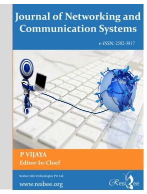
Vol.7 No.4 Oct 2024

Journal of Networking and Communication Systems

Received 12 August, Revised 16 August, Accepted 3 October



Customer Segmentation in E-commerce using Weighted Kernel Fuzzy C-Means and KL-Mumford Shah Loss

N.Senthil Murugan

Assistant Professor, Department of Information Technology
J.P.College of Engineering
sendhil28@gmail.com

Abstract: Nowadays, customer segmentation is considered a well-established marketing strategy followed for the management of customer relationships for increasing profitability and delivering customer-centric services. In customer segmentation, customers of similar characteristics are categorized into the same groups. Various clustering algorithms are employed for identifying the purchase behavior of customers to increase profits, but these techniques encountered shortcomings in simplifying the decision-making process while utilizing the large database. In this research, the Weighted Kernel Fuzzy C-Means+ Kullback–Leibler -Mumford Shah Loss (WKFCM+ KL-Mumford Shah Loss) is designed for the segmentation of customers in Electronic Commerce (E-commerce). At first, the input customer data is accumulated from the database for customer segmentation. Here, the catalog portioning is executed and then, customer segmentation is performed for each partitioning using WKFCM with KL-Mumford Shah Loss Thereafter, the segmented feature matrix is formed for the users by considering the relative category, and the customer segmentation is performed finally using Deep Embedded Clustering (DEC). In addition, the superiority of WKFCM+ KL-Mumford Shah Loss is validated by comparing with prevailing segmentation schemes. The WKFCM+ KL-Mumford Shah Loss attained clustering accuracy, random coefficient, and Hubert index of 93.677%, 89.765%, and 90.876% respectively.

Keywords: Customer Segmentation, Clustering Algorithms, E-Commerce, Customer Relationship Management, Weighted Kernel Fuzzy C-Means

Nomenclature

Abbreviation	Expansion
BDA	Big Data Analytics
BI	Business Intelligence
K-means-PSO-ANN	K-means-Particle Swarm Optimization- Artificial Neural Network
PCA+GMM	Principal Component Analysis+Gaussian Mixture Model
MTS	Multivariate Time Series
CH	Calinski–Harabasz
FCM	Fuzzy C-Means

1. Introduction

Nowadays, E-commerce is growing vigorously worldwide with the progress of various E-commerce platforms, like JD.com, Tmall, Amazon, and so on [5]. E-commerce grows continuously due to the advancement of internet technologies and it continuously grows with the increasing development of new technologies [6]. E-commerce platforms are becoming more popular and are implemented in almost every business area to promote products to customers and support marketing online [4]. Nowadays, the E-commerce system acts as a new shopping channel due to the movement of various shopping outlets to the Internet. The conversion of more customers and businesses to E-commerce tends to continuously change the shopping styles of customers, regardless of the business-to-customer or business-to-business transactions [6]. Thus, E-commerce plays a vital role in ensuring stimulating economic growth, supply of living materials, and ensuring basic production. E-commerce platforms must gain the attention of more customers, which also meets the customer needs to get more customers. The diversification of customer needs varies based on the various products and services preferred by different customers [5]. Thus, it is important to understand the customers and the interaction patterns among online companies to increase the growth of the online market. In recent years, for profit-making, companies often set goals and perform

https://doi.org/10.46253/jnacs.v7i4.a2

different validation to adjust marketing strategies and manage customer relations to achieve these goals [6].

Customer segmentation is considered the fundamental strategy performed to understand the behavior of customers across different industries [9]. It involves the categorization of customers into various groups by considering the particular preferences and needs of the customer as well as their shared characteristics [9]. Customer segmentation is frequently applied in strategic marketing to provide the idea of dividing heterogeneity into homogeneous forms [7]. In the customer segmentation framework, BDA and BI are employed for marketing both established and new products and to discern the prospective customer profile. Meanwhile, the techniques encountered a lack of contextual framework and influenced decision marketing strategic issues [2]. Also, the classical customer segmentation techniques affected the breaking of marketing strategies and did not effectively respond on time to huge customer data [8]. Thus, the unlabelled data are converted into groups called clusters based on the manual work and increasing amount of data for segmentation. The clustering techniques analyze and segment the characteristics of members who are similar to each other and differ from members of other clusters [10]. The clustering analysis categorizes the E-commerce customers based on the customer's interested brand, specific product, purchasing habit, or spending habit. The clustering approach segments the customers by processing the collected data [4]. Meanwhile, the clustering approaches encounter some gaps in the existing literature. Most of the algorithms used for customer segmentation did not consider the operation efficiency of the algorithm while selecting the clustering algorithm [5].

This paper presents the WKFCM+KL-Mumford Shah loss approach for customer segmentation in E-commerce. Initially, the input multi-category data is accumulated from the database for customer segmentation. Then, the accumulated multi-category data is subjected to catalog partitioning, and customer segmentation for each partitioned data is performed using WKFCM with the KL-Mumford Shah loss function. Following this, the formation of the user-to-segment matrix is performed. Further, the final customer segmentation is executed using DEC for the analysis of customer data.

The primary contribution of the research is summarized as follows,

- ***Developed WKFCM+KL-Mumford Shah loss for segmentation of customers:*** The WKFCM+KL-Mumford Shah loss is designed for customer segmentation in E-commerce. Here, the customer segmentation for each portioned category is performed using WKFCM by considering the KL-Mumford Shah loss function. Further, the final segmentation is done using the DEC.

The article is organized in a subsequent way as follows; section 2 explicates the brief review of prevailing literature employed for customer segmentation. In section 3, the designed approach for customer segmentation is portrayed. In addition, section 4 elucidates the results and discussions followed, and section 5 shows the conclusion with future work of the article.

2. Motivation

Customer segmentation plays a crucial part in the marketing industry, and it helps organizations meet and understand the demands of consumers for increasing their services and managing customer relationships. The prevailing techniques employed for customer segmentation effectively analyze the values of customers accurately and meet the timely needs of customers. However, the prevailing techniques utilized historical order of customer data that cannot reflect the customer behavioral performance fully during feature selection, which was also time-consuming. Also, the efficiency of algorithmic operations was not considered by prevailing clustering algorithms. Thus, it encourages to design of a clustering technique for the segmentation of customers in E-commerce.

2.1. Literature Review

Alghamdi, A., [1] developed a K-means-PSO-ANN approach for customer segmentation to reveal customer satisfaction. In this model, the customers were segmented based on online reviews, and the segmentation analysis was performed by utilizing hybrid deep learning approaches. This model attained minimum errors during customer segmentation. Meanwhile, this approach failed to incorporate text mining techniques for the analysis and segmentation of customers. John, J.M., *et al*. [2] designed a PCA+GMM for customer segmentation to increase the decision-making process. Here, the customer values were quantified by utilizing the Recency, Frequency, and Monetary (RFM) framework. This model effectively handled mislabelled data samples and overlapping issues during customer segmentation under less computational time. However, this approach failed to apply the kernel function to increase the linear separability of the output of the GMM cluster. Wang, S., *et al*. [3] introduced Dynamic Customer Segmentation by combining Length, Recency, Frequency, Monetary, and Satisfaction (LRFMS) and MTS (DCSLM) for dynamic customer segmentation to maintain the profitability of the group. Here, the dynamic time wrapping was incorporated with the MTS clustering approach based on complexity-invariant dissimilarity, shape-based distance, and multiple dimensions. This model significantly provided more detailed data descriptions and positioned the customer more accurately. However, this approach was not successful in utilizing customer information, like browsing history, reputation, and geographic location for customer segmentation. Tabianan, K., *et al*. [4] established SAPK + K-Means-based clustering technique for the segmentation of customers in E-commerce. Here, the relationship among three clusters, like categories, products, and types was taken under consideration to develop highly profitable segments. This technique determined the behaviors of customers more accurately under a low error rate, but it failed to employ deep learning-based techniques for the segmentation of customers. Wu, Z., *et al*. [5] devised the Recency, Frequency, Money, Add to Cart, and Add Favourites (RFMCV) approach for the selection of segmentation features to segment customers in E-commerce. Here, the initial clustering center was selected by utilizing the K-means++ algorithm, and the optimal number of clustering was determined using the CH index. This approach effectively stimulated the consumption potential by maintaining the important profit source in the E-commerce platform. Meanwhile, this model failed to classify customers by considering more features of consumer behavior and also failed to utilize density-based clustering, hierarchical clustering, and other clustering techniques to increase performance.

2.2. Challenges

The limitations of baseline customer segmentation techniques are enlisted as follows,

- The K-means-PSO-ANN model employed in [1] failed to utilize the large database to better identify the segments and estimate the effects with high precision during the segmentation of customers. Also, this approach was not successful in determining its effectiveness based on prediction accuracy.
- The PCA+GMM approach used in [2] was not successful in providing accurate insights for the segmentation of customers by utilizing sophisticated machine learning algorithms like spectral clustering to validate the behavior of customers.
- The DCSLM technique utilized in [3] significantly reduced the time complexities for recording dynamic customer behavior during customer segmentation. However, this model failed to obtain comprehensive customer information for identifying customer behaviors in various areas.
- The RFMCV technique used in [5] failed to handle data inaccuracies or uncertainty problems by utilizing fuzzy similarity consumptions for the segmentations of customers. Also, it was not successful in maintaining the important profit sources for customer segmentation in the E-commerce platform.
- The baseline approaches employed for customer segmentation were not successful in understanding the customer purchase behavior to increase profitability and provide customer-centric service by considering the operation efficiency of the clustering algorithm.

3. Designed Weighted Kernel Fuzzy C-Means+ KL-Mumford Shah Loss (WKFCM+KL-Mumford Shah loss) for customer segmentation

This paper presents WKFCM+KL-Mumford Shah Loss for the segmentation of customers in E-commerce. The input multi-category data is initially accumulated from the database [11] and then the input data is subjected to catalogue partitioning, where the customers are grouped based on the categories of the products they purchase. Thereafter, customer segmentation using WKFCM [12] is executed for each partitioned data by considering the KL-Mumford Shah loss function according to the finer details of their purchase. Here, the KL-Mumford Shah loss function is designed by incorporating KL divergence loss [13] and the Mumford-Shah loss function [14]. Then, the user-to-segment matrix is constructed from the segmented customers. Finally, customer segmentation is performed using the analysis of customer data using DEC [15]. Further, the abstract view of the WKFCM+KL-Mumford Shah loss model for customer segmentation is displayed in Fig. 1.

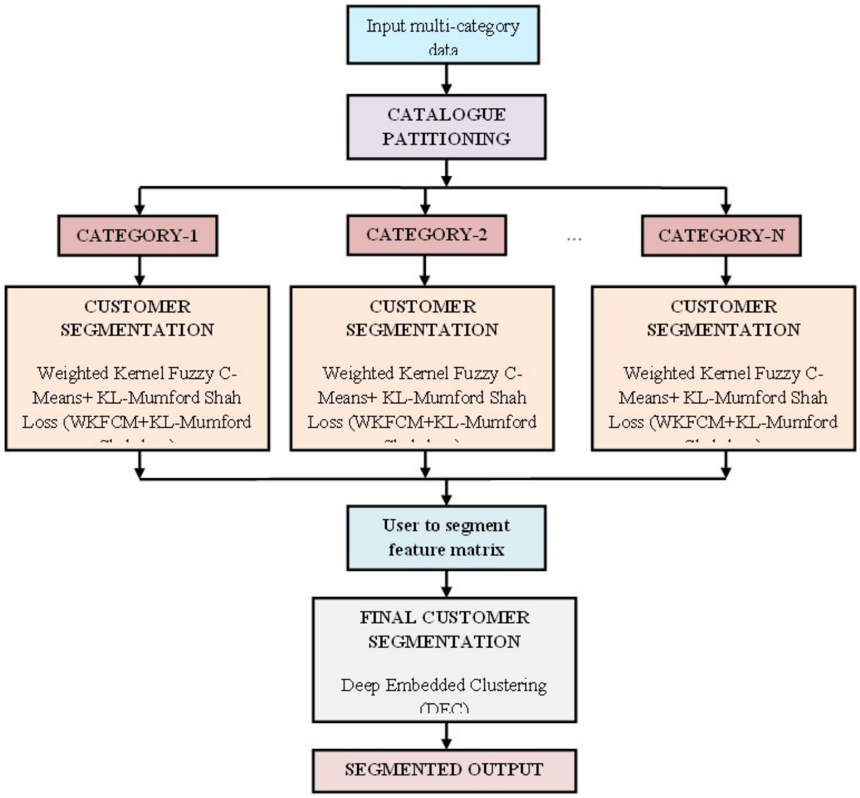


Fig. 1. *Abstract view of WKFCM+KL-Mumford Shah loss model for customer segmentation*

3.1. Multi-Category Data Acquisition

At first, the multi-category data is accumulated from the E-commerce Customer Behavior and Purchase dataset [11] for customer segmentation in E-commerce, and is expressed as,

$$C = \left[C_1, C_2, C_3, ..., C_E, ..., C_P \right] \tag{1}$$

Where, the total dataset considered for the segmentation of customers is represented as C, C_E represents the E^{th} multi-category data taken as input for customer segmentation and the total available multi-category data is indicated as P.

3.2. Catalogue Partitioning

In catalog partitioning, the customers are divided into distinct groups based on the characteristics of the products they purchase to increase the customer experience and marketing strategies. Catalog partitioning plays a crucial role in customer segmentation in E-commerce, as it helps to enhance marketing tactics and customer engagement. In this research, the input multi-category data C_E is partitioned into different types of categories to understand the preferences and needs of customers in providing suitable services. Here, the input multi-category data C_E is split based on category as per the type of products for each customer. For instance, let us consider the total number of X customers $G = \left[G_1, G_2, G_3, ..., G_X \right]$, where each customer buys products from N categories, and is given by $P = \left[P_1, P_2, P_3, ..., P_N \right]$. Thus, the catalog partitioning of each customer is delineated in Fig. 2.

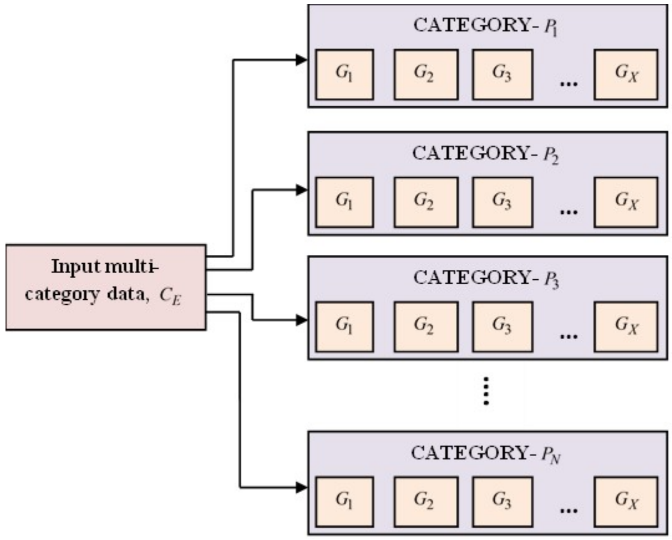


Fig. 2. *Catalogue partitioning of customer*

For analysis, only 4 customers and 3 groups are considered here. The customers are grouped into various categories based on the products purchased.

3.3. Customer Segmentation for Each Partition

Customer segmentation is the process of dividing customers into segments or subgroups by considering the relationships between the items and their behavior patterns in various categories. Here, the customer segmentation is executed for each partition using WKFCM [12] based on the KL-Mumford Shah loss function, where the KL-Mumford Shah loss function is the incorporation of KL divergence loss [13] and Mumford-Shah loss function [14]. The WKFCM is effective in handling high-dimensional data and is robust to outliers. The process performed during the customer segmentation for each partitioning is demonstrated as follows,

Customer segmentation using WKFCM based on KL-Mumford Shah loss function:

The WKFCM [12] model is the improved version of classical FCM by modifying the fuzzy cluster optimization, the membership degrees, and the kernel function. The FCM with kernel function $W(a,b)$ is expressed as,

$$W(a,b) = \left[\Phi(a), \Phi(b)\right] \tag{2}$$

Here, a non-linear function that is utilized to map low-dimensional to high-dimensional feature space is represented as Φ, and the inner product is signified as $(.)$. Further, a denotes the data and b stipulates the cluster. Then, the Gaussian kernel function is used to modify the vector and is given by,

$$W(a,b) = \exp\left(-\|a-b\|^2 / \alpha^2\right) \tag{3}$$

Where the characteristic parameter of kernel functions is denoted by α. The WKFCM extraction algorithm with objective function is designated as,

$$K_g = \sum_{r=1}^{X} \sum_{s=1}^{Y} u_{rs}^{g} * L \tag{4}$$

Where, X indicates the number of clusters, u_{rs}^{g} designates the membership value, and Y is the number of data points, the KL-Mumford Shah loss function is denoted by L. The membership value is designated as,

$$u_{rs} = \frac{\left(1 - W(a_s, w_r)\right)^{-1/(g-1)}}{\sum_{t=1}^{X} \left(1 - W(a_s, w_t)\right)^{-1/(g-1)}} \tag{5}$$

Here, g stipulates the weighting exponent, and w_r designates the cluster centroid given by,

$$w_r = \frac{\sum_{r=1}^{Y} u_{rs}^{g} W(a_s, n_r) a_s}{\sum_{r=1}^{Y} u_{rs}^{g} W(a_s, n_r)} \tag{6}$$

Further, the KL-Mumford Shah loss function is the incorporation of KL divergence loss [13] and Mumford-Shah loss function [14]. The expression of KL-Mumford Shah loss function is expressed as,

$$L = \beta_1 (L_{MSL}) + \beta_2 (L_{KL}) \tag{7}$$

From the expression, the scaling factor is indicated as β_1 and β_2, the Mumford-Shah loss function is symbolized as L_{MSL}, and the KL divergence loss is signified as L_{KL}.

Mumford-Shah loss function: The Mumford-Shah loss function [14] is used to perform computer vision tasks that help the network utilize unlabeled data as elements for training. The expression of the Mumford-Shah loss function is given by,

$$L_{MSL} = \sum_{u}^{k} \sum_{v \in \Omega} F_{uv} \left\| Y_u - Z_v \right\|^2 + \delta \sum_{u}^{k} \sum_{v \in \Omega} \left| \nabla F_{uv} \right| \tag{8}$$

Here, Z_v indicates average voxel intensity, F_{uv} represents the categorized data, Y_u designates the actual data, the approximate total variant is denoted by $\left|\nabla F_{uv}\right|$, and δ signifies parameters that balance the influence.

KL divergence loss function: The KL divergence loss function [13] is used to compute the distance among two probability distributions, which is expressed as,

$$L_{KL} = Min \sum_{u} \sum_{v} M_{uv} \log\left(\frac{M_{uv}}{N_{uv}}\right) \tag{9}$$

Where, the sharpened clustering distribution is represented as N_{uv}, and the actual clustering distribution is given by M_{uv}.

To perform customer segmentation, the WKFCM considers the input data and then, the cluster centroids are found using the standard FCM, which are fed to the WKFCM as input. Thereafter, using equation (6), the membership values are found. The process is repeated till the maximum iteration is reached and the cluster centroids are updated using equation (7). Thus, each customer is segmented into groups for each category by considering imbalanced clustering, which is displayed in Fig. 3. Here, each customer is categorized into groups based on their behavior patterns. For instance, in the category P_1 the customer G_1 and G_2 in each partition are clustered in the segment J_1, whereas in the group J_2, the customers G_3 as well as G_4 are grouped. Similarly, the customer G_1, G_3 in each partition are clustered in the segment J_1, and the customer G_2 as well as G_4 are clustered in the group J_2 under a category P_2. Further, in the category P_3, the customers G_4 G_3 are clustered in the group J_1, whereas the customers G_2 G_1 are clustered in group J_2.

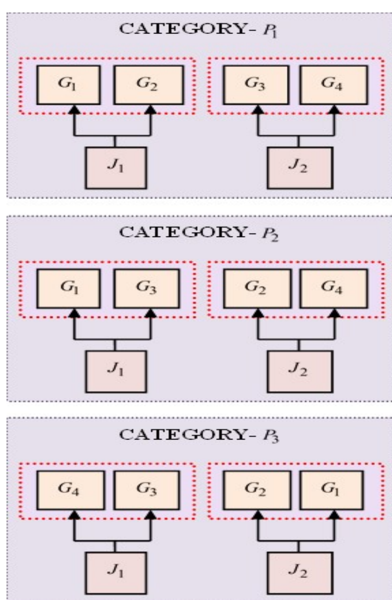


Fig. 3. *Customer segmentation for each partition*

3.4. Formation of the user to Segment Feature Matrix

The segmented feature matrix is created to determine the behaviors, preferences, and trends of customers within each segment for providing targeted marketing efforts. During the formation of the segmented feature matrix, the segmented customers obtained for each partition are arranged as per the customer group and category. The segmented feature matrix S formed for the segmented customer-based ln their categories is delineated in Table 1.

Table 1. *Segmented feature matrix formation*

CUSTOMER	CATEGORY		
	P_1	P_2	P_3
G_1	J_1	J_1	J_2
G_2	J_1	J_2	J_2
G_3	J_2	J_1	J_1
G_4	J_2	J_2	J_1

3.5. Final customer segmentation using DEC

Once the segmented feature matrix S is created, the customers are finally segmented based on their categories into groups. Here, the customer segmentation is performed finally using DEC [15], and the segmentation process performed in DEC is explicated below,

The DEC [15] is considered an unsupervised clustering technique, which is used to significantly map the lower-dimensional feature space from the created segmented matrix by iteratively optimizing the clustering objective. The DEC is less sensitive and helps to assign clusters and learns feature representation using deep neural networks by offering the advantages of superior performance, high scalability, and feature representations. The DEC utilizes non-linear mapping by transforming the data to partition the files, once the created segmented feature matrix is fed into it. Here, the partitioning is performed under two steps, namely parameter initialization as well as optimization. At first, a deep autoencoder is used and the clustering is performed in the second step, where the steps are repeated continuously between the auxiliary target distribution calculation and KL divergence. Assume, a set of p points $\{q_s \in Q\}_{s=1}^{p}$ is partitioned into d clusters. The DEC uses non-linear mapping $\gamma_a : Q \to V$ for transforming data into a latent feature space V. The learnable parameters a and $\{\rho_w\}_{w=1}^{d}$ are estimates for initiating the clustering process, here the cluster centroid is signified as ρ.

Clustering Using KL Divergence

In DEC, the cluster refinement and soft assignment are altered using non-linear mapping γ_a and original estimate of the $\{\rho_w\}_{w=1}^{d}$ to perform unsupervised clustering. Then, the function γ_a is modified and the cluster centroids are refined by employing an auxiliary target distribution. The steps are continuously followed till satisfying the convergence condition and the clustering process performed is briefly demonstrated in the following steps.

(1) Soft assignment: The student's t-distribution is used as the kernel in soft assignment for the estimation of similarity among the centroid ρ_w and embedded point h_s, where the probability of assigning samples I_{sw} is expressed as,

$$I_{sw} = \frac{\left(1 + \left\| h_s - \rho_w \right\|^2 / \varepsilon\right)^{-\frac{\varepsilon+1}{2}}}{\sum \left(1 + \left\| h_s - \rho_{w'} \right\|^2 / \varepsilon\right)^{-\frac{\varepsilon+1}{2}}} \tag{10}$$

here, the degree of freedom is given as ε and after embedding $z_i = \gamma_a(q_s) \in Q$ corresponds to $I_s \in V$.

(2) Minimization using KL divergence: The auxiliary target function is used to iteratively refine the clusters by learning high-confidence assignments. Here, the model is trained by matching the target to the soft assignment. The KL divergence loss is measured using the equation (7).

Later, the values are normalized using the frequency per cluster and are designated as,

$$M_{sw} = \frac{N_{sw}^2 / D_s}{\sum_{w'} N_{sw'}^2 / D_{s'}} \tag{11}$$

where the frequencies of the soft cluster are signified as $D_s = \sum_s N_{sw}$, N_{sw} is the probability of assigning the sample s to the cluster w Hence, the DEC increased low confidence predictions and effectively segmented the customers finally. Fig. 4 shows the final segmented customers, where the customers G_1 and G_2 are segmented under group J_1 and customer G_3 as well as G_4 are segmented under group J_2 .

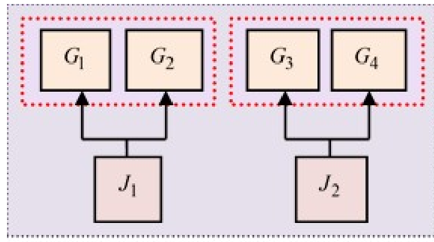


Fig. 4. *Final segmented customers*

4. Results and Discussion

The outcomes obtained from the experiment by the WKFCM+KL-Mumford Shah loss model used for customer segmentation in E-commerce and the discussions followed are demonstrated below,

4.1. Experimental Set-Up

The WKFCM+KL-Mumford Shah loss technique designed for the segmentation of customers is implemented using a Python tool.

4.2. Dataset Description

The multi-category data accumulated for customer segmentation is taken from the E-commerce Customer Behavior and Purchase dataset [11]. This database is a synthetic database created by utilizing the Faker Python library. This database possesses purchase history and various aspects of customer behavior collected from the digital marketplace. The database is designed mainly for predictive modeling and data analysis in the area of E-commerce. This database is highly suitable for trend analysis, recommendation systems, market basket analysis, and prediction of customer churn. The column information of the database mainly comprises churn, returns, payment method, amount purchased, quantity, price of the product, category of product, purchase date, gender, age, name, and ID of the customer.

4.3. Performance Metrics

The efficacy of the WKFCM+KL-Mumford Shah loss model designed for customer segmentation in E-commerce is validated using different evaluation measures, which are delineated as follows,

(i) Hubert index: The Hubert Index is used to measure the clustering solution values to access the separation between the clusters using various sets of data points, and the expression of the Hubert index is given by,

$$Hubert\,index = \frac{1}{B}\sum_{g}^{f-1}\sum_{e=g+1}^{f} A_{1ge}A_{2ge} \tag{12}$$

here, B represents total pairs of points, the total data points are signified as f , as well as A_1 and A_2 indicates the objects of certain classes.

(ii) Clustering accuracy: The clustering accuracy [16] is the accurately assigned data points of true classes from the total classes, which is given by,

$$Clustering\ accuracy = \frac{1}{B}\sum_{c}^{B}\eta\left(k_c, map\left(l_c\right)\right) \tag{13}$$

where the ground truth label of the data point is indicated as k_c , η designates the Dirac delta function, and l_c indicates clustering of the point.

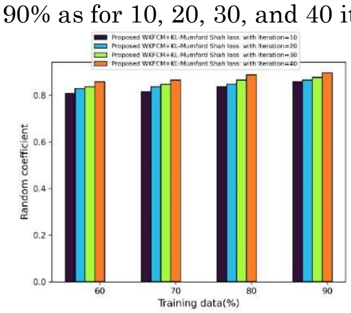
(iii) Rand coefficient: The Rand coefficient is used to accurately estimate the pairs of data points clustered with each other by comparing the clustering outcome with the partitioned ground truth, where the expression of Rand coefficient is given by,

$$Rand\ coefficient = \frac{\mu_1 + \mu_4}{\mu_1 + \mu_2 + \mu_3 + \mu_4} \tag{14}$$

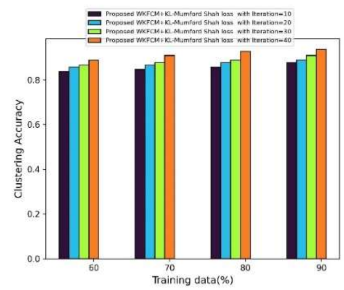
Here, the true positive is signified as μ_1, false positive is indicated by μ_3, a false negative is represented as μ_4, and μ_2 denotes true negative.

4.4. Performance Analysis

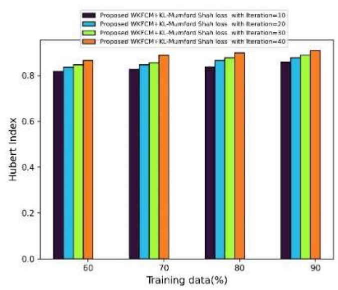
Fig. 5 depicts the performance valuation of WKFCM+KL-Mumford Shah loss utilized for customer segmentation by altering training data. In Fig. 5(a), the validation of WKFCM+KL-Mumford Shah loss by utilizing the Rand coefficient is demonstrated. Here, the WKFCM+KL-Mumford Shah loss observed a Rand coefficient of 0.857, 0.865, 0.876, and 0.897 with 90% training data for iterations 10, 20, 30, and 40. The analysis of WKFCM+KL-Mumford Shah loss in terms of clustering accuracy is depicted in Fig. 5(b). For training data of 90%, the clustering accuracy of 0.876, 0.888, 0.908, and 0.936 is recorded by WKFCM+KL-Mumford Shah loss for 90% training data, and iteration 10, 20, 30, and 40. Further, Fig. 5(c) shows the validation of WKFCM+KL-Mumford Shah loss employing the Hubert Index. For 90% of training data, the Hubert index of 0.857, 0.876, 0.888, and 0.908 is measured by WKFCM+KL-Mumford Shah loss for training data of 90% as for 10, 20, 30, and 40 iterations.



(a)



(b)



(c)

Fig. 5. *Performance evaluation of WKFCM+KL-Mumford Shah loss utilizing (a) Rand coefficient, (b) Clustering accuracy, and (c) Hubert index*

4.5. Comparative Techniques

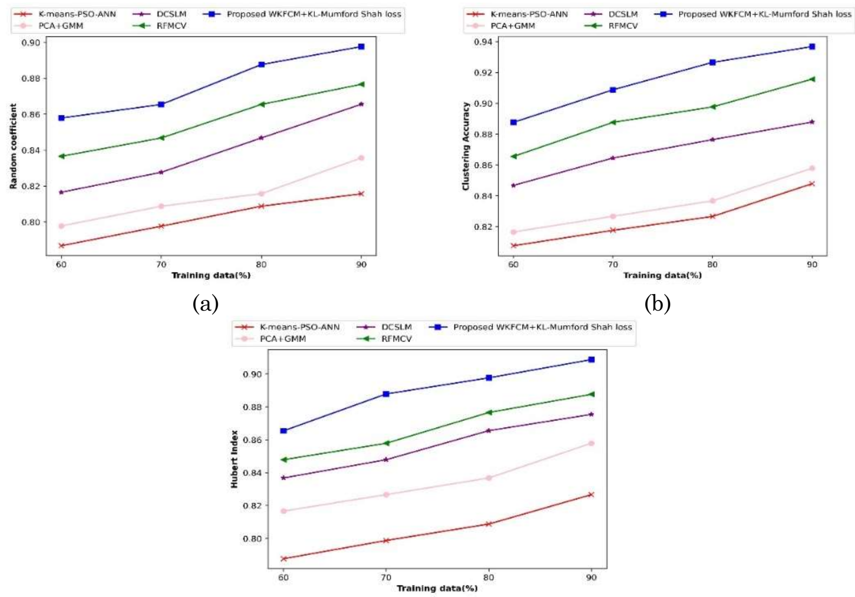
To quantitatively validate the effectiveness of WKFCM+KL-Mumford Shah loss, the performance of WKFCM+KL-Mumford Shah loss is validated with other baseline segmentation techniques. Here, the baseline schemes, such as K-means-PSO-ANN [1], PCA+GMM [2], DCSLM [3], and RFMCV [5] are considered.

4.6. Comparative Analysis

The supremacy of WKFCM+KL-Mumford Shah loss designed for the segmentation of customers in E-commerce is demonstrated in Fig. 6.

Here, the valuation of different customer segmentation models by utilizing the Rand coefficient is depicted in Fig. 6(a). The Rand coefficient of 89.765% is observed by WKFCM+KL-Mumford Shah loss for training data of 90%, whereas the Rand coefficient measured by existing techniques, like K-means-PSO-ANN is 81.567%, PCA+GMM is 83.567%, DCSLM is 86.555%, and RFMCV is 87.666%. The results show that the WKFCM+KL-Mumford Shah loss outperforms other existing schemes with the maximum Rand coefficient. Here, the WKFCM+KL-Mumford Shah loss observed a maximum performance of 9.13% than the K-means-PSO-ANN model. In Fig. 6(b), the validation of the WKFCM+KL-Mumford Shah loss model is introduced in this research, and other existing segmentation schemes are utilized for customer segmentation. From the graph, it is determined that the maximum clustering accuracy of 93.677% is measured by WKFCM+KL-Mumford Shah loss for 90% training data than other existing techniques employed for comparison. Likewise, the existing schemes recorded clustering accuracy of 84.788% by K-means-PSO-ANN, 85.789% by PCA+GMM, 88.786% by DCSLM, and 91.566% by RFMCV. The results show that the

WKFCM+KL-Mumford Shah loss outperforms the existing RFMCV approach with a superior performance of 2.25%. In addition, the comparative evaluation of various approaches used for the segmentation of customers in E-commerce utilizing the Hubert index is depicted in Fig. 6(c). Here, the WKFCM+KL-Mumford Shah loss measured Hubert index of 90.876% for 90% training data, where the prevailing approaches, like K-means-PSO-ANN, PCA+GMM, DCSLM, and RFMCV obtained Hubert index of 82.665%, 85.789%, 87.544%, and 88.765%. It is noticed that the WKFCM+KL-Mumford Shah loss attained superior results of 5.60% than the PCA+GMM approach employed for comparison.



(a) (b) (c)

Fig. 6. *Performance evaluation of WKFCM+KL-Mumford Shah loss utilizing (a) Rand coefficient, (b) Clustering accuracy, and (c) Hubert index*

4.7. Comparative Discussion

The experimental results obtained by the WKFCM+KL-Mumford Shah loss approach introduced for customer segmentation in E-commerce and existing customer segmentation schemes are summarized in Table 2. The results show that the WKFCM+KL-Mumford Shah loss obtained superior performance with R and coefficient of 89.765%, clustering accuracy of 93.677%, and (c) Hubert index of 90.876% for 90% training data. Similarly, the Rand coefficient measured by baseline segmentation techniques, like K-means-PSO-ANN is 81.567%, PCA+GMM is 83.567%, DCSLM is 86.555%, and RFMCV is 87.667%. The clustering accuracy recorded by existing techniques, like K-means-PSO-ANN, PCA+GMM, DCSLM, and RFMCV is 84.788%, 85.789%, 88.786%, and 91.566%, where the techniques also measured Hubert index of 82.665%, 85.789%, 87.544%, and 88.765%. The results proved that the WKFCM+KL-Mumford Shah loss outperforms existing schemes employed for comparison during the segmentation of customers in E-commerce. The WKFCM significantly handled non-linear data by utilizing kernel functions for mapping data in high-dimensional space, which also captured complex patterns for the analysis of behaviors and preferences of customers.

Table 2. *Comparative discussion*

Metrics (%)	**K-means-PSO-ANN**	**PCA+GMM**	**DCSLM**	**RFMCV**	**Proposed WKFCM+KL-Mumford Shah loss**
Rand Coefficient	81.567	83.567	86.555	87.666	89.765
Clustering accuracy	84.788	85.789	88.786	91.566	93.677
Hubert index	82.665	85.789	87.544	88.765	90.876

5. Conclusion

In recent years, customer segmentation has been the most common practice followed in various industries like tourism, education, marketing, and so on. Customer segmentation helps organizations to understand the requirements of various demands of consumers for increasing the decision-making processes. In this article, customer segmentation in E-commerce industries is executed using the WKFCM+KL-Mumford

Shah loss model. Here, the multi-category data is initially acquired from the database and is subjected to catalog partitioning. Later, the portioned data is allowed for customer segmentation using WKFCM by considering the KL-Mumford Shah loss function. Then, the user segmented matrix is constructed for the segmented customers and the customer segmentation in E-commerce is finally performed for the validation of customer data using DEC. Moreover, the superiority of WKFCM+KL-Mumford Shah loss is validated by comparing with existing schemes employed for comparison. Here, the WKFCM+KL-Mumford Shah loss attained superior performance of clustering accuracy of 93.677%, Rand coefficient of 89.765%, and Hubert index of 90.876% than other baseline segmentation models. In future research, advanced deep learning techniques will be utilized for the segmentation of customers in E-commerce.

Compliance With Ethical Standards

Conflicts of interest: Authors declared that they have no conflict of interest.

Human participants: The conducted research follows ethical standards and the authors ensured that they have not conducted any studies with human participants or animals.

References

[1] A. Alghamdi, "A hybrid method for customer segmentation in Saudi Arabia restaurants using clustering, neural networks and optimization learning techniques", Arabian Journal for Science and Engineering, vol.48, no.2, pp.2021-2039, 2023.
[2] J. M. John, O. Shobayo, and B. Ogunleye, "An exploration of clustering algorithms for customer segmentation in the UK retail market", Analytics, vol.2, no.4, pp.809-823, 2023.
[3] S. Wang, L. Sun, and Y. Yu, "A dynamic customer segmentation approach by combining LRFMS and multivariate time series clustering", Scientific Reports, vol.14, no.1, pp.17491, 2024.
[4] K. Tabianan, S. Velu, and V. Ravi, "K-means clustering approach for intelligent customer segmentation using customer purchase behavior data", Sustainability, vol.14, no.12, pp.7243, 2022.
[5] Z. Wu, L. Jin, J. Zhao, L. Jing, and L. Chen, "Research on Segmenting E-Commerce Customer through an Improved K-Medoids Clustering Algorithm", Computational Intelligence and Neuroscience, vol.1, pp.9930613, 2022.
[6] R. S. Wu, and P. H. Chou, "Customer segmentation of multiple category data in e-commerce using a soft-clustering approach", Electronic Commerce Research and Applications, vol.10, no.3, pp.331-341, 2011.
[7] A. Saxena, A. Agarwal, B. K. Pandey, and D. Pandey, "Examination of the Criticality of Customer Segmentation Using Unsupervised Learning Methods", Circular Economy and Sustainability, pp.1-14, 2024.
[8] X. Li, and Y. S. Lee, "Customer Segmentation Marketing Strategy Based on Big Data Analysis and Clustering Algorithm", Journal of Cases on Information Technology (JCIT), vol.26, no.1, pp.1-16, 2024.
[9] A. Afzal, L. Khan, M. Z. Hussain, M. Z. Hasan, M. Mustafa, A. Khalid, R. Awan, F. Ashraf, Z. A. Khan, and A. Javaid, "Customer Segmentation Using Hierarchical Clustering", In Proceedings of 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), pp. 1-6, 2024.
[10] H. J. Wilbert, A. F. Hoppe, A. Sartori, S. F. Stefenon, L. A. Silva, and V. R. Q. Leithardt, "Using Clustering for Customer Segmentation from Retail Data", 2023.
[11] E-commerce Customer Behavior and Purchase Dataset taken from "https://www.kaggle.com/datasets/shriyashjagtap/e-commerce-customer-for-behavior-analysis", accessed on October 2024.
[12] K. Wisaeng, "Retinal blood-vessel extraction using weighted kernel fuzzy C-means clustering and dilation-based functions", Diagnostics, vol.13, no.3, pp.342, 2023.
[13] Y. Liu, K. Liang, J. Xia, S. Zhou, X. Yang, X. Liu, and S. Z. Li, "Dink-net: Neural clustering on large graphs", In Proceedings of International Conference on Machine Learning, pp. 21794-21812, 2023.
[14] F. Yousefirizi, I. Shiri, J. H. O, I. Bloise, P. Martineau, D. Wilson, F. Bénard, L. H. Sehn, K. J. Savage, H. Zaidi, and C. F. Uribe, "Semi-supervised learning towards automated segmentation of PET images with limited annotations: application to lymphoma patients", Physical and Engineering Sciences in Medicine, pp.1-17, 2024.
[15] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis", In Proceedings of International conference on machine learning, pp. 478-487, 2016.
[16] K. Zhan, J. Shi, J. Wang, H. Wang, and Y. Xie, "Adaptive structure concept factorization for multiview clustering", Neural computation, vol.30, no.4, pp.1080-1103, 2018.