

# Study on the Web Courses Resources Recommendation System based on WP Algorithm

<http://dx.doi.org/10.3991/ijet.v11i06.5666>

Zhang Hua<sup>1,2</sup>, Zheng Shi-jue<sup>2</sup> and Xu Hong<sup>2</sup>

<sup>1</sup> Fuyang Vocational and Technical College, Fuyang, Anhui, China

<sup>2</sup> Huazhong Normal University, Wuhan, China

**Abstract**—To solve the problem of low WEB course resources utilization rate and help users quickly find the high-quality course resources, a WEB course resources recommendation system based on WP algorithm was established. In this system, course resources were automatically classified using WP algorithm and a course quality evaluation model based on user implicit evaluation was also set up. The experimental results showed that the method had a very good classification effect and it could effectively narrow the scope of the resources searched by users and improve resource search quality.

**Index Terms**—Web Courses Resources, Recommendation System, WP Algorithm.

## I. INTRODUCTION

Along with the further development of educational information system, digital resources are increasingly abundant and have become important parts of teaching. To make the high-quality teaching resources widely shared, the "Excellent Courses Construction Project Subject to the Teaching Quality and Educational Reform Project for Institutions of Higher Learning" was launched by the Ministry of Education in 2003. Under the project 14,348 national-level excellent undergraduate courses and 5,924 higher vocational courses have been established so far. Meanwhile, rich course resources have been accumulated, as the course construction is successively implemented by the institutions of higher learning in all provinces. Currently, in course resources website construction, tree structure is primarily considered. Trees are independent of each other; the leaf nodes (i.e. courses) within the same tree have no direct link between each other and also are very difficult to search by the WEB PageRank algorithm and the data sharing is poor. As a result, users do not find the presence of their demand in rich course resources. For this reason, the construction of the WEB course resources recommendation system based on data mining technology is an effective way to improve the utilization ratio of course resources.

Recommendation system, based on mass data mining provides relevant information and services according to user's demands. It is mainly applied to the fields such as e-commerce, library resources, and network news now. In references [1-4], the system architecture, algorithm and design of e-commerce recommendation system were studied respectively. In references [5-6], the application of data mining to the field of digital books was discussed. In references [7-9], the application of recommendation system to network news was discussed. In reference [10], a

recommendation system based on semantic net was established and applied to agriculture and finally good results were achieved. The current recommendation system is divided into two types: the behavior-based recommendation system and the content-based recommendation system. In the behavior-based recommendation system, comments from users are demanded, and therefore the input data sources are restricted. In references [11-13], recommendation system based on implicit comments and collaborative filtering was raised, and a certain effect was played. In the content-based recommendation system, the accurate keywords provided by the users in the field of demand are demanded. Thus, this system has been meaningless for the vast new users who desire to quickly find the core data of some field. How to recommend high-quality course resources to users according to the user's fuzzy demand is of important application value. In this paper, course characteristic matrix was established based on course contents. The courses were classified using WP algorithm, recommendation model based on intra-class course quality evaluation was designed and finally an experimental process based on WEB course resources clustering was given. The experimental result showed that the clustering result of this method was with better stability and accuracy in comparison with that of other algorithms.

## II. WEB COURSE RESOURCES RECOMMENDATION SYSTEM BASED ON WP ALGORITHM

Content-based project recommendation method was applied to this system. WEB course resources were collected through machine learning, and then the themed characteristics of courses were refined. It also clustered according to course's similarity and the high-quality course resource information focused by users was recommended by combining with the quality of course resources. The system mainly contains offline and online parts, as shown in Fig.1. In Part 1, a course characteristics model was established through the regular machine learning of WEB course resources, the textual analysis of resource contents, the characteristics extraction and then the courses were clustered. This part was the basic of the system and its operation was time-consuming, and offline operation helped improving the system performance efficiency. In Part 2, classes were identified according to the contents input by the users and then the corresponding course resources recommendation list of output according to identification rules. The details of each functional module are introduced below.

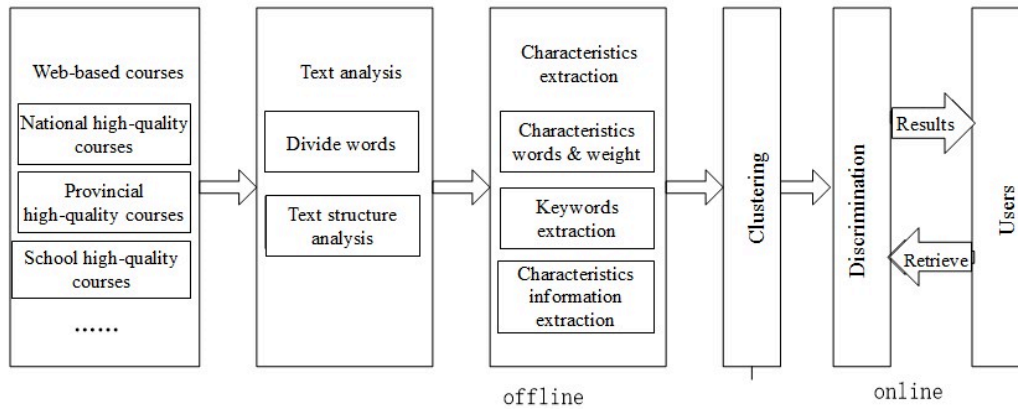


Figure 1. Recommendation system structure

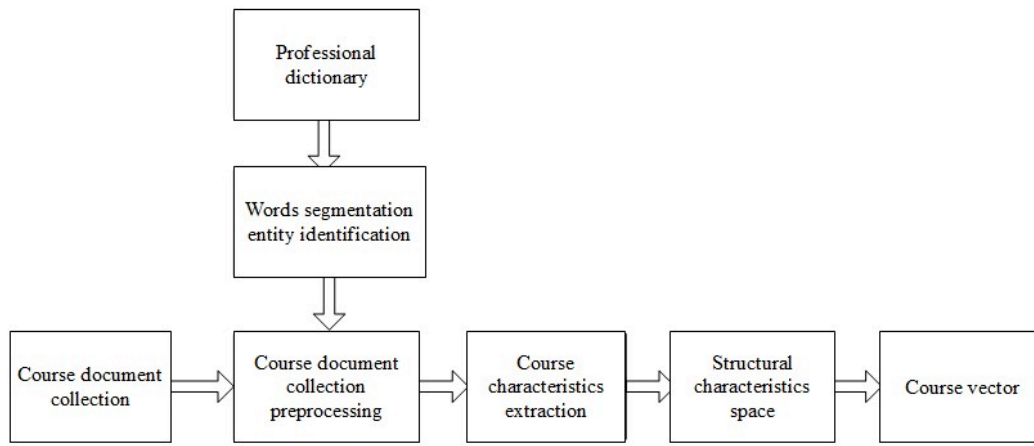


Figure 2. Process of data processing

A. Text preprocessing

The users of WEB courses are numerous and constantly changing, but the quantity and contents of WEB courses remain relatively stable. Therefore, the project recommendation method based on related contents was applied. This method is only related to course contents and does not depend on users' behavior. Moreover, the update frequency of the course similarity matrix calculated by the method is low and the machine learning tasks are reduced. The contents of course websites were acquired first through "crawler". Then the documents of course characteristics were established. After that the document collection of course websites was sorted, converted and standardized. Finally keyword classification, course characteristics extraction and course characteristics matrix establishment were implemented by combining with professional dictionary. The specific process is shown in Fig.2.

B. Course resource pool model

The definition of course resource pool model was presented here. According to the Vector Space Model (VSM), course was defined as  $c_i = (f_j, w_{ij}) (i = 1 \dots n, j = 1 \dots m)$ , in which  $f_j$  (frequency) was the  $j$ th characteristic of course;  $w_{ij}$  was the weight of course  $i$  under characteristic  $j$ . If a characteristic item was contained by course, its frequency was taken and otherwise its value was 0. Then, course characteristics matrix was established as shown in table 1.

TABLE I.  
TABLE 1: COURSE CHARACTERISTIC MATRIX

Characteristic item / Course No.	$f_1$	$f_2$	$f_3$	...	$f_j$	...	$f_m$
$c_1$	$w_{11}$	$w_{12}$	$w_{13}$	...	$w_{1j}$	...	$w_{1m}$
$c_2$	$w_{21}$	$w_{22}$	$w_{23}$	...	$w_{2j}$	...	$w_{2m}$
$c_3$	$w_{31}$	$w_{32}$	$w_{33}$	...	$w_{3j}$	...	$w_{3m}$
$\vdots$	...	...	...	...	...	...	...
$c_i$	$w_{i1}$	$w_{i2}$	$w_{i3}$	...	$w_{ij}$	...	$w_{im}$
$\vdots$	...	...	...	...	...	...	...
$c_n$	$w_{n1}$	$w_{n2}$	$w_{n3}$	...	$w_{ni}$	...	$w_{nm}$

C. Course clustering based on WP algorithm

**Definition 1** (Pearson's correlation coefficient): Pearson's correlation coefficient  $r$  is used for measuring the degree of linear correlation between variables  $r \in [-1, 1]$  was established, suggesting the correlation was stronger and variables were in perfect positive correlation if  $r$  was larger. Its design formula is as follows:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

In equation (1),  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ,  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ , in which  $x_i$  and  $y_i$  were any variables;  $r$  was the correlation coefficient between any variables; if  $r=0$ , linear correlation did not exist between variables; if  $r$  was larger, the correlation was stronger.

**Definition 2** (sum of squares of deviations): It refers to the sum of the squares of deviations between each item and average.  $n$  samples were divided into  $k$  classes:  $G_1, G_2, \dots, G_k$ .  $X_i^{(l)}$  was used for expressing the  $i$ th sample in  $G_i$ ;  $n_i$  was the number of samples in  $G_i$ ;  $\bar{X}^{(l)}$  was center of  $G_i$ . Then, the sum of squares of deviations of samples in  $G_i$  [14, 15, 16] is as follows:

$$S_i = \sum_{l=1}^{n_i} (X_i^{(l)} - \bar{X}^{(l)}) (X_i^{(l)} - \bar{X}^{(l)}) \quad (2)$$

The between-class sum of squares of deviations of  $k$  classes is as follows:

$$S = \sum_{i=1}^k S_i = \sum_{i=1}^k \sum_{l=1}^{n_i} (X_i^{(l)} - \bar{X}^{(l)}) (X_i^{(l)} - \bar{X}^{(l)}) \quad (3)$$

**Definition 3** (WP algorithm): In the characteristic Vector Space Model of course document  $C$ , the higher the weight of characteristic word  $F$ , the more important  $F$  became to  $C$ ; the first  $N$  preset numbers were chosen from the descending order list of  $C$ 's characteristic words to reduce the characteristic dimensions of the document. Pearson's correlation measure similarity between courses was chosen; courses were clustered using Ward's method [17][18].

The course clustering steps based on WP algorithm are as follows:

1. topN items were chosen from the descending order list of  $C$ 's characteristic words to reduce the characteristic dimensions
2. The similarity matrix between courses was calculated according to formula (1)
3. Each course was seen as one class and there were a total of  $n$  classes. Two courses with the maximal similarity were incorporated into one class, and then there were a total of  $n-1$  classes
4. The distance between new class and other classes was calculated
5. The sum of squares of deviations would increase when each class was narrowed and two classes with the minimal increase of the sum of squares of deviations were chosen and incorporated
6. Step (4) was repeated until all courses were incorporated into one class

#### D. The generation of the recommendation list

Through the above clustering operation, tens of thousands of courses collected from the Internet were clustered according to the course contents and there were still numerous courses within each class. All courses within the same class might not be browsed by users. How were high-quality courses chosen from each class? The course quality evaluation model based on implicit evaluation was established. Course quality was decided by the course service condition for users, and therefore, user implicit evaluation data was required to collect. The characteristic

actions (e.g. access time and frequency) of users to access to course websites were collected from servers, and then the course quality evaluation model was defined based on the above implicit evaluation data. Bipartite graph containing  $n+m$  nodes was defined:  $n$  was the number of users, and  $m$  was the number of courses; if user  $k$  accessed to course  $i$ ,  $k$  and  $i$  were linked to generate edge  $e_{ki}=1$ , and otherwise  $e_{ki}=0$ , as shown in Fig.3. User course bipartite graph adjacency matrix was constructed:  $u_k$  was user  $k$ ,  $c_i$  was course  $i$ , and  $e_{ki}$  meant user  $k$  accessed to course  $i$ . According to the principle of energy diffusion [19, 20, 21], the resources of each course were first allocated to each user accessing to the course, and each user then allocated the obtained resources to all courses he was involved in [22, 23] The formula for calculating the weight  $w_{ij}$  of resource allocation between courses is as follows:

$$w_{ij} = \frac{1}{D_j} \sum_{k=1}^n \frac{e_{ik} e_{jk}}{D_k}, (i, j = 1 \dots m, k = 1 \dots n) \quad (4)$$

Where,  $D_j$  was the degree of course  $j$ ;  $D_k$  was the degree of user  $k$ . Thus,

$$e_{ik} = \begin{cases} 0, & e_{ik} \notin W \\ 1, & \text{else} \end{cases} \quad (5)$$

$f_k = (e_{k1}, e_{k2}, \dots, e_{km})$  was ordered to express the initial resources allocation by user  $k$  to  $m$  courses;  $f'_k$  was the ultimate resources allocation of user  $k$  to  $m$  courses.  $f'_k = W f_k$  was established, and then the recommendation list was output according to the descending order of  $f'_k$  value.

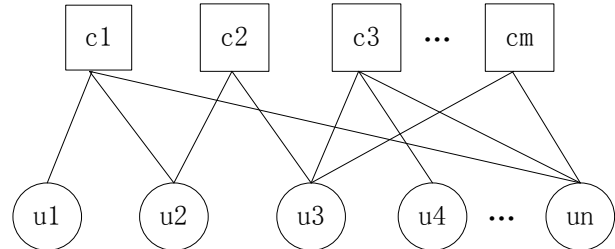


Figure 3. User-Course bipartite graph

### III. EXPERIMENTAL RESULTS AND ANALYSIS

#### A. Experimental methods and data

10 types of WEB course resources such as Parts Processing, Food Inspection, Advanced Mathematics, Logistics Technology and Practice, Obstetrics and Gynecology, Garden Plant Cultivation Techniques, Display System Design, Fundamental Nursing Technology, Interactive Media Design, and Aesthetics Foundation. These were acquired from the Internet using ROST Content Mining as the experimental data and 10 courses were contained in each type of resources. Lenovo Intel Core (i7 4510U, CPU frequency 2GHz, and memory 4Gb) was used as the experiment machine. Text mining software ROSTCM and data mining software SPSS20.0 were chosen as the experimental software. In this experiment, the experimental results of different algorithms were compared with comparison method, and the experimental effectiveness was verified using verification method.

B. The experimental process

Four steps were included in the experimental process. First, course contents were acquired. All HTML files of the given course websites were captured by ROST Content Mining software, and then analyzed and converted to textual files. Finally, that incorporated into one large textual file. Second, the characteristics of each course were abstracted. The words in the large textual file were segmented. After that the file was cleaned according to the limiting conditions such as enabling filtering word list, enabling retaining word list, enabling merger word list and outputting no monosyllabic words. Finally, the word frequency in the file was statistically analyzed and also arranged in a descending order. The frequencies of the first  $N$  characteristic words were chosen as the course characteristic items and this method was used for analyzing the characteristic items of the 100 courses in the 10 types of resources. Third, the courses library characteristic matrix was established. The above course characteristic items were incorporated into the courses library characteristic matrix using VC++ programming, as shown in Fig.4. Fourth, courses were clustered. For the courses matrix, courses were clustered using K-means, centroid clustering—cosine, the farthest neighbors—Pearson's correlation, the farthest neighbors—cosine, and Ward—cosine algorithms first and then the WP algorithm, and the result is shown in table II.

C. Result analysis

K-means algorithm, first chose  $k$  courses from  $n$  courses as the initial clustering centers and then calculated the distance between the rest courses and  $k$  centers, and then chose the nearest center for clustering. But its choice on the initial  $k$  centers was of strong subjectivity. Table II shows that the clustering result of the K-means algorithm was the worst: in class one, original 10 objects turned into 89, which gravely deviated from the actual situation. WP algorithm classified each course as one class and incorporated the two courses with the maximal similarity into one class. After that calculated the distance between new class and other classes, and later chose and incorporated the two classes with the minimal sum of squares of derivations until all courses were incorporated into one class. In this method, the characteristics of each course were manifested and two courses with the most similar characteristics were primarily clustered together. In the face of tens of thousands of different types of WEB course resources, specifically knowing the total number of their types was impossible. But clustering them together was possible. Concerning about this characteristic, unsupervised clustering operation was more suitably applied and WP algorithm just fit well because it obtained the better clustering results than other algorithms. But K-means algorithm violated this characteristic. Because it first artificially determined courses as  $k$  classes and then classified the rest

TABLE II. CLUSTERING RESULTS

Number of cases / Class No. / Clustering algorithm	1st	2nd	3 <sup>rd</sup>	4 <sup>th</sup>	5th	6 <sup>th</sup>	7th	8th	9th	10th
k-means	89	2	2	1	1	1	1	1	1	1
centroid clustering—cosine	54	5	11	5	5	2	8	3	5	2
the farthest neighbors—Pearson's correlation	17	5	6	5	23	6	6	13	9	10
the farthest neighbors—cosine	17	5	4	5	18	6	17	9	9	10
Ward—cosine	9	28	9	8	10	6	7	8	10	5
WP algorithm	10	9	8	27	10	6	7	8	10	5

...	V25	V26	V27	V28	V35	V36	V37	V38	V39	V40	V41	V56	V57	V58	V59	V60	V61	V62	V63	V64	V65	V66	V70	...
...	17	19	118	5	52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...
...	4	1	11	1	7	0	0	0	0	0	7	43	14	44	1	2	0	2	0	1	0	0	10	...
...	0	0	2	0	1	0	0	0	0	1	0	34	25	92	13	17	5	23	0	0	0	0	0	...
...	4	4	2	6	0	0	0	3	3	0	1	7	1	5	1	0	0	1	0	0	0	0	12	...
...	13	0	71	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...
...	0	0	2	5	0	1	0	0	0	0	0	2	4	2	5	2	3	0	0	0	0	0	5	...
...	0	4	0	0	0	0	0	0	0	0	2	26	9	30	2	0	0	2	0	0	0	1	12	...
...	0	0	4	3	3	0	1	1	0	0	6	25	5	12	3	3	2	7	0	0	0	0	0	...
...	0	0	0	0	0	0	0	0	0	0	2	11	2	5	0	1	0	0	0	4	0	0	0	...
...	0	0	2	0	40	0	0	4	0	14	9	3	1	0	3	0	0	1	0	0	0	0	0	...
...	3	1	9	3	3	0	0	1	0	1	3	14	3	16	0	0	0	1	0	0	0	0	2	...
...	36	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...
...	0	0	0	0	0	0	0	0	0	1	0	4	15	40	2	1	3	13	0	0	0	0	0	...
...	4	1	5	0	3	0	0	0	0	0	7	3	11	3	1	0	4	1	0	0	1	2	...	
...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...
...	4	0	8	1	0	0	0	1	3	0	1	3	5	12	1	0	1	0	0	1	0	0	0	...
...	0	0	0	0	0	0	0	0	0	0	0	18	3	46	0	2	0	0	0	0	0	0	0	...
...	15	0	30	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...
...	0	0	0	0	7	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	...
...	0	0	0	3	0	0	0	0	0	4	0	1	1	3	0	1	0	0	0	0	0	0	3	...
...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...
...	2	2	8	11	2	0	0	1	0	0	0	0	0	0	1	0	0	3	0	0	0	0	0	...
...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...
...	1	0	4	0	4	0	0	0	0	1	2	8	2	6	1	0	0	3	0	2	0	0	0	...
...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...

Figure 4. Course matrix

into  $k$  classes. But the  $k$  value was difficult to determine. It did not beforehand known how many classes of WEB course resources were necessarily divided into. Its constant iteration later was only a gradual adjustment and failed to change the initial subjectivity too. However, in WP algorithm, a reasonable number of categories were obtained through the automatic incorporation of classes. Table II shows that the result of WP algorithm was with better stability and accuracy in comparison with those of other clustering algorithms. The effectiveness of the clustering result was a standard of measuring a clustering algorithm. The degree of the consistency between clustering result and artificial classification result could be examined using Pearson's correlation coefficient and variance analysis.

In WP algorithm, the correlation between courses was measured using Pearson's correlation coefficient. Pearson's coefficient reflected the degree of the correlation between courses. If the classification effect was good, Pearson's coefficient between courses in the same class was greater. But Pearson's coefficient between classes was smaller. Any class was chosen. For example, 5 cases in the tenth class were chosen for testing the Pearson's coefficient and the experimental result is as shown in table III:  $p$  value in significance test was 0 ( $p < 0.01$ ), and thus the null hypothesis was rejected and there was a significant correlation between variables at the level of 0.01. The degree of correlation between variables was measured by effect size. As shown in table III, the correlation coefficient of V93 was 1 and conformed to the actual situation;  $r$  value of V93 with V94, V97, V98, and V101 was 0.54, 0.611, 0.293, and 0.611 respectively.  $r$  estimate value provided by Cohen was  $\pm 0.1, \pm 0.3, \pm 0.5$ , corresponding to small, medium and large effect sizes respectively. Thus, the correlation coefficients of V93 with V94, V97, V98, and V101 were corresponding to strong positive correlation in the actual situation. From the above analysis, it is known that the degree of correlation between courses in the same class was high.

In WP algorithm, the Ward method was used as the clustering method and the clustering effect of the Ward method was tested using  $F$  statistics. The total variation

was  $SST = SSA + SSE$ , in which  $SSA = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2$  was used for expressing the sum of squares of deviations between groups, and  $SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$  was the sum of squares of deviations in the same group.  $F$  statistics was the ratio between the average between-class sum of squares and the average intra-class sum of squares, as follows:

$$F = \frac{SSA / (k - 1)}{SSE / (n - 1)} \quad (6)$$

Thus, the greater the between-class sum of squares of deviations, the greater  $F$  value was. Similarly, the greater the intra-class sum of squares of deviations, the smaller  $F$  value was. The experimental result of variance analysis is as shown in table IV: the null hypothesis was rejected ( $p < 0.01$ ), indicating the between-group difference was significant. ANOVA effect size was

$$\eta^2 = \frac{\text{sum of squares between classes}}{\text{Total sum of squares}}$$

and the values of the small, medium and greater effect sizes agreed by Cohen were corresponding to 0.01, 0.06 and 0.14, respectively. The values of effect size in table IV show that the difference between classes was corresponding to very great effect.

#### IV. CONCLUSION

To change the problem of low WEB course resources utilization rate and help users quickly find the high-quality course resources, a WEB course resources recommendation system based on WP algorithm was established. In this system, course resources were automatically classified using WP algorithm and a course quality evaluation model based on user implicit evaluation was also set up. The experimental results showed that the method played a very good classification effect and could effectively narrow the scope of the resources searched by users and improve resource search quality.

TABLE III.  
PETERSON INSPECTION

		V93	V94	V97	V98	V101
V93	Pearson correlation	1	.541**	.611**	.293**	.376**
	Significant (both sides)		0	0	0	0
V94	Pearson correlation	.541**	1	.515**	.365**	.323**
	Significant (both sides)	0		0	0	0
V97	Pearson correlation	.611**	.515**	1	.275**	.206**
	Significant (both sides)	0	0		0	0
V98	Pearson correlation	.293**	.365**	.275**	1	.355**
	Significant (both sides)	0	0	0		0
V101	Pearson correlation	.376**	.323**	.206**	.355**	1
	Significant (both sides)	0	0	0	0	

\*\* . Significant correlation (both sides) at the level of .01

PAPER  
STUDY ON THE WEB COURSES RESOURCES RECOMMENDATION SYSTEM BASED ON WP ALGORITHM

TABLE IV.  
ANOVA

		Sum of squares	df	MS	F	Significance	Effect size
v1	Between-class	1296.99	9	144.11	3.89	.000	0.28013
	Intra-class	3333.00	90	37.03			
	Sum	4630.00	99				
v2	Between-class	5285.25	9	587.25	9.61	.000	0.49015
	Intra-class	5497.78	90	61.08			
	Sum	10783.04	99				
v3	Between-class	319.86	9	35.54	3.98	.000	0.28516
	Intra-class	801.84	90	8.90			
	Sum	1121.71	99				
v4	Between-class	7749.65	9	861.07	34.38	.000	0.7747
	Intra-class	2253.90	90	25.04			
	Sum	10003.56	99				
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

REFERENCES

- [1] YU X, "Distributed data mining-based e-commerce recommendation system", *Computer Systems & Applications*, no.11, PP.33-37, cn.2009.
- [2] HUANG Y, "Application of improved apriori algorithm in e-commerce recommendation system", *Computer & Digital Engineering*, vol.40, no.8, pp.35-38, cn, 2012.
- [3] WANG Z.DENG L.SHI W, "Research on the application of the data mining technique to the e-commerce recommender system", *Microelectronics & Computer*, vol.24, no.4, pp.197-199, cn, 2007.
- [4] SHI R.YAO T, "Recommendation system design based on web data mining", *Microcomputer Applications*, vol.28, no.11, pp.35-36,cn, 2012.
- [5] CAI W, "University library website personalized recommendation system based on data mining", *Information Research*, no.3,pp.88-90,cn, 2011.
- [6] DING X, "Research on book intelligent recommendation system based on data mining" *Information Studies: Theory & Application*, vol.33, no5, pp.107-110, cn, 2010.
- [7] CHENG H.CHENG W, "Web news recommendation based on multiple topic tracking", *Journal of Computer Applications*, vol. 31, no.9, pp.2426-2428, cn, 2011.
- [8] PENG F.QIAN X, "Personalized recommendation system for news based on user concern", *Application Research of Computers*, vol. 29, no.3, pp.1005-1007, cn, 2012.
- [9] WU Y.WANG X.DING Y et al, "Adaptive on-line web topic detection method for web news recommendation system", *Acta Electronica sinica*, vol. 38, no.11, pp.2620-2624, cn,2010.
- [10] WANG J, "On agricultural learning resource recommendation system based on semantic web", *Computer Applications and Software*, vol. 30, no.8, pp.233-235, cn, 2013.
- [11] TUO J.WANG S.LI X et al, "Framework and implementation of recommender system in great course sharing system", *Computer Engineering and Design*, vol. 27, no.17, pp.3119-3122, cn, 2006.
- [12] Jon Herlocker. Joseph A. Konstan. John Riedl, "An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms",*Information Retrieval*.no.4,pp.287-310, 2002. <http://dx.doi.org/10.1023/A:1020443909834>
- [13] Balabanovic M. Shoham Y. Fab: Content-Based collaborative recommendation, *Communications of the ACM*, vol.40, no. 3, pp.66-72, 1997. <http://dx.doi.org/10.1145/245108.245124>
- [14] JIA L.ZHANG Z.WANG T et al, " Study of "Touch pitch and you will be defiled" based on the clustering method", *Neijiang Science & Technology*, no.11, pp.73-74,cn, 2012.
- [15] Xin Li, "An online blog reading system by topic clustering and personalized ranking", *ACM Transactions on Internet Technology (TOIT)*, 2009.
- [16] Luro F. Laigret F. Bove JM et al, "DNA amplified fingerprinting, A useful tool for determination of genetic origin and diversity analysis in citrus", *Hort Science*, vol. 30, no.5, pp.1063-1067,1995.
- [17] J. Vesanto. E. Alhoniemi, "Clustering of the self-organizing map", *IEEE Trans Neural Network*, vol.11, no.3, pp.586-600, 2000. <http://dx.doi.org/10.1109/72.846731>
- [18] Yu Jinghua.Yang Weiquan, "Multivariate Statistical Analysis and Application", *Guangzhou, Zhongshan University publishing House*, cn, 2006, pp.176-177.
- [19] WANG Q.DUAN S, "Improved recommendation algorithm based on bipartite networks", *Application Research of Computer*, vol. 30,no.3 , pp.771-774,cn, 2013.
- [20] YANG B.ZHAO P, "Review of the art of recommendation algorithms", *Journal of Shanxi University (Nat.Sci.Ed.)*, vol.34, no.3, pp.337-350,cn, 2011.
- [21] L. Qing. M. K. Beyong, "Clustering Approach for Hybrid Recommender System", *Proceedings IEEEWIC International Conference on Web Intelligence (WI'03)*, 2003. <http://dx.doi.org/10.1109/WI.2003.1241167>
- [22] ZHANG X.JIANG S, "Personalized recommendation algorithm based on weighted bipartite network", *Journal of Computer Applications*, vol. 32, no. 3, pp.654-657, cn, 2012.
- [23] Tao Zhou etc, "Bipartite network projection and personal recommendation", *Physical Review*, E76 (4):0466115-0466115, 2007.

AUTHORS

**ZHANG Hua** is with the Department of Electronic Engineering and Information Technology, Fuyang Vocational and Technical College, Fuyang, Anhui 236031, China (e-mail:fyzyzh@126.com).

**ZHENG Shi-jue** is with the Department of Computer Science, Huazhong Normal University, Wuhan 430079, China (e-mail:Zhengsj@mail.cnu.edu.cn).

**XU Hong** is with the Department of Computer Science, Huazhong Normal University, Wuhan 430079, China (e-mail: 3792578@qq.com).

Submitted 20 March 2016. Published as resubmitted by the authors 16 May 2016. This paper was aided by the National Natural Science Fund Project (No.61370108), the Quality Engineering Project of Anhui Province (No.2013gxxk120) and the Teaching Research Project of Hubei Province (No.2013096).