Course Recommendation Mechanism Based on Content Recommendation Method – Illustrated by the Example of "NetEase Cloud Classroom"

Min Tan
Maoguo Wu
SHU-UTS SILC Business School, Shanghai University, China

Doi:10.19044/esj.2018.v14n19p167 URL:http://dx.doi.org/10.19044/esj.2018.v14n19p167

Abstract

This paper analyzes NetEase Cloud Classroom in various aspects, including development and business environment of MOOCs, technological maturity, and the algorithm of the recommendation model and its feasibility. Among these aspects, this paper mainly gives a detailed analysis and explanation of the content-based course recommendation. The technology of course recommendation is relatively mature at present. Meanwhile, the content-based recommendation mechanism is also proved to be feasible to some extent. However, the drawbacks of the recommendation mechanism itself can still bring risks and restrictions to the sustainable development of NetEase Cloud Classroom.

Keywords: Online Education, MOOCs, Course Recommendation, NetEase Cloud Classroom, Content-Based Recommendation

1. Background

With the continuous development of computing and network technology and their continuous penetration in many fields, traditional education has begun to shift to online education. Compared with traditional education, online education changes students' learning styles, which allows learners to acquire knowledge at any time, any place, and on any platform. Meanwhile, learners can choose what they really want to learn based on their own needs.

According to the *White Paper on Household Education Consumption of 2012 in China*, educational consumption is the most important part of household consumption, which accounts for 1/7 of the income of the Chinese quintiles and the proportion is expected to continue to grow. In addition, 31.7% of white-collar workers expressed that they would like to participate in vocational education to enrich and enhance themselves because of bottlenecks

in their work. In this group, no less than 56% of them have made huge investment in vocational training each year. And those on-the-job personnel account for more than 57% of the group who plan to learn online. From a holistic point of view, the scale of online education users in China is increasing year by year.

Under this background, massive open online courses (MOOCs) came into being. It emphasizes the openness of knowledge and the scale of students, as well as the social interaction between teaching and learning. Based on the teaching philosophy of lifelong education, MOOCs regard education as a basic human right and aims to provide more people with free learning opportunities and resources, knowledge and skills improvement, and promote personal and social development. Therefore, it is widely supported by the government and society.

At present, various kinds of MOOCs platforms have been set up. In worldwide, MOOCs platforms include enterprise-operated platforms such as Coursera, Udasity, Iversity, and Udemy, which are often profitable institutions that finance and operate through cooperation with governments and venture capital; it also includes the college alliance operating platform like edX, Open2Study, NovoED and FutureLearn, which are mainly in alliance with universities, and most of them also cooperate with government with non-profit operation model. In China, although the development of online education is relatively late and immature compared with Western developed countries, many online education platforms have also emerged, such as 'the smart tree', Muchanet, Superstar MOOC., NetEase Cloud Classroom *etc.*Their operation mainly relies on financing and the company's own investment, and also cooperate with the government and universities. Those Chinese MOOCs platforms also incorporate the excellent curriculum resources of major universities as elite universities in Western developed countries do.

The open online education platform provides learners a wealth of course resources and convenience for learning, but to a certain extent, it still has some problems. A large number of curriculum resources are presented at the same time, and learners cannot readily obtain the courses that are useful or of interest to themselves. Their true needs cannot be satisfied or even overwhelmed by useless information, which is called 'Information overload'. Under such circumstances, the mode that 'people seeking information' is no doubt much more difficult than the mode that 'the right information is already prepared for people'. If people can make full use of the data accumulated on platforms such as users' access and registration information, then they can analyze users' browsing preferences and learning habits to forecast the user's need of learning and form the individualized precise delivery, which can greatly improve the efficiency of online education and solve the problem of 'information overload'. The emergence of intelligent recommendation

algorithms and systems is undoubtedly the best way to deal with this phenomenon.

2. Technology Maturity Analysis

Technology maturity analysis is basically analyzed through data mining and recommendation system and personalized recommendation system. The recommendation system is to extract the resources that the user may be interested in or have demand from the mass information according to the interest characteristics of different users, and then conduct the recommendation. Therefore, the application of the recommendation system must be based on data mining. Data mining refers to extract the unknown and implicit but potentially valuable information or knowledge from fuzzy, noisy, large-scale, random, and incomplete data.

In Western developed countries, the idea of personalized service has long prevailed. In 1995, the personalized navigation system 'Web Watcher' was created at Carnegie Mellon University, followed by Stanford University which launched the personalized recommendation system 'LIRA'. Since then, the study of personalized recommendation services began. In 1996, 'Web Watcher' was upgraded to 'Personal Web Watcher'. In the same year, the well-known Internet company Yahoo launched the personalized portal 'My Yahoo!'.

Research on personalized recommendation technology in China started relatively late. Since Tsinghua University proposed a personalized recommendation based on multi-agent hybrid intelligence in 2000 to an agent-based personalized information filtering system 'Open Bookmark', and then to the application of the recommendation system in Douban website and 360doc, China has officially initiated the study of personalized services. At present, China's recommendation technology is mainly used in e-commerce, digital libraries, websites and other fields. However, at present, research on personalized recommendation in China still needs to be based on classical methods and techniques.

3. Common Recommendation Algorithms

Common recommendation algorithms include collaborative filtering, association rules, and content-based recommendation. Collaborative filtering algorithms generate recommendation based on similarities or common experiences among users, and it is the mainstream of today's recommendation technologies. The advantage of a collaborative filtering system is that it can discover new information for users that they may be interested in, but the problem of data sparsity and low scalability of the system might hinder the effectiveness of this algorithm.

The association rule algorithm is mainly used to analyze the correlation between different projects. There are two modules, i.e., 'search for high frequency item set' and 'find association rule from high frequency item set' which are included in this algorithm, so the study of it usually has supportiveness and confidence, these two indexes, as its indicators. But the 'cold start' problem is still a very typical defect of this method.

Content-based recommendation algorithm mainly uses the similarity between the content of the resource and the users' interests to filter information and data, and then provides advice to customers by automatically matching customer preferences with product content. This is also the most widely used method for personalized recommendations. This paper analyzes this algorithm in detail.

4. Method Design

NetEase Cloud Classroom is an online skill learning platform created by NetEase. It establishes cooperation with a number of education and training institutions. Its courses cover more than a dozen categories such as software application, IT, foreign language studying, and professional certification and so on. Among the numerous algorithms, the hybrid algorithm which mixes content recommendation and collaborative filtering is the basis for most course recommendations. The recommendation system of NetEase's cloud classroom is also based on this hybrid algorithm. This paper focuses on the part of content-based recommendation algorithms.

Information filtering and retrieval are the main theoretical basis for content recommendation. At present, there are two methods based on content recommendation. One is a model-based method, which means that collecting and integrating previous data of users to form the dataset, and then generating a recommendation model; the other one is a heuristic method, which means defining the similarity calculation formula of the project by experiences firstly, and then using the actual results to verify the calculation results and optimizing the formula. This paper focuses mainly on the heuristic approach.

The principle of the content-based recommendation mechanism is to match the user's interest preferences with the curriculum's content, and it includes the following three steps:

1) Acquire the characteristics of all courses within the course platform to form a course feature vector (this paper focuses on the course text content which includes title, profile, source, and speech draft to do the feature extraction)
2) Get the user's interest on the course selection and form a user preference vector
3) Match the similarity between the two and make a course recommendation based on the similarity

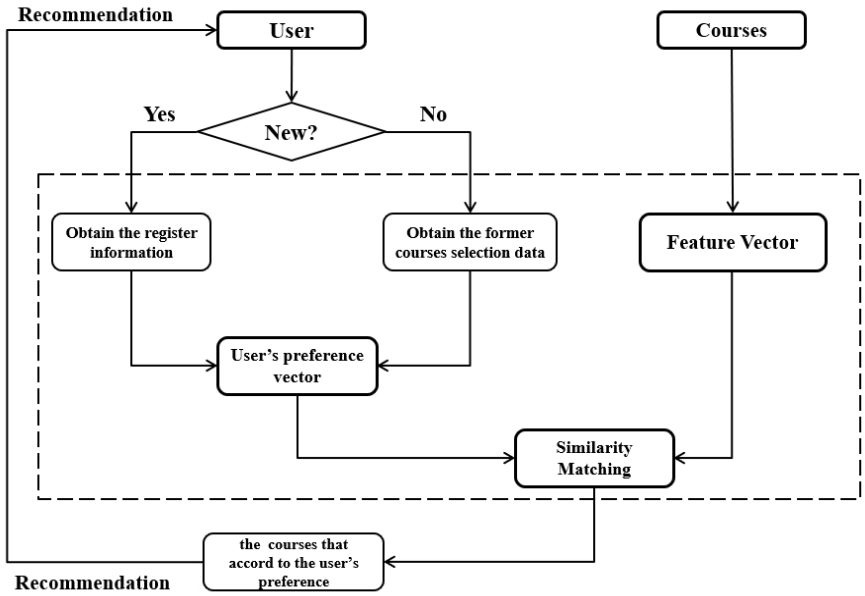

Figure 1 Principle of Content-Based Course Recommendation

Afterwards, form the feature vector of the course. All courses in the platform are defined as a set D: D = $\{d_j|d_1, d_2,..., d_n\}$, where *n* is the total number of courses provided by the platform. The words appearing in the content of all the course texts are defined as set T = $\{t_u|t_1, t_2..., t_u\}$. That is to say, on the platform, there are *n* courses that need to be processed and a total of *u* words are included. The j^{th} article is defined as a phrase set $d_j=\{w_{kj}|w_{1j},w_{2j},\ldots,w_{uj}\}$, where W_{kj} is the weight of the phrase t_k in the text content of the j^{th} course, *u* is the total number of words in the course document.

When the weight is greater, the more important the word is and it can represent the course more. Therefore, if one wants to use key words to represent the course content, the key is to calculate the weight of these words. At this point, the TF-IDF method can be applied for weight calculation. TF indicates the frequency of a feature in a project, and the higher frequency it is, the more representative the feature is in the project; IDF is the reverse document word frequency, indicating that when a feature is high frequency in all projects, the smaller the ability to use this feature to distinguish different items. Thus, IDF can be used to exclude some nonsensical words such as auxiliaries. The specific process is as follows:

Find the frequency TF_{kj} of the keyword t_k in the course document d_j:

$$TFkj = \frac{tf_{kj}}{n_j}, \qquad (1)$$

where tf_{kj} is the frequency of t_k in the course document d_j, n_j is the total number of words in the course document d_j.

Find the inverse document word frequency IDF_k of keyword t_k in all courses on the platform:

$$IDF_k = \log \frac{n}{n_k}, \quad (2)$$

where *n* is the number of courses in the course platform; n_k is the number of courses in the course platform that contain the word t_k.

Get the weight of the keyword t_k in the course document C_j:

$$wk_{kj} = TF_{kj} \cdot IDF_k = \frac{tf_{kj}}{u} \cdot \log \frac{n}{n_k}. \quad (3)$$

As a result, the weights of different words in the same course are shown in Table 1.

Table 1 Weight Table of Words in A Course

	t_1	t_2	t_3	……	t_u
d_j	w_{1j}	w_{2j}	w_{3j}	……	W_{uj}

According to the weight table, the feature vector of course D_j is $m = \{w_{1j}, w_{2j}, w_{3j}, \ldots, W_{uj}\}$. Since not every word appears in every course, w_{kj} may be zero in this feature vector.

Form the users' preference vectors. For the old users, since the system has already established a course selection database exclusively for old users during their use of the course platform, the system only needs to mine their preference information in the database. Therefore, dig the old user's course selection path is the first step to get the path set. And the user's course selection path set A_i is included in the set D of all courses provided by the platform.

$$Ai = \{Ck|C1, C2, \ldots, Cn\}, \quad (4)$$

where i is the i^{th} user, and C_k is the k^{th} course selected by the user.

The next step is to extract interest features from all courses selected by the user, which also uses the TF-IDF method to extract the features of those courses. The users' preference feature weight table is shown in Table 2.

Table 2 Weight Table of the Features of Users' Preference Courses

	t_1	t_2	t_3	……	t_u
C_1	w_{11}	w_{21}	w_{31}	……	w_{u1}
C_2	w_{12}	w_{22}	w_{32}	……	w_{u2}
C_3	w_{13}	w_{23}	w_{33}	……	w_{u3}
……	……	……	……	……	……
C_n	w_{1n}	w_{2n}	w_{3n}	……	w_{un}
Average Weight	w_1	w_2	w_3	……	w_u

According to the average weight, the old user's course selection preference vector is $I = \{w_1, w_2, w_3, \ldots, w_u\}$. Since the courses selected by the

users are different, not all feature words will appear in the path selected by the user, so w_k may be 0.

For the new users, the platform will ask them to complete their own relevant information when they register, including their age, gender, and interests. The system can directly obtain their favorite course feature information and form a course preference feature set. Because the new user's preference information is judged by his own self-assessment, the system cannot give weights to that directly. Therefore, in order to facilitate the similarity matching process, the model will give the value of 1 if the user has a preference of the course and 0 if the user does not have to form the new user's course selection preference vector.

For similarity matching and course recommendation, cosine similarity method is used to calculate the similarity of these two vectors.

$$\mathrm{Sim}(m,i) = \cos(m,i) = \frac{\sum_{k=1}^{u} w_{kj} w_k}{\sqrt{\sum_{k=1}^{u} w_{kj}^2}\sqrt{\sum_{k=1}^{u} w_k^2}}. \qquad (5)$$

Since the range of cosine is [-1, 1], the closer the cosine value is to 1, the higher the degree of similarity between the two vectors, and the higher the user's preference for the course. And then, the system will recommend the course to the user.

The feasibility of the method is tested through algorithm practice based on content recommendation for old users. Considering the data complexity caused by too many courses, this paper chooses 22 courses in different fields at random in the NetEase Cloud Classroom, and assumes that these courses are the only courses currently available in the NetEase Cloud Classroom. Then set up the vector set of all the courses in the platform, D = {1000110002,... 10022}. This algorithm involves extracting effective vocabulary from all the textual information contents, and the TF-IDF method is used to calculate the weights of these vocabularies. In order to simplify the process, this paper selects 21 keywords to represent the vocabulary set of all the courses, T = {office, management, psychology, ... , success, skill}. Based on the TF-IDF method, the weights given to the text content are shown in Figure 2. The value of each row represents the feature vector *m* of each course.

Courses in NetEase Cloud Classroom（assume that there are only 22 courses）												
No.	Name	Office	Management	Phychology	Relationship	Music	Language	Improvement	Market	Culture	Application	Thinking
10001	The Skills of MS Word that you don't know	0.17	0.09	0	0.02	0	0.07	0.12	0	0.01	0.22	0.01
10002	Excel Class for new workplace	0.16	0.1	0	0.01	0	0.09	0.1	0	0	0.2	0.01
10003	Premiere Basic Course: One day learn video	0.13	0.03	0	0.06	0.04	0.01	0.13	0	0	0.21	0
10004	Learn spoken English with foreign teachers	0.11	0.05	0.02	0.11	0	0.13	0.12	0	0.04	0.14	0
10005	PPT class for new career	0.16	0.07	0	0.01	0	0.06	0.11	0	0	0.23	0.03
10006	Learn to draw in 10 minutes	0.07	0.03	0.01	0.12	0	0	0.09	0	0	0.23	0.04
10007	Python combat: A week to crawl Web pages	0.16	0.06	0	0.01	0	0.07	0.14	0	0.01	0.21	0.03
10008	The Great God of PS	0.13	0.07	0	0.03	0	0.06	0.11	0	0	0.2	0.02
10009	10 Basic courses for singing beginners	0.04	0	0.02	0.14	0.25	0.01	0.02	0	0	0.13	0
10010	8 Week Fitness Class	0.1	0.13	0.03	0.05	0	0	0.11	0	0	0.15	0
10011	Mind Mapping Time Management	0.16	0.13	0.09	0	0	0	0.08	0	0	0.2	0.02
10012	Communication Phychology	0.11	0.05	0.13	0.07	0	0.01	0.07	0	0.01	0.13	0.1
10013	Microeconomics	0	0.06	0	0.01	0	0.05	0.11	0.13	0.05	0.15	0.07
10014	Digital marketing: Entering the brand of wisdom	0.06	0.14	0.01	0	0	0.01	0.1	0.15	0.02	0.12	0.08
10015	C Language Programming Advanced	0.11	0.03	0	0	0	0.18	0.13	0	0	0.24	0.09
10016	American Culture	0	0.1	0.06	0.04	0.07	0.04	0	0.09	0.13	0	0.07
10017	Logistic expression	0.07	0.04	0.04	0.09	0	0.03	0.13	0	0	0.15	0.1
10018	Appreciation and application of cosmetics	0	0	0	0.1	0	0	0.17	0	0	0.3	0
10019	The operation system of computer	0.09	0.1	0	0.03	0	0.1	0.11	0	0	0.2	0.04
10020	Luo Yonghao's Way of speaking	0	0	0.05	0.11	0	0.08	0.13	0	0	0.11	0.09
10021	Easy reading: Speaking like Ted	0	0	0.07	0.05	0	0.06	0.1	0	0.01	0.2	0.07
10022	K-Line theory-a detailed explanation of stock	0	0.02	0	0	0	0	0.08	0.18	0.01	0.22	0.02

Figure 2(a) Feature Vector of 22 Courses

No.	Name	Life	Fitness	Quick	Economics	Speaking	Wisdom	Stock	Reading	Success	Skill
10001	The Skills of MS Word that you don't know	0.02	0	0	0	0	0	0	0.02	0.07	0.18
10002	Excel Class for new workplace	0.01	0	0.03	0.01	0	0	0.02	0	0.09	0.17
10003	Premiere Basic Course: One day learn video	0.03	0	0.04	0	0	0	0	0	0.12	0.2
10004	Learn spoken English with foreign teachers	0.01	0	0	0	0.02	0	0	0	0.07	0.18
10005	PPT class for new career	0	0	0.07	0	0	0	0	0	0.11	0.15
10006	Learn to draw in 10 minutes	0.05	0	0.04	0	0	0	0	0	0.07	0.25
10007	Python combat: A week to crawl Web pages	0.01	0	0.03	0	0	0	0	0	0.09	0.18
10008	The Great God of PS	0.04	0	0.01	0	0	0	0	0	0.13	0.2
10009	10 Basic courses for singing beginners	0.03	0	0.04	0	0	0	0	0	0.12	0.2
10010	8 Week Fitness Class	0.11	0.17	0	0	0	0	0	0	0.11	0.04
10011	Mind Mapping Time Management	0.01	0	0	0	0	0.03	0	0	0.13	0.15
10012	Communication Phychology	0.01	0	0	0	0.05	0.04	0	0	0.07	0.15
10013	Microeconomics	0.09	0	0	0.13	0	0	0	0	0.04	0.11
10014	Digital marketing: Entering the brand of wisdom	0.04	0	0	0.12	0	0	0	0	0.06	0.09
10015	C Language Programming Advanced	0	0	0	0	0	0	0	0	0.01	0.21
10016	American Culture	0.13	0	0	0.1	0	0.1	0	0	0.07	0
10017	Logistic expression	0	0	0	0	0.09	0.05	0	0	0.03	0.18
10018	Appreciation and application of cosmetics	0.15	0	0	0	0	0	0	0	0.05	0.23
10019	The operation system of computer	0.04	0	0	0	0	0.03	0	0	0.07	0.19
10020	Luo Yonghao's Way of speaking	0	0	0	0	0.11	0.05	0	0.05	0.09	0.13
10021	Easy reading: Speaking like Ted	0.01	0	0	0	0.15	0.02	0	0.05	0.09	0.12
10022	K-Line theory-a detailed explanation of stock	0	0	0	0.11	0	0	0.13	0	0.08	0.15

Figure 2(b) Feature Vector of 22 Courses

Assume X is an old user. There have been 6 course selection records on this platform, and its path set is $A_x = \{10003, 10012, 10018, 10021, 10011\}$. Extracting and summarizing the weight of each feature word of all courses selected by X, and based on that to calculate the average weight of each word, then form the course selection preference vector *i* of X. Just as the highlighted rows in Figure 3.

Former Course Selection of X															
No.	Name	Office	Management	Phychology	Relationship	Music	Language	Improvement	Market	Culture	Application	Thinking
10003	Premiere Basic Course: One day learn video	0.13	0.03	0	0.06	0.04	0.01	0.13	0	0	0.21	0
10012	Communication Phychology	0.11	0.05	0.13	0.07	0	0.01	0.07	0	0.01	0.13	0.1
10018	Appreciation and application of cosmetics	0	0	0	0.1	0	0	0.17	0	0	0.3	0
10021	Easy reading: Speaking like Ted	0	0	0.07	0.05	0	0.06	0.1	0	0.01	0.2	0.07
10010	8 Week Fitness Class	0.1	0.13	0.03	0.05	0	0	0.11	0	0	0.15	0
10011	Mind Mapping Time Management	0.16	0.13	0.09	0	0	0	0.08	0	0	0.2	0.02
	Average Weight of User's preference	0.08	0.06	0.05	0.06	0.01	0.01	0.11	0.00	0.00	0.20	0.03
	Square of the Average Weight	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.04	0.00

No.	Name	Life	Fitness	Quick	Economics	Speaking	Wisdom	Stock	Reading	Success	Skill
10003	Premiere Basic Course: One day learn video	0.03	0	0.04	0	0	0	0	0	0.12	0.2
10012	Communication Phychology	0.01	0	0	0	0.05	0.04	0	0	0.07	0.15
10018	Appreciation and application of cosmetics	0.15	0	0	0	0	0	0	0	0.05	0.23
10021	Easy reading: Speaking like Ted	0.01	0	0	0	0.15	0.02	0	0.05	0.09	0.12
10010	8 Week Fitness Class	0.11	0.17	0	0	0	0	0	0	0.11	0.04
10011	Mind Mapping Time Management	0.01	0	0	0	0	0.03	0	0	0.13	0.15
	Average Weight of User's preference	0.05	0.03	0.01	0.00	0.03	0.02	0.00	0.01	0.10	0.15
	Square of the Average Weight	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02

Figure 3 Preference Vector of the Old User X (Highlighted)

After obtaining the feature vectors of each course in the platform and the preference vector of the X, use cosine similarity method to calculate the similarity degree by matching the individual courses and users' preferences.

Name	Similarity	$\sum_{k=1}^{u} w_{kj}^2$	$\sum_{k=1}^{u} w_k^2$	$\sum_{k=1}^{u} w_{kj} w_k$
The Skills of MS Word that you don't know	0.922	0.143	0.105	0.113
Excel Class for new workplace	0.902	0.132	0.105	0.106
Premiere Basic Course: One day learn video	0.947	0.141	0.105	0.115
Learn spoken English with foreign teachers	0.870	0.117	0.105	0.097
PPT class for new career	0.913	0.140	0.105	0.110
Learn to draw in 10 minutes	0.928	0.154	0.105	0.118
Python combat: A week to crawl Web pages	0.925	0.140	0.105	0.112
The Great God of PS	0.949	0.137	0.105	0.114
10 Basic courses for singing beginners	0.658	0.158	0.105	0.085
8 Week Fitness Class	0.795	0.120	0.105	0.089
Mind Mapping Time Management	0.917	0.138	0.105	0.110
Communication Phychology	0.892	0.100	0.105	0.091
Microeconomics	0.707	0.104	0.105	0.074
Digital marketing: Entering the brand of wisdom	0.689	0.105	0.105	0.072
C Language Programming Advanced	0.820	0.172	0.105	0.110
American Culture	0.301	0.093	0.105	0.030
Logistic expression	0.893	0.110	0.105	0.096
Appreciation and application of cosmetics	0.887	0.207	0.105	0.131
The operation system of computer	0.927	0.126	0.105	0.107
Luo Yonghao's Way of speaking	0.794	0.100	0.105	0.081
Easy reading: Speaking like Ted	0.849	0.114	0.105	0.093
K-Line theory-a detailed explanation of stock specifications	0.680	0.146	0.105	0.084

Figure 4 Course Similarity Matching

After getting similarity data, rank the courses according to the value of similarity, and then select the top 5 courses to recommend to X. If there is a course that has been already selected by X in the top 5 courses, it will skip this course and continue to recommend according to the rank.

No.	Name	Similarity Ranking
10001	The Skills of MS Word that you don't know	6
10002	Excel Class for new workplace	9
10003	Premiere Basic Course: One day learn video	2
10004	Learn spoken English with foreign teachers	13
10005	PPT class for new career	8
10006	Learn to draw in 10 minutes	3
10007	Python combat: A week to crawl Web pages	5
10008	The Great God of PS	1
10009	10 Basic courses for singing beginners	21
10010	8 Week Fitness Class	16
10011	Mind Mapping Time Management	7
10012	Communication Phychology	11
10013	Microeconomics	18
10014	Digital marketing: Entering the brand of wisdom	19
10015	C Language Programming Advanced	15
10016	American Culture	22
10017	Logistic expression	10
10018	Appreciation and application of cosmetics	12
10019	The operation system of computer	4
10020	Luo Yonghao's Way of speaking	17
10021	Easy reading: Speaking like Ted	14
10022	K-Line theory-a detailed explanation of stock specifications	20

Figure 5 Course Recommendation to Old User X (Highlighted)

In this practice, 'The great God of PS', 'Premiere Basic Course: one day learn video', 'Learn to draw in 10 minutes', 'The operation system of computer', 'Python combat: A week to crawl Web pages'. Since the data used in this practice is real data, we seek feedback from the user after the completion of the recommendation, and the user believes that the recommended result is accord to his personal preferences to a high extent.

For a new user Y, the system cannot get the weight of the feature word in his course preference vector because there is no information about his course selection in the system, so according to the registration information, this paper uses 1 to represent that Y has a preference to the course and 0 to represent that he does not have.

Preference Information of New User Y（0: do not like it，1: like it）										
Office	Management	Phychology	Relationship	Music	Language	Improvement	Market	Culture	Application	Thinking
1	1	1	1	0	1	1	0	0	1	1

Life	Fitness	Quick	Economics	Speaking	Wisdom	Stock	Reading	Success	Skill
0	1	0	0	0	1	0	1	1	1

Figure 6 Preference Information of a New User

Based on the data, this paper uses the same method to get the similarity between the existing courses and the preferences of the new users.

Name	Similaity	Similarity Ranking
The Skills of MS Word that you don't know	0.710	9
Excel Class for new workplace	0.709	10
Premiere Basic Course: One day learn video	0.657	15
Learn spoken English with foreign teachers	0.753	4
PPT class for new career	0.690	12
Learn to draw in 10 minutes	0.642	16
Python combat: A week to crawl Web pages	0.703	11
The Great God of PS	0.711	8
10 Basic courses for singing beginners	0.474	20
8 Week Fitness Class	0.714	7
Mind Mapping Time Management	0.740	6
Communication Phychology	0.816	1
Microeconomics	0.517	19
Digital marketing: Entering the brand of wisdom	0.574	17
C Language Programming Advanced	0.668	14
American Culture	0.436	21
Logistic expression	0.760	3
Appreciation and application of cosmetics	0.518	18
The operation system of computer	0.749	5
Luo Yonghao's Way of speaking	0.780	2
Easy reading: Speaking like Ted	0.682	13
K-Line theory-a detailed explanation of stock specifications	0.414	22

Figure 7 Course Recommendation to New User Y (Highlighted)

Therefore, 'communication psychology', 'Luo Yonghao's way of speaking', 'logistic expression', 'learn spoken English with native-speaker teachers', and 'operation system of computer' will be recommended to Y. In this practice, three of Y's options coincide with the recommended courses, indicating that the algorithm is also available.

5. Conclusion and Further Thoughts

From a practical point of view, this method is feasible to a certain extent.

Firstly, the content-based curriculum recommendation method uses weight index to represent the characteristics of the curriculum and the preferences of the users, which avoids the phenomenon that the curriculum features are not obvious and the users select such courses just by accident rather than based on their real interests. Secondly, when facing new users without path data, this method also can process the similarity matching according to the user's self-evaluation of their preferences to recommend curriculum. Thus, there is no 'cold start' problem. Furthermore, the method recommends curriculum based on the degree of the match between the content of the course and the user's preferences, so long as the users provide their own preference information, the course recommendation can be obtained, which means that the problem of 'data sparsity' does not exist.

However, even content-based recommendation can avoid the problem of 'cold start' and 'data sparsity', this algorithm method itself has some inevitable defects. First of all, TF-IDF method is used as a method to characterize text, which is the base of weighting the characteristics and processing the content-based recommendation method. Although some texts are used to represent the course under TF-IDF method, they still cannot represent the whole courses. So, in the feature extraction process of curriculum, content-based recommendation method cannot overcome the difficulty of extracting features to form the absolutely precise character vectors. Secondly, content-based recommendation matches the project features mainly based on the user's own setting or the path trajectory left on the platform during his use, which actually does not dig out the user's potential interests and cannot provide a more comprehensive recommendation service for the customer in the future. If other algorithms are not combined with content-based recommendation method when put into practice, it may bring risks to the daily operations of those platforms such as NetEase's Cloud Classroom.

References:

1. Han J, Pei, J, Yin Y. *Mining frequent patterns without candidate gener*ation. Data Mining and Knowledge Discovery. 2004; 8: 53-87

2. Jianguo L, Tao Z, Binghong W. *A recommendation system model based on content and collaborative filtering integration*. Natural Science Progress. 2009;1
3. Moon S, Russell G. J. *Predicting product purchase from inferred customer similarity: An autologistic model approach*. Management Science. 2008; 54(1): 71-82
4. Peng X. *A large data perspective analysis of learning changes--an interpretation and enlightenment of the report on education and learning through educational data mining and learning analysis*. Distance Education. 2013; 6: 11-17
5. Yukselturk E, Ozekes S, Türel Y K. *Predicting dropout student: An application of data mining methods in an online education program*. European Journal of Open, Distance and e-Learning. 2014; 17(1): 118-133
6. Yunyun C. *The operation and characteristics of online courses in foreign high-quality Universities--Based on the MOOCs curriculum of edX platform*. Modern Educational technology. 2014; 9
7. Zhongjun L, Qihai Z, Qinghong S. *Research of personalized recommendation system*. Computer Science. 2009; 12