应选择此前最后一个变化率超过10%所对应的窗口长度值作为SSA分解的参数。基于这一准则，本文最终确定的最优窗口长度为5。

### 2.2.2 特征分量的选择

基于第2.2.1节选取的窗口长度对该电商的用户访问数据进行奇异谱分解，将得到的特征值进行降序排列，特征值变化曲线如图3所示。

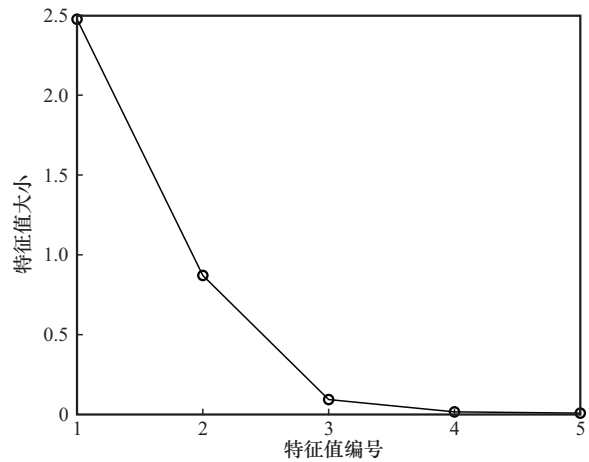


图3 特征值变化曲线

由图3可知，特征值越大意味着对应的特征分量含有的有用信息越多，特征值小意味着对应的特征分量含有的有用信息越少，特征值1和特征值2远大于其他特征值，而特征值3虽然相对较小但数值并未趋近于零，因此难以直接判断是否应该保留该特征分量。为解决这一问题，本文采用斯皮尔曼相关系数分析方法，考虑极大特征值对应的特征分量通常包含最丰富的有用信息，本文将其作为参考特征分量，通过计算该参考特征分量与其他各特征分量之间的斯皮尔曼相关系数，并结合特征值变化曲线的分布特性，实现对特征分量的选取。

若参考特征分量与其他特征分量的斯皮尔曼相关系数越大，说明二者相关性越强，意味着该特征分量与有用信息的相关性越强，说明该分量含有的有用信息不可忽略，需要选取该分量；若斯皮尔曼相关系数较小，说明相关性很弱，则舍去。计算分解得到特征分量与参考分量间的斯皮尔曼相关系数，斯皮尔曼相关系数变化曲线如图4所示。

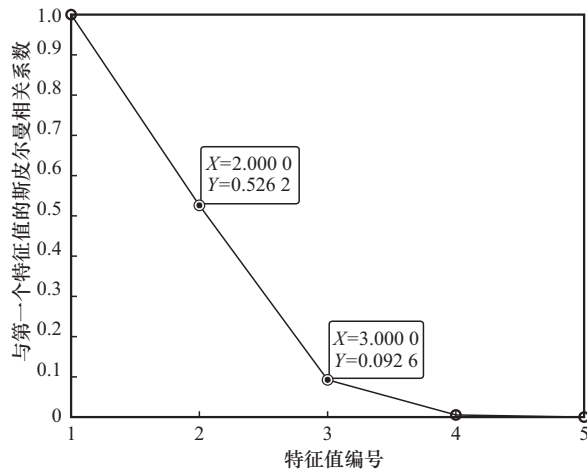


图4 斯皮尔曼相关系数变化曲线

由图4可知, 由于特征值1为极大特征值, 所以与本身的斯皮尔曼相关系数为1, 特征值2与特征值1的斯皮尔曼相关系数为0.5262, 可视为有中等强度的相关性<sup>[15]</sup>。对于特征值3而言, 与特征值1的斯皮尔曼相关系数小于0.1000, 可视为弱相关或无关, 可以舍去, 所以特征值4和特征值5也可舍去。对特征值1对应的特征分量进行快速傅里叶变换 (fast Fourier transform, FFT), 这里采样频率单位为次/天, 特征值1对应特征分量的频谱图如图5所示。

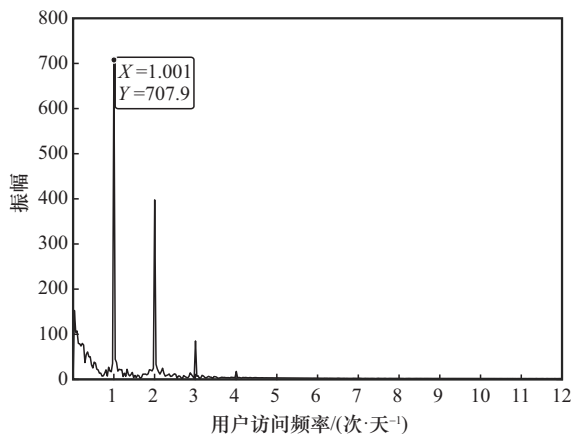


图5 特征值1对应特征分量的频谱图

根据图5所示, 特征值1对应的特征分量频谱显示出能量峰值集中在特定频率, 该峰值对应的频率约为1次/天。因此, 可以认为特征值1所

对应的特征分量为日周期分量, 可以反映该电商用户全天整体的活跃趋势。同样对特征值2对应的特征分量进行FFT, 特征值2对应特征分量的频谱图如图6所示。

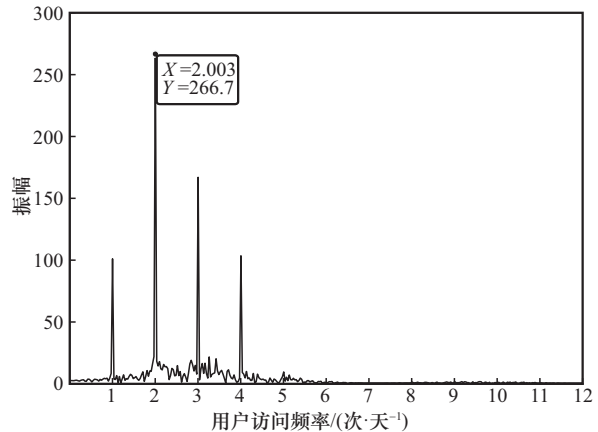


图6 特征值2对应特征分量的频谱图

由图6可知, 特征值2对应的特征分量也为周期分量, 该峰值对应的频率约为2次/天, 可视为半日周期分量, 可以反映该电商用户一天内的短周期波动趋势。

### 2.3 基于K-means聚类的用户分群

经过第2.2节提取的日周期分量、半日周期分量为去噪后的分量, 将这两个分量输入K-means聚类中, 输入前进行Z-score标准化, 目的是消除量纲差异性, 确保网站用户访问数据的聚类结果更科学、可解释。Z-score标准化的计算式如式(2)所示。

$$z = \frac{x - \mu}{\sigma} \quad (2)$$

其中,  $x$ 为输入的周期分量原始值,  $\mu$ 为周期分量的均值,  $\sigma$ 为周期分量的标准差。

将日周期分量、半日周期分量进行标准化后输入到K-means算法中, 采用手肘法来确定合适的K值来提升聚类的准确性, 手肘法的核心思想是随着聚类数K的增大, 对应样本数据的划分会逐渐精细, 每个簇的聚合度会不断增大, 误差平方和 (sum of squares due to error, SSE) 会逐渐

减小。SSE 的表达式如式 (3) 所示。

$$SSE = \sum_{i=1}^K \sum_{n \in C_i} |x - \mu_i|^2 \quad (3)$$

其中,  $K$  为聚类数,  $C_i$  表示第  $i$  个类别,  $x$  为  $C_i$  中的样本点,  $\mu_i$  为  $C_i$  中所有样本的均值。

输入分量对应的 SSE 随  $K$  值增大的变化趋势如图 7 所示。

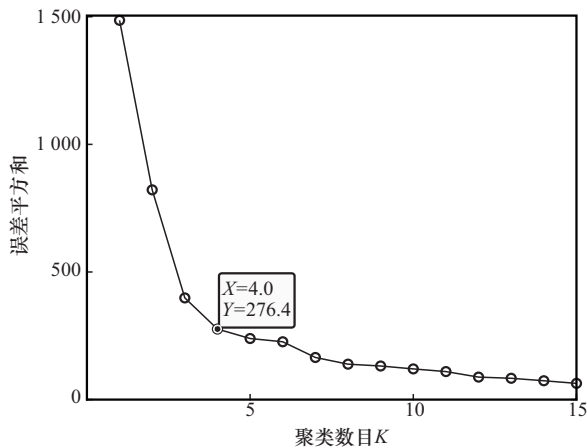


图 7 输入分量对应的 SSE 随  $K$  值增大的变化趋势

SSE 先急剧减小,  $K$  增大到一定程度后, SSE 的减小幅度会变小然后趋于稳定, 取此时变化曲线的拐点处的  $K$  值作为最佳值, 这里选取 4, 经过改进的 SSA-Kmeans 算法的用户分群聚类结果如图 8 所示。

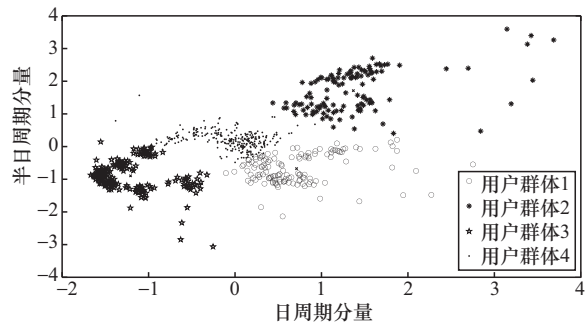


图 8 经过改进的 SSA-Kmeans 算法的用户分群聚类结果

由图 8 可知, 以日周期和半日周期为二维坐标, 将用户划分为 4 个象限, 用户群体 1 基本位于第 4 象限内, 用户群体 2 基本位于第 1 象限内, 用户群体 3 基本位于第 3 象限内, 用户群体 4 大多

集中于坐标原点处, 有一部分位于第 2 象限内。

经过 Z-score 标准化后, 横纵坐标的正负值表示样本点相对整体均值的偏离方向。样本点的横坐标大于 0, 说明该样本点在日周期分量上高于平均值; 横坐标小于 0, 说明该样本点在日周期分量上低于平均值。样本点的纵坐标大于 0, 说明该样本点在半日周期分量上高于平均值; 纵坐标小于 0, 说明该样本点在半日周期分量上低于平均值。

由上述分析可知, 用户群体 1 的日周期特性高于均值, 半日周期特性低于均值, 为全天平稳高活跃的用户群; 用户群体 2 的日周期特性和半日周期特性均高于均值, 为全天高活跃且早晚峰突出的用户群; 用户群体 3 的日周期特性和半日周期特性均低于均值, 为全天平稳低活跃的用户群; 用户群体 4 中的用户大多集中于原点, 无显著的全天活跃度偏离, 也无异常的日内波动, 为稳定型普通用户群, 还有一部分用户的日周期特性低于均值, 半日周期特性高于均值, 为全天低活跃但早晚峰突出的用户群。基于上述对于该电商用户群体的分析, 可进行分类并制定相应的用户策略, 用户分群及用户策略见表 1。

由表 1 可知, 根据该电商用户访问数据的周期特征, 将该电商的用户划分为 4 个典型群体, 并针对每个群体制定差异化的运营策略。对强周期性的用户群体 2 采取“精准推送”策略, 对弱周期性的用户群体 3 实施“唤醒策略”, 对用户群体 1 实施“需求预测”策略, 对用户群体 4 实施“短时促销”策略。上述分析框架为电商平台的精细化运营提供了实用参考。

### 2.4 算法对比

为验证本文所提的改进 SSA-Kmeans 算法在提升电商用户分群聚类效果方面的有效性和必要性, 本文将其与两种聚类方法进行对比: 一种是直接将 SSA 分解后所有特征分量全量输入进行聚类, 另一种是先对数据进行小波变换 (wavelet



表1 用户分群及用户策略

用户群体	归属象限	日周期	半日周期	用户画像	用户策略
用户群体1	第4象限	高	低	固定时间集中购物者, 目标明确型用户	集中促销、快速结账、需求预测
用户群体2	第1象限	高	高	规律性强的上班族或学生, 高频购物用户	精准推送、个性化推荐、定时优惠
用户群体3	第3象限	低	低	低频/随机用户或机器人	唤醒策略、内容营销、排除异常
用户群体4	无(原点周围)	中	中	有购物需求时才会访问, 但不集中在特定时间, 需求驱动型用户	兼顾引导规律性和激发需求
	第2象限	低	高	夜班工作者或自由职业者, 冲动消费型用户	短时促销、多样化推荐、夜间优惠

transformation, WT) 降噪处理后再进行聚类。具体而言, 本文对用户访问数据进行4层小波分解, 得到的4个细节分量既包含重要细节信息, 也含有噪声。随后, 通过设定固定阈值的方法来实现去噪, 并对阈值大小进行调整, 以获取较好的去噪效果。最终, 将去噪后的4个细节分量作为特征输入 *K*-means 算法中。同时, 本文还将该数据进行 SSA 分解, 但不进行特征分量的选取, 而是直接将得到的所有特征分量输入 *K*-means 算法中进行聚类。

本文引入戴维斯堡丁指数 (Davies-Bouldin index, DBI) 作为量化评估指标, 以更直观地说明使用上述3种方法在聚类效果上的优劣。该指数通过综合考量类内紧密度和类间分离度, 能够有效地评估网站用户访问数据的聚类质量。DBI 的核心思想是计算各聚类簇之间最大相似度的平均值: 指数值越小, 表明聚类效果越理想; 反之则表明聚类效果欠佳。由于对比的3种方法得到的聚类结果均为4个簇, 首先需要计算每个簇的类内离散度  $S_i$ , 计算式如式(4)所示。

$$S_i = \frac{1}{|C_i|} \sum_{x \in C_i} \|x - \mu_i\|, i \in [1, 4] \quad (4)$$

其中,  $|C_i|$  表示簇  $C_i$  内的样本数量,  $\mu_i$  是簇  $C_i$  内的质心,  $x$  为簇  $C_i$  内的样本。

接下来计算簇间相似度  $R_{ij}$ , 计算式如式(5)所示。

$$R_{ij} = \frac{S_i + S_j}{\|\mu_i - \mu_j\|}, i \in [1, 4], j \in [1, 4] \quad (5)$$

其中,  $S_j$  是簇  $C_j$  的类内离散度,  $\mu_j$  是簇  $C_j$  的质

心。最大簇间相似度  $D_i$  如式(6)所示。

$$D_i = \max_{i \neq j} R_{ij}, i \in [1, 4], j \in [1, 4] \quad (6)$$

DBI 为4个簇的最大相似度的平均值, 如式(7)所示。

$$DBI = \frac{1}{4} \sum_{i=1}^4 D_i \quad (7)$$

DBI 值对比见表2。

表2 DBI 值对比

方法	DBI
SSA-Kmeans	1.095 1
WT-Kmeans	0.755 2
改进的 SSA-Kmeans	0.688 0

由表2可知, 运用本文所提的改进 SSA-Kmeans 算法进行聚类, DBI 值低于其他两种方法。SSA-Kmeans 算法将所有分量直接输入 *K*-means 聚类, 未对噪声进行抑制, 使聚类结果欠佳, 从侧面说明了本文所提的改进 SSA-Kmeans 方法对于提升聚类效果的有效性和必要性。WT-Kmeans 方法的去噪效果依赖于阈值选取, 阈值过大, 会使更多小波系数被置零, 导致原始信号过度平滑, 丢失重要细节与特征; 阈值过小, 使噪声抑制效果不佳, 本文通过多次实验手动选取阈值, 但难以确保所选阈值为最优值, 因此效率较低, 且难以保证达到最佳的噪声抑制效果。

综上所述, 本文所提的改进 SSA-Kmeans 算法能有效地抑制原始用户访问数据中因网络波

动、异常用户行为等因素产生的噪声干扰，显著地提升了聚类效果，为用户群体划分提供了更精准的依据。

### 3 结束语

本文聚焦传统SSA算法在处理网站用户访问数据时存在的不足，即窗口长度及特征分量须手动选取，导致效率低下。为解决这一问题，提出了一种改进的SSA-Kmeans算法。该算法先确定合适的窗口长度，再从SSA分解结果中提取关键特征分量并开展聚类分析，显著地提升了网站用户群体的聚类效果。本文以某电商APP的用户访问数据集为研究对象，验证了改进SSA-Kmeans算法的可行性与优越性。未来研究可进一步拓展数据来源，将企业内部平台网站的用户访问量数据纳入分析范畴，构建更全面的用户行为数据集。通过对多源数据的深度挖掘与精准分析，依据网站用户群体的划分特征，制定差异化的运营推广策略，并科学规划系统的建设与发展。这不仅能够更好地满足不同用户群体的需求，提升用户体验，还将增强企业在数字化时代的市场竞争力，推动企业的可持续发展。

### 参考文献：

- [1] 伍青生, 余颖, 郑兴山, 等. 精准营销的思想和方法[J]. 市场营销导刊, 2006(5): 39-42.  
Wu Q S, Yu Y, Zheng X S, et al. Thoughts and methods of precision marketing[J]. Marketing Guide, 2006(5): 39-42.
- [2] 张祖平, 沈晓阳. 基于深度学习的用户行为推荐方法研究[J]. 计算机工程与应用, 2019, 55(4): 142-147.  
Zhang Z P, Shen X Y. Research on user behavior recommendation method based on deep learning[J]. Computer Engineering and Applications, 2019, 55(4): 142-147.
- [3] 苏舒菲, 蔺聪. 京东手机用户评论的情感分析及聚类分析[J]. 现代计算机, 2024, 30(21): 141-145.  
Su S F, Lin C. Emotion analysis and clustering analysis of Jingdong mobile phone user comments[J]. Modern Computer, 2024, 30(21): 141-145.
- [4] 葛青龙. 基于改进K-means聚类分析法算法的电商网站用户行为特征建模与分析[J]. 自动化与仪器仪表, 2024(12): 262-266.  
Ge Q L. Modeling and analysis of user behavior characteristics on e-commerce websites based on improved K-means clustering analysis algorithm[J]. Automation & Instrumentation, 2024(12): 262-266.
- [5] 杨泽丞, 张振亚, 王萍. 基于聚类的用户用电事件辨识模型[J]. 淮北师范大学学报(自然科学版), 2024, 45(2): 63-68.  
Yang Z C, Zhang Z Y, Wang P. An approach to model for the identification of electrical event based on clustering analysis[J]. Journal of Huaibei Normal University (Natural Sciences), 2024, 45(2): 63-68.
- [6] 戴圆圆. 基于用户需求分类的差异化营销研究[D]. 广州: 华南理工大学, 2017.  
Dai Y Y. Research on differentiated marketing based on user's demand classification[D]. Guangzhou: South China University of Technology, 2017.
- [7] 徐良军, 张笑第, 王立军, 等. 基于聚类分析的用户分类和用电行为分析[J]. 山西电力, 2016(4): 23-27.  
Xu L J, Zhang X D, Wang L J, et al. Analysis of user classification and electricity consumption behavior based on cluster analysis[J]. Shanxi Electric Power, 2016(4): 23-27.
- [8] 郭鹏超, 邓洪高, 孙希延, 等. 基于奇异谱分析与ARIMA-LSTM的滑坡位移预测[J]. 计算机仿真, 2024, 41(12): 523-530.  
Guo P C, Deng H G, Sun X Y, et al. Landslide displacement prediction based on singular spectrum analysis and ARIMA-LSTM[J]. Computer Simulation, 2024, 41(12): 523-530.
- [9] 吕泽富, 钟阳阳, 王攀, 等. 基于奇异谱分析和图注意力网络的测井曲线补全研究[J]. 地球物理学进展, 2025, 40(4): 1788-1799.  
Lyu Z F, Zhong Y Y, Wang P, et al. Logging curves prediction based on singular spectrum analysis and graph attention networks[J]. Progress in Geophysics, 2025, 40(4): 1788-1799.
- [10] Sharma A, Ojha S, Shelke A, et al. Scanning acoustic microscopy for biomechanical characterization of reindeer antler using singular spectral analysis[J]. Bone, 2025, 196.
- [11] 苗明瑞, 熊京华, 冯律航, 等. 基于多源数据与K-means聚类的地域特征关联纸质包装色彩应用研究[J]. 中国造纸, 2023, 42(3): 95-103.  
Miao M R, Xiong J H, Feng L H, et al. Application research of regional feature correlation packaging color based on multi-source data and K-means clustering[J]. China Pulp & Paper, 2023, 42(3): 95-103.
- [12] 孔钰婷, 谭富祥, 赵鑫, 等. 基于差分隐私的K-means算法优化研究综述[J]. 计算机科学, 2022, 49(2): 162-173.



- Kong Y T, Tan F X, Zhao X, et al. Review of K-means algorithm optimization based on differential privacy[J]. Computer Science, 2022, 49(2): 162-173.
- [13] 侯月姣, 李青, 王晓军, 等. 基于K-means聚类算法的远程学习者效果分析[J]. 北京邮电大学学报(社会科学版), 2011, 13(1): 104-109.
- Hou Y J, Li Q, Wang X J, et al. Effect analysis of distance learners' action based on K-means clustering algorithm[J]. Journal of Beijing University of Posts and Telecommunications (Social Sciences Edition), 2011, 13(1): 104-109.
- [14] 吴广建, 章剑林, 袁丁, 等. 基于K-means的手肘法自动获取K值方法研究[J]. 软件, 2019, 40(5): 167-170.
- Wu G J, Zhang J L, Yuan D, et al. Automatically obtaining K value based on K-means elbow method[J]. Software, 2019, 40(5): 167-170.
- [15] 刘娟, 彭剑虹, 曹伟, 等. 县域城市质量效益指数与城市综合竞争力相关性研究—基于斯皮尔曼秩相关系数的实证分析[J]. 质量与认证, 2024(12): 63-65.
- Liu J, Peng J H, Cao W, et al. Correlation study between the quality and benefit index of county-level cities and the comprehensive competitiveness of cities[J]. China Quality Certification, 2024(12): 63-65.

#### [作者简介]



雷雨霄(1999-), 男, 中移物联网有限公司数智化部产品开发工程师, 主要研究方向为用户访问量预测与用户行为特征分析。