Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Efficacy Improvement of Convolutional Neural Network-based Artificial Intelligence Recommendation Model in Cross-border E-commerce Marketing

Juan Liu[1,✉], Hyoungtae Kim[2]

[1] *School of International Communication, Communication University of China, Nanjing, Nanjing, Jiangsu, 211172, China*
[2] *Endicott College, Woosong University, Daejeon, 34606, Korea*

ABSTRACT
This paper optimizes the K-means clustering algorithm based on the RFM model improved by the entropy weight method and then using the distribution between the samples, and adopts the combination of both density and distance to accurately classify the cross-border e-commerce customers. Finally, the capsule network recommendation model is used as the benchmark model, and the CCN4SR model is designed to accurately recommend goods to customers. The results show that cross-border e-commerce customers are categorized into five-star to one-star customer groups, which focus on "return on investment, pursuit of social value, the pursuit of cost-effective, the pursuit of low prices, while having their own different consumer preferences". The capsule network outperforms CNN on both training and test sets, and its precision, recall and F1 value are above 92% on the test set, which shows that the capsule network is well adapted in the field of implicit feedback recommendation.

Keywords: Entropy weighting method, RFM model, K-means clustering, capsule network, cross-border e-commerce

1. Introduction

Cross-border e-commerce refers to the sale of goods to overseas markets through the Internet while the enterprise operates at home. With the development of globalization, cross-border e-commerce is becoming the choice of more and more enterprises [1-2]. However, to be successful in the international market, it is not enough to rely on quality products, and appropriate marketing and promotion strategies are also crucial. Artificial intelligence, as a representative of the new generation of

✉ Corresponding author.
E-mail address: 15850699921@163.com (J. Liu).
Received 20 February 2024; Revised 15 April 2024; Accepted 20 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-507

technology, is gradually applied to cross-border e-commerce, providing more efficient marketing strategies for enterprises [3-6].

Artificial intelligence can help cross-border e-commerce enterprises more accurately locate the target market and customer groups through the analysis of massive data. By collecting and analyzing data such as users' browsing behavior, purchase history, and search keywords, AI is able to construct detailed customer profiles [7-9]. These portraits not only include basic demographic information, such as age, gender, geography, etc., but also cover more in-depth characteristics of consumption preferences, interests, brand loyalty, etc. [10-11]. Based on accurate customer profiling, AI enables personalized marketing content recommendation. Instead of sending out one-size-fits-all advertising emails or push messages to all customers, it provides personalized product recommendations, preferential activities and related information for each customer according to their unique preferences and needs [12-15]. Through AI algorithms, e-commerce platforms can analyze customers' current browsing behavior and historical purchase data in real time, predict the products that customers are likely to be interested in, and display the relevant recommendations to customers at the right time [16-18]. This kind of personalized recommendation can not only improve the customer's purchase conversion rate, but also enhance the customer's satisfaction and loyalty to the brand [19-20].

Literature [21] emphasized the wide application of technologies such as big data, AI, and LOT in cross-border e-commerce, and reviewed the roles played by these technologies in cross-border e-commerce, and proposed an intelligent and integrated cross-border e-commerce platform with a conceptual design. Literature [22] reveals the international trade development trend of smart trade, and points out the obvious advantages of AI in obtaining information, reducing costs, improving efficiency, etc., and AI is having a significant impact on the world division of labor and international trade. Literature [23] based on the huge low-value density cross-border e-commerce consumer behavioral data, designed a precision marketing system applicable to processing user behavioral data to provide reference for enterprises. Literature [24] introduces the information personalization algorithm based on AI cross-border e-commerce platform, implements big data technology on Hadoop platform, uses Map to realize the task decomposition, and adopts Reduce to aggregate the decomposed multi-task processing results to get the processing results, so as to realize the recommendation of products for users. Literature [25] puts forward a proposal for the promotion and application of intelligent prediction of customer satisfaction in cross-border e-commerce, aiming to help enterprises predict customer satisfaction, so as to make informed product development decisions, and provide a theoretical basis for intelligent decision-making in e-commerce product development. Literature [26] explored the big data platform structure and business value of cross-border e-commerce. It not only discusses the multiple applications of big data in cross-border e-commerce such as cross-border affiliate marketing and mobile + social network marketing, but also provides suggestions for cross-border e-commerce platforms and enterprises. Literature [27] proposed a construction plan of precision marketing management system, aiming to realize the precision marketing of cross-border e-commerce platform by using big data analysis and mining means. And the effectiveness of the system is verified through practice, which provides technical support for precision marketing by constructing user profiles. Literature [28] constructed a precision marketing system for cross-border e-commerce with the help of data mining technology. The system design, architecture and other aspects are defined, and the deployment strategy and implementation plan of the system are formulated. Based on the analysis of typical cases, literature [29] examines the current problems of Chinese cross-border e-commerce enterprises in the process of intelligent operation and puts forward suggestions, aiming to promote the rapid development of Chinese cross-border e-commerce enterprises.

This paper proposes a cross-border e-commerce customer value segmentation method based on the entropy power method to improve the RFM model, which utilizes the entropy power method to determine the weights of the three scoring indicators of the RFM model, calculate the scores and evaluate the cross-border e-commerce customer value. Subsequently, the optimal K-mean clustering

model is selected for cluster analysis of cross-border e-commerce customer value using the profile coefficient as a criterion. From the relationship between the samples, the sample density is used to optimize the K-means algorithm, and the clustering effectiveness index is used to compare the accuracy of the optimized K-means algorithm clustering. The clustered customers are segmented and then a convolutional capsule network is used to recommend products to cross-border e-commerce users, and the effectiveness improvement is verified on the relevant dataset.

2. Cross-border e-commerce user segmentation modeling

2.1. Construction of RFM e-commerce customer segmentation model improved by entropy weight method

2.1.1. Traditional RFM models. The traditional RFM model [30] includes three dimensions: proximity R, frequency F, and purchase amount M. R represents the span from the time of the customer's most recent purchase to a certain cut-off time, reflecting the degree of customer activity; F represents the frequency of the customer's purchases during the statistical time, which is used to evaluate the customer's loyalty; and M represents the amount of the customer's purchases during the statistical time, which is used to evaluate the customer's purchasing power. By assigning specific weights to the three indicators R, F and M, each customer value segmentation index score and comprehensive score can be measured, and the higher the comprehensive score, the greater the customer value is considered to be.

2.1.2. Entropy weight method. Entropy weight method is a kind of evaluation theory based on information entropy, which can objectively determine the index weight of the evaluated object and is often used in the evaluation of complex system problems. In this paper, the entropy weight method is used to determine the weights of the three index scores of RFM model, and the specific calculation process is as follows:

Record the score data of RFM model indicators as $IS = \left(is_{ij}\right)_{m\times n}$, where m represents the number of samples, n represents the number of indicators, and the j th indicator value of the ith sample is recorded as $i_{ij}\left(i = 1, 2, \cdots, m; j = 1, 2 \cdots n\right)$.

Calculate the weight of the ith sample's indicator value p_{ij} under the j th indicator with the following formula:

$$p_{ij} = is_{ij} \Big/ \sum_{i=1}^{m} is_{ij} \tag{1}$$

Record the information entropy of the j st indicator as E_j:

$$E_j = -k\sum_{i=1}^{m} p_{ij} \ln p_{ij}, \text{ where } k = 1/\ln m \tag{2}$$

The weight of the j st indicator is noted as w_j, then:

$$w_j = \left(1 - E_j\right) \Big/ \sum_{j=1}^{n}\left(1 - E_j\right) \tag{3}$$

Based on the weights of the indicators obtained from the above equation, the customer value composite score can be calculated and recorded as S:

$$S = IS * W \tag{4}$$

2.1.3. E-commerce Customer Value Segmentation Model with Improved RFM. This paper innovatively proposes the e-commerce customer value segmentation method based on the entropy power method to improve [31] RFM model, the model construction steps are shown in Figure 1, the method contains the following seven steps.

1) Collect and clean e-commerce customer transaction data. Export relevant transaction data from the e-commerce trading platform, and pre-process the data with desensitization, missing values, outliers, etc., to ensure data reliability and accuracy.

2) Set up sub-boxes. Generally, the transaction data of R, F and M indicators are divided into 5 sub-boxes, and each level is assigned 1-5 points respectively.

3) Calculate the raw scores of R, F, M indicators. According to the results of the sub-box and the transaction data of each customer, calculate the original score of R, F, M indicators for each customer.

4) Determine the weights of the R, F, M indicator scores using the entropy weighting method and calculate the indicator scores. Taking the original index score obtained in step 3 as the base data, the entropy weighting method introduced above is used to calculate the weights of the three index scores and multiply them respectively by the original score obtained in step 3 to obtain the index score based on the entropy weighting method.

5) Construct the K-mean model and calculate the contour coefficient value. Considering that the RFM model has three indicators, each indicator score exists greater than (equal to) the average value, less than the average value of two cases, a total of $2 \times 2 \times 2 = 8$ cases, so the maximum number of clusters is 8, and the minimum number of clusters is 2. Therefore, we constructed seven K-mean models, and computed the profile coefficient value of each clustering model respectively.

6) Select the number of clusters corresponding to the maximum contour coefficient value to build the optimal K-mean clustering model. Comparing the contour coefficient values obtained in step 5, the number of clusters corresponding to the largest contour coefficient is used as the category parameter of the clustering model to establish the K-mean clustering model to determine the categorization of each sample.

7) Analyze customer characteristics based on clustering results. Analyze the characteristics of different categories of customers, compare the value of different customer groups, and propose differentiated marketing strategies.

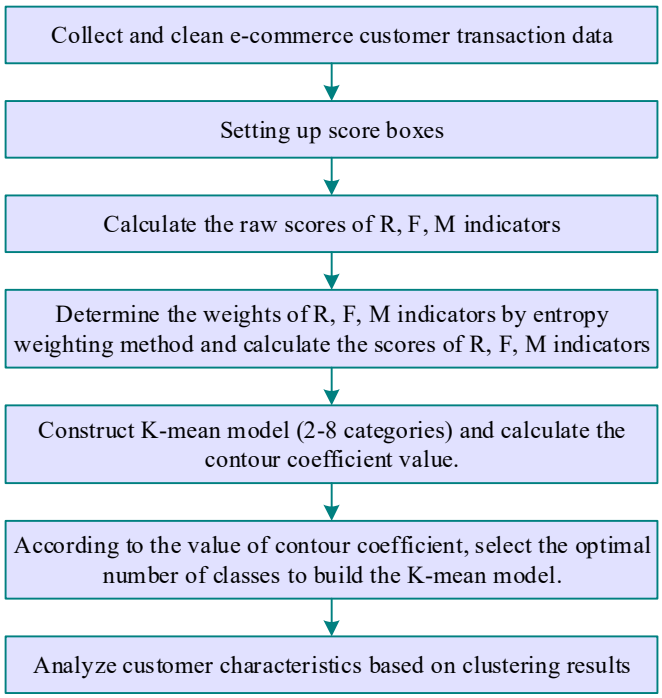

Fig. 1. Model build steps

2.2. Optimization algorithm and results of K-means clustering algorithm

2.2.1. K-means algorithm. In cluster analysis, K-means clustering is one of the most common data mining algorithms, which is a division-based clustering method proposed by Macqueen. The K-means algorithm is fast in clustering, simple and fast in operation, but there are some defects in the clustering process, such as relying on the random selection of the initial clustering centers, being highly susceptible to the influence of outliers, and unstable clustering results. The algorithm usually uses the Euclidean distance as a measure of similarity between two objects to divide the clustering results, and its basic idea is to choose any k initial clustering center, calculate the Euclidean distance between the remaining data objects and the clustering center, find the closest distance to the $k-1$ clustering objects, and continuously update the iterative clustering center until the sum of squares of the errors (SSE), that is, the criterion function converges, and get the clustering results, which are denoted as $C=\{C_1, C_2, \cdots, C_k\}$.

Assuming a dataset $U \in \{x_p\}(p=1,2,\cdots,n)$ with n m -dimensional attributes, denoted $c_i(i=1,2,\cdots,k)$ as k clustering centers, and each clustering center c_i with m -dimensional attributes, denoted $c_{ij}(j=1,2,\cdots,m)$, the Euclidean distance of each object x_p from each clustering center c_i is defined:

$$Dist(x_p, c_i) = \sqrt{\sum_{j=1}^{m}(x_{pj} - c_{ij})^2} \tag{5}$$

Error sum of squares definition:

$$SSE = \sum_{i=1}^{k}\sum_{x \in C_i}\left|Dist(x, c_i)\right|^2 \tag{6}$$

2.2.2. Optimized K-means algorithm. K-means, as one of the unsupervised learning algorithms, does not need to know the clustering categories in advance and is able to cluster unidentified objects. The Euclidean distance is utilized as a similarity measure, which yields that the smaller the distance between clusters of the same class, the higher their similarity, and the larger the distance between clusters of different classes, the lower their similarity. This method will change the closeness and dispersion of class clusters in the clustering process because of the influence of extreme values, which reduces the accuracy of the whole clustering. In this paper, from the relationship between the samples, the first use of high density instead of distance mean, using formula (7) to calculate the nearest-neighbor density to select the higher density of samples in the dataset as the first initial clustering center c_1, and the rest of the objects are divided into the categories of the identified clustering centers.

$$\bar{X}^* = \frac{1}{n}\sum_{p=1}^{n} x_p \tag{7}$$

where $x_p(p=1,2,\cdots,n)$ denotes n a m -dimensional dataset.

Second, among the remaining objects that are not classified, the Euclidean distance is used to find the next temporary clustering centroid c_k furthest from c_1 and clustered using the following equation:

$$c_i = \max_{1<p<q<n}\left\{Dist(x_p, x_q)\right\} \tag{8}$$

$$Den_\lambda(x_{p_i}, c_i) = \left\{x_{p_i} \middle| Dist(x_{p_i}, c_i) \le \lambda\right\} \tag{9}$$

where, $Den_\lambda\left(x_{p_i}, c_i\right)\left(p_i = 1,2,\cdots,n_i\right)$ indicates all data objects whose distance from the clustering center c_i is less than λ; n_i indicates the number of objects in $Den_\lambda\left(x_{p_i}, c_i\right)$; A is an arbitrary normal number.

The formula for λ is:

$$\lambda = A * Dist\left(x_{p_i}, \bar{X}^*\right) \tag{10}$$

Calculate the density of this temporary clustering center using the following formula, and search for the data object closest to the average of its density as the updated clustering center; and so on iteratively until all initial clustering centers are obtained. The bipartite clustering result of all data objects is obtained, denoted as $C' = \left\{C_1', C_2', \cdots, C_k'\right\}$. Eq:

$$\overline{Den_\lambda\left(x_{p_i}, c_i\right)} = \frac{1}{n_i} Den_\lambda\left(x_{p_i}, c_i\right) \tag{11}$$

K-means clustering algorithm is usually arbitrary selection of k clustering centers, usually with a certain degree of randomness. This paper proposes a method to optimize the selection of the initial clustering center, using the sample distribution information, select the point with the highest density as the initial clustering center, which effectively solves the problem that the clustering falls into the local optimum due to the interference of human factors or the influence of extreme values; restricts the clustering object to be within λ, which is more accurate and far away from the interference of the noise points; and in the final clustering result, it is able to meet the condition that the clusters of the same class have the highest similarity degree and the clusters of different classes have the lowest similarity degree, which ensures the stability of clustering. The final clustering result can satisfy the condition of the highest similarity degree of clusters of the same class and the lowest similarity degree of clusters of different classes, which ensures the stability of clustering.

2.2.3. Clustering accuracy of optimized K-means algorithm. In order to verify whether the optimization algorithm can correct the K value, this study uses the make_blobs function to generate the data set, according to the 7 data centers, the number of samples is 45000 to generate the data set, and the random seed is set to 200. Then the elbow-point method is used to obtain the a priori K value of the interval, and the values are selected for clustering. After correcting the K-value using the improved algorithm, the data were then clustered. The contour coefficients obtained by using the elbow point method are shown in Fig. 2. The elbow point is not obvious in the graph, and the empirical K value interval is between [4,7].

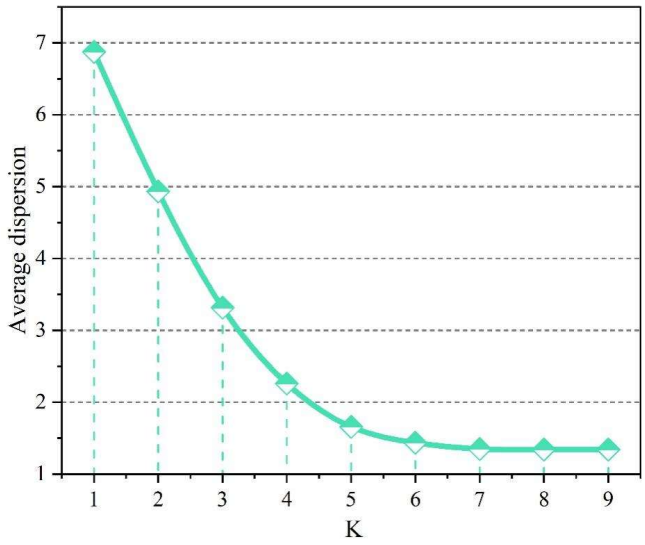

Fig. 2. Contour coefficient map

K-means clustering was performed according to K=4. The clustering results for K=4 are shown in Fig. 3. It can be found that there are isolated points in the clustering results, and some edge points are classified into the wrong category, which are caused by the wrong selection of K value.

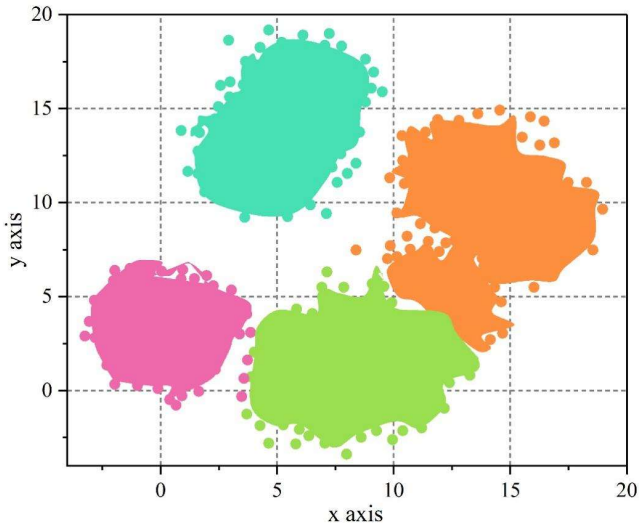

Fig. 3. Clustering results of K=4

Using the improved algorithm to correct the K value, the clustering result of K=8 is shown in Fig.4,the corrected K value is 8,the corrected K value clustering result is shown in Fig.4. This shows that the improved algorithm is more accurate than the elbow-point method, which effectively finds the K-value and verifies the accuracy of the algorithm.

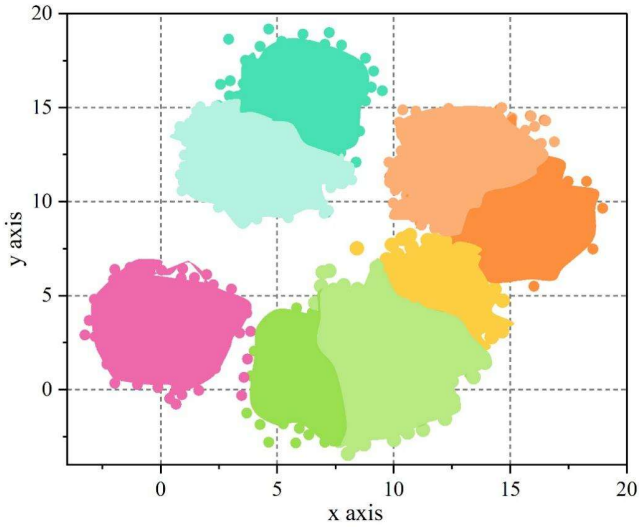

Fig. 4. Modified k value clustering results

3. Improving the effectiveness of K-means clustering method for customer segmentation

3.1. RFM Scores by Customer Segments

The 15972 consumption records of 1926 customers are classified into 8 sub-categories by the RFM model, which means that the number of customer categories classified according to the traditional

RFM method is too many, which is not conducive to the proposal of the subsequent marketing optimization scheme. Therefore, clustering is carried out on the basis of RFM to achieve the purpose of scientific and reasonable classification of these 1926 customers. In this paper, two indicators are used to determine how much K value is in line with the practical requirements, and the clustering effect is the best. One of the indicators is the average distance within a class, and the other is the average distance between classes. Generally speaking, the smaller the intra-class average distance and the larger the inter-class average distance, the better the clustering effect.

In this paper, "a cross-border e-commerce" background retrieval of 1926 customers will be subdivided into 8 categories, the RFM score of each subdivided customer group is shown in Table 1. From the table, it can be seen that the general value of the largest number of customers, followed by high-value customers, their RFM scores were 116,947 and 78,197; and focus on retaining the smallest number of customers, their RFM score is only 14,596.

Table 1. RFM scores of each segment of the customer base

Customer category	MS	FS	RS	RFM score
High value customer	349	350	354	78197
Potential customer	127	124	126	21369
General customer	163	338	165	30878
General development customer	172	172	340	69804
General value customer	465	527	527	116947
Focus on maintaining customers	215	217	106	20729
Key development customer	227	114	225	46904
Focus on retaining customers	208	84	83	14596

3.2. Clustering and Analysis of Segmented Customer Groups

3.2.1. Determination of K-value in clustering process. The number of customer categories classified according to the traditional RFM method above is too many, which is not conducive to the proposal of the subsequent marketing optimization scheme. Therefore, K-means clustering is carried out on the basis of RFM to achieve the purpose of scientific and reasonable categorization of these 1926 customers.

The determination of K value in the clustering process is shown in Figure 5. The results show that as the number of clusters becomes larger, the average distance within a class becomes smaller and smaller, and the corresponding average distance between classes also becomes larger and larger. However, we will find a very significant feature, that is, when the value of K is taken as 5, both the intra-class average distance and the inter-class average distance have an inflection point, that is, both the intra-class average distance and the inter-class average distance are no longer so obvious whether they are shrinking or enlarging the numerical value of the change, especially the inter-category average distance, when the value of K is greater than 5, there is no longer a change in the value of the value of the value of a fixed at 0.9512. 0.9512. Then, from the point of view of intra-class average distance as well as inter-category average distance, the final number of clusters in this paper, K, is taken as 5.

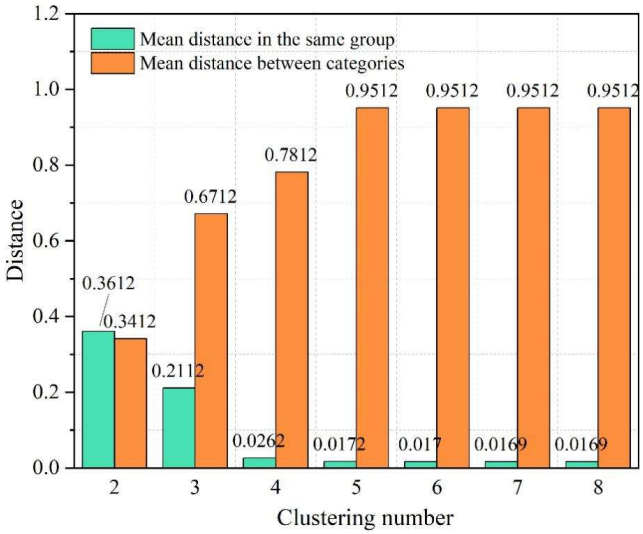

Fig. 5. The k value is fixed in the clustering process

When the value of K is 5 the changes in both the mean distance within the category and the mean distance between the categories tend to relatively stable values, the number of cases in the clusters and the percentage of the results are shown in Table 2. It has been seen to show that the number of people in the five categories are 266, 419, 373, 505 and 363, and their percentages in order are 13.81%, 21.75%, 19.37%, 26.22% and 18.85%, respectively. It can be seen that the improved K-means clustering results in K = 5, the number of people in the five categories accounted for a uniform proportion, the clustering effect is good.

Table 2. The number of cases in the cluster and the proportion of the cases

Cluster name	Clustering number	The number of cases in the cluster is the ratio(%)
Class 1	266	13.81
Class 2	419	21.75
Class 3	373	19.37
Class 4	505	26.22
Class 5	363	18.85

The customer basic information as well as transaction records of the five categories of segmented customer groups are compared with the RFM values to get the customer behavioral characteristics as well as consumption preferences. The characteristics of each category of segmented customer groups are shown in Table 3. The results show that Cluster 3 is high-value customers, Cluster 5 is key retention customers, Cluster 4 is key development customers, Cluster 1 is key retention customers, and Cluster 2 is general value customers. Among them, high-value customers have the highest purchase amount of 10,341 yuan; key retention customers have the highest purchase frequency (38 times); key development customers have the highest percentage (26.22%), and key retention customers have the shortest shopping intervals (12 days); while general value customers have the lowest purchase frequency (9 times) and purchase amount (1,022 yuan), and the longest purchase intervals (46 days).

Table 3. The characteristics of each category of segmentation customers

Cluster name	Customer type	Customer ratio(%)	Purchase frequency (times)	Purchase gold (yuan)	Time interval (day)
Class 3	High value customer	13.81	21	10341	14
Class 5	Focus on maintaining customers	21.75	38	8271	23
Class 4	Key development customer	19.37	19	6819	28
Class 1	Focus on retaining customers	26.22	11	3317	12
Class 2	General value customer	18.85	9	1022	46

3.2.2. Analysis of five major categories of customer segments:

1) Overview of high-value customer groups: the age of most of these customer groups is between 30 and 50 years old, with most of them being male customers, and they have a high level of education, with nearly 80% of them having received undergraduate education or above. Their jobs are relatively stable and decent, and their social status is high. They usually work in state-owned enterprises and public institutions, and most of them are national public officials. Therefore, they have a stable source of income and strong purchasing power, with an annual salary of more than 120,000 yuan. The proportion of this customer group is not too high, accounting for about 20%, or 19%. High-value customers generally focus on purchasing high-end bank financial products. In this paper, this type of customer group is referred to as the five-star customer group.

2) Overview of Key Maintained Customer Groups: The age of this part of the customer group is mostly between 28 and 45 years old, with most of the customers being female, and the education level is also higher, with most of the academic backgrounds above the college level, and the jobs are also more stable and decent, including civil servants, employees of listed companies of local leading enterprises, and those who are working at the bank, and thus they also have a stable income, with an annual salary of about 80,000 to 120,000 yuan, which also accounts for about 20% of the sample size. Roughly about 20% of the sample size, the specific percentage is 18.08%. This customer segment has a significant characteristic of pursuing quality life, and has a large number of transaction records in "All Shop", only that the average amount of goods purchased is less than that of the high-value customer group, and the main shopping categories are medium- to high-end cell phones, white goods and other electronic products, medium- to high-end bank finance, as well as medium- to high-end cosmetics, medium- to high-end clothes, etc., which basically cover all kinds of products and services. Their main shopping categories are medium- to high-end mobile phones, white goods and other electronic products, medium- to high-end bank finance, and medium- to high-end cosmetics, medium- to high-end clothes, etc., which basically cover the categories of daily life goods. Most of the commodities consumed also belong to higher gross profit related products. In this paper, this type of customer group is called the four-star customer group.

3) Overview of key development customer groups: most of these customer groups are between the ages of 25 and 35, mostly female customers, with a high level of education, most of them with an educational background of more than a bachelor's degree, and generally work in local small and medium-sized enterprises or open their own stores to do small business. This type of customer group probably accounts for about 30% of the sample size. In this paper, this kind of customer group is called three-star customer group.

4) Overview of key retention customer groups: the age of most of these customer groups is between 20 and 30 years old, and most of them are female customers, mainly employees of small and medium-sized enterprises, and the proportion of key retention customers is 10.08%, which is not too big in relative terms. In this paper, this kind of customer group is called two-star customer group.

5) Overview of general value customer groups: the customer groups in this sector are mostly between the ages of 25 and 36, mainly male customers, relatively speaking, the level of education is also lower, and most of the personnel's educational background is at the level of vocational high school

and below. Most of them are working in micro and small enterprises or have no fixed occupation, so their source of income is not stable, and the proportion of this type of customers is about 22.83%. In this paper, this type of customer group is called one-star customer group.

4. Cross-border e-commerce marketing based on convolutional capsule networks

4.1. Capsule Networks

1) Concept of capsule network. Capsule network [32] is used to solve the "invariance" problem of convolutional neural network, the direction of the capsule vector indicates the position, direction, size, color and other information of the entity. The relationship between low level features and high level features can be expressed by the following equation:

$$v_{j|i} = W_{ij} \cdot u_i \tag{12}$$

where u_i represents the low level gel table, w_{ij} represents the relationship between the low level capsule and the high level capsule, and $v_{j|i}$ represents the high level capsule predicted from the low level capsule.

2) Dynamic routing algorithm. The dynamic routing process of the capsule network is shown in Fig. 6. It describes the process of two low-level capsules v_1, v_2 through the iterative dynamic routing mechanism to get the high-level capsule v^1.

The purpose of the softmax function in the dynamic routing algorithm is to normalize each term, defined as follows:

$$soft\max\left(b_{ij}\right) = \frac{\exp\left(b_{ij}\right)}{\sum_k \exp\left(b_{ik}\right)} \tag{13}$$

The Squash function is a nonlinear squeezing function that aims to make the modulus of the vector in the range of 0 to 1. The formula is defined as:

$$Squash\left(s_j\right) = \frac{\left\|s_j\right\|^2}{1+\left\|s_j\right\|^2} \cdot \frac{s_j}{\left\|s_j\right\|} \tag{14}$$

Dynamic routing is continuously updated b_{ij} during the iteration process, which makes it possible to obtain a more accurate relationship between the lower and higher level capsules after each iteration than the previous one.

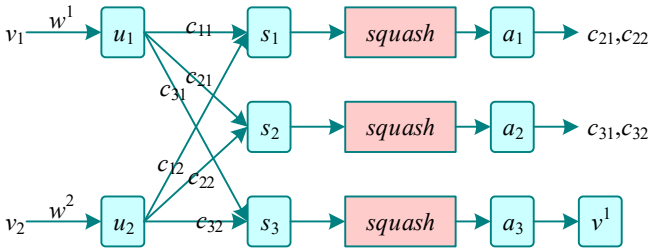

Fig. 6. The dynamic routing process of the capsule network

3) Sequence Recommendation. Sequence recommendation models the user-item interaction as a dynamic sequence and uses sequence dependencies to capture the user's current and recent preferences, and the sequence recommendation system is represented by a function that maximizes the following:

$$R = \arg\max f(S) \tag{15}$$

where f is the utility function that outputs the sorting score of candidate items, $S = \{i_1, i_2, \cdots, i_{|S|}\}$ represents the sequence of interactions between the user and the item, each interaction in S is denoted as $i_j = \langle u, a, v \rangle$, u in the tuple denotes the user, a denotes the user's behavior (clicking, purchasing, adding to cart, etc.), and v represents the item that is interacted with by the user. R represents the sequence of items generated based on the sorting score.

4.2. CCN4SR Design

The structure of CCN4SR is shown in Fig. 7.CCN4SR consists of the following components: embedding layer, convolutional capsule layer and prediction layer. To better model the sequence task, for each user u, the user sequence is denoted as $S^u = \left(S_1^u, \cdots, S_{|S^u|}^u\right)$, where S_i^u denotes the ith item in S^u that interacts with the user u. The CCN4SR extracts every $|L|$ consecutive item as an input, i.e., $|L| = \left(S_J^u, \cdots, S_{J+|L|-1}^u\right)$, and takes the next $|T|$ consecutive item as the target to be predicted.

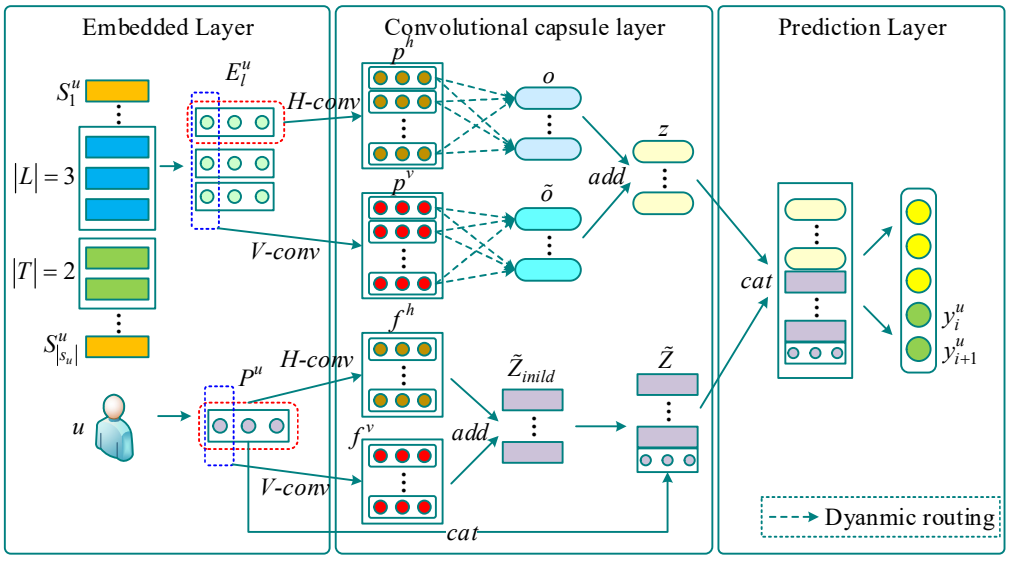

Fig. 7. CCN4SR model frame diagram

1) Embedding layer. The input to the embedding layer is a continuous sequence of items which embeds each item as a vector. The embedding containing *L* items is denoted as:

$$E_l^u = \begin{bmatrix} e^{S_J^u} \\ e^{S_{J+1}^u} \\ \vdots \\ e^{S_{J+|L|-1}^u} \end{bmatrix} \tag{16}$$

where $E_l^u \in R^{|L|\times d}$ is the embedding of the lrd sequence of user u and the embedding of user u is denoted as $P^u \in R^d$.

2) Convolutional capsule layer. The convolutional capsule layer consists of two parts: sequence pattern feature extraction part and user feature extraction part.CCN4SR uses horizontal and vertical capsule networks to learn sequence pattern features and attribute features of sequence items.CCN4SR uses both horizontal and vertical convolutional capsule networks to learn more specific and finer-grained user features, and then obtains the complete long-term user preferences through residual concatenation.

The horizontal capsule network has n horizontal convolutional kernel $H^k \in R^{1\times d}$, where $1 \le k \le n$. H^k slides from the top to the bottom of matrix E with the purpose of interacting with all the horizontal dimensions of item S_h^u of matrix E_l^u, where $J \le h \le J + |L| - 1$, the result of the interaction of the hth convolutional kernel is represented as:

$$c_h^k = \varnothing c\left(E_h \square H^k\right) \tag{17}$$

where $\varnothing c(\cdot)$ is the activation function for horizontal convolution, $\square$ denotes the inner product operation, and E_h is the submatrix consisting of the h rows of E_l^u. c_h^k denotes the result of the horizontal convolution of H^k and E_h, and the final convolution result of H^k is denoted as:

$$c^k = \left[c_J^k c_{J+1}^k \cdots c_{J+|L|-1}^k\right] \tag{18}$$

Then, the final convolution result of each horizontal convolution kernel is initialized as a primary capsule, and the set of primary capsules p^h is denoted as:

$$p^h = \left[c^1, c^2, \cdots, c^n\right] \tag{19}$$

There are n primary capsules in p^h. The primary capsules are fed into the capsule network as low level capsules. Then the fuzzy representation of the high level capsule is generated by feature mapping operation and the fuzzy representation of the ird high level capsule q_i is given by the following equation:

$$q_i = W_h \cdot p_i^h \tag{20}$$

where p_i^h is the ind primary capsule, W_h is the mapping matrix to represent the relationship between the lower and upper capsules, and the set of fuzzy upper capsules q^h is denoted as:

$$q^h = \left[q_1, q_2, \cdots, q_n\right] \tag{21}$$

The final explicit representation of the high-level capsules is computed by an iterative dynamic routing mechanism. Taking q_i as an example, first, the candidate vector R_j for the jth high-level capsule is obtained by weighting the coupling coefficients C_{ij} of the ind low-level capsule and the jrd high-level capsule, C_{ij} given by the following equation:

$$C_{ij} = soft\max\left(b_{ij}\right) \tag{22}$$

where b_{ij} is the consistency score between the ind low-level capsule and the jrd high-level capsule, and b_{ij} is initialized to 0. The candidate vector R_j is given by the following equation:

$$R_j = \sum_1^i C_{ij} \cdot q_i \tag{23}$$

The high level capsule is further represented by the probability of the existence of the attribute by compressing its mode length to between 0 and 1 after the squash function, which is given by the following equation:

$$o_j = \text{squash}\left(R_j\right) = \frac{R_j}{\left\|R_j\right\|} \cdot \frac{\left\|R_j\right\|^2}{1+\left\|R_j\right\|^2} \tag{24}$$

After each iteration of dynamic routing, b_{ij} is updated by the following equation:

$$b_{ij} = b_{ij} + o_j \cdot q_i \tag{25}$$

After dynamic routing, the output O is a collection of clear capsules containing M high levels, which is denoted as:

$$O = \left[o_1, \cdots, o_j, \cdots, o_M\right] \tag{26}$$

Vertical Capsule Network: each item may have multiple attributes which can better describe the short-term preferences of the user.CCN4SR uses a vertical capsule network to model the characteristics of each attribute in the item. It contains $\tilde{n}$ convolution kernel $\tilde{H}^k \in R^{L\times 1}$, where $1 \le k \le \tilde{n}$. $\tilde{H}^k$ slides from left to right on matrix E, initializes the convolution result as primary capsule p^v, inputs p^v into the vertical capsule network, and finally computes the high-level clear capsule representation of the vertical capsule network as $\tilde{O}$. CCN4SR represents short-term preferences by fusing $\tilde{O}$ and O by bitwise summation, and short-term preference Z is given by the following equation:

$$Z = \varnothing Z\left(\tilde{O} \cdot W_{vh}\right) + O \tag{27}$$

where W_{vh} is a trainable weight matrix, $\varnothing Z\left(\cdot\right)$ is the activation function, and Z represents the user's short-term preferences.

User feature extraction: the user feature extraction layer has n horizontal convolutional kernel $V^k \in R^{1\times d}$ and $\tilde{n}$ vertical convolutional kernels $\tilde{V}^k \in R^{1\times 1}$, where ∇^k is able to learn the features of each dimension (attribute) of P^u, and V^k is able to learn the features of all dimensions in P^u. The result of the convolution is denoted as f^h and f^v; then, CCN4SR generates more specific and fine-grained user features by fusing f^h and f^v by bitwise summation, which is given by the following equation:

$$\tilde{Z}_{initial} = \varnothing \tilde{Z}_{initial}\left(f^v \cdot W_{\tilde{v}h}\right) + f^h \tag{28}$$

where $W_{\tilde{v}h}$ is a trainable weight matrix, $\varnothing \tilde{Z}_{initial}$ is the activation function, $\tilde{Z}_{initial}$ represents more specific and fine-grained user features, and finally, CCN4SR uses residual linking operation to ensure the completeness of the user's long term preference features, and the final result of the long term preference is given by the following equation:

$$\tilde{Z} = \begin{bmatrix} Z_{initial} \\ P^u \end{bmatrix} \tag{29}$$

where $\left[\cdot\right]$ is the cascade operation and $\tilde{Z}$ denotes the user's long-term preference.

3) Prediction Layer. In this layer, firstly Z and $\tilde{Z}$ are spliced to generate the combined preference representation of the user and then further the prediction is generated for the user and the prediction process is given by the following equation:

$$y^u = \overline{W}\begin{bmatrix} Z \\ \tilde{Z} \end{bmatrix} + \overline{b} \tag{30}$$

where $\overline{b}$ and $\overline{W}$ are the bias terms and weight matrices of the prediction layer. y_i^u denotes the probability that user u interacts with item i.

4) Network Training. The objective function of CCN4SR is defined as the binary cross-entropy loss:

$$Loss = \Sigma_u \Sigma_{J \in Y^u} \sum_{i \in D_i^u} -\log\left(\sigma\left(y_i^u\right)\right) + \sum_{o \neq i} -\log\left(1 - \sigma\left(y_i^{(u,t)}\right)\right) \tag{31}$$

where $Y^u = \left\{1, \cdots, \left|S^u\right| + 1 - L\right\}$ is the value of J in $|L|$. CCN4SR also considers the next $|T|$ consecutive objective items, where $D_i^u = \left\{S_{J+|L|}^u, S_{J+|L|+1}^u, \cdots, S_{J+|L|+|T|}^u\right\}$. $\sigma(\cdot)$ is the sigmoid function. Secondly, CCN4SR optimizes the objective function using the Adam optimizer, which can automatically adjust the learning rate of each parameter during the training process.

4.3. Model Evaluation and Parameter Adjustment

4.3.1. Selection of assessment indicators. Based on the characteristics of the dataset used in this paper, it is difficult to measure the effectiveness of the model by the method of predicting the scoring error because the user behavior data in the dataset are all implicit feedback data. At the same time, after sample screening, the proportion of positive and negative samples in the current dataset has been equalized. Predicting whether a user buys a certain product essentially belongs to a dichotomous problem, combining the above three points, it was decided to adopt the calculation of precision rate, recall rate and F1 value as the evaluation index of the algorithmic model.

Research object: the dataset is a total of 4G, which contains the behavioral data of nearly 1 million users between October 1, 2023 and September 30, 2024, with a total of more than 100 million pieces of behavioral data. After data preprocessing, 15,972 consumption records of 1,926 customers were sampled for analysis. In the dataset, each behavioral data corresponds to a user ID, a product ID, a product class ID, a behavior, a timestamp, and the product ID corresponds to only one product class ID.

When we predict whether a user purchases an item on November 1-3, we may get four scenarios, i.e., predicted purchase and actually purchased, predicted purchase and actually did not purchase, predicted no purchase and actually purchased, predicted no purchase and actually did not purchase. The prediction results of the model are shown in Table 4 below: by discussing the above scenarios, we obtain the confusion matrix that can determine whether a user purchases an item or not, and based on the confusion matrix, we can introduce the computation of the precision rate, recall rate, and F1 value.

Table 4. The prediction of the model

Results	Purchase (positive example)	Unpurchased (negative example)
Purchase (positive example)	Real example (TP)	False case (FN)
Unpurchased (negative example)	False example (FP)	True negative (TN)

Precision rate that is, when we are predicting, we determine how many of the results predicted as purchases are the result of real purchases, the formula is as follows:

$$\Pr ecision = \frac{TP}{TP + FP} \tag{32}$$

Recall i.e. when we are predicting, how much of the results of their purchases we predict are accurate in the results of real purchases, i.e:

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{33}$$

The $F1$-value is a synthesis of both precision and recall, and since it takes a range of values between $(0,1)$, it also facilitates the evaluation of the effect between models, and when the value of $F1$ is closer to 1, it proves that the model predicts better, $F1$-value Eq:

$$F1 = \frac{2 \times \mathrm{Pr}\,ecision \times \mathrm{Re}\,call}{\mathrm{Pr}\,ecision + \mathrm{Re}\,call} \tag{34}$$

4.3.2. Adjustment of model parameters. In the field of deep learning, the model often does not converge at all, partially converge and all converge but not good, for the above three cases, we need to categorize and analyze the problems. When not converging at all, we need to consider whether the values in the dataset are poor or have anomalies, whether they are generalized when constructing features, and whether the amount of values should not be too large or too small. When all converge but poorly, we only need to adjust one of the parameters and keep the other parameters unchanged, the way of manual parameter adjustment needs to be changed according to the actual situation. When partially converged, this situation is more complex, there may be a variety of parameters are not ideal state, this time you can prioritize from the activation function, loss function to start.

4.3.3. Capsule network prediction results. The capsule network training results are shown in Figure 8: From the figure, we can see that the capsule network is relatively slow throughout the training iteration process, and the convergence speed is much slower compared to the CNN network. We set the maximum number of iterations to 130. In the training set, the capsule network iteration number reaches the peak of accuracy around 124 times and tends to level off, we choose the result of the 124th iteration as the model accuracy, at this time, the accuracy of the model is 98.09%. For the test set, after we input the data of the test set, we also select the result of the 124th iteration as the accuracy result of the test set, and finally the precision rate, recall rate, and F1 value are as follows.

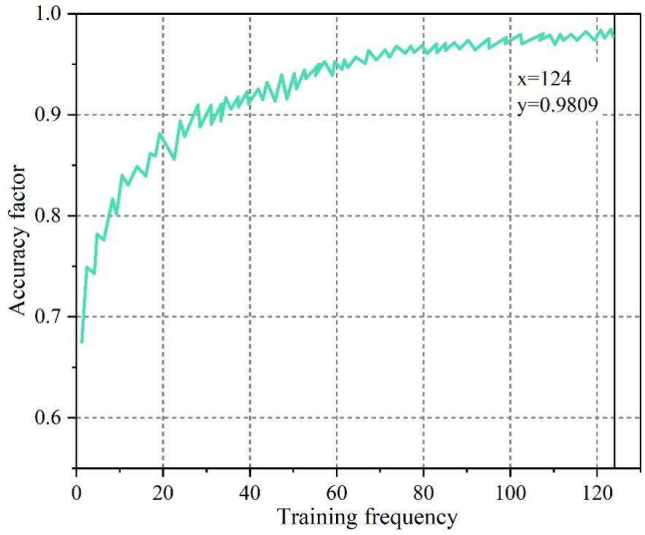

Fig. 8. Network training results of capsules

The comparison of experimental results is shown in Table 5 below. The results show that the capsule network outperforms the CNN on both the training set and the test set. The capsule network achieves 11.99%, 8.5%, and 11.96% higher precision, recall, and F1 values on the test set than the CNN on the

test set, respectively. The dynamic routing algorithm in the capsule network makes the information as much as possible preserved in the process of transmission, and this connection allows better access to the implicit information present in the features and mining deeper feature relationships. It can be seen that capsule networks have better adaptability in the field of implicit feedback recommendation, compared to the more widely used CNN, which has a better performance effect.

Table 5. Comparison of experimental results of different models

Data set	Model	Precision(%)	Recall(%)	F1(%)
Training set	CNN	81.25	88.34	82.57
	Capsule network	93.75	95.88	93.76
Test set	CNN	80.15	86.43	80.19
	Capsule network	92.14	94.93	92.15

4.3.4. Getting a list of product recommendations. The preference for a commodity will reflect the user's demand for a certain commodity or a certain type of commodity, and the user's behavioral characteristics can reflect the user's preference for the commodity. By assigning the user's behavior, the user's rating of the commodity can be obtained. Since the user's behavior is divided into four kinds of browsing, collection, add purchase, purchase, will be interested in the weight assigned to 1, 2, 3, 4. These four kinds of behavior on the user's final purchase behavior of the impact of the different, according to the results of the conversion rate of the user's behavior as shown in the preceding, browsing behavior of the conversion rate is the lowest, add the purchase of the next behavior, the collection of behavior conversion rate is relatively high. Through the weighted assignment of user-commodity behavior, the user-commodity scoring matrix can be obtained. For the missing values, we fill them with "0" values. Obtain TOP-6 similar users:

After constructing the user-item scoring matrix, the similarity can be calculated for the users within the scoring matrix. Generally speaking, the metrics used to calculate similarity are two kinds of Pearson coefficient and cosine similarity, and the formulas for Pearson coefficient and cosine similarity are as follows:

Pearson coefficient:

$$Sim(a,b)=\frac{\sum_{i\in I_{ab}}\left(R_{ai}-\overline{R_a}\right)\left(R_{bi}-\overline{R_b}\right)}{\sqrt{\sum_{i\in I_{ab}}\left(R_{ai}-\overline{R_a}\right)^2}\sqrt{\sum_{i\in I_{ab}}\left(R_{bi}-\overline{R_b}\right)^2}} \tag{35}$$

Cosine similarity:

$$Sim\left(A,B\right)=\frac{\sum_{i=1}^{n}\left(A_i\times B_i\right)}{\sqrt{\sum_{i=1}^{n}\left(A_i\right)^2}\times\sqrt{\sum_{i=1}^{n}\left(B_i\right)^2}} \tag{36}$$

By calculating the similarity between users, a user rating similarity matrix can be obtained. After obtaining the similarity between users, the nearest similar neighbors to the target user can be obtained, and here the top-6 neighbors that are most similar to the user are selected for the subsequent generation of the product recommendation list. The TOP-6 similar users of the target user are shown in Table 6.

Table 6. Target user TOP-6 similar user

user ID	Similar user ID of TOP-6					
41	1235	887	194	540	926	781
263	1446	1054	704	281	622	1700
676	23	853	2208	1355	1823	997
796	1085	740	873	669	939	826
1224	597	1849	920	584	940	1202
1508	1469	510	965	984	918	1028
…	…					

Getting a list of product recommendations:

We choose to get the 6 users who are most similar to the target user as the variable for generating the product list. By obtaining the purchase records of the 6 similar neighbors, we can determine the preferences of the similar neighbors, which can also react to the preferences of the target user, so we can recommend the commodities in the purchase records, to the target user, so that we can realize the commodity recommendation. The user commodity recommendation set is shown in Table 7: the user as the target can realize the commodity recommendation to the user, and the commodity as the target can realize the accurate promotion of the commodity. When it is necessary to promote commodities, it is possible to query the user commodity recommendation set to obtain potential users who are interested in the target commodities, and by advertising to these users, it is possible to save the advertising cost of merchants, and improve the turnover rate of commodities and the satisfaction of users to the e-commerce platform.

Table 7. Results recommended by user products

user ID	User recommendation set					
41	7248	10146	13553	9896	14380	14447
263	8277	8311	10606	14572	14704	9025
676	13676	15065	9701	1384	1494	12655
796	8917	13554	10469	14689	13651	7788
1224	15243	8358	5746	1018	2938	15259
1508	9613	9415	10805	15347	13796	14511
…	…					

5. Conclusion

In this paper, after improving the e-commerce customer value segmentation model of RFM, the sample density optimized K-means algorithm is used to divide the cross-border e-commerce customer group into five categories, and then the convolutional capsule network is used to make accurate recommendation of goods to the divided customers.

1) The clustering results determined by the optimized K-means algorithm are more accurate than the traditional clustering algorithm. In customer segmentation, when the value of K is taken as 5, the change in the value of the average distance within the class and the average distance between the classes of the samples, whether shrinking or expanding, is no longer so obvious, and at this time, its value is fixed at 0.9512, which ultimately determines the number of clusters (K=5).

2) According to the customer's consumption frequency, consumption amount and consumption interval time, the customer population of the platform is divided into five categories of customer segments, and at the same time summarizes the consumption behavior characteristics of these five categories of customer segments. Five-star customer groups focus on return on investment, four-star

customer groups pursue social value, three-star customer groups pursue cost-effective, two-star and one-star customers have their own different consumption preferences while pursuing low prices. These five categories of segmented customer groups basically encompass the customer groups of cross-border e-commerce.

3) On the test set, the capsule network outperforms CNN by 8.5%-11.99% in terms of precision, recall and F1 value. The predicted day user is taken as the target user, the TOP-6 users who are most similar to the target user are obtained, and the list of goods that have been purchased by the similar users is recommended to the user, forming a recommended goods recommendation column and realizing the process of accurate recommendation of goods.

References

[1] He, Y., & WANG, J. (2019). A panel analysis on the cross border e-commerce trade: Evidence from ASEAN countries. *The Journal of Asian Finance, Economics and Business*, 6(2), 95-104.

[2] Kawa, A. (2017). Supply chains of cross-border e-commerce. *Advanced Topics in Intelligent Information and Database Systems 9*, 173-183.

[3] Tu, Y., & Shangguan, J. Z. (2018). Cross-border E-commerce: A new driver of global trade. *Emerging Issues in Global Marketing: A Shifting Paradigm*, 93-117.

[4] Hazarika, B. B., & Mousavi, R. (2021). Review of cross-border e-commerce and directions for future research. *Journal of Global Information Management* (JGIM), 30(2), 1-18.

[5] Wang, Y. (2018, January). Research on marketing adaptability of cross-border e-commerce under different cultural symbols. *In 2017 7th International Conference on Education and Management* (ICEM 2017) (pp. 626-628). Atlantis Press.

[6] Tian, M. (2019). The In-Depth Marketing Environment of Cross-Border E-Commerce Experience Stores. *Ekoloji Dergisi*, (107).

[7] Jin, L., & Chen, L. (2024). Exploring the Impact of Computer Applications on Cross-Border E-commerce Performance. *IEEE Access*.

[8] Sakhvidi, M. Z., & Saadat, R. (2024). A Study of Artificial Intelligence and E-Commerce Economy. *International journal of industrial engineering and operational research*, 6(3), 139-155.

[9] Zhu, J. (2021, October). Application analysis of artificial intelligence technology in the development of cross-border E-commerce. *In 2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture* (pp. 1112-1116).

[10] Zhu, Q., Ruan, Y., Liu, S., Yang, S. B., Wang, L., & Che, J. (2023). Cross-border electronic commerce's new path: from literature review to AI text generation. *Data Science and Management*, 6(1), 21-33.

[11] Chen, K., Chen, J., & Ahmed, M. (2023, November). Digital Analysis of Information on Cross-Bordere-Commerceplatform Guide Platforms Based on Artificial Intelligence Algorithms. *In International Conference on Cognitive based Information Processing and Applications* (pp. 261-269). Singapore: Springer Nature Singapore.

[12] Guo, C., & Zhang, X. (2022). Research on Intelligent Customization of Cross‐Border E‐Commerce Based on Deep Learning. *Mathematical Problems in Engineering*, 2022(1), 4211616.

[13] Tang, Y. M., Chau, K. Y., Lau, Y. Y., & Zheng, Z. (2023). Data-intensive inventory forecasting with artificial intelligence models for cross-border e-commerce service automation. *Applied Sciences*, 13(5), 3051.

[14] Wang, W. (2023). A IoT-based framework for cross-border e-commerce supply chain using machine learning and optimization. *IEEE Access*.

[15] Chen, Y., Li, M., Song, J., Ma, X., Jiang, Y., Wu, S., & Chen, G. L. (2022). A study of cross-border E-commerce research trends: Based on knowledge mapping and literature analysis. *Frontiers in Psychology*, 13, 1009216.

[16] Cheng, J., & Zhang, J. (2024, January). Intelligent Pricing Model for Cross-border E-commerce Based on Artificial Intelligence. *In 2024 IEEE 7th Eurasian Conference on Educational Innovation* (ECEI) (pp. 258-261). IEEE.

[17] Hoque, M. R., & Bashaw, R. E. (Eds.). (2020). Cross-border E-commerce Marketing and Management. *IGI Global*.

[18] Xu, Y., He, D., & Fan, M. (2024). Antecedent research on cross-border E-commerce consumer purchase decision-making: The moderating role of platform-recommended advertisement characteristics. *Heliyon*, 10(18).

[19] Yuwen, H., Guanxing, S., & Qiongwei, Y. (2022). Consumers' perceived trust evaluation of cross-border e-commerce platforms in the context of socialization. *Procedia Computer Science*, 199, 548-555.

[20] Zhang, Y., Feng, Y., Zhou, W. J., Ye, Y., Tan, M., Xiao, R., ... & Yu, J. (2024, March). Multi-Domain Deep Learning from a Multi-View Perspective for Cross-Border E-commerce Search. *In Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 8, pp. 9387-9395).

[21] Ilmudeen, A. (2021). Big data, artificial intelligence, and the internet of things in cross-border E-commerce. *Cross-border E-commerce marketing and management*, 257-272.

[22] Li, L., Wang, Y., & Zhang, Y. (2021, January). Analysis on the application of artificial intelligence in cross-border E-commerce. *In 6th Annual International Conference on Social Science and Contemporary Humanity Development* (SSCHD 2020) (pp. 667-670). Atlantis Press.

[23] Chen, J., & Chunqiong, W. U. (2021, April). The design of cross-border E-commerce recommendation system based on big data technology. *In 2021 6th International Conference on Intelligent Computing and Signal Processing* (ICSP) (pp. 381-384). IEEE.

[24] Wu, W., & Xu, Y. (2022, December). Information Personalized Recommendation Algorithm of Cross-Border E-Commerce Shopping Guide Platform Based on Multi-Objective Optimization. *In 2022 Euro-Asia Conference on Frontiers of Computer Science and Information Technology* (FCSIT) (pp. 198-200). IEEE.

[25] Guo, C., & Zhang, X. (2024). Intelligent Prediction of Cross-Border E-Commerce Customer Satisfaction Using Deep Learning Embeddings. *IEEE Access*.

[26] Zhao, X. (2018, October). A Study on the Applications of Big Data in Cross-border E-commerce. *In 2018 IEEE 15th International Conference on e-Business Engineering* (ICEBE) (pp. 280-284). IEEE.

[27] Zhang, H. (2023, October). Formulation and Implementation of Precision Marketing Strategy for Cross-border E-commerce under Big Data Technology. *In 3rd International Conference on Internet Finance and Digital Economy (ICIFDE 2023)* (pp. 238-244). Atlantis Press.

[28] Zhang, H. (2024, April). Precision Marketing System for Cross-Border E-commerce Based on Data Mining Technology. *In 2024 IEEE 4th International Conference on Electronic Communications, Internet of Things and Big Data* (ICEIB) (pp. 578-582). IEEE.

[29] Li, Y. (2022). Predictive Analysis of User Behavior Processes in Cross-Border E-Commerce Enterprises Based on Deep Learning Models. *Security and Communication Networks*, 2022.

[30] Victor Hugo Antonius & Devi Fitrianah. (2024). Enhancing Customer Segmentation Insights by using RFM + Discount Proportion Model with Clustering Algorithms. International Journal of Advanced Computer Science and Applications (IJACSA)(3),

[31] Guohui Li,Haonan Deng & Hong Yang. (2024). A multi-factor combined traffic flow prediction model with secondary decomposition and improved entropy weight method. Expert Systems With Applications(PA),124424-124424.

[32] Nian Wang,Aitao Yang,Zhigao Cui,Yao Ding,Yuanliang Xue & Yanzhao Su. (2024). Capsule Attention Network for Hyperspectral Image Classification. Remote Sensing(21),4001-4001.