# DESIGN AND IMPLEMENTATION OF MULTIMEDIA FILE MANAGEMENT SYSTEM UNDER THE BACKGROUND OF BIG DATA BASED ON COLLABORATIVE RECOMMENDATION ALGORITHM

F.G. YAO†

† *Network and Teaching Resources Center, Chongqing University of Education, Chongqing, China.*
*euniceGjF@yahoo.com*

***Abstract*— With the rapid expansion of the market scope of e-commerce, people's choice of shopping channels is not limited to physical stores, and purchases through e-commerce platforms, but the differences in user group behaviors make the content pushed by e-commerce platforms not suitable for every user. In order to further improve the efficiency and practicability of Internet users using e-commerce platform and save users' search time for content of interest, this paper designs a personalized recommendation system belonging to users and conducts empirical research. The empirical results show that the collaborative recommendation algorithm is used for data hypothesis and model calculation simulation test. This paper designs the multimedia file management system to meet the user's needs.**

***Keywords*— Collaborative recommendation; Collaborative filtering algorithm; Management system.**

## I. INTRODUCTION

In recent years, with the development of network technology, the networking of multimedia technologies and networked multimedia applications have been cooperating with each other and multimedia technologies have gradually demonstrated the characteristics of diversification and digitization (Sang *et al.*, 2017). Unlike images, video, audio, etc., characters, numerical data, etc., these media contain non-format data and the amount of data is huge (Ma *et al*., 2016). As far as we know, the multimedia database is a new type of database technology that embodies the access, storage, retrieval and management of multimedia data (Thorat *et al.*, 2015). With the rapid expansion of the market scope of e-commerce, people's choice of shopping channels is not limited to physical store purchases and it prefers to purchase through e-commerce platforms (Naji *et al.*, 2017).

This not only saves on time, but also facilitates the comparison of the same kind of cost-effectiveness. However, the pushed content is not suitable for each user in the different user groups and behavioral differences (Xiaojun *et al*., 2017). However, there is still a certain gap in the construction and application of domestic multimedia resource libraries relative to foreign countries (Lim *et al.*, 2016). The multimedia file system should be a multi-faceted multimedia digital information resource library. Its media representations include views, models, controllers, etc. Through co-inference testing, data are based on cosine similarity and correlation similarity and modified cosine similarity (Zarzour *et al.*, 2018). Then it creates a memory model based on the user's history on the network platform and it is filtered and designs a multimedia file management system that meets the user's needs (Lin *et al*., 2015).

## II. STATE OF THE ART

With the rapid increase in the amount of information on e-commerce platforms, how to help users quickly find the products of interest and to solve the lack of "shopping" role in similar entity stores (Brown *et al.*, 2018). The personalized recommendation system will recommend possible product information to the user according to the behavior of the users (Fernández-Pousa, 2018). The user's unique behavioral habits may be browsing the history of the product and evaluating the behavior of the product (explicit) or purchasing behavior (implicit) and so on (Lim *et al*., 2016). This technology has been applied to many large websites both at home and abroad on the Internet to make the recommended content more intelligent. At present, many e-commerce websites have already used recommendation systems. Personalized recommendation algorithms are usually divided into content-based recommendations and collaborative recommendations (Fernández-Pousa, 2018).

The content-based recommendation filtering is to recommend information to the user based on the relationship between the content of the information and the interest of the user, but its disadvantage is information that cannot analyze the content. It cannot handle complex and difficult to express concepts. Collaborative recommendation is also called social information filtering. It is based on the user's past behavior. A user or item that has similar behavior to a specified user or item is found in the user or program group and a possible behavioral prediction for the specified user or item is formed based on the behavior of similar users or items. The concept of collaborative filtering was first proposed in 1992 and applied to the Tapestry system. It is one of the earliest and most successful recommendation algorithms.

At present, some researches in the field of distance education also mention the personalized recommendation of teaching resources according to the characteristics of interests and hobbies of the students.

However, although some systems provide customized curriculum functions for students, there are still some shortcomings in personalized needs.

## III. METHODOLOGY

### A. A Personalized Recommendation System

Collaborative recommendation algorithms mainly include memory-based collaborative recommendation and model-based collaborative recommendation. Memory-based collaborative recommendation, this method first obtains neighboring users or projects with similar interests by statistical methods and performs optimal calculation based on neighbors. It mainly includes user-based collaborative recommendation and item-based collaborative filtering. The user-based collaborative filtering makes recommendations to the target user based on the rating data of nearest neighbors with similar scores. Item-based collaborative filtering recommendation depends on the similarity of the project for recommendation. Model-based collaborative recommendation, this method first uses a large number of historical scoring matrices to train and build the model and then uses this model to recommend the forecasting project offline. A variety of machine learning methods can be used to build models, such as Bayesian networks, clustering algorithms, neural networks and latent semantic retrieval.

For the user-based collaborative filtering recommendation algorithm, this method is based on the assumption that if the user's ratings for some items are similar and their ratings for other items may also be similar. The algorithm predicts the target user's rating of an item based on the rating of an item by the nearest neighbor (a number of similar users) of the target user and finally recommends the top N items with the highest rating to the target user. This algorithm can be divided into the following three steps.

The first is to construct a user-item evaluation matrix and collect user's rating and evaluation behaviors of the project and clean and tidy up and finally form a user evaluation matrix for the project. Where $U=\{u_1,u_2,\ldots,u_m\}$ represents m users; $I=\{i_1,i_2,\ldots,i_n\}$ represents n items, thus forming an m*n user-item evaluation matrix R(m,n).rp,q represents Userp's rating of item Itemq, see the following Table 1.

**Table 1.** Item score sheet.

| User/ Item | $i_1$ | $i_2$ | … | $i_q$ | … | $i_n$ |
|---|---|---|---|---|---|---|
| $u_1$ | $r_{1,1}$ | $r_{1,2}$ | … | $r_{1,q}$ | … | $r_{1,n}$ |
| $u_2$ | $r_{2,1}$ | $r_{2,2}$ | … | $r_{2,q}$ | … | $r_{2,n}$ |
| … | … | … | … | … | … | … |
| $u_p$ | $r_{p,1}$ | $r_{p,2}$ | … | $r_{p,q}$ | … | $r_{p,n}$ |
| … | … | … | … | … | … | … |
| $u_m$ | $r_{m,1}$ | $r_{m,2}$ | … | $r_{m,q}$ | … | $r_{m,n}$ |

The nearest neighbor set is formed, firstly, the similarity between the users is calculated and then the nearest neighbor set of the target user is searched according to the similarity between the users. There are many ways to measure the similarity between user up and user uq(labeled as sim(up,uq)),mainly including cosine similarity, correlation similarity and modified cosine similarity.

Cosine similarity: The similarity between two users is measured by the cosine of the angle between the two vectors.

$$sim(u_p,u_q)=\cos(u,u)=\frac{\vec{u}_p\cdot\vec{u}_q}{\|\vec{u}_p\|\times\|\vec{u}_q\|}=-\frac{\sum_{i=1}^{n}r_{p,i}\times r_{q,i}}{\sqrt{\sum_{i=1}^{n}(r_{p,i})^2}\times\sqrt{\sum_{i=1}^{n}(r_{q,i})^2}} \quad (1)$$

Correlation similarity: The correlation coefficient proposed by Pearson measures the similarity between users. Among them, rp,k and rq,k represent the score values of the user up and the user uq for the item k(k is the element of the scoring item set Ipq). $\bar{r}_p$ and $\bar{r}_q$ refer to the user's up and user uq average ratings for all items, respectively.

$$sim(u_p,u_q)=-\frac{\sum k\in I_{pq}(r_{p,k}-\bar{r}_q)}{\sqrt{\sum k\in I_{pq}(r_{p,k}-\bar{r}_q)^2}\times\sqrt{\sum k\in I_{pq}(r_{p,k}-\bar{r}_q)^2}} \quad (2)$$

Modified cosine similarity: In the cosine similarity measurement method, different users' rating scale problems are not considered and the modified cosine similarity measurement method improves the above-mentioned defects by subtracting the user's average score of the items. Among them, Ipq denotes a set of items evaluated by the user up and Ip denotes a set of items evaluated by the user uq.

$$sim(u_p,u_q)=-\frac{\sum k\in I_{pq}(r_{p,k}-\bar{r}_q)(r_{q,k}-\bar{r}_q)}{\sqrt{\sum k\in I_p(r_{p,k}-\bar{r}_q)^2}\times\sqrt{\sum k\in I_q(r_{p,k}-\bar{r}_q)^2}} \quad (3)$$

Then, a recommendation object is generated and based on the user's rating of the nearest neighbor of the item, a predicted value of the user's evaluation of the target item is obtained and a Top-N recommendation set is obtained. We chose the most similar k users to the target user and scored their predictions. Among them, ik represents candidate items in the candidate set. sim(up,uq))is the similarity between the user up and uq and uk is the nearest neighbor set of the target user up.

$$P(i_k)=\bar{r}_p+\frac{\sum u_q\in U_k sim(u_p,u_q)\cdot(r_{q,k}-\bar{r}_q)}{\sum u_q\in U_k sim(u_p,u_q)} \quad (4)$$

The degree of preference of the user up to the unrated item is predicted by calculation and sorted in descending order and the first N items are selected as the final recommendation set and returned to the target user up.

### B. The MVC Architecture

Using the MVC (Model-View-Controller model-view-controller) architecture to design web applications. MVC is a software design specification that separates business logic and data display through models, views and controllers.

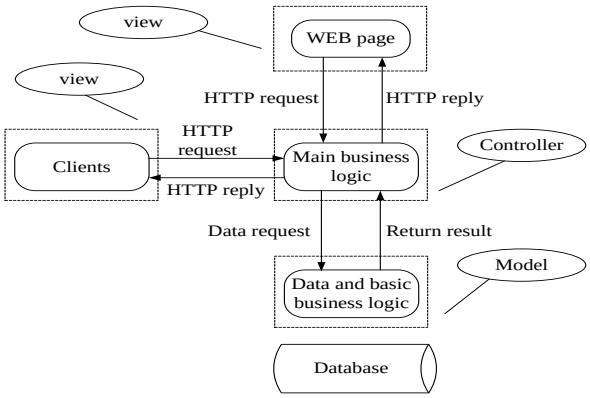


**Figure 1.** Shows the hierarchical architecture of the system.

A view is an application interface that interacts with the user to construct a friendly user interface and simple validation.

For example: the user interface style and password validation and other functions. An MVC system can have multiple views, which serve as a display output for data and a processing platform for user interaction. This part of the code is mainly composed of Html web pages, Javascript scripts and Phpxxx.php code.

Models are rules that deal with data and business. For example, user registration, login, query, and modification, are all implemented here. A model can provide data for multiple views. The model's implementation code can also be reused by multiple views, thus greatly reducing code reproducibility. This part of the code consists of the Php basic functions defined in the util.inc base class file.

The controller is used to receive the user's input and call the model and view to process the user's request. When the page button is clicked to send the request, the controller itself does not do any processing. It simply receives the request and decides which model to call to process the request, and then determines which view to use to display the returned data. This part of the code is mainly composed of Phpxxx_do.php code.

## IV. RESULT ANALYSIS AND DISCUSSION

### A. The Algorithm Test

The MovieLens data set (http://www.movielens.org/) was used in the experimental test, which is currently the most commonly used data set to evaluate the quality of recommended algorithms. It was created and maintained by the GroupLens research team at Minneaphdon University in the United States. The statistical work lasted for 7 months and the data was cleaned.

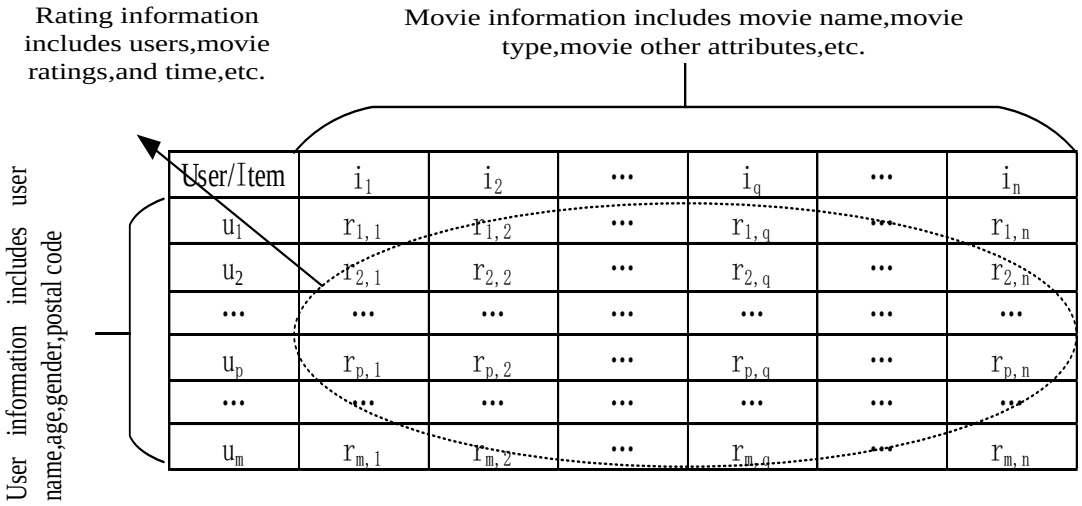


**Figure 2.** Schematic diagram of the MovieLens evaluation matrix structure.

Among them, the ml-100k data set includes 943 users' 100,000 comments on 1682 movies. The experiment needs to divide the data into two parts, which is 80 % of the training set, 20 % of the test set and two sets of base and test collection files.

By counting the distribution of data in the file, it selects the user 655 with the second highest total score. She evaluated a total of 685 data and there were 490 evaluations in the training u4.base file. There were 195 evaluations in the u4.test file, which basically met the test conditions of 80 % of the training set.

The experiment uses a user-based collaborative filtering algorithm and predicts the unevaluated items of the user 655 according to the input-process-output steps.

First, it analyzes the impact of the number of nearest neighbors on MAE evaluation.

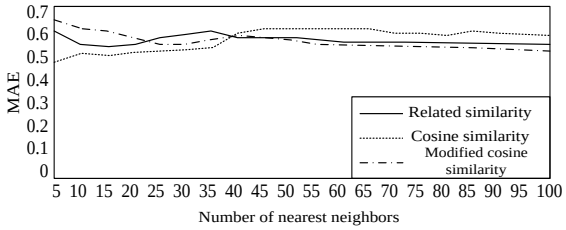


**Figure 3.** Influence of number of nearest neighbors on MAE in three similarity methods.

Using the three similarity functions to calculate separately, when the number of nearest neighbors is found to be 55, the MAE results of the correlation algorithms tend to be flat. Seeing the above figure for the experimental results.

•Then analyze the effect of Top-N number on MAE evaluation.

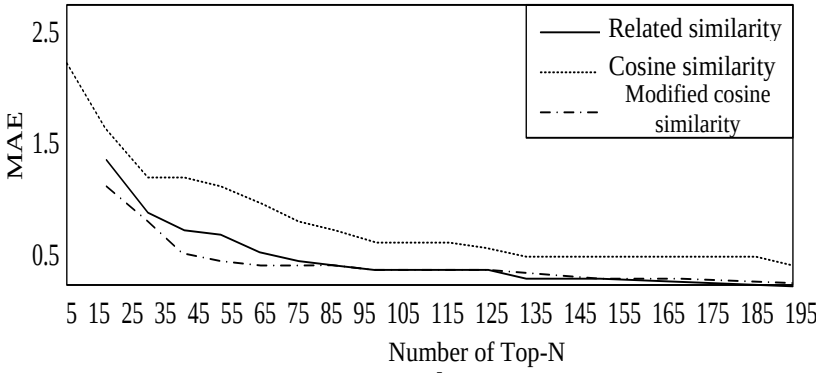


**Figure 4.** Effect of the number of recommended items in the three similarity methods on MAE.

Three similarity functions are used with the aim of calculate separately and attention is paid to the direct influence of Top-N recommendation on the results. When Top-N is greater than 35, the MAE results of each correlation algorithm are gradually flattened. Seeing the above figure for the experimental results. The improvement method I proposed is: The user-item evaluation matrix is screened, for example, the user can sort according to the behavior time according to the behavior of the last rating or purchase and only take the score data in the most recent period of time. In this way, a large amount of scoring data can be converted into some targeted and effective information and then the nearest neighbors are recommended.

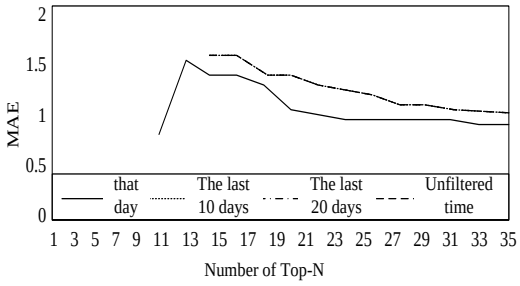


**Figure 5.** Effect of the cosine similarity method after filtering nearest neighbors on time for MAE.

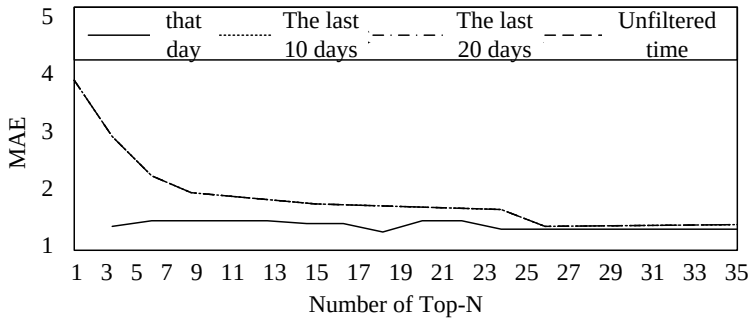


**Figure 6.** Effect of the similarity method after filtering nearest neighbors on time for MAE.

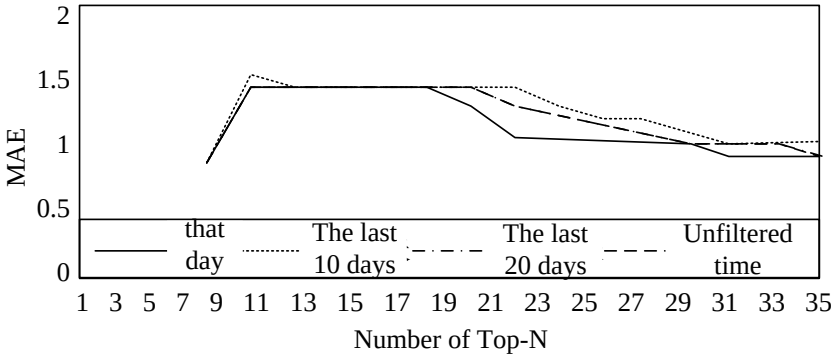


**Figure 7.** Effect of modified cosine similarity method after filtering nearest neighbors on MAE.

According to the above analysis, three similarity functions were used to calculate and the data was filtered according to the data of the end of the test, the last 10 days and the last 20 days. The test results are shown in the figure above.

Since the items to be evaluated in the test set are determined, it may not necessarily be the recommended result in the algorithm output Top-N. But it can still be seen from the result curve that the algorithm is more effective than others in the limited recommended quantity.

### B. System Application

The "i-magazine" refers to the related theories of the e-Journal service system reference evolution model (a kind of software life cycle model) to analyze and design and then uses the RUP iteration method for system development.

The evolution model actually belongs to the iterative development method and is mainly used for development projects that cannot completely define user requirements in advance. According to the original needs of users to quickly build a prototype to achieve basic user interaction with the system.

The prototype (alpha version program) provides a small-scale user trial and further clarify the requirements through discussion and communication with the user for the prototype, so as to truly grasp what the user needs of the software product looks like. After at least two iterations of the developed software system (beta version of the program) can provide users with a wide range of trials. In fact, each iteration adds a definable and manageable function to the entire system. In fact, the evolution model can be seen as iteratively executing multiple" waterfall models".

The system adopts the development mode combining BS(PHP+AJAX) and CS(C#+XML) architecture, and provides a complete application system for users with Windows Server IIS website server and SQL Server database server as the platform.

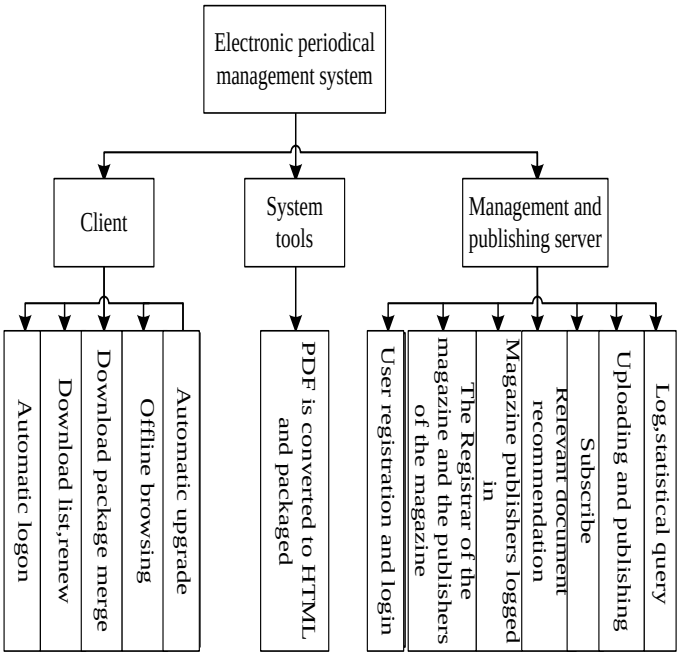


**Figure 8.** System architecture.

For the registration and login of the users, ordinary users, magazine publishers and system administrators register and log in the system. The interface is as follows:

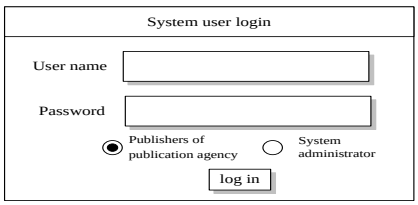


**Figure 9.** Shows the administrator login screen.

Ordinary users, magazine publishers and system administrators can register and log in the system. After logging in, the system administrator first registers the magazine and then registers the magazine user (i.e., publishes the personnel account) and then informs the magazine of the publisher's account. Magazine publishers can use this account to log in to the system. Ordinary users can directly register and log in to the system. For the upload of e-magazines, the uploading of e-magazines and the release of information. Magazine publishers can upload e-magazines to servers and publish information on servers. After the new magazine is released, the program will automatically scan the subscription table. If a user subscribes to this magazine, then adding a corresponding record in the download table for users to download and read. For e-magazine subscription, the interface is as follows:

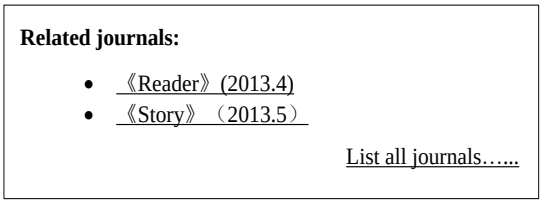


**Figure 10.** Shows the related journal list interface after login.

Users can subscribe to e-magazines online. After the user logs in and click “Subscribe to this book” to complete the subscription. If the user is not logged in, the prompt “You have not logged in yet” is given. If the posting person or system administrator clicks “Subscribe to this book”, the message “You have not logged in yet” is displayed (i.e., treated as a non-login user). If the user logs in to the system, a collaborative recommendation related magazine based on the user behavior is displayed in the page. After the user clicks on an item in the related journal list, the user enters the latest interface of the magazine (i.e., the subscription interface of the magazine). If the user is not logged in, only the magazine's past journals are displayed and they are reversed in chronological order. The first five items are displayed by default. The user can also click the More button to show all journals. If the user subscribes to a newly published journal, the default user will always subscribe to the next journal; if the user subscribes to a published journal, only the journal of this period can be downloaded.

For the related functions of the client, the main functions include functions such as automatic login of the client, offline browsing, viewing and downloading of subscription information and the client can automatically log in, browse offline, automatically upgrade, view subscription information and so on. After the user opens the client, he or she can automatically log in to the system (user can also enter the user name and password to log in), select magazine browsing, view subscription records and so on. When the client opens, it automatically checks if it wants to upgrade. There are two types of automatic login: automatic login to the system and automatic login to the client only. Among them, the former includes the latter, that is, when the user selects "automatically log in to the system", it is also required to automatically log in to the client. This is equivalent to setting a password for the client. This facilitates users who use public machines and also allows users to consume valuable bandwidth while browsing offline.

The left side of the "Reading Magazine" interface is a tree view, and the right side is a Web control. The tree view shows the organization structure of the files in the storage directory. The Web control shows the profile of the magazine selected by the user. The tree view of "Reading Magazine" shows the directory structure of the magazine. The Web control shows the content of the magazine selected by the user. A hidden Web control is embedded in each of the login interface and the main dialog interface (i.e., the carrier of the "my magazine" interface) for interacting with the server.

When the user logs in to the system through the login interface, the server-side login authentication interface is opened using the hidden Web control. If successful, the user enters the My Journal interface. If the user fails, the browser is invoked to open the failed page. After logging in to the system or opening the "My Download" screen, user can use the hidden Web control to open the server-side download information interface to obtain the user's latest download information. Users can download subscribed e-magazines.

The user can see the progress of the download and can resume it by clicking resume. You can pause or delete the download task. When the client just logs in to the server or receives a server query, it sends a request to the server to see if there is a new subscription. After the user confirms the download or the user continues the previous download, the client sends a download request to the server and the server starts the download after verification. If the content is an advertisement, the user does not need to confirm and download directly. If the downloaded file is split, the client automatically merges them.

## V. CONCLUSIONS

This service application system has been developed and implemented through a basis on the principles of starting from the magazine agency and serving the general users. It provides translation and uploading of

format for e-magazines by magazines and functions of ordinary users (such as registration, login, subscription, download and reading). The user-item evaluation matrix constructed by the system's own "subscription information table" is used and then the collaborative recommendation application module is invoked and finally a user-based related document recommendation function is formed. We learned from the text that collaborative hypotheses and model calculations are performed through the collaborative recommendation algorithm and based on the project based on the needs of the users and the use of function data is measured and analyzed. Some users were selected to match the historical data of the server and the results were calculated and analyzed. It was concluded that the optimized use of the e-journal management system was more reasonable. Finally, it has reached an empirical conclusion and designed a multimedia file management system that meets the needs of users.

## ACKNOWLEDGEMENTS

The present job has been supported by: the project 'Future School (infant education)' of National Center for Schooling Development Program of China (Nº. CSDP18 FC3204); Scientific Research Platform of Chongqing University of Education (No. 2017XJPT07); Scientific Research Program of Chongqing University of Education (No. KY2016TZ02).

## REFERENCES

Brown, T., S. Du, H. Eruslu, and F. J. Sayas, "Analysis of models for viscoelastic wave propagation," *arXiv preprint*, 1802.00825 (2018).

Fernández-Pousa, C.R., "Perfect phase-coded pulse trains generated by Talbot effect," *Applied Mathematics and Nonlinear Sciences*, **3(1)**, 23-32 (2018).

Lim, C., M. Kang and S. Lee, "Design and implementation of a collaborative team-based cloud storage system," *The Journal of Supercomputing*, **72(9)**, 3428-3442 (2016).

Lin, L.L., Y.T. Song Y.X. Tang, Q.Q. Du, and Y.P. Gong, "Implementation and application study on 3D collaborative design for CFETR based on ENOVIA VPM", *Fusion Engineering and Design*, **100(1)**, 198-203 (2015).

Ma, T., L. Guo, M. Tang, and Y. Tian, "A collaborative filtering recommendation algorithm based on hierarchical structure and time awareness," *IEICE TRANSACTIONS on Information Systems*, **99(6)**, 1512-1520 (2016).

Naji, H.M. and G.A. Al-Sultany, "Collaborative filtering recommendation using personalized page rank algorithm with new personalized parameters," *Journal of Engineering Applied Sciences*, **12(16)**, 4108-4112 (2017).

Sang, A. and S.K. Vishwakarma, "Design and implementation of collaborative filtering approach for movie recommendation system," *Intnl. Journal of Comp. Applications*.., **167(12)**, 18-24 (2017).

Thorat, B.P., R.M. Goudar and S. Barve, "Survey on collaborative filtering, content-based filtering and hybrid recommendation system", *Intnl. Journal of Comp. Applications*., **110(4)**, 31-36 (2015).

Xiaojun, L., "An improved clustering-based collaborative filtering recommendation algorithm", *Cluster Computing*, **20(2)**, 1281-1286 (2017).

Zarzour, H., F. Maazouzi, M. Soltani, and C. Chemam, "An improved collaborative filtering recommendation algorithm for Big Data", *C.I.I.A.: 6th IFIP TC 5 International Conference*, 660-668 (2018).

**Received: December 15th 2017**
**Accepted: June 30th 2018**
**Recommended by Guest Editor**
**Juan Luis García Guirao**