

A Hybrid Recommendation Algorithm Adapted in Integration of Informatization and Industrialization for Industrial Enterprises

Laisong Kang¹, Shifeng Liu² and Daqing Gong³

School of Economics & Management, Beijing Jiaotong University, Beijing, China
¹13120623@bjtu.edu.cn; ²shfliu@bjtu.edu.cn; ³gongtuipigua@163.com

Abstract

As the rapid development of integration of informatization and industrialization (IoII), information overload become a serious problem in the knowledge platform for IoII. To address this issue, this paper introduces a hybrid framework based on the assessment system of IoII and user learning behavior. First, using the assessment specification on IoII for industrial enterprises, we establish the similarity model of IoII; then, the similarity model of user behavior is built based on three kinds of learning behaviors in the knowledge platform; at last, after studying the advantages and disadvantages of the two models, this paper proposes a linear fusion framework combining both models. With several experiment conducted, we get the optimal parameters in the framework, and the experimental results show that the proposed framework can achieve the better recommendation quality.

Keywords: Recommendation algorithm; collaborative filtering; learning behavior; IoII

1. Introduction

Industrialization and informatization are the two most important characteristics of modern society. Studies in theory indicate that the development of information industry can promote technology and management innovation and improve productivity of traditional industries [1-2]. In the trend of industrial convergence, the integration development of informatization and industrialization is chosen as a strategy to narrow the gap of industrialization and informatization between China and developed countries. Therefore, a knowledge platform for IoII was built to server.

As the rapid development of IoII, the exponential growth of the available case in the platform causes the information overload problem, which refers that user cannot quickly and accurately locate the case they need. However, recommendation system solve this problem easily by mining user behavior data and information push-delivery. Goldberg [3] proposed the thought of collaborative filtering for the first time in 1992. Now, collaborative filtering recommendation algorithm has become an important kind of method for recommendation system, which was widely used and studied due to its academic and commercial value.

2. Related Work

As our algorithm is based on the assessment of IoII and user implicit preferences, we will review related works from three different research areas: assessment Index system for IoII, implicit rating and collaborative filtering recommendation algorithm.

When it comes to the assessment of IoII, Liu J regulated “assessment specification on IoII for industrial enterprises” [4] based on previous works of many enterprises, industry associations and research institutions. In this specification, the principle of scientific, effectiveness, operability and scalability was proposed. Zhang J. and Zheng J [5]

proposed that the index system for integration of informationization and industrialization should be consistent in multi-angle, multi-dimensional and multi-method and provided method based on the previous theories of informationization measurement. Li S, Dong S and Zhou J [6] studied the influence of information technology on enterprise forming and competitiveness and have shown that the ability of enhancing enterprise competitiveness from information technology is measured by cumulative investment index. The authors established an enterprise competitiveness system dynamic model based on the presentation layer, factors layer and decision-making layer. As a method of explicit rating, the problem of data-sparsity and cold-start will limit the recommendation quality, thus integrating the implicit rating is proposed to address this issue.

Implicit rating is based on observable behaviors exhibited by a user. Nichols [7] first surveyed a list of useful behaviors. Oard and Kim [8-9] categorized these behaviors into four sets: annotation, examination, retention and reference. Koren [10] has shown that is possible to get better performance by using binary implicit user feedback data and proved this implicit user behavior model increases the prediction accuracy. Kostkova and Madle [11] found that recommender systems can learn from the essential role of qualitative behaviors for understanding user preferences.

Based on above researches, some scholars developed collaborative filtering recommendation algorithm to solve the problem. Hidasi and Tikk [12] proposed to use tensor to do the context-aware recommendation. In general, there are three types of integration method: 1. contextual pre-filtering method; 2. context post-filtering method; 3. contextual modeling method. In contrast to the previous two methods (pre and post-filtering), the contextual modeling method uses all the contextual and user-item information simultaneously to make predictions. Kim and El Saddik [13] reveal that the exploitation of social tagging is beneficial for recommender systems that provide users with suggestions about interesting communities that a user may want to join. Pirolli and Kairam [14] research to what extent it is possible to infer users' expertise regarding topics from their browsing behavior in social tagging systems. Aligon J., Gallinucci and Golfarelli [15] found that the whole sequence of queries belonging to an OLAP session is valuable and proposed a recommendation approach stemming from collaborative filtering by treating sessions as first-class citizens.

The lesson learned from this literature review is that assessment system on IoII is rarely used in recommendation and seldom combined with implicit rating model. Based on the assessment of IoII for industrial enterprises and user learning behavior, the framework proposed in this article overcomes these limitations. We organize this paper as following: In section 3, we define the problem and use the assessment system of IoII to build user similarity model. Then, the user learning behavior information is integrated into collaborative filtering to address the cold-start problem and the data sparsity problem. In section 4, we demonstrate the evaluation metric and experiment results. Finally, we conclude the whole study.

3. Recommender Algorithm

3.1. Problem Definition

Formally, our problems is formulated as follows. We denote $U = \{u_1, u_2, \dots, u_m\}$ the set of users, $C = \{c_1, c_2, \dots, c_m\}$ the set of cases and $I = \{i_1, i_2, \dots, i_m\}$ the set of assessment indexes on IoII. Therefore, the relationships between users and indexes are denoted by a $U \times I$ matrix, called user-index rating matrix. Every entry r_{ui} represents the value that user $u \in U$ rated index $i \in I$. Then, let $P_{u,c}^{index}$ be a recommending function that measures the preference of user u on case c based on the user-index rating matrix. Similarly, we get function $P_{u,c}^{case}$ based on the user-case implicit rating matrix. The two functions $P_{u,c}^{index}$ and $P_{u,c}^{case}$ lead to a joint function $P_{u,c}$, which uses both

explicit rating and implicit preference. Then given a user u and an case list L , we will rank the cases in L according to $P_{u,c}$ and select top N cases as the recommending items for user u . As show in Equation (1), where R is the recommendation result.

$$\forall u \in U, R = \arg \text{TopN} \left(P_{u,c} \right)_{c \in L} \quad (1)$$

3.2. Similarity Model Based on the Assessment of IoII for Industrial Enterprises

While industrial enterprises in different stage of IoII have different feature in competition ability, economic benefit and social benefit, the case fit for the industrial enterprises is different. Therefore, the similarity model based on the assessment of IoII for industrial enterprises effectively show which kind of case the enterprise is interest in. This paper analyzes the development process of IoII for industrial enterprises and proposed the similarity model based on the assessment of IoII for industrial enterprises. The model described in detail in the Tab.1 was constructed by three dimensions of production, management and value chain.

Table 1. Similarity Model Based On Assessment of loii for Industrial Enterprises

First Level Indicator	Second Level Indicator	Indicator Symbol	Collect Items
Implementation of CNC	ratio of CNC production equipment account for production equipment	i_1	number of CNC production equipments
			number of production equipments
Implementation of ERP	application of material requirement planning	i_2	no partial all
	application of purchasing planning	i_3	no partial all
	application of master production planning	i_4	no partial all
	application of sales planning	i_5	no partial all
	application of financial budget planning	i_6	no partial all
	application of human resources planning	i_7	no partial all
Implementation of MES	ratio of automatic production scheduling workshops account for production workshops	i_8	number of automatic production scheduling workshops
			number of production workshops
	ratio of production process monitoring workshops account for production workshops	i_9	number of production process monitoring workshops
			number of production workshops
ratio of equipment state monitoring workshops account for production workshop	i_{10}	number of equipment state monitoring workshops	
		number of production workshops	
Implementation of SCM	application of collaborative management in supply chain	i_{11}	no partial all
	application of supply chain execution system	i_{12}	no partial all
Implementation of E-commerce	ratio of e-commerce sales account for total sales	i_{13}	amount of e-commerce sales
			amount of total sales
	ratio of e-commerce purchase account for total purchase	i_{14}	amount of e-commerce purchase
			amount of total purchase

Given two users u, v , $\text{int_sim}(u, v)$ which based on the assessment of IoII for

industrial enterprises is formally defined as follows:

$$\text{int_sim}(u,v) = \frac{\vec{u}_I \times \vec{v}_I}{\|\vec{u}_I\| \times \|\vec{v}_I\|} = \frac{\sum_{k=i_1}^{i_m} u_k \cdot v_k}{\sqrt{\sum_{k=i_1}^{i_m} u_k^2} \sqrt{\sum_{k=i_1}^{i_m} v_k^2}} \quad (2)$$

3.3. Similarity Model Based on User Learning Behavior

Generating recommendatory case requires to capture user preference. There exist basically two main user profiling strategies: explicit and implicit. Explicit user modeling requires users to manually provide information about their tastes. The main problem with explicit preferences is the extra burden they place on users and psychological tendency which users have to rate the preferences. Implicit user modeling relies on data mining techniques to automatically extract user preferences from their actions. Therefore, it does not impose any extra effort on users and looks like more objective.

In the knowledge platform for IoII, the learning behavior of user can be divided into three types: browse, enshrine and download. Browse refers to the behavior that user retrieves information by reading the case. Enshrine refers to the behavior that user add case to "my favorite". Download refers to the behavior that user download case to the local. As show in Equation (3), this paper establish a combined implicit rating model based on the above three types of learning behavior. Where $P(u, c)$ represent the rating on case c from user u , $P_b(u, c)$ represent user rating from browse, $P_e(u, c)$ represent user rating from enshrine and $P_d(u, c)$ represent user rating from download. α, β, γ are three parameters which determine the weights of different types of rating.

$$P(u, c) = \alpha * P_b(u, c) + \beta * P_e(u, c) + \gamma * P_d(u, c) \quad (3)$$

Through analyze the browse behavior of users, we find that the more users browse the case, the higher rating the case get. Meanwhile, we also observe the fact that classic case which is frequently browsed cannot fit various requirements of users. Therefore, as show in Equation (4), user rating from browse is weighted by the browse-frequency inverse-user-frequency (BF-IUF). $BF(u, c)$ is the browse-frequency on case c for user u , $F(u, c)$ is the browse count on case c for user u . $IUF(U, c)$ is the inverse-user-frequency for case c , C_u represents the set of cases learned by user u .

$$BF(u, c) = \frac{F(u, c)}{\sum_{c_i \in C_u} F(u, c_i)}$$

$$IUF(U, c) = \log \frac{|U|}{|\{u \in U : c \in C_u\}|} \quad (4)$$

$$P_b(u, c) = BF \cdot IUF(u, c, U) = BF(u, c) \times IUF(U, c)$$

In practice, if a user u add a case to "my favorite" which indicates that the user is highly interested in the case, a positive mark to the case will be given by the user. Motivated by this, we model user rating from enshrine by a binary function as Equation(5).

$$P_e(u, c) = \begin{cases} 1, & u \text{ enshrine } c \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Though one case may be downloaded many times for one user compare to just one enshrine, there is no essential difference between the two behaviors. According to this, we model user rating from download by Equation(6).

$$P_d(u, c) = \begin{cases} 1, & u \text{ download } c \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

Just like depicted in section 3.2, we can define the user learning behavior similarity model by Equation(7).

$$\text{beh_sim}(u, v) = \frac{\vec{u}_c \times \vec{v}_c}{\|\vec{u}_c\| \times \|\vec{v}_c\|} = \frac{\sum_{k=c_1}^{c_m} u_k \cdot v_k}{\sqrt{\sum_{k=c_1}^{c_m} u_k^2} \sqrt{\sum_{k=c_1}^{c_m} v_k^2}} \quad (7)$$

3.4. Neighbor Selection Algorithm and Recommendation Generating

Once the user similarity is modeled, we can select the neighbor set for users. For every user u , we select the ones whose similarity is greater than the chosen threshold as neighbors and we will get two neighbor sets because we have two different similarity model.

As show in Equation(8), I_u is the neighbor set based on the rating matrix of IoII for industrial enterprises, $int_sim(u, v)$ represents the user similarity based on the assessment of IoII for industrial enterprises and φ is a threshold between 0 and 1.

$$I_u = \{v | v \in U \wedge int_sim(u, v) > \varphi\} \quad (8)$$

As show in Equation(9), S_u is the neighbor set computed based on user learning behavior, $beh_sim(u, v)$ represents the user similarity based on user learning behavior and η play the same role as φ .

$$S_u = \{v | v \in U \wedge beh_sim(u, v) > \eta\} \quad (9)$$

We find that Equation (8) is simple and generally used. However, it is sensitive to the problem of data-sparsity and cold-start. Meanwhile, relying on complex user similarity modeling, Equation (9) can easily tackle these two problems. To take advantage of both models, we first make predictions using I_u and S_u respectively, then combines the predictions linearly by Equation(10). Where \bar{p}_u represent the average rating of user u and θ is the weight of predictions based on I_u . Then, top N cases will be selected as the recommending cases for target user u .

$$P_{u,c} = \bar{p}_u + \theta \cdot \frac{\sum_{v \in N_u} int_sim(u, v) \cdot (p_{v,c} - \bar{p}_v)}{\sum_{v \in N_u} int_sim(u, v)} + (1 - \theta) \cdot \frac{\sum_{v \in N_u} beh_sim(u, v) \cdot (p_{v,c} - \bar{p}_v)}{\sum_{v \in N_u} beh_sim(u, v)} \quad (10)$$

As we have already given all the algorithm, the pseudocode for this paper is sketched in following.

Input: users set U , cases set C , indexes set I , user-index rating matrix $R_{U,I}$, user-behavior information matrix $R_{U,C}$

Output: recommendatory case list R

//calculate the rating on case c for user u

function implicitRating ():

for each $u \in U, c \in C$ *do*

$$P(u, c) \leftarrow \alpha * \text{BF-IUF}(u, c, U) + \beta * (r_{u,c}^e ? 1 : 0) + \gamma * (r_{u,c}^d ? 1 : 0)$$

//calculate the distance of user u

function distance ():

for each $u_k \in \vec{u}$ *do*

$$dis_u \leftarrow dis_u + u_k * u_k$$

$$dis_u \leftarrow \text{sqr}(dis_u)$$

//calculate the int_sim for user u

function int_sim ():

for each ($v \in U$) & ($v \neq u$) *do*

for each ($u_k \in \vec{u}$) & ($v_k \in \vec{v}$) *do*

$$dP \leftarrow dP + u_k * v_k$$

for each $v_k \in \vec{v}$ *do*

$$dis_v \leftarrow dis_v + v_k * v_k$$

$$dis_v \leftarrow \text{sqr}(dis_v)$$

$$u.\text{int_sim}(v) \leftarrow dP / dis_u * dis_v$$

// acquire the neighbor set based on int_sim

function intNeighbor ():

for each ($v \in U$) & ($v \neq u$) *do*

if ($u.\text{int_sim}(v) > \phi$)

add $I_u \leftarrow v$

//calculate the beh_sim for user u ()

function beh_sim ():

the same as *function* int_sim (), we get $u.\text{beh_sim}(v_k)$

// acquire the neighbor set based on beh_sim

function behNeighbor ():

the same as *function* intNeighbor (), we get S_u

// acquire the predictions

function predictions ():

for each ($w \in I_u$) || ($w \in S_u$) *do*

for each $c \in w.\text{rating} []$ *do*

$$P_{u,c} \leftarrow \bar{p}_u + \theta * I_u.w.\text{rating}(c) + (1 - \theta) * S_u.w.\text{rating}(c)$$

// generate recommendation

function topN ():

for $i = 1$ *to* N *do*

$$R(i) \leftarrow \max(P_{u,c})$$

$$P_{u,c}.\text{remove}(R(i))$$

4. Experimental Analysis

4.1. Data Sources

Relying on the project of knowledge platform for IoII, the dataset is acquired from questionnaire for industrial enterprise in Beijing. As the dataset is provided in 2014, it reflects current status of industrial enterprise in Beijing faithfully. Thus the experiment results is in practice. The dataset contains ratings rated by 655 industrial enterprises on 16 indexes of IoII, as well as the learning behavior records of 1194 cases. The density of the index matrix is 4.93%, while the learning behavior matrix is 3.28%.

4.2. Performance Evaluation Metrics

Considering the overfitting problem, we use the five-fold cross-validation to evaluate the performance of this algorithm and calculate the average of five results as the final result. MAE (Mean Absolute Error) is the most common method for evaluating the accuracy of a recommender algorithm by comparing the numerical prediction values against user raw ratings. As show in Equation(11), p_i and r_i , represent, respectively, the predicted and practical rating, N denotes the number of users.

$$\text{MAE} = \frac{\sum_{i=1}^N |p_i - r_i|}{N} \quad (11)$$

4.3. Results and Discuss

In This section, we are expended to answer three questions. Frist, what impact will α, β, γ have on the user learning behavior model? Second, how do the two neighbor selection algorithm benefit each other in improving prediction accuracy? Third, as we got three models for recommendation in section 3.4, which one performs the best? All tests are conducted on a 64-bits Intel Core i3 @ 2.53 GHz, with 6 GB RAM, running Windows 8 pro. We develop the program with Java on the IDE (Integrated Development Environment) of myeclipse.

To show the impact of α, β, γ , we set θ to 0, that is to say the predictions based on I_u is excluded. Then, we set φ to 0.1, N to 8, and respectively vary the range of α, β, γ from 0 to 1 with a step value of 0.1. Accordingly, we plot the MEA – α, β, γ curve. As sketched in Figure 1, MAE follows the Cycle that decrease first and then increase with the variation of α, β, γ . It implies that the most positive influence on prediction accuracy is browse($\alpha = 1, \beta = 0, \gamma = 0$), the second is download($\alpha = 0, \beta = 0, \gamma = 1$) and the last is enshrine($\alpha = 0, \beta = 1, \gamma = 0$). We get the best performance when $\alpha = 0.5, \beta = 0.3, \gamma = 0.2$ on our experiment dataset, which means the weight of rating on browse, enshrine and download is 0.5,0.3,0.2.

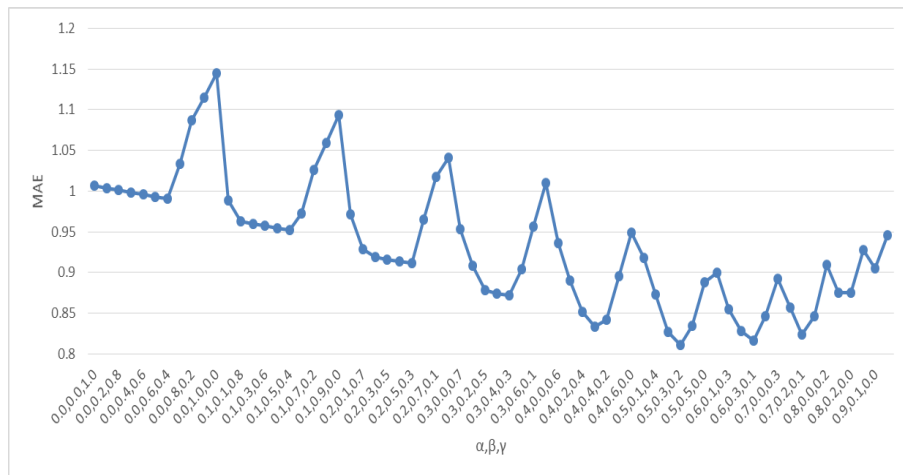


Figure 1. Impact of Two Neighbor Selection Algorithm on Prediction Accuracy

With the second question, these two neighbor selection algorithms were combined to make predictions. Parameter θ balances the effect of ratings on IoII and user learning behavior. If $\theta = 0$, the final predictions is only influenced by the neighbor selection algorithm based on user learning behavior. On the contrary, if $\theta = 1$, only the ratings on IoII is used to compute neighbor set for users. In other cases, we combine the predictions based on the two neighbor sets to get the final predictions. We set φ to 0.1, η to 0.5 and then vary the range of N . Accordingly, we plot the MEA – θ curve. As sketched in Figure 2, MAE increases as we increase the value of N . However, we find that the best performance always appear at $\theta = 0.7$, which may indicate that the rating information is more important than user behavior information.

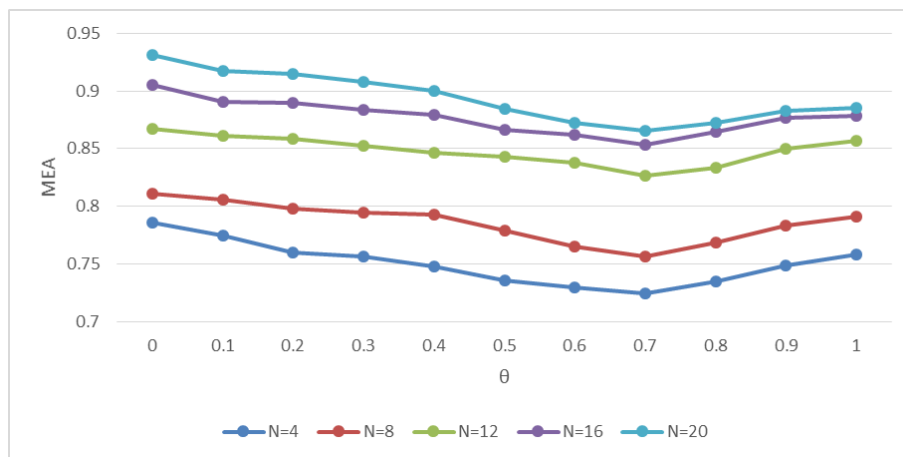


Figure 2. Impact of User Learning Behavior Model on Prediction Accuracy

To compare the three recommendation algorithms respectively, we set all parameters to the optimal, *i.e.* $\alpha = 0.5, \beta = 0.3, \gamma = 0.2, \varphi = 0.1, \eta = 0.5, \theta = 0.7$ and take different values on N . As depicted in Table 2, Int represents the recommendation algorithm based on the assessment of IoII for industrial enterprises. Beh represents the recommendation algorithm based on user learning behavior and Hyb represents the hybrid recommendation algorithm. Obviously, Hyb is evidently superior and Int exhibits the better performance compared with Beh.

Table 2. Prediction Accuracy of Three Recommendation Algorithms

	N=4	N=8	N=12	N=16	N=20
Int	0.7582	0.7910	0.8565	0.8785	0.8855
Beh	0.7861	0.8117	0.8676	0.9054	0.9310
Hyb	0.7250	0.7570	0.8268	0.8530	0.8654

5. Conclusions

As the rapid development of IoII, users in the platform for IoII cannot quickly and accurately locate the case they need. In this paper, a hybrid framework is established based on both the assessment system of IoII and user learning behavior. With the information of assessment of IoII for industrial enterprises and user learning behaviors, we build two user similarity models respectively. Then, a linear fusion framework which inherits advantages of both is proposed. The experiment results show that our framework improves the recommendation quality significantly. However, the similarity model based on user learning behavior may have great impact on the effectiveness of the framework and more specific user similarity models need to be developed. Further study may explore more learning behavior to deepen our modeling of user similarity. In addition to the user learning behavior information, mining rating information based on IoII is also valuable to enhance the predictions accuracy.

Acknowledgments

The study is supported by National Natural Science Foundation (grant numbers 71132008, 71390334). Thank Beijing Municipal Science & Technology Commission for providing the dataset.

References

- [1] L. Min, J. Wang and Q. Luo, “Software Industry Integration Process of Informatization and Industrialization”, 2015 International Conference on Education Technology Management and Humanities Science, Atlantis Press, (2015).
- [2] D. Wu and Y. Sun, “A Study on the Evolution Model of the Integration of Informatization and Industrialization”, Control and Decision Conference, the 26th Chinese. IEEE, (2014). pp. 4226-4229.
- [3] D. Goldberg, D. Nichols, BM Oki and D. Terry, “Using Collaborative Filtering to Weave an Information Tapestry”, Communications of the ACM, vol.35, no.12, (1992), pp.61 -70.
- [4] GB/T 23020-2013, Assessment Specification on IoII for Industrial Enterprises [S].
- [5] J. Zhang and J. Zheng, “Measure Theory System of the Integration of Infomationization and Industrialization”, Information Science, vol.1, no.006, (2013).
- [6] S. Li, S. Dong and J. Zhou, “Growth Model of Production Enterprise Competitiveness Based on Informatization”, Information Science, vol.2, no.010, (2013).
- [7] N. David, “Implicit Rating and Filtering”, 5th DELOS Workshop on Filtering and Collaborative Filtering (ERCIM), Budapest, Hungary, (1997), pp.31-36.
- [8] D. Oard, “Implicit Feedback for Recommender Systems”, In Proceedings of the AAAI Workshop on Recommender Systems, (1998).
- [9] D. Oard and J. Kim, “Modeling Information Content Using Observable Behavior”, In Proceedings of the 64th Annual Conference of the American Society for Information Science and Technology, (2001).
- [10] Y. Koren, “Factor in the Neighbors: Scalable and Accurate Collaborative Filtering. ACM Transactions on Knowledge Discovery from Data, vol.4, no.1, (2010), pp.1–24.
- [11] P. Kostkova and G. Madle, “What Impact Do Healthcare Digital Libraries Have? An Evaluation of National Resource of Infection Control at the Point Of Care Using The Impact-ED Framework”, International Journal on Digital Libraries, vol.13, no.2, (2013), pp.77–90.
- [12] B. Hidasi and D. Tikk, “Fast Als-Based Tensor Factorization For Context Aware Recommendation From Implicit Feedback in Machine Learning and Knowledge Discovery in Databases”, Flach P., Bie T., Cristianini N., Eds. Springer Berlin Heidelberg, (2012).
- [13] H.N. Kim and El Saddik, “A.Exploring Social Tagging for Personalized Community Recommendations. User Modeling and User-Adapted Interaction”, vol.23, no. (2-3), (2013), pp. 249-285.
- [14] P. Pirolli, S. Kairam, “A Knowledge-Tracing Model of Learning from a Social Tagging System. In:

- Brusilovski,P., Chin, D. (eds.) User Modeling and User-Adapted Interaction. Special Issue on Personalization in Social Web Systems, (2012).
- [15] J. Aligon, E. Gallinucci and M. Golfarelli, “A Collaborative Filtering Approach for Recommending OLAP Sessions”, Decision Support Systems, vol.69, (2015), pp.20-30.

Authors



Laisong Kang has received the B.B.M degree in School of Economics and Management, Beijing Jiaotong University. Now he is a Master Candidate of Information Management.



Shifeng Liu was born in 1970 in Hebei, Baoding. Now, he majors in information management and serves as doctoral tutor in school of economics and management, Beijing Jiaotong University.



Daqing Gong has graduated as a doctor from Beijing Jiaotong University. Now, he majors in the application of simulation in resources and environmental economics.