Research of Recommendation Algorithm Based on Q&A Forum of Online Learning

Junmin Ye¹, Qiang Wang¹, Zhifeng Wang², and Rong Zhao² ¹School of Computing, Central China Normal University, Wuhan, China ²School of Educational Information Technology, Central China Normal University, Wuhan, China Email: zfwang@mail.ccnu.edu.cn

Abstract—The forums have become the most general ways for people to study and communicate nowadays, and both the level of users' interactions and the efficiency of solving problems will influence the effect of users' learning on the forums. In this paper, in order to solve the problems that how to find users who have similar tastes and interests rapidly and how to get a quick answer when users have agent questions, a framework has been proposed to apply the recommendation technology to the forums. For the empirical research, the data from an authentic forum is used, and finally we find that this recommendation technology can do improve the forum's service efficiency and it is not only useful for reality uses of forum system but also provide research references on how to improve the use efficiency of form

Index Terms—forum question and answer analysis, learning analysis, recommendation technology, data mining

I. INTRODUCTION

Information technology develops rapidly in the contemporary, the scale of network information resources is growing exponentially, but the huge amount of information is far more than the effective processing ability of users, which seriously affects the user's choice and decision. This problem is called Information Overload [1]. To solve the Information Overload problem, information technology experts have studied a variety of information processing technology, including information filtering technology, information retrieval technology, information mining technology, information recommendation technology. Information filtering technology is a kind of information processing technology which can reduce the size of client information by filtering and removing unrelated information resources in the network. It is mainly used to maintain the network order [2]. While information retrieval technology is to provide personalized search results for users according to their needs to reduce the information size of the clients, it is mainly used in search engines [3]. Data mining technology explores the potential law of the data through data mining to improve the efficiency of decision making, and it is mainly used in the field of big data analysis [4]. Information recommendation technology push content satisfied

Manuscript received August 8, 2017; revised February 2, 2018.

personality and characteristic needs based on the user's personalized analysis results, it is mainly used in various types of e-commerce sites [5] and is the most widely used software applications currently.

The Internet in the process of promoting education informatization has also brought the problem of information overload to education area. There are accumulated a large number of educational data in many MOOC platform, forums and even closed campus learning platform, but learners and managers feel helpless on how to using these data. In the forum system, for example, the forum provides a public platform for users to discuss, communicate and seek for help. However, as the operating time of forums grows, the number of users in the forum will increase, and the daily growth of data of the forum will increase day by day. That is the forum data grows exponential, with the content classification or retrieval technology can not follow its development, it will lead to users cannot quickly from the massive data resources to obtain the necessary resources and cannot quickly find like-minded users from the massive users to communicate, and then the learning process of users is impacted. At present, the application of recommended technology is mainly concentrated in the field of business, Taobao, Amazon, Jingdong and other large e-commerce platform with Youtube, Tencent, Youku and other large video playback sites, they are using different recommended techniques to solve the information overload problem to improve the experience sense and using satisfaction of users. So how to learn the recommended technology application experience from the business areas to enhance the user experience and satisfaction of forum has become a very meaningful research topic.

In the field of forum research, the foreign researchers put forward a number of modeling and algorithm ideas. Fu *et al.* [6] proposed a problem recommendation algorithm (BMFC-ARM) based on the construction of "biological mechanism driving feature". This algorithm provides the proper replication to relevant questions in forum through the eigenvector of the response sequence based on the information of the questioner and the respondent. Seifzadeh *et al.* [7] proposed a new calculation method of user similarity, which constructs the user contact graphs by mining the relationship between questions and answers and then divides the user sets into multiple levels through the minimum spanning tree algorithm. It also applies this algorithm to the recommendation system based on collaborative filtering and the accuracy of the algorithm is studied by using the real data in large BBS system. Peng *et al.* [8] proposed a personalized recommendation model based on user search behavior and resource attributes, which mines user's personality traits from user search behavior and recommends users the resources in line with their interests using collaborative filtering algorithm based on Top-N. These are the more advanced forum Q&A recommended technology, their research base on different point of entry: Q&A content, the relationship of user's Q&A and user search behavior with resource attributes.

Currently, the domestic research work mainly focus on the combination of theoretical research and practical research. Hudson et al. [9] study the new post classification recommend technology of forums deeply, it analyzes the new post by the historical data in forum and recommend it to the proper classification, they also study how to continued optimizing dynamic recommendation performance of the forum during its update process. Iyengar et al. [10] apply collaborative filtering recommended technology to the post recommendation of forum, they get predict score of a post by calculate weighted score of adjacent users' scores on the post of, according to the user's interest quickly recommended the post for them, thereby reducing the cost of user's search posts. Aciar et al. [11] conducts a comprehensive analysis with the user's degree of activity, the user's degree of interest and professional authority, they recommend for users to interested and have the ability to answer questions combined with their characteristics, aimed at improving user participation at the same time improving question solving efficiency. Wang et al. [12] study the personalized question recommendation on the network Q&A system adopting recommended technology based on the relevance of the rules deeply, they find the associated question collection with users according to the characteristics of the frequently questions user asked, present a suitable questions set to be resolved as user login, achieve the online Q&A function of virtual "one on one" in the system. In general, the domestic research is more in line with the practical application of the cognitive [13]-[19], such as through the score to arrange posts [20]-[24] by posting the external information as a judge the quality of the indicators and so on.

II. RESEARCH METHODOLOGY

A. Bayesian Text Classification

Bayesian classification algorithm [25] is based on the Bayesian theorem, the basic principle of this algorithm is to correct the prior probability to get the posterior probability based on the probability of probability, and then do the pattern classification according to the size of the posterior probability, the classification results is the category the maximum value of the posterior probability corresponds to. The priori probability refers to the probability obtained from previous experience or data analysis. The conditional probability is the probability that another event occurs when an event occurs. The posteriori probability refers to a correction probability of priori probability that closer to the probabilistic estimate result based on new information or new data.

B. Recommendation Algorithm

Recommendation technology is an information processing technology that to solve the information overload problem mainly through narrowing the user's choice and slowing down their pressure of choice based on user needs. The main principle is that the system will get a recommended result based on the user's personalized features and specific needs through the recommendation algorithm, ant the system could accelerate the choice or decision process of users when they cannot choose or make a quick decision facing a large number of information.

As shown in Fig. 1, the content-based recommendation algorithm is the earliest recommendation algorithm and is mainly used in the information retrieval system in the early time. Its basic principle is digging the association between the user's operational data and the characteristics of the item based on the past operational data information of users, then learning the user's personality characteristics through the decision tree, neural network, vector representation method and other methods, filtering the candidate collection according to characteristics of users and candidate, ultimately getting the personalized recommendation results meeting the personality characteristics and specific needs of the users. In the case of a merchandise recommendation in a merchandise sales website, the content-based recommendation algorithm recommends a product that matches the user's preference mainly based on their operations of previous browsing of the merchandise, the addition of a shopping cart, the purchase, the non-browsing, and the like. The different operations of the goods represent the preference degree of users on the product, so that we could present the user's preference from their behavior of different goods and the characteristics of product and then recommend the likely love product to users.



Figure 1. Content based recommendation algorithm.

The recommendation algorithm based on collaborative filtering is the most widely used recommendation algorithm nowadays. The basic principle is to realize the recommended function by calculating user's prediction score based on similarity (or item similarity) of users and feedback evaluation of users. In the case of a merchandise recommendation in a merchandise sales website, the recommendation algorithm based on collaborative filtering considers that the users' interests are similar whose personality is similar, that is, they like the similar goods. According to this rule, we can calculate the similarity of users and then recommend to the target user the favor product of the user whose personality is similar with it. Similarly, collaborative filtering based on the recommended algorithm considers that the user's personality characteristics are stable, so the goods users liking have common features. According to this law, we can calculate the similarity of products and then recommend to the target users the products with the same characteristics of the products they like. According to the different calculation object, the recommendation algorithm based on collaborative filtering is mainly divided into user-based collaborative filtering algorithm and item-based collaborative filtering algorithm.

III. PROPOSED RECOMMENDATION MODEL

The varieties and amount of the data in forum is big, in order to narrow the scope of the candidate in recommendation process improve and to the recommendation efficiency, as shown in Fig. 2, we design a system in this paper do pretreatment classification of the new post question adopting the Bayesian classification algorithm before recommendation questions and answers. Then we select one category set of question to analyze, and get the recommendation result adopting the hybrid algorithm based on the content recommendation algorithm and the collaborative algorithm based on collaborative filtering. At last, the system will push the recommendation result to the user for their choice.



Figure 2. Forum recommended technical schematic.

The algorithm implementation process has three steps: 1) Question classification. Set the data in the system as a training set and extract the characteristics of each category, then make a comparative analysis of between the new question and the other category data, the new post belongs to the category which get the maximum match degree. 2) Analysis recommended. The interaction between users and questions can indirectly reflect the user's personality characteristics and the interaction between questions can reflect the relationship of questions. So in this paper, we get the appropriate user to communicate and the user to answer the new question according to the user-based recommendation algorithm, and get the similar question which has the similar semantics with new question according to the projectbased collaborative filtering recommendation algorithm. 3) Recommend the results. Users could operate the recommendation result based on their own needs after getting the results, they can scan the historical answer of the recommend question and they can also post quick question to the recommended respondents to achieve the one to one communication.

A. User Recommendation Model

1) Basic assumption

Assuming that U= { \cdots ;u_i, \cdots } represents all forum users, u_i is any user in collection U, the subscript i is the user ID. Q= { \cdots ;q_j, \cdots } represents all questions in forum, q_j is any question in collection Q, the subscript j is the question ID. K= { \cdots ; k_g, \cdots } represents the type of question in the forum, k_g is any question type in collection K, the subscript g is the type ID. A= { \cdots ; a_{i,j,g}, \cdots } represents the type of all the responses in the forum, a_{i,j,g} is any response in collection A, the subscript i is the user ID, the subscript j is the question ID ,the subscript g is the question type ID. That is, the record a_{i,j,g} means the user u_i answer the question q_j whose type is k_g. Then make t_{i,j,g} express endorsing vote, f_{i,j,g} express against vote, Availability_{i,j,g} express the user's breadth of knowledge for the set, width_i express the knowledge width of user i.

 $\begin{array}{l} A_i = \{ \ \cdots, \ a_{i,j,g}, \cdots \} \ \text{represents all the responses of user } i \\ \text{in the forum, } A_{ij} = \{ \ \cdots, \ a_{i,j,g}, \cdots \} \ \text{represents all the responses for question type } k_g \ \text{of user } i \ \text{in the forum. } n_{ij} \ \text{is the sum of responses in collection } A_{ij}, \ \text{depth}_{ij} \ \text{is the knowledge depth of user } i \ \text{in question type } k_g. \end{array}$

Simi_{ii'} represents the similarity between user i and user i', it is measured by the common coverage of the knowledge breadth of the two. Last, set width_i represents the knowledge width of user i, width_i.size indicating its size; set width_{i'} represents the knowledge width of user i', width_{i'}.size indicating its size; set comm(width_i, width_{i'}) represents the intersection set of knowledge width of user i and user i', comm(width_i, width_{i'}).size indicating its size. *Fundamental Model*

Calculate the Availabilityi,j,g of each ai,j,g.

When $t_{i,j,g}=0$, Availability_{i,j,g} = 0

When
$$t_{i,j,g} \neq 0$$
, Availability_{i,j,g} = $\frac{t_{i,j,g} - f_{i,j,g}}{t_{i,j,g}}$ (1)

Calculate the knowledge widthi of each user i.

$$width_i = \{\cdots, k_g, \cdots\} (a_{i,j,g} \in A_i)$$

$$(2)$$

Calculate depthij of each user i.

$$depth_{i,g} = \frac{\sum_{a_{i,j,g} \in A_{i,g}} Availability_{i,j,g}}{n_{i,g}}$$
(3)

Calculate the Simili' of user i and user i'.

$$simi_{i,i'} = \frac{comm(width_i, width_i).size}{width_i.size + width_i.size - comm(width_i, width_i).size}$$
(4)

Set the recommended threshold of selection results to get the recommended results.

Assume that t is the selection threshold of the similarity degree of the users, s is the selection threshold of the knowledge depth of the user. Include the users whose user similarity is greater than the selection threshold into communication user set, and whose knowledge depth greater than its selection threshold into ask-for-help user set, as shown in Fig. 3.



Figure 3. User recommendation algorithm.

B. Problem Recommendation Model

1) Basic assumptions

a) Cosine similarity calculation related hypothesis.

Assume that $P = \{t_1, t_2, \dots, t_m\}$ is the words collection of candidate question participle, *m* is the sum of words in the collection; $Q = \{s_1, s_2, \dots, s_n\}$ is the words collection of new question participle, *n* is the sum of words in the collection; $A = \{a_1, a_2, \dots, a_k\}$ is the union collection of *P* and *Q*, *k* is the sum of words in collection A.

Assume that $C1_i$ is the semantic score of word element a_i in the candidate question, $C2_i$ is the semantic score of

word element a_i in the new question, The semantic fraction vector corresponding to the candidate problem and the semantic score vector corresponding to the new problem are $\vec{P}_i(C1_1, \dots, C1_k)$ and \vec{Q} $(C2_1, \dots, C2_k)$.

b) Common sub-string similarity calculation related hypothesis.

For the two given sequences $X = (x_1, x_2, \dots, x_m)$ and $Z = (z_1, z_2, \dots, z_k)$, when and only when there exist an incremental subsequence (i_1, i_2, \dots, i_k) makes $z_j = x_{i_j} (1 \le i_j \le m)$ for all $j = 1, 2, \dots k$, that means Z is the sub-sequence of X. If the sequence Z is the sub-sequence of both X and Y, we call Z the common sub-sequence of X and Y; the longest common sub-sequence is the longest common sub-sequence of the two sequences in the common sub-sequence.

Set $P = \{t_1, t_2, \dots, t_m\}$ is the preprocessed candidate problem string, m is the length of string P; $Q = \{s_1, s_2, \dots, s_n\}$ is the preprocessed new problem string, n is the length of string Q. Set L(m, n) represents the length of the longest common subsequence of $P = \{t_1, t_2, \dots, t_m\}$ and $Q = \{s_1, s_2, \dots, s_n\}$. The twodimensional table S(m, n) shows the state change during the solution, S(i, j) indicating the search state at the time of L(i, j) calculation.

The initial state is set as follows:

$$L(0,0) = L(0,j) = L(i,0) = 0 \qquad (1 \le i \le m, 1 \le j \le n)$$

$$S(0,0) = S(0,j) = S(i,0) = 0 \qquad (1 \le i \le m, 1 \le j \le n)$$
(5)

Dynamic programming functions and state changes are as follows.

$$L(i, j) = -\begin{bmatrix} L(i-1, j-1) + 1 & t_i = s_j, i \ge 1, j \ge 1 \\ \max\{L(i, j-1), L(i-1, j)\} & t_i \ne s_j, i \ge 1, j \ge 1 \end{bmatrix}$$
(6)
$$S(i, j) = -\begin{bmatrix} 1 & t_i = s_j \\ 2 & t_i \ne s_j, L(i, j-1) \ge L(i-1, j) \\ 3 & t_i \ne s_j, L(i, j-1) < L(i-1, j) \end{bmatrix}$$
(7)

2) Fundamental model

a) Set the cosine similarity and the common substring similarity selection threshold.

b) Calculate the cosine similarity of each candidate question and new question according to the semantic score of each candidate question and new question.

similarity₁ = cos(
$$\vec{P}, \vec{Q}$$
) = $\frac{\sum_{i=1}^{k} Cl_i C2_i}{\sum_{i=1}^{k} Cl_i^2 \sum_{i=1}^{k} C2_i^2}$ (8)

c) Calculate the common sub-string similarity of each candidate question and the new question according to the common sub-string of the candidate question and the new question.

$$similarity_2 = \frac{L(m,n)}{\max\{m,n\}}$$
(9)

d) Select candidate question that satisfies the cosine similarity selection threshold and the common substring similarity selection threshold into the recommended set.

The designed problem recommended algorithm is shown in Fig. 4.



Figure 4. Problem recommended algorithm.

IV. RECOMMENDATION SYSTEM DESIGN FOR ONLINE LEARNING

As shown in Fig. 5, the BBS Forum based on the recommended technology designed and realized in this paper adopts the standard B/S architecture which is a three layers system structure of the browser, Web server and database server. The browser is the presentation layer, it is used to display the user interface of system, accept the input data or instructions of users and display the output data or instructions of system; Web server is the application logic layer, it is used to response to the request issued by browser, execute the Servlet controller program according to the call request of application server and return the results to the browser to display to users; database is the data layer, storage a variety of data used in in system.

As shown in Fig. 6, the system adopts the MVC-based design pattern to alleviate the data pressure of the web server in the application logic layer (web server layer) through the JSP + Servlet + JavaBean. The JSP layer is used to design and generate the HTML user interface for the browser display, it mainly complete the functions of

user operation, data input and data output, corresponding to the view in MVC framework. Servlet layer used to accept the submission request of user in the front page, it completes the different business request processing functions by calling different JavaBean and forward the results to the relatively view, corresponding to the controller in MVC framework. The JavaBean is used to complete the background database operation, which corresponds to the model in the MVC framework and which can be divided into the entity layer and the business layer. The entity layer mainly performs the modeling function of each entity in the database. The business layer has the direct access to the database and mainly completes the connection function of the database and the operation function of adding, deleting, changing and checking the records in the database.



Figure 5. B/S architecture of Recommendation System.

The BBS forum based on recommended technology in this paper mainly includes six modules: registration module, data collection module, data display module, analysis recommendation module, user interaction module and administrator calibration module.



Figure 6. MVC architecture flow chart.

As shown in Fig. 7, the registration module is mainly to provide an interface to the system for ordinary users and administrators, including user registration, user login and administrator login three functions; data collection module is mainly used to collect the record data of user's question and answer to provide the data base for the recommended function, mainly including new questions, questions replication, answer evaluation of three functions; data display module mainly to provide browsing functions for users, including five functions: all questions list browsing, questions details browsing, questions replication browsing, my questions list browsing; analysis recommended module is the core of the recommended system, its mainly work is recommending the appropriate question replication and similar question set to users according to their needs,

including user recommendation and question recommendation the two main functions; the user interaction module mainly provides a quick interaction between the users, including two functions of the quick question and message processing. The administrator correction module mainly realizes the function of question type correction.



Figure 7. System module structure diagram

V. EXPERIMENTAL RESULTS AND ANALYSIS

This paper collects the real data in the forum of ZHIHUI to do the empirical research. The experimental data mainly includes the candidate user set, the target user, the new question and the candidate question set. In this case, we select 10 users to constitute the candidate user set from user set and 1 user as the target user by using random sampling method. Then we set the new question is "How to choose and prepare for the postgraduate entrance examination?". According to the classification result of question from classification experiment, we randomly select 10 similar candidate question data from the data set to compose candidate question set. Last, we realize recommending appropriate user sets and problem sets to the user by analyzing the experimental data using the recommended techniques described in the previous section.

A. Problem Text Classification Experiment

According to the experimental data used in this paper, the content of the target question is "How to choose and prepare for the post-graduate entrance examination?". Classify the target problem by Bayesian Chinese text categorization algorithm. The classification results are shown in Table I.

Category	Prior Probability	Posterior Probability
Traffic	0.7717	0.0046
Sports	0.0675	0.0083
Military	0.1222	0.0055
Medicine	0.1286	0.0049
Politics	0.1029	0.0073
Education	0.1286	0.1071
Environment	0.1158	0.0008
Economy	0.1061	0.0055
Computer	0.1511	0.0028

According to the experimental results, the posteriori probability of the education class is the largest in the calculation result of the posterior probability of each classification. Therefore, the Bayesian text classification algorithm is used to classify the text as an educational category.

B. User set Recommended Experiment

According to the experimental data used in this paper, the candidate user set includes 10 users selected randomly. Then select the question data and the response data related to each user from the data set, calculate the knowledge breadth, the knowledge depth and the similarity of each user. Set the calculation result of the knowledge width of target user is {traffic, politics, economy}.

ID	Politics	Education	Environment	Computer	Traffic
13	0.5900	0.3940	0.6200	0.8790	0
2	0.4020	0.1920	0.9560	0.3000	0.7830
26	0.4440	0.3900	0.8190	0.2680	0
34	0.1700	0	0	0.1500	0.3000
41	0	0	0	0.7800	0.4880
54	0	0	0.888	0.3820	0
58	0	0	0.831	0.5200	0
59	0.3860	0	0.364	0.4860	0
8	0	0	0	0.7180	0.7860

TABLE II. USER KNOWLEDGE DEPTH IN EACH CATEGORY

As shown in Table II, the results indicate that the user 34 and the user 41 have the highest degree of interest matching with the target user according to the user similarity. According to the user's knowledge depth in each classification can indicate user's the ability and recognition of solving such question, if the question to be resolved belongs to political category, it will recommend the user 13 to answer the question; if the question to be resolved belongs to education category, it will recommend the user 13 to answer the question; if the question to be resolved belongs to environment category, it will recommend the user 13, the user 2, the user 26, the user 54, the user 58 to answer the question; if the question to be solved belongs to the computer category, it will recommend the user 13, the user 41, the user 58 The user 8 to answer the question; if the question to be resolved belongs to the economy category, it will recommend the user 2, the user 8, the user 9 to answer the question; if the question to be resolved belongs to the traffic category, it will recommend the user 34, the user 41, the user 58 to answer the question.

C. Problem set Recommended Experiment

According to the experimental data used in this paper, the content of the target question is "How to choose and prepare for the post-graduate entrance examination?". The candidate question set includes 10 educational questions selected randomly. Then calculate similarity between the candidate problem and the target problem using the cosine similarity method and the common substring similarity method respectively. And to the judge whether the candidate question is added to the recommendation set according the result of two similarity calculation method.

ID	Candidate	Cos-sim	Sub- sim	Result
1	Q1	0.5431	0.0952	Not recommended
2	Q2	0.6863	0.3636	Recommended
3	Q3	0.4089	0	Not recommended
4	Q4	0.4089	0	Not recommended
5	Q5	0.6202	0.2222	Recommended
6	Q6	0.5605	0.1666	Not recommended
7	Q7	0.3846	0.0833	Not recommended
8	Q8	0.7169	0.4444	Recommended
9	Q9	0.4042	0	Not recommended
10	Q10	0.3859	0.1111	Not recommended

TABLE III. SIMILARITY CALCULATION RESULTS

As shown in Table III, in view of the similarity algorithm based on cosine theorem taking the semantic processing of Chinese text into account, but the common string similarity algorithm completely adopts the matching method of the string pattern. Therefore, when the question recommendation set is obtained, we get an initial question recommendation set according to the cosine similarity: set the cosine similarity selection threshold is 0.5, that is, we include the question in the preliminary question recommendation set whose cosine similarity result is higher than the selected threshold problem; then the preliminary question recommendation set is filtered to the final problem recommendation set according to the common substring similarity degree: set the selection threshold of the common substring similarity is 0.2, that is, we include the question in the final question recommendation set whose common substring similarity result is higher than the selected threshold problem. The similarity calculation of the target problem and each question of the candidate question are shown in the above table.

According to the experimental results, firstly we analyze the candidate questions according to the cosine similarity. In the candidate problem, the similarity degree of problem 1, question 2, question 5, question 6 and question 8 is more than the selection threshold of 0.5, so these questions are included in the preliminary question recommendation set. Secondly, we analyze the question the preliminary question recommendation in set according to the common substring similarity, the similarity degree of question 2, question 5 and question 8 is higher than the selection threshold of 0.2, so these questions are included in the final question recommended set. The user can solve the target question by looking at the answer details of question 2, question 5, or question 8.

VI. CONCLUSION

In order to solve the problem of information overload in the forum, this paper puts forward the forum recommended technology based on the recommended technology. This technology realizes recommend user set to users which has similar interest, and we could recommend the proper question answering set and question set to users for their question. On this basis, we also put forward the user recommendation algorithm and the problem recommendation algorithm in the forum according to the characteristics of the forum data. Then we do the experimental research using the real forum data to process the whole recommended flow, and design a corresponding forum recommended system. This is not only effectively improve the efficiency of the forum, but also strengthen the exchange between the forum users, effectively solving the lack of links between users in the forum and the problem of inefficient questions.

Nevertheless, we need to do some improvement and innovation work for this forum recommendation technology; there also exist some salient problems in the specific application as fowling: 1) The data in forums is diversity, we need to further study how to process multimedia data in the forum like video and audio; 2) We know that different classification algorithms have different effects on different types of data classification. Similarly, different recommendation algorithms have different effects on different types of data classification. So one of the important follow-up research of this paper is comparing and analyzing the forum text classification of different classification algorithms and the forum recommendation function performance of the different recommendation algorithms.

ACKNOWLEDGMENT

This work was supported by the Projects of the Social Science and Humanity Fund of the Ministry of Education (No. 15YJA880095), the National Natural Science Foundation of China (No. 61501199), the National Social Science Fund project (No.17BTQ061), the selfdetermined research funds of CCNU from the colleges'basic research and operation of MOE (No. CCNU17GF0003), the Hubei Provincial Natural Science Foundation (No. 2017CFB683), the Specific Funding for Education Science Research by Self-determined Research Funds of CCNU from the Colleges' Basic Research and Operation of MOE (No. CCNU17KYZHSY21), and the National Science and Technology Ministry (No. 2015BAK33B02).

REFERENCES

- J. Eppler, Martin, and J. Mengis, "The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines," *IEEE Engineering Management Review*, vol. 38, no. 1, pp. 3-3, 2010.
- [2] F. Luo, G. Ranzi, X. Wang, and Z. Y. Dong, "Social information filtering based electricity retail plan recommender system for smart grid end users," *IEEE Transactions on Smart Grid*, pp. 1-1, 2017.
- [3] H. Sun and S. A. Jafar, "The capacity of private information retrieval," *IEEE Transactions on Information Theory*, vol. 63, pp. 4075-4088, 2017.
- [4] A. Yassine, S. Singh, and A. Alamri, "Mining human activity patterns from smart home big data for health care applications," *IEEE Access*, vol. 5, pp. 13131-13141, 2017.
- [5] Q. Zhao, C. Wang, P. Wang, M. Zhou, and C. Jiang, "A novel method on information recommendation via hybrid similarity," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1-12, 2017.
- [6] H. Fu, Z. Niu, C. Zhang, J. Ma, and J. Chen, "Visual cortex inspired CNN model for feature construction in text analysis," *Frontiers in Computational Neuroscience*, vol. 10, 2016.
- [7] S. Seifzadeh and A. Miri, "Similarity measure based on low-rank approximation for highly scalable recommender systems," *IEEE Trustcom/BigDataSE/ISPA*, pp. 66-71, 2015.
- [8] X. Peng, S. Huang, and Z. Niu, "An study on personalized recommendation model based on search behaviors and resource properties," in *Proc. International Conference on Information Engineering and Computer Science*, 2010, pp. 1-4.
- [9] N. Hudson, P. K. Chilana, X. Guo, J. Day, and E. Liu, "Understanding triggers for clarification requests in communitybased software help forums," in *Proc. IEEE Symposium on Visual Languages and Human-Centric Computing*, 2015, pp. 189-193.
- [10] M. Iyengar, A. Sarkar, and S. Singh, "A Collaborative Filtering based model for recommending graduate schools," in *Proc. International Conference on Modeling, Simulation, and Applied Optimization*, 2017, pp. 1-5.
- [11] S. V. Aciar, G. I. Aciar, C. A. Collazos, and C. S. Gonz alez, "User recommender system based on knowledge, availability, and reputation from interactions in forums," *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, vol. 11, pp. 18-22, 2016.
- [12] L. Wang, B. Wu, J. Yang, and S. Peng, "Personalized recommendation for new questions in community question answering," in *Proc. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2016, pp. 901-908.
- [13] F. Ali, D. Kwak, P. Khan, S. H. A. Ei-Sappagh, S. M. R. Islam, D. Park, *et al.*, "Merged ontology and SVM-Based information extraction and recommendation system for social robots," *IEEE Access*, vol. 5, pp. 12364-12379, 2017.
- [14] M. Borg, K. Wnuk, B. Regnell, and P. Runeson, "Supporting change impact analysis using a recommendation system: An industrial case study in a safety-critical context," *IEEE Transactions on Software Engineering*, vol. 43, pp. 675-700, 2017.
- [15] S. Deng, D. Wang, Y. Li, B. Cao, J. Yin, Z. Wu, et al., "A recommendation system to facilitate business process modeling," *IEEE Transactions on Cybernetics*, vol. 47, pp. 1380-1394, 2017.

- [16] M. Jüttner, J. Falk, and W. M. Rucker, "A neural network based recommendation system for solvers and preconditioners for systems of linear equations," *IEEE Transactions on Magnetics*, vol. 53, pp. 1-4, 2017.
- [17] S. Li, J. Wen, F. Luo, M. Gao, J. Zeng, and Z. Y. Dong, "A new QoS-Aware web service recommendation system based on contextual feature recognition at server-side," *IEEE Transactions* on Network and Service Management, vol. 14, pp. 332-342, 2017.
- [18] Y. Mo, J. Chen, X. Xie, C. Luo, and L. T. Yang, "Cloud-Based mobile multimedia recommendation system with user behavior information," *IEEE Systems Journal*, vol. 8, pp. 184-193, 2014.
- [19] Y. S. Rawat and M. S. Kankanhalli, "ClickSmart: A context-aware viewpoint recommendation system for mobile photography," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, pp. 149-158, 2017.
- [20] Z. Tian, T. Jung, Y. Wang, F. Zhang, L. Tu, C. Xu, et al., "Real-Time charging station recommendation system for electric-vehicle taxis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, pp. 3098-3109, 2016.
- [21] C. Verma, M. Hart, S. Bhatkar, A. Parker-Wood, and S. Dey, "Improving scalability of personalized recommendation systems for enterprise knowledge workers," *IEEE Access*, vol. 4, pp. 204-215, 2016.
- [22] J. D. West, I. Wesley-Smith, and C. T. Bergstrom, "A recommendation system based on hierarchical clustering of an article-level citation network," *IEEE Transactions on Big Data*, vol. 2, pp. 113-123, 2016.
- [23] D. Zhang, T. He, Y. Liu, S. Lin, and J. A. Stankovic, "A carpooling recommendation system for taxicab services," *IEEE Transactions on Emerging Topics in Computing*, vol. 2, pp. 254-266, 2014.
- [24] C. Zou, D. Zhang, J. Wan, M. M. Hassan, and J. Lloret, "Using concept lattice for personalized recommendation system design," *IEEE Systems Journal*, vol. 11, pp. 305-314, 2017.
- [25] K. Hyun-Chul and Z. Ghahramani, "Bayesian gaussian process classification with the EM-EP algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1948-1959, 2006.



Junmin Ye is a full professor in the School of Computing, Central China Normal University, Wuhan, Hubei. His research interests include software structure design, learning analytics, and big data application.



Zhifeng Wang is the corresponding author of this paper. He received the BEng degree in Electronic Engineering from China University of Geosciences in 2008, and the PhD degree in Electronic Engineering from South China University of Technology, China, in 2013. He was a joint training PhD student in the computer science department of Carnegie Mellon University, Pittsburg, PA, during 2010 to 2011. He is now an associate professor in

the School of Educational Information Technology of Central China Normal University. His research interests include learning analytics, data mining, machine learning, and learning theory.