Media and Communication Research (2023)
Clausius Scientific Press, Canada
DOI: 10.23977/mediacr.2023.040904
ISSN 2523-2584 Vol. 4 Num. 9

Design of Accurate Placement Method for Film and Television Advertisements Based on Digital Twin and Data Mining

Shouliang Lai[1,a], Min Sun[1,b], Binrong Huang[1,c,*]

[1]*Collage of Packaging Design and Art, Hunan University of Technology, Zhuzhou, Hunan, China*
[a]*laishouliang@hut.edu.com*, [b]*10437@hut.edu.com*, [c]*14016@hut.edu.com*
[*]*Corresponding author*

Keywords: Digital Twin, Data Mining, Precise Placement of Film and Television Advertisements, Recommendation Prediction Algorithm

Abstract: With the development of economy, the Internet has developed rapidly as a new economic platform. At present, the Internet has become the best carrier of advertising and has huge commercial value. However, the extensive placement of advertisements not only brings troubles to Internet users, but also wastes the cost of advertisers. It cannot achieve the corresponding publicity effect, and cannot stimulate the consumer psychology of users, so the rate of return is low. This paper takes the network platform as the environment and data mining technology and digital twins as the technical support to study the dynamic clustering analysis advertising delivery algorithm and the collaborative filtering advertising delivery algorithm. This paper conducts experiments with KDD CUP 2012-Track 2 as the main dataset. The results show that the dynamic clustering analysis based on the K-means algorithm can greatly reduce the classification prediction error of users and film and television advertisements, and can better improve the accuracy of film and television advertisements. The error rate greater than 0.5 accounted for only 11.4% of the total number of advertisements. In the collaborative filtering algorithm, the performance of WSO is significantly better than that of the Slop One algorithm, and the error value becomes smaller when the number of test days is 15. At this time, the average MAE value of WSO is 0.916.

1. Introduction

The digital twin model is an information mirror model and a virtual-real interconnection model, which exists in the whole life cycle of physical objects and co-evolves with them. Through continuous in-depth research, the concept of digital twin has been applied in many fields, such as simulation, factory equipment assembly and so on. With the continuous development of IoT technology and the advent of intelligence and informatization, more and more industrial products, design patterns and industrial equipment have been empowered, and digital twins have also been expanded throughout the product life cycle. The precise placement of advertisements is to predict user preferences through algorithmic operations based on a given user and the content on the web page, and place the most relevant advertisement content.

Most of the traditional advertising alliances have carried out in-depth research on the tracking of advertisers' advertising effects, and the choice of advertising content depends on the personal interests and experience of website owners. The effect is often determined by the content selected by the website owner, which is highly subjective and unpredictable. Therefore, precise marketing methods are needed, and the advertising push algorithm is integrated to automatically select the advertising delivery strategy. This requires relying on the technical means of data mining and collaborative filtering algorithms to establish a personalized advertising recommendation system and customize personalized products for the market. When it comes to advertisement placement, it means placing the advertisements that users have the most needs and are most likely to be interested in at the right time and at the right place. It not only reduces the interference to the user's browsing experience, but also may solve the user's needs in a timely manner by pushing advertisements, thereby enhancing the user's favorable impression of the website and realizing a virtuous circle.

According to the given user and platform content, the calculation advertisement can obtain the most matching advertisement through calculation and carry out precise targeted delivery. This paper studies this, and the main innovations are: First, this paper improves the K-means clustering algorithm according to the number of iterations and uses the sum of squares of errors as the threshold, which improves the classification accuracy of the algorithm. Second, this paper compares the performance of the WSO algorithm and the Slop One algorithm. In the algorithm, the concept of rating similarity and user credibility is introduced in this paper, and users who may be classified into the wrong set are eliminated. Third, this paper evaluates the performance of the algorithm with a variety of evaluation indicators.

2. Related Work

Digital twin technology can reason based on information content, combining current conditions with expected processes, while predicting future possibilities based on information content. To achieve efficient implementation of digital twin technology, Slot M proposes a purpose-driven approach. A purpose-determining approach can be facilitated by following the 3P characteristics: purpose, perspective, and priority [1]. In Roy A's research, he proposed a hybrid optimization method for big data mining and provided a heuristic method to determine consumer preferences, even in the case of Arrow's paradox in big data analysis. Both methods play a central role in analyzing CRM cases, and experiments provide a comparison of their efficiency with traditional methods [2]. Users prefer real-world scenarios expressed in the form of positive feedback, such as thumbs up on YouTube, which makes the cold-start problem more challenging for users. To address the cold-start problem in the case of only positive feedback, Tomeo P proposes to utilize data other than user preferences through a specialized hybrid recommendation method. He studied some recommendation patterns based on graph and matrix factorization [3]. Liu Y mainly studies the CTR prediction model in the video recommendation scenario. In order to make full use of the user's historical behavior and user feedback, PMN uses an attention mechanism to obtain the user's historical behavior representation and user preference representation from the original input. Finally, he introduced a user preference baseline to address the inconsistency of different user ratings [4]. Wang X's research proposes a hybrid collaborative filtering model based on an improved K-means clustering algorithm to address the information expiration problem in traditional collaborative filtering algorithms. In order to further alleviate the problem of data sparseness, he proposed a method to calculate users' emotional tendencies by analyzing the features of user reviews, mining users' attitudes towards review items, and extending the rating matrix [5]. Tan Z designed an MPM algorithm to efficiently capture users' contextual preferences by using a graphical model and a new pruning strategy. He uses the user's common preference to build a belief system, and combines the Bayesian method to obtain the user's Top-K preference rule. Finally, the experimental results show that the MPM algorithm is far more efficient in mining all rules than the three classical or cutting-edge mining algorithms [6]. According to the characteristics of social networks, Ge J proposes an intelligent data management mechanism based on relationship strength to achieve reliable prediction in online social networks. Second, he identified network structure properties and user interest preferences as important factors affecting the link prediction process in online social networks. By designing a friend recommendation model, he incorporates the user's relationship information and interest preference features into the community detection algorithm, further improving the prediction process [7].

3. Accurate Advertising Model

3.1 Advertising Process

Advertising is a means of publicity and publicity to widely transmit publicity information through media for specific needs. Whether or not economic benefits are the ultimate goal, it is hoped that a better publicity effect will be obtained. This paper mainly focuses on commercial advertisements, and the starting point is the precise placement of film and television advertisements. In order to achieve the ideal advertising effect, in addition to accurately expressing advertising information for different groups, the advertisements placed also need to establish a brand image and guide consumption. Therefore, most advertisements are dynamically and target-oriented, and it is necessary to save key information and conduct long-term monitoring when users browse advertisements, analyze users' Internet browsing behaviors, and realize regular advertisements [8-9]. Figure 1 shows the process of ad placement.

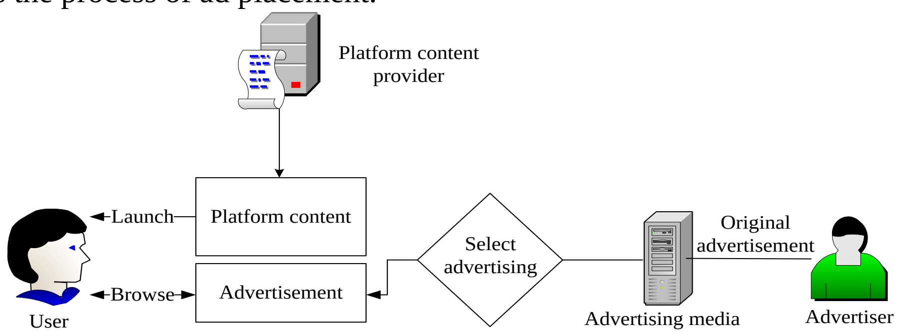


Figure 1: Advertising process

3.2 Online Advertising Value Chain Model

The subjects in the advertising value chain include advertisers, advertising providers, advertising content and audiences, and accurate advertising services are generally provided by online advertising publishers. The value chain structure of online advertising is shown in Figure 2. Advertisers are businesses that need to promote their content on the Internet. Advertisers generally entrust third-party companies to design advertisements for products, and provide online advertisement publishers with pictures, texts, videos and other forms of advertisement content. Audience refers to the recipients of information dissemination. Online advertising publishers refer to possessing online promotion resources and providing online marketing services to advertisers [10]. The advertiser plays the advertiser's product advertisements on its own relevant websites, and charges the advertiser according to the marketing effect such as the number of views and clicks of the advertisement.

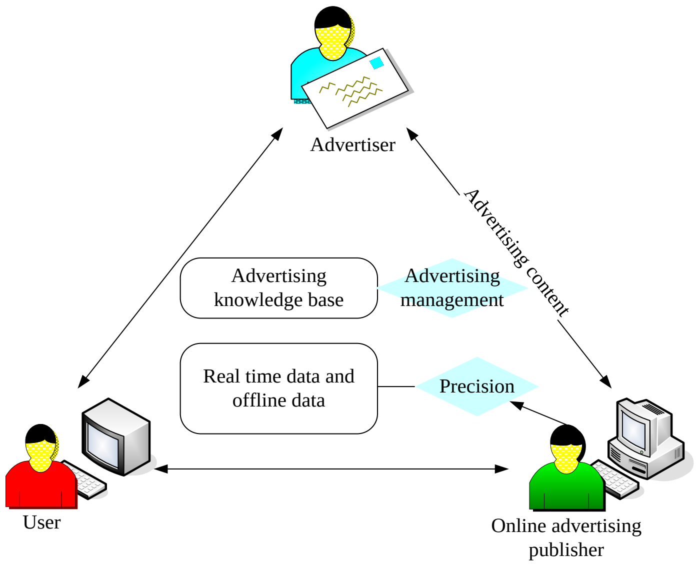


Figure 2: Online Advertising Value Chain Model

The delivery of ads and their management are in most cases the responsibility of online ad publishers. Therefore, the entity will charge advertisers according to the final publicity effect. There are six main charging modes, including click mode, CPM mode, access mode, time mode, transaction mode and hybrid mode.

4. Digital Twin Framework and Data Mining Technology

4.1 The Concept of Digital Twin

Digital twin is also known as digital mapping, digital mirroring. A digital twin is a digital model of a physical entity object that visualizes, diagnoses, and predicts the current or future state of a physical entity in real time through measurement, simulation, and data analysis. It iteratively optimizes the behavior of physical objects through optimization and instruction regulation, as well as the interaction between related digital models [11-12]. Big data provides the ability to calculate, store and process shared data, then digital twin is to map physical model to digital model, and the two can exchange data to improve the performance of physical entities. A conceptual diagram of the digital twin is shown in Figure 3.

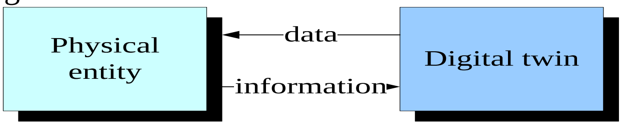


Figure 3: Digital twin concept map

4.2 Digital Twin Maturity Model

The digital twin is not only a mirror image of the physical world, but also receives real-time information from the physical world, and in turn drives the physical world in real time, and evolves into a prophet and a forerunner of the physical world. This evolution process is called maturity evolution, that is, the growth and development of a digital twin will go through the process of evolution of models such as digitization, interaction, prophet, foresight, and shared intelligence, as shown in Figure 4.

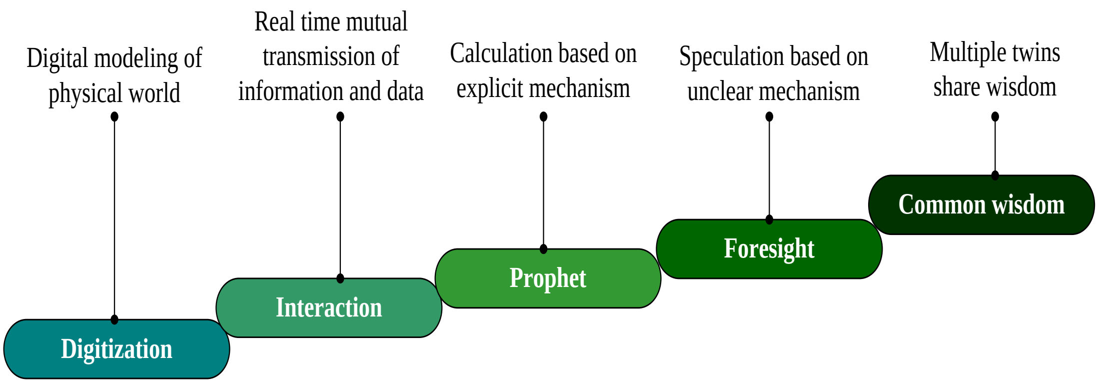


Figure 4: Digital Twin Maturity Model

The digitization process is to realize the digitization of the physical world through modeling. It needs to express the physical objects as digital models that can be recognized by computers and networks. The interaction process mainly refers to the real-time dynamic interaction between digital objects and objects in the physical world. The prophetic process is the process of using simulation techniques to make dynamic predictions about the physical world. The harbinger process predicts the future of the digital twin based on incomplete information and unclear mechanisms. The mutual intelligence process realizes intelligent information exchange and sharing between different digital twins through cloud computing [13].

4.3 Data Mining Technology

Data mining technology is based on big data and cloud computing technology. The amount of big data is very large and the categories are rich. Data preprocessing, data interaction visualization and data mining technology are the core technologies of big data. Data preprocessing organizes and transforms the obtained massive data. The cloud data center reasonably plans and schedules various processing tasks to optimize resource utilization. It dynamically deploys the information provided to the user, which requires analysis of the tasks from the user. The main characteristics of big data are: large amount, variety, low value and high speed [14-15], as shown in Figure 5.

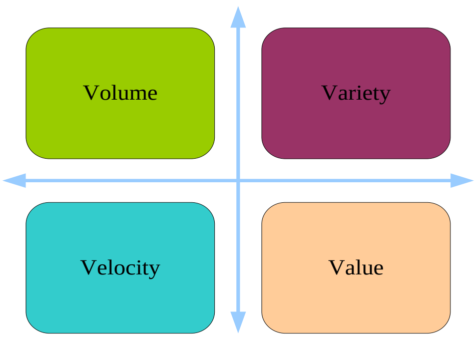


Figure 5: Big data characteristics

The purpose of data mining comes from application requirements, and its mining process is composed of multiple steps, which involve different mining methods and related technologies in different fields. At present, the standard process of big data mining modeling includes six stages, namely business understanding, data understanding, data preparation, modeling, evaluation and application. The specific mining process model is shown in Figure 6.

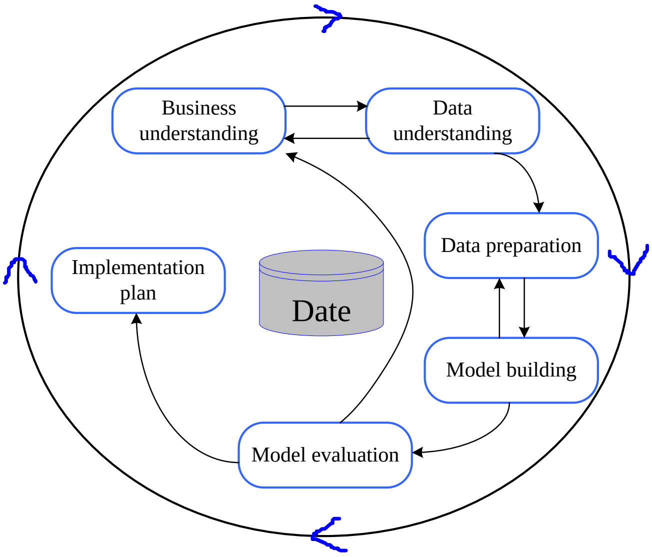


Figure 6: Big data mining process

5. Advertising Recommendation Algorithm Based on Digital Twin and Data Mining

The basic algorithms of data mining include four types, namely clustering, classification, association and regression.

5.1 K-means Algorithm

K-means algorithm is one of the most widely used basic segmentation algorithms in cluster analysis. This method can cluster according to user value, subdivide audiences, and then adopt different market strategies for different audiences to optimize the allocation of customer resources. The dynamic clustering algorithm can divide audiences with high similarity into a class of massive data information, while audiences with low similarity can be divided into different categories [16]. In addition, the dynamic clustering algorithm is used to adjust the advertising strategy, which can accurately capture the changes in the needs of advertising audiences and improve the pertinence of advertising.

Suppose the dataset to be clustered is:

$$A = (x_1, x_2, \dots, x_n) \tag{1}$$

The k cluster centers are:

$$C = (c_1, c_2, \dots, c_k) \tag{2}$$

The Euclidean distance between two individuals in the K-means algorithm is the square root Formula of the sum of the squares of the difference between the K variable values of the two individuals. x_i and x_j represent data objects with 2 q-dimensional attributes.

$$D(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{iq} - x_{jq})^2} \tag{3}$$

$$x_i = (x_{i1}, x_{i2}, \dots, x_{iq}) \tag{4}$$

$$x_j = (x_{j1}, x_{j2}, \dots, x_{jq}) \tag{5}$$

n is the total number of sample objects, and the Formula for the average distance of all sample points is:

$$\mathrm{Mid(S)} = \frac{2}{\mathrm{n(n-1)}} * \sum_{\mathrm{i \neq j, i、j=1}}^{\mathrm{n}} \mathrm{D(x_i, x_j)} \tag{6}$$

Take the squared error criterion function as the objective function:

$$\mathrm{E} = \sum_{\mathrm{i=1}}^{\mathrm{k}} \sum_{\mathrm{j \in N_i}} \left\| \mathrm{x_j - c_j} \right\|^2 \tag{7}$$

The basic steps of the algorithm are: First, select k samples as the initial center point. Second, the samples outside the center point are classified, and each sample is grouped into the class closest to the center point. The center point of the class will be updated, the value will become the current mean, and all samples will be classified, then repeat the previous step until the termination condition is met. One is based on the number of iterations to terminate, and the other is based on the degree of offset of the center. The iteration can be terminated if one of the two termination conditions is satisfied [17].

Improve the K-means algorithm: The cluster is iterated multiple times, recording the number of times each element x_i is divided into the JTH cluster c_j in each iteration, and then calculating the distance of any element x_i to each central point in the next iteration. Assuming that there are k cluster centers, the Formula for calculating the distance to the JTH cluster center for each element $\mathrm{x_i}$ in the MTH iteration is:

$$D_p(x_i, c_j) = \left(1 - \frac{r_{ij} \div k}{\sum_{m=1}^{k} t_{im}}\right) * D(x_i, c_j) \tag{8}$$

Where x_i is the ith element, c_j represents the centroid of the jth class, and t_{im} represents the number of times the ith element is assigned to the MTH class in the iteration.

To check the performance of the clustering algorithm, the K-means algorithm and the improved K-means algorithm were written in Java language and tested using the iris dataset, which is a public dataset. Both algorithms are run multiple times, and the initial center point is not exactly the same here. Some data are given here for a brief description, as shown in Table 1.

Table 1: Degree of clustering

Test result No	K-means	iK-means
1	38(2,36)	44(9,35)
	50(0,50)	51(1,50)
	62(14,48)	55(15,40)
2	38(2,36)	37(1,36)
	50(0,50)	50(0,50)
	62(14,48)	63(14,49)
3	38(1,37)	37(0,37)
	50(0,50)	54(4,50)
	62(13,49)	59(12,47)
4	46(5,41)	49(11,38)
	50(0,50)	50(0,50)
	54(9,45)	51(12,39)

Option 1 is used to represent the performance of traditional K-means. For the improved algorithm, it can be seen from the test results that the performance is not stable, sometimes better than the traditional algorithm, and inferior to the traditional algorithm when the probability is small. Dispersion and clustering are used as indicators for evaluation. $\mathrm{C_n}$ is used to represent the number of elements in the current set, $\mathrm{C_{n1}}$ is the number of the largest class, and $\mathrm{C_{n2}}$ is the number of the lesser class.

The dispersion Formula is:

$$\mathrm{Dis} = \frac{C_{n2}}{C_n} * 100\% \tag{9}$$

The clustering degree Formula is:

$$\mathrm{Clu} = \frac{C_{n1}}{C_n} * 100\% \tag{10}$$

The relative pros and cons of the algorithms are shown in Table 2.

Table 2: Comparison of advantages and disadvantages of algorithms

Type	Dispersion		Clustering degree
K-means	9%		91%
Comparison algorithm	advantage	inferiority	Flat
iK-means	8%	6%	84%

Based on the test results of the K-means algorithm, the agglomerated Agnes clustering algorithm was tested with the iris data set. The results are shown in Figure 7(a), the iris data set has a total of 150 elements, and the ideal classification result should be averagely divided into three clusters, but from the test results, Agnes does not reach the ideal state. Analysis may be due to a problem in one or several merging processes, the types of several clusters are mixed, resulting in 12 data being wrongly allocated. In order to further judge the difference between elements and improve the classification accuracy, the sum of squares of errors in K-means is used as the threshold, and the threshold interval is shown in Formula (11). If the sum of squared errors of a cluster is greater than the threshold, the cluster is decomposed.

$$\mathrm{M} = \left[\frac{\mathrm{SSE}}{\mathrm{k}}, \frac{\mathrm{SSE}}{\mathrm{k}-1}\right] \tag{11}$$

The test results of the improved algorithm are shown in Figure 7(b). According to the results, it can be seen that the clustering results have been optimized, but these optimizations come at a cost of time. The time to decompose the clusters that exceed the threshold and the time to run the main program increases each time.

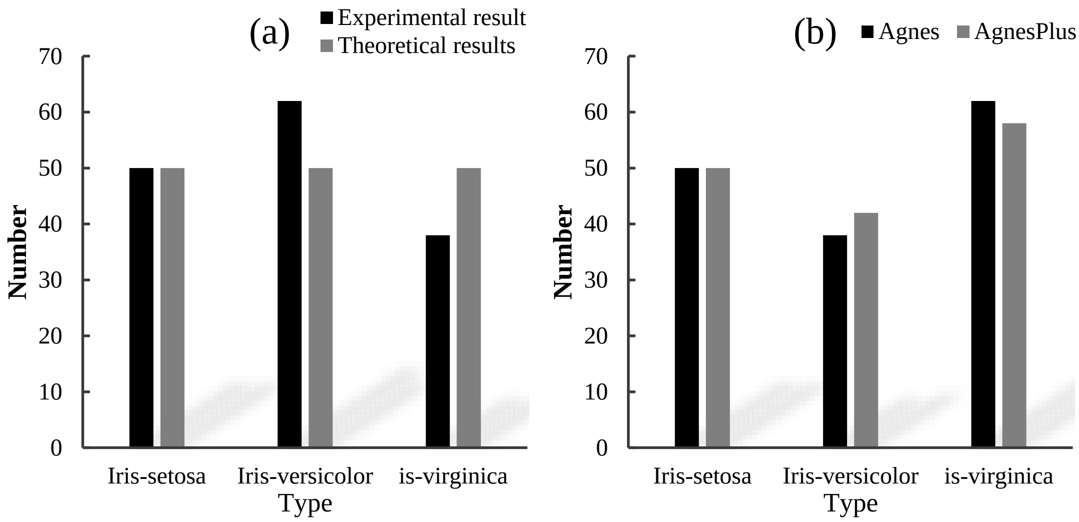


Figure 7: Test results

5.2 Collaborative Filtering Algorithm

Collaborative filtering algorithm is a recommendation based on users with the same interests, which also requires big data analysis. It generally consists of three steps. The first step is to analyze user behavior data, and form a matrix table of user behavior data and item scores, as shown in Table 3. The second step is to form a set of similar items or a set of neighboring users, which is divided by users' attention to advertisements. Both users 1 and 4 have scored advertisements 1, 2, and 3. Therefore, if the two users have the same point of interest, they can be divided into neighboring users, and then according to the average score of advertisements, those with similar scores can be divided into similar advertisements. The third is to perform user feature matching and recommend items [18].

Table 3: User - Ad Ratings Table

User	Advertisement 1	Advertisement 1	Advertisement 1	Advertisement 1	Advertisement 1
1	3	2	3	Null	0
2	4	Null	Null	3	1
3	2	5	Null	Null	3
4	1	4	3	Null	Null
5	Null	4	5	0	Null

Here we mainly introduce the Slope One algorithm, which is an algorithm that only focuses on users, and does not care about the similarity between items or between users. It also requires the use of a scoring matrix. Assuming that two items m and n are scored by three users A, B, and C, respectively, the results are shown in Table 4. User B does not rate item n. According to the Slope One algorithm, the predicted value of user B's rating for item n can be calculated from Formula (12), which is 4.5.

$$\mathrm{P}(\mathrm{B_n}) = \mathrm{B_m} + \mathrm{Md} \tag{12}$$

$$\mathrm{Md} = \frac{1}{2}[(\mathrm{A_n} - \mathrm{A_m}) + (\mathrm{C_n} - \mathrm{C_m})] \tag{13}$$

Where $\mathrm{S_n}$ and $\mathrm{S_m}$ represent the user's ratings of n and m, respectively. Md is the average difference between the ratings of users A and C for items m and n.

Table 4: Example of Slop One algorithm

	m	n
A	3	5
B	4	Null
C	4	3

Define $S_{m,n}(x)$ as the set of all users scoring items m and n, then the average deviation of item j to item i is:

$$Dev_{n\to m} = \sum_{\mu \in S_{m,n}(x)} \frac{\mu_n - \mu_m}{count(S_{m,n}(x))} \tag{14}$$

$count(S_{m,n}(x))$ represents the number of elements of $S_{m,n}(x)$, and let Q_m be the item set that has been rated, then the recommended rating expression for predicting user μ for item n is:

$$Prediction(\mu_n) = \frac{\sum_{m \in Q_m}(Dev_{n\to m} + \mu_m)}{count(Q_m)} \tag{15}$$

The Slop One algorithm does not take into account the influence of the number of participating users on the scoring results, so with the WSO algorithm, the number of all scoring users of the project is added to the Slop One algorithm, and the final average is weighted [19]. In order to group users' interests more accurately, the concepts of scoring similarity and user credibility can also be introduced when generating nearest neighbors. When the score similarity and user credibility are negative, it means that the similarity between the two users is low, and they are removed from the nearest neighbor set, and then the predicted value is calculated.

$$Num_{n\to m} = count(S_{m,n}(x)) \tag{16}$$

$$Prediction^{WSO}(\mu_n) = \frac{\sum_{m\in S(\mu)-\{n\}}(Dev_{n\to m}+\mu_m)*Num_{n\to m}}{\sum_{m\in S(\mu)-\{n\}} Num_{n\to m}} \tag{17}$$

5.3 Evaluation of Recommendation Algorithms

The quality of different recommendation algorithms can be evaluated by using appropriate evaluation metrics. Commonly used recommendation algorithm evaluation indicators are: accuracy, coverage, diversity and so on.

For an object that has not been selected by the user, the final result has four cases, namely: recommended by the system and liked by the user, recommended by the system but not liked by the user, not recommended by the system but liked by the user, and not recommended by the system and not liked by the user. These four results are represented by N1-N4 respectively [20].

For a certain user u, the recommendation accuracy of the system is the proportion of the objects that the user is interested in among the N objects recommended by the system:

$$P_u(N) = \frac{N1}{N1+N2} \tag{18}$$

Let M represent the number of users, and calculate the average of the accuracy rates of all users in the system. The final overall recommendation accuracy of the system is:

$$\mathrm{P(N)} = \frac{\sum_u P_u(N)}{\mathrm{M}} \tag{19}$$

For a user u, the recall rate is:

$$Recall_u(N) = \frac{N1}{N1+N3} \tag{20}$$

The overall recovery rate of the system is:

$$Recall(N) = \frac{\sum_u Recall_u(N)}{\mathrm{M}} \tag{21}$$

6. Test, Evaluation and Application of the Precise Placement Method of Film and Television Advertisements

6.1 Experimental Environment and Experimental Data

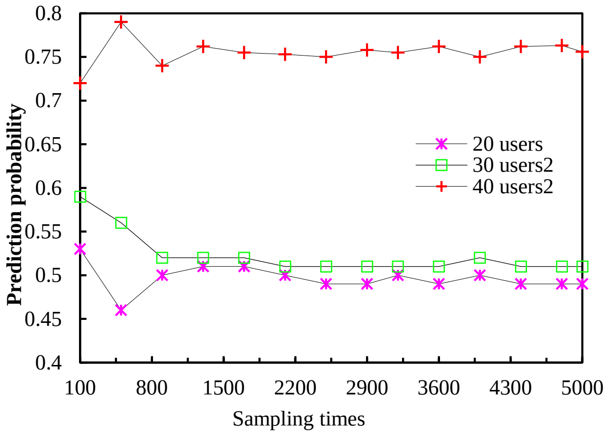
The whole experimental platform is developed with B/S architecture, and user requests are received and transmitted through the browser, and processed by the background server. The system mainly uses Python language to write programs and MongoDB to store data. The server is located in the remote computer room and communicates with the machines in the local area network through the public network. The hardware environment used in the experiment: IBM server with Web program and mysql database deployed, and the operating system is Ubuntu 11.10.

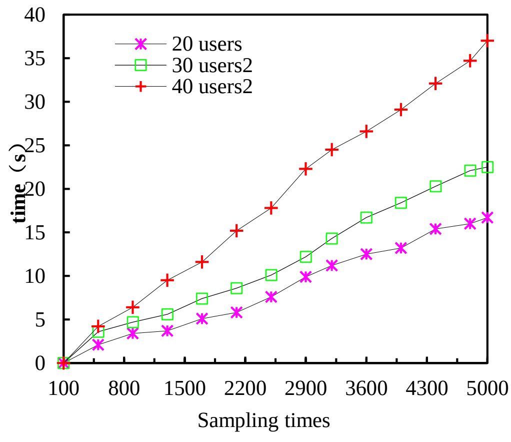
Dataset 1 is the data in KDD CUP 2012-Track 2. The total data set reaches 10GB, with a total of more than 600,000 ads, 200 million ad impressions, and 9 million ad clicks. The raw data contains many attributes, such as basic user information (gender, age, identification ID), the number of times the advertisement is presented, the number of times the user clicks, the ID of the advertisement and the advertiser, etc.

In addition to the data, the experiments in this paper also extract advertising data from other sources as dataset 2. It comes from multiple advertisers, with a total of 1888 advertising content. The user extracted 1,000 relatively active users, and the effect data were mainly the user's clicks on these contents and the follow-up attention behavior. And in the collaborative filtering algorithm experiment, a 5-point scoring method is adopted. Every time a user clicks on an advertisement, 1 copy will be recorded, and if the user stays on the advertisement page for more than 1 minute, 2 points will be recorded. When the user continues to pay attention to the advertising content or product, 5 points will be awarded.

6.2 Experimental Results of Film and Television Advertising Based on Dynamic Cluster Analysis

In the experiments in this section, KDD CUP 2012-Track 2 is used as the main dataset. The modeling performance of the algorithm is analyzed, including the convergence of the K-means algorithm and the efficiency of inference classification. Select data containing 20, 30, and 40 users, and test the change of the classification prediction probability value with the increase of sampling times. The results are shown in Figure 8(a). It can be concluded that with the increase of sampling times, the classification results can quickly converge to a stable value under different user numbers. The efficiency index of classification prediction also selects data containing 20, 30, and 40 users, and tests the change of the time required with the increase of sampling times. The results are shown in Figure 8(b). It can be seen that the execution time of the algorithm increases in a linear trend with the sampling times under different user numbers.





(a) Changes in classification predicted probability values (b) Changes in required time

Figure 8: K-means clustering prediction efficiency

Then 12,652 users and advertisements were extracted from the test dataset, and the predicted classification results were compared with the ground-truth values. Perform segmentation statistics on all prediction results errors, and the number of advertisements whose classification prediction errors are lower than 0.1, 0.1-0.2,. 0.4-0.5, and greater than 0.5 are shown in Figure 9. It can be seen that the cluster classification prediction method can obtain relatively accurate results, but in actual situations, the historical click records of advertisements are relatively sparse and the click rate value is small, and the error still needs to be further reduced.

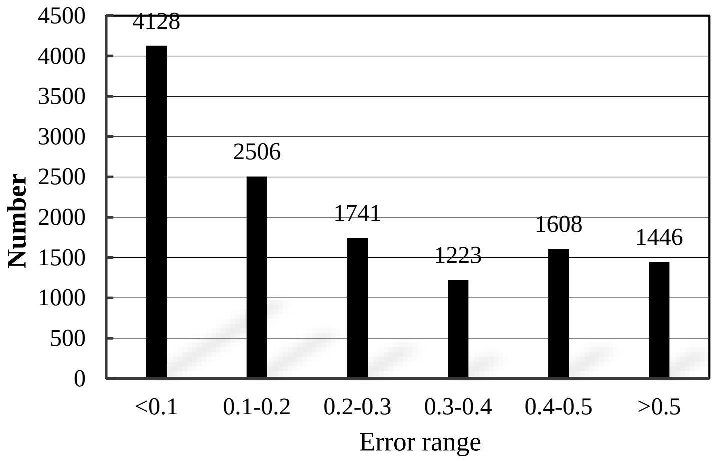


Figure 9: Predicted classification error distribution

6.3 Experimental Results of Film and Television Advertising Based on Collaborative Filtering Algorithm Analysis

In the experiments in this section, the WSO algorithm is compared with the traditional Slop One algorithm, and the data set is the data set 2 mentioned in this paper. During the experiment, the advertisements corresponding to the advertisement score data table were divided into 5 parts, and the test set and the training set were divided according to 1:4. The number of neighbors is set to 10, 20, 30 and 40 respectively, and the statistical test days are 10, 15, 20 and 30 days for the MAE values (mean absolute error value) of the WSO algorithm and the Slop One algorithm. The experimental results are shown in Figure 10. As can be seen from the figure, as the number of user neighbors in the experimental data increases, the trend of MAE decreases as a whole, indicating that the recommendation accuracy is improved. And the accuracy of the WSO algorithm is higher than that of the Slop One algorithm, the mean absolute error value is smaller, and the recommendation accuracy is higher. The recommended test days here is 15 days, because when T=15, the decrease of MAE is more obvious in the experiments of several neighbors.

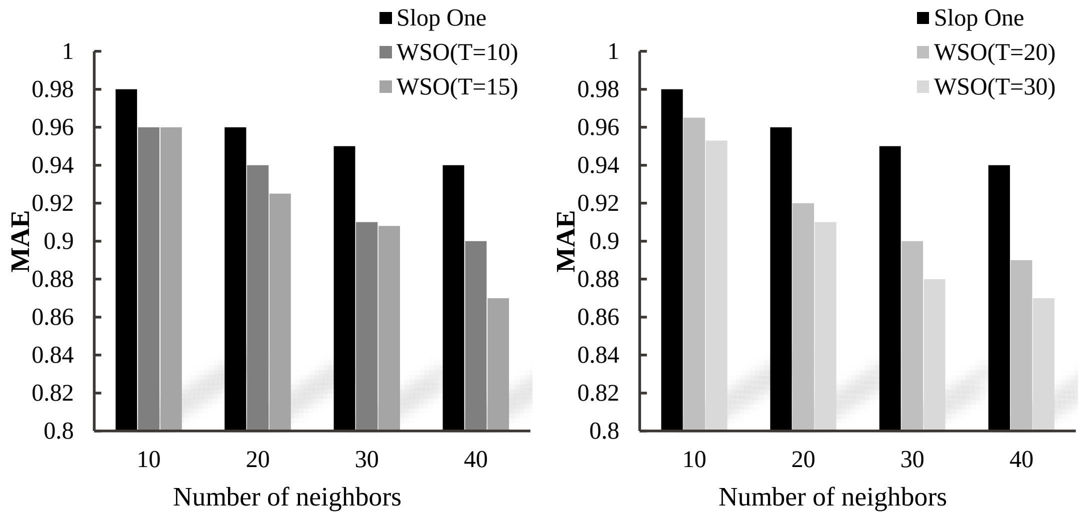


Figure 10: MAE Comparison of Collaborative Filtering Advertising Algorithms

7. Discussion

The main research direction of this paper is the design of the accurate placement method of film and television advertisements. With digital twin and data mining as the main technical and theoretical support, two feasible recommendation schemes for film and television advertisements to improve the accuracy of advertisement placement are constructed. The article starts with relevant research, summarizes the relevant research contents related to big data mining technology and user preference recommendation, especially for user preference prediction in social networking, and summarizes the previous research experience and research focus. Secondly, taking the precise model of advertisement placement as the main content, it describes the general process of advertisement placement, the value chain model of online advertisement and other related contents. Then there is an overview of the digital twin framework and data mining technology. Its main contents include the basic concepts of digital twins, the maturity model of digital twins, the basic characteristics of big data, the basic process of big data mining, etc., which provide a theoretical reference for the algorithm conception that follows. Next is the core content of the article, that is, the advertising recommendation algorithm based on digital twins and data mining, which focuses on the clustering algorithm and collaborative filtering algorithm. In the content of the clustering algorithm, the performance of the algorithm is compared using the iris data. The compared algorithms include K-means algorithm, Agnes algorithm and AgnesPlus algorithm. The content of the collaborative filtering algorithm mainly introduces the basic idea and principle of the algorithm. This paper focuses on the Slop One algorithm and the WSO algorithm. The last part is the test comment and application part of the precise placement method of video advertisements. In this paper, several classical advertising recommendation algorithms and their optimization algorithms mentioned are tested and applied. In the comparative experiment of K-means clustering algorithm, this paper tests the prediction accuracy and efficiency of the algorithm. The results show that the algorithm has better prediction efficiency and prediction classification accuracy. In the experiments based on the collaborative filtering algorithm, this paper mainly compares the performance of the Slop One algorithm and the WSO algorithm, and studies the optimal prediction days of the recommendation algorithm. The results show that the error value of WSO is significantly smaller than that of the Slop One algorithm, and the optimal test days for WSO is 15 days.

8. Conclusion

The advertising delivery technology of dynamic cluster analysis and the collaborative filtering advertising delivery technology are both an attempt to accurately deliver video advertisements. Dynamic clustering analysis is to use the clustering method in data mining to subdivide audiences, and then adopt different business strategies to accurately place film and television advertisements, tap potential audience groups, and improve advertising benefits. The collaborative filtering advertising delivery technology focuses more on the user's interest in the type of advertising, and also pays attention to the user's neighbors to implement advertising and improve the click-through rate of advertising. The research in this paper is largely theoretical and simulation, and has not been used for actual advertising. It is hoped that the system can be further studied in the next step.

Acknowledgement

2022 Hunan Provincial Social Science Achievement Evaluation Committee: Aesthetic Evolution and Response Strategies in Network Chaos - Virtual Aesthetics Construction (XSP22YBC600); 2022 Key Research Project of Hunan Provincial Department of Education: Virtual Design and Digital Communication of Huxiang Culture (22A0416).

References

[1] Slot M, Lutters E. Digital twinning for purpose-driven information management in production. Procedia CIRP, 2021, 100(3):666-671.

[2] J Wa ta da, Roy A, Vasant P. Preference Identification Based on Big Data Mining for Customer Responsibility Management. International Journal of Intelligent Technologies and Applied Statistics, 2020, 13(1):1-24.

[3] Tomeo P, I Fernández-Tobías, Cantador I, et al. Addressing the Cold Start with Positive-Only Feedback Through Semantic-Based Recommendations. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2017, 25(Suppl. 2):57-78.

[4] Liu Y, Yang T, Qi T. An Attention-Based User Preference Matching Network for Recommender System. IEEE Access, 2020, 8(99):41100-41107.

[5] Wang X, Dai Z, Li H, et al. Research on Hybrid Collaborative Filtering Recommendation Algorithm Based on the Time Effect and Sentiment Analysis. Complexity, 2021, 2021(2):1-11.

[6] Tan Z, H Yu, Wei W, et al. Top-K interesting preference rules mining based on MaxClique. Expert Systems with Application, 2020, 143(Apr.):113043.1-113043.18.

[7] Ge J, Shi L L, Liu L, et al. Intelligent Link Prediction Management Based on Community Discovery and User Behavior Preference in Online Social Networks. Wireless Communications and Mobile Computing, 2021, 2021(1):1-13.

[8] Xiao C, Qiao Y, Xia K, et al. Research on Tourist Routes Recommendation Based on the User Preference Drifting Over Time. International Journal of Computer Systems Science & Engineering, 2018, 33(2):95-103.

[9] Janiijevi S, Petrovi O, Stefanovi M. Sales prediction on e-commerce platform, by using data mining model. Serbian Journal of Engineering Management, 2020, 5(2):60-76.

[10] Tadeja S K, Seshadri P, Kristensson P O. AeroVR: An immersive visualisation system for aerospace design and digital twinning in virtual reality. Aeronautical Journal -New Series-, 2020, 124(1280):1-21.

[11] Hiskey T. Digital Twinning: The Future of Manufacturing?. Software World, 2017, 48(6):15-16.

[12] Berry C. Screening Post-1989 China: Critical Analysis of Chinese Film and Television. The China quarterly, 2017(230):553-554.

[13] D Zaborski, Ali M, Eyduran E, et al. Prediction of Selected Reproductive Traits of Indigenous Harnai Sheep under the Farm Management System via various Data Mining Algorithms. Pakistan Journal of Zoology, 2019, 51(2):421-431.

[14] Deng B, Xu J, Wei X. Tourism Destination Preference Prediction Based on Edge Computing. Mobile Information Systems, 2021, 2021(1):1-11.

[15] Sharma P K, Rathore S, Park J H. Multilevel learning based modeling for link prediction and users' consumption preference in Online Social Networks. Future Generation Computer Systems, 2017, 93(APR.):952-961.

[16] Wang Q, Liu F, Huang P, et al. A Hierarchical Attention Model for CTR Prediction Based on User Interest. IEEE Systems Journal, 2020, 14(3):4015-4024.

[17] Chen H, Sun H, Cheng M, et al. A Recommendation Approach for Rating Prediction Based on User Interest and Trust Value. Computational Intelligence and Neuroscience, 2021, 2021(9):1-9.

[18] Castillo-Soriano M, Canavati-Espinosa A, Maldonado-Flores D I. O imaginário suburbano e o Mass Media: um reflexo de sua construo e desmontagem na gerao do chamado Baby Boom nos Estados Unidos (1946-1974). Revista Perspectivas, 2021, 6(1):6-23.

[19] Alvarezdionisi L E. Envisioning Skills for Adopting, Managing, and Implementing Big Data Technology in the 21st Century. International Journal of Information Technology & Computer Science, 2017, 9(1):18-25.

[20] Colombus C. BCE Monitors Hybrid Ops with Tektronix. TV technology, 2018, 36(11):34-34.