User Segmentation-Based Multimodal Feature Analysis and Service Performance Enhancement for Digital Community Consumers: A Postprint

Authors: Li Canyao, Wei Wei, Liu Xiaoli, Zhou Linxing, Wang Shuai

Date: 2023-05-08T00:00:00+00:00

Abstract

[Purpose/Significance]Conducting multimodal feature analysis and service effectiveness enhancement for digital community consumers provides new perspectives for intelligence-empowered online community construction and new impetus for relevant departments to deploy digital decision-making.[Method/Process]By integrating community characteristics, we construct data dimensions for consumer segmentation, achieve segmentation through secondary aggregation of 24 indicator data within these dimensions, and establish parameter, decision variable, and function tables, thereby analyzing consumer multimodal features and realizing the enhancement of digital consumption service effectiveness based on these features.[Results/Conclusion]Empirical analysis results demonstrate that the proposed model can generate reasonable and effective segmentation results, thereby enabling inter-cluster feature differentiation and analysis of inter-cluster penetration and drift phenomena; the segmentation results reveal six consumer groups: key, central, special, dormant, churned, and general clusters. The vast majority of clusters exhibit user penetration phenomena, while only the general user cluster experiences inter-cluster drift; the service effectiveness enhancement model indicates that the groups of greatest concern and value are the central and key clusters.

Full Text

Preamble

Multi-Modal Feature Analysis and Service Efficiency Enhancement of Digital Community Consumers Based on User Clustering

LI Canyao[1], WEI Wei[1], LIU Xiaoli[1], ZHOU Linxing[2*], WANG Shuai[2]
(1. Guangxi China Tobacco Industry Co., Ltd., Nanning 530001; 2. School of

Cultural Heritage and Information Management, Shanghai University, Shanghai 200444)

[**Purpose/Significance**] Multi-modal feature analysis and service efficiency improvement for digital community consumers can provide new perspectives for building digitally empowered online communities and offer new impetus for relevant departments to deploy digital decision-making. [**Method/Process**] This study constructs data dimensions for consumer clustering based on community characteristics, performs secondary aggregation on 24 indicator data under these dimensions to achieve clustering, and constructs parameter, decision variable, and function tables to analyze consumer multi-modal features. Based on these features, digital consumption service efficiency enhancement is realized. [**Results/Conclusions**] Empirical analysis results demonstrate that the proposed model can generate reasonable and effective clustering outcomes, enabling inter-group feature differentiation and analysis of penetration and drift phenomena between groups. The clustering results reveal six consumer groups: focus, center, special, sleeping, loss, and general groups. Most groups exhibit user penetration, while only the general user group shows inter-group drift. The service efficiency enhancement model indicates that the center and focus groups are the most valued.

Keywords: User Clustering; AP-DBSCAN; Multi-Modal Features; Digital Community; Digital Consumption
CLC Number: G250
Document Code: A
Article ID: 1002-1248

1 Introduction

In recent years, digital consumption has opened a new pattern of intelligent life for residents, defined as the practice of promoting and selling products and services through digital communities [1]. The planning outline explicitly proposes the Digital China strategy, aiming to seize the opportunities of the era, accelerate digital transformation to foster new industrial modalities, and expand the digital economy as a new engine for economic growth [2]. As the main participants in digital community activities and the primary subjects of digital information consumption [4], consumers have become the focal point for service optimization. How to conduct service optimization centered on consumers has gradually become a research hotspot. However, digital product suppliers have not truly understood consumer groups, making it difficult to meet actual needs through digital community consumption activities due to the lack of refined operation models that match preferences. This study addresses digital community consumers through clustering, analyzes multi-modal features, and conducts service efficiency enhancement analysis, providing new solutions for digital intelligence empowerment and new perspectives for Digital China construction.

2 Literature Review

2.1 Types and Content of Digital Community Consumption Activities

Digital community consumption activities can be categorized into four types based on orientation: product and technology, regional, media, and digital categories. Regional classification divides into domestic [5] and overseas [6] based on administrative jurisdiction. Media classification uses activity dissemination carriers, including software and hardware. Software includes internet+ platforms [7], APPs or communities [8], and ecosystem-based operations [9]. Hardware covers wearable devices under VR technology [10] and new intelligent terminals [11]. Product and technology categories target commodity or brand promotion, conducting digital 铺设 based on product function realization [12], information integration [13], and digital assets [14]. Digital categories specifically employ high-precision technologies such as AI and eye-tracking to propose digital consumption promotion strategies like SEO or PPC for brand strategic innovation, establishing dynamic information feedback channels [15].

Research on digital communities mainly falls into four aspects: vitality stimulation and operation strategies [16], focusing on analyzing digital activity operation mechanisms; precision content dissemination and user interpretation [17], focusing on refined digital content recommendation mechanisms and consumer analysis; talent cultivation and capacity building [18], focusing on human resource strategies under national digitalization strategies; and frontier theory and technology applications [19,20], focusing on analyzing application scenarios of cutting-edge theories and technologies.

2.2 User Portrait and Clustering Research

User portraits originated in e-commerce, with definitions covering three aspects: data collection as the prerequisite, business correlation to reflect strong relevance, and data mining as the foundation for user interpretation [21]. Portrait implementation research divides into four categories: user preference-oriented [22], emotion-oriented [24], theme-oriented [25], and user behavior-oriented [26] solutions. The technology has been widely applied in the library and information science field [27].

User clustering derives from the user portrait concept, representing a secondary aggregation concept extended to consumption activities. For enterprises, clustering is a cost-reducing and efficiency-improving method for user analysis [28], capable of summarizing group characteristics and multi-modal information such as relationships and behavior preferences. Representative research includes Zhan Zhangfan's [29] user clustering for product knowledge push, which formally defines clustering models and designs clustering schemes based on clustering algorithms through preference diffusion and feature extraction. Methods mainly divide into mathematical modeling and machine learning. The former designs specialized algorithms, such as Song Song's [30] algorithm for URL data. The latter uses existing algorithms in machine learning models for automatic clustering, such as Wu Yanling and Sun Siyang's [31] adaptive natural gradient algorithm for optimal clustering.

2.3 Online Service Efficiency Improvement Research

As a new service mode in the mobile internet era, online services differ significantly from mature models in interaction levels, support modes, and experience perception, particularly as online consumption activities often involve massive instant messaging and interactions [32]. How to guarantee online service quality and level to enhance user perceived satisfaction has become an urgent issue. Existing research divides into five categories by orientation: management efficiency improvement [33], service efficiency optimization [34], environmental efficiency improvement [35], layout adjustment [36], and user value orientation [37].

---

3 Methodology

3.1.1 Clustering Indicator System Establishment

Multi-modal feature analysis refers to the process of multi-source heterogeneous data fusion and feature mining. By observing domestic and international user portrait construction characteristics, indicators can be divided into user basic information [38], social conditions [39], and Users Generated Contents (UGC) [40]. Table 1 shows user portrait indicators and corresponding meanings. Among them, user region, UGC themes, and keywords require data processing to transform into portrait labels and clustering bases.

To effectively reflect digital community users' original state and enhance clustering scientificity, we selected 4 first-level indicators for analysis, as shown in Table 2 . Group preference dimensions and data processing include 5 second-level indicators requiring data processing: activity participation time, activity frequency relative index, consumption frequency relative index, interest matrix magnitude, and activity tendency score. The interest matrix magnitude refers to the potential energy level, while activity tendency score uses the Logistic regression formula for calculation.

3.1.2 Clustering Process and Implementation

The first step in clustering is user stratification—dividing users into several non-overlapping parts based on characteristics to reflect their needs. However, stratification methods easily overlook behavioral data gaps and 割裂 user interactions, making it difficult to globally interpret the causes of user participation phenomena. Therefore, this study achieves equivalent stratification through portrait construction, then aggregates portrait data to realize clustering.

The clustering process design based on the AARRR (Acquisition, Activation, Retention, Revenue, Referral) model [45] is shown in Figure 1 [Figure 1: see original paper]. The process involves: preprocessing user raw data and mapping it to the indicator system; using visualMap to construct label-associated datasets from the indicator system; applying the AP (Affinity Propagation) algorithm [46] to aggregate label data and realize portraits; and performing DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering on aggregated data to achieve clustering. After clustering, groups are named using the AARRR model, with the largest user base group designated as the general group.

3.1.4 Inter-Group Penetration and Drift Phenomena

User group penetration refers to some users in a group tending toward other groups, while drift refers to all users in a group tending toward other groups, causing the original group to gradually dissolve without generating new groups. Social Network Analysis (SNA) [47] is commonly used to study relationships among social members and their impact on the overall network structure. This study employs the group interaction discrimination method proposed by Hu Changping et al. [48] for network construction and relationship interpretation.

From the user group co-occurrence perspective, label data from group data builds links between potentially related labels to form networks. In relationship interpretation, the number of associated labels serves as the co-occurrence relationship strength metric. For penetration phenomenon identification, the branch and bound method is primarily used. For drift user identification, the Relpscost algorithm [49] in the open-source solver SCIP is employed.

3.2 Digital Community Consumption Service Efficiency Enhancement

Referencing Meng Xiuli et al.'s method [50], this study constructs parameters and decision variable functions (Table 3) based on user clustering results and inter-group characteristics to determine the service efficiency enhancement model and derive optimal behaviors and equilibrium conditions.

3.2.1 Parameter and Decision Variable Function Establishment Digital consumption comprises a three-layer network structure of suppliers, consumers, and digital communities. The principle includes two main lines: information flow and commodity flow. Information flow involves suppliers (α) sending demands to digital communities (β) after considering costs, while consumers (γ) decide whether to participate based on their own factors. Commodity flow involves consumers (γ) consuming time or other costs to obtain services or products related to activities.

Assuming the probability of negative influence factors is Pj, the average perceived service quality is st, and the actual quality is st-1, the model describes the cost-benefit relationships among the three parties. The parameter and decision variable functions include: consumer activity cycles, activity participation consumption numbers, digital community platform numbers, supplier activity initiation numbers, cost coefficients, benefit coefficients, and various interaction metrics between parties.

3.2.2 Service Efficiency Enhancement Model Construction Modeling assumptions include: activities initiated by suppliers will be joined by consumers within a certain time; perceived service quality results from joint supplier-digital community actions; all three network members are rational decision-makers maximizing their own interests; and cost functions are continuous and differentiable. The conceptual model aims to find optimal behaviors and equilibrium conditions for suppliers, digital communities, and consumers.

The optimization problem for digital communities and consumers is convex. Using the projection method for variational inequalities, we obtain solutions for decision variables and Lagrange multipliers in constraints. AdaBoost algorithm is applied to transform efficiency enhancement into optimal planning and classification problems. Figure 3 [Figure 3: see original paper] shows the AdaBoost efficiency improvement process, where weak models are trained iteratively, and test set prediction errors update sample weights to form a final ensemble model for benefit and loss discrimination of clustered user sets.

---

4 Empirical Analysis

Under Digital China strategic requirements, traditional enterprises are adapting to diversified and upgrading digital consumption trends. The tobacco industry has implemented market trend tracking, user precise portraits, and consumer demand mining. This study uses Guangxi China Tobacco Industry Co., Ltd. as an example, constructing a new dataset from its digital community consumption platform. After filling missing values, removing duplicates, and standardizing dimensions, 55,200 users' 138,400 activity records from January 2021 to July 2022 were randomly selected for empirical analysis.

4.1.1 Clustering Results

Using the methods described in Sections 3.1 and 3.2, portraits were constructed and DBSCAN algorithm aggregated them into clusters. When the number of clusters was fixed at 6, the clustering silhouette coefficient indicated good effect with inter-group cosine similarity below 0.4. Table 4 shows each cluster's specific morphology across dimensions.

Cluster 1, mainly in South China, shows strong activity participation, knowledge dissemination themes, high consumption probability and frequency, reflecting strong consumption willingness. Cluster 2's UGC themes focus on activity effects, with high profit scores in impact dimensions and high implicit consumption levels, indicating rational consumption capacity. Cluster 3, mostly in North China with evening activity times, shows high consumption frequency and index, entertainment attributes, and strong consumption capacity but low fission K-value and profitability, reflecting low stickiness. Cluster 4 has high consumption levels and explicit-implicit values, indicating key maintenance value. Cluster 5's UGC keywords relate to winning prizes, showing strong attraction to rewards. Cluster 6 expresses stable user status with high expectations for the enterprise.

4.1.2 Naming Basis

After clustering, the variational method constructs a digital traffic funnel for naming. Using rewritten AARRR models based on harmonic mean values, six groups were identified: center (~17%), focus (~18%), special (~5%), sleeping (~9%), loss (~30%), and general groups, constituting all digital community activity users.

4.2.1 Commonality Analysis

To identify target group interactions, co-occurrence networks were constructed with frequency threshold and repulsion factor set to 0.1, 10, and 80. Centrality analysis shows center and sleeping groups as most and least active respectively. Center groups should be fully utilized while sleeping groups require guidance and activation. Centrality polar coordinates reveal resource occupation ratios, with sleeping groups showing minimal sector area and average values.

4.2.2 Characteristic Analysis

Each cluster has distinct characteristics for service efficiency strategies. Center clusters are stable users with high consumption probability and interest magnitude—analyze their sunk costs and increase consumption frequency. Focus clusters are most valuable—implement VIP service strategies and refined operations. General clusters are the only drift phenomenon—enhance their drift toward center/focus while avoiding drift to sleeping/loss. Special clusters easily convert to loss—improve participation willingness through interest attraction and emotional guidance. Sleeping clusters also easily become loss—avoid penetration and strengthen activity associations. Loss clusters, though weakly connected, constitute ~30% and require targeted recovery strategies from a business value perspective.

4.3 Practical Strategies for Service Efficiency Enhancement

Normalized empirical data corresponding to Table 3 decision variables were used to build a binary tree under optimal behavior and equilibrium conditions. Figure 4 [Figure 4: see original paper] shows the optimal state for digital consumption: starting from root node judgment of general groups, promote user drift to non-general controllable groups. For consumers, optimal equilibrium occurs when perceived benefits create strong emotional expectations and consumption stickiness. For digital communities, early activity stages should create strong appeal, activate sleeping users, and build communication effect evaluation models. Mid-stage should focus on recommendation feedback and content layout improvement. Late-stage should evaluate overall communication effects and retain high-value data.

Figure 5 [Figure 5: see original paper] shows the service efficiency improvement path: initial stage determines activity goals and target audiences; mid-stage considers enterprise promotion ROI, user cost-return, and digital community benefits; late-stage focuses on content feedback and layout adjustment. The improvement tree indicates users can achieve optimal equilibrium states, and through comprehensive consumption data analysis, service efficiency can be enhanced for suppliers, consumers, and digital communities.

---

5 Conclusion

This study establishes a digital community consumer clustering indicator system, achieves clustering through DBSCAN after AP-based portrait aggregation, and accurately identifies center, focus, special, sleeping, loss, and general user groups. Most groups exhibit user penetration, while only general groups show drift. The service efficiency enhancement model based on three-layer network equilibrium provides practical strategies for pre, mid, and post activity stages. The study's limitation is that the model's applicability to multi-source heterogeneous data requires further testing, and clustering granularity can be improved. Future research should continuously update product layouts, adjust development trends, and build harmonious, mutually beneficial digital ecosystems through relationship expansion and digital twin interactions.

References

[1] DENG M. Research on consumer-centered digital marketing strategy[J]. China market, 2022(27): 134-136.
[2] China Education and Research Network. The 14th Five-Year Plan: Accelerating digital development and building Digital China[EB/OL]. [2022-11-02]. https://www.edu.cn/xxh/zt/lhxxh/202103/t20210315_{2084688}.shtml.
[3] CHEN Y Q, SUN H, GE W R. Research on clothing brand marketing strategy in digital economy[J]. Management and administration, 2021(11): 51-55.
[4] ZHOU Y, BAI W L. System design of public information service policy content[J]. Information studies: Theory & application, 2013, 36(10): 10-15.
[5] DING Y. Research on digital marketing strategy of agricultural products under the background of "internet plus"[J]. Shanxi agricultural economy, 2022(14): 157-159.

[6] LI J, WU Y, XIAO J J. The impact of digital finance on household consumption: Evidence from China[J]. Economic modelling, 2020, 86: 317-326.
[7] GUO C Y. Digital marketing of Chinese podcasts in pan-media era - A case study of small universe App[J]. New media research, 2021, 7(17): 47-49.
[8] WANG J. Strategies for commercial banks to optimize digital marketing mode in the digital age[J]. China market, 2021(29): 119-120.
[9] BAO L, JIANG Z Y, XI K Y. Development of digital marketing: Evolution from era 1.0 to era 4.0[J]. Shandong textile economy, 2022, 39(2): 19-22.
[10] DENG Z W. Research on the construction of competence model of digital marketing talent[D]. Hangzhou: Zhejiang Gongshang University, 2022.
[11] FUCHS C. The digital commons and the digital public sphere: How to advance digital democracy today[J]. Westminster papers in communication and culture, 2021, 16(1): 9-26.
[12] DONG Y F, TAN R H, ZHANG P, et al. Product redesign using functional backtrack with digital twin[J]. Advanced engineering informatics, 2021, 49(3): 101.
[13] YOON S, MCCLEAN S T, CHAWLA N, et al. Working through an "infodemic": The impact of COVID-19 news consumption on employee uncertainty and work behaviors[J]. Journal of applied psychology, 2021, 106(4): 501-517.
[14] CUMMINGS B. Digital assets add new concerns to estate planning[J]. Journal of financial planning, 2022, 35(8): 19.
[15] XU T. Digital marketing strategy based on new technology - Taking All-Saints as an example[J]. Inner Mongolia coal economy, 2020(22): 87-89.
[16] KAPKAEVA N, GURZHIY A, MAYDANOVA S, et al. Digital platform for maritime port ecosystem: Port of hamburg case[J]. Transportation research procedia, 2021, 54: 909-917.
[17] HU Z Y. A study on the causes and governance of ethical chaos in digital marketing in China - Based on an interview with digital marketing professionals[J]. Contemporary communication, 2018(5): 80-84.
[18] GAO P P. Construction and practice of virtual simulation training center for digital marketing specialty under VR technology[J]. Marketing circles, 2022(1): 65-67.
[19] SKARE M, SORIANO D. How globalization is changing digital technology adoption: An international perspective[J]. Journal of innovation & knowledge, 2021, 6(4): 222-233.
[20] MIAO Y X. Analysis on influencing factors of online retail based on marketing macro-environment theory[J]. China market, 2022(5): 125-127.
[21] LIU Q L. The basis, principle, methodology (model) and application of user portrait[EB/OL]. [2022-11-06]. https://zhuanlan.zhihu.com/p/140104236.
[22] WANG S. Study on user portrait and clustering of online health community in the context of public health emergencies[J]. Information science, 2022, 40(6): 98-107.
[23] WANG X P. Construction of user portrait based on video big data[J]. Video engineering, 2017, 41(6): 20-23.
[24] LI X G, XIAO S Q, LI S S, et al. Research on user profiles of Xiaomi community based on knowledge behavior[J]. Journal of library and information

science in agriculture, 2021, 33(8): 4-12.
[25] LI D. Text mining model for virtual community user portrait based on social network analysis[J]. Tehnicki vjesnik - Technical gazette, 2019, 26(4): 1145-1151.
[26] MIN T, CAI W X. Portrait of decentralized application users: An overview based on large-scale Ethereum data[J]. CCF transactions on pervasive computing and interaction, 2022, 4: 124-141.
[27] CHEN Y B, HE J S, WEI W, et al. A multi-model approach for user portrait[J]. Future Internet, 2021, 13(6): 147.
[28] JIAN S Q, LI Q H, QIN Y Q. User group analysis based on K-means[J]. Modern computer, 2017(29): 29-31.
[29] ZHAN Z F. Research on product design knowledge push based on the user group[D]. Nanchang: Nanchang Hangkong University, 2016.
[30] SONG S. Mobile Internet user clustering based on the analysis of URL[D]. Baoding: Hebei University, 2013.
[31] WU Y L, SUN S Y. Clustering method of library users based on Markov model[J]. Information science, 2021, 39(11): 167-172.
[32] CHEN C Q. Research on customer relationship management of online travel agency - With Tuniu network as example[D]. Nanjing: Nanjing Agricultural University, 2017.
[33] WANG W, ZHANG X X, GUO S L, et al. Construction of user demand model of mobile professional virtual community based on grounded theory[J]. Information science, 2022, 40(6): 169-176.
[34] ZHANG B H, WAN Q L, WU J. Analysis of optimizing business climate by online government services: "Digital dividend" and "digital gap" effects[J]. Chinese public administration, 2021(4): 70-75.
[35] YU L J. Research on "Internet +smart service" optimizing digital reading accurate promotion model[J]. The library journal of Henan, 2021, 41(9): 4-6.
[36] WANG Y, WU R Q. Research on user portrait of digital cultural resource service in public library[J]. Library and information service, 2021, 65(16): 42-55.
[37] SUN Y, CHAI R Q. An early-warning model for online learners based on user portrait[J]. Ingénierie des systèmes d'information, 2020, 25(4): 535-541.
[38] WANG X, WEI X, MA J, et al. User portrait technology and its application scenario analysis[C]//BDE 2021: The 2021 3rd international conference on big data engineering, New York: Association for Computing Machinery, 2021: 64-69.
[39] LI P, GU D L, ZOU T, et al. Exploration of potential reading interest groups in social sciences of undergraduates based on social network analysis: A case study of China pharmaceutical university[J]. Journal of library and information science in agriculture, 2021, 33(3): 78-89.
[40] SHI G L, ZHANG M, ZHENG W W. Advances in applied research of formal concept analysis in folksonomy[J]. Library and information service, 2014, 58(9): 136-142.
[41] ZHANG S P. An evaluation of international consumption center cities competitiveness in Yangtze River Delta urban agglomeration[J]. Journal of commercial economics, 2022(10): 44-49.

[42] WANG H, WANG J L. The empirical study on night consumption tendency of Beijing residents: Based on Logit regression model[J]. Urban problems, 2021(2): 75-83.
[43] MENG X L, WU Y F, LIU B. Multi-phase crowdsourcing logistics service quality optimization considering delay insurance[J/OL]. China management science: 1-15[2022-11-08]. DOI:10.16381/j.cnki.issn1003-207x.2021.1807.
[44] LIN Z, CHAO M G. Research on acquisition of authors sources of sci-tech journals based on AARRR model[C]//ICCIR 2021: Proceedings of the 2021 1st international conference on control and intelligent robotics, New York: Association for Computing Machinery, 2021: 193-196.
[45] ZHAN J J. Research on affinity propagation clustering algorithm for probabilistic undirected graph model[D]. Nanning: Guangxi University, 2017.
[46] HU C P, HU J M, DENG S L. Analysis of network users group interaction and research for service based on the web 2.0[J]. Journal of library science in China, 2009, 35(5): 99-106.
[47] GASSE M, CHETELAT D, FERRONI N, et al. Exact combinatorial optimization with graph convolutional neural networks[C]//Advances in neural information processing systems 32 (NeurIPS 2019), Vancouver Canada: MIT Press, 2019: 15580-15592.
[48] NAGURNEY A, DANIELE P, SHUKLA S. A supply chain network game theory model of cybersecurity investments with nonlinear budget constraints[J]. Annals of operations research, 2017, 248(1/2): 405-427.

Note: Figure translations are in progress. See original paper for figures.