WALAILAK JOURNAL
http://wjst.wu.ac.th Information Technology

Hybrid Approach for a Knowledge Recommender Service: A Combination of Item-Based and Tag-Based Recommendation*

Winyu NIRANATLAMPHONG[1] and Worasit CHOOCHAIWATTANA[2,*]

[1]*Department of Interactive Design and Game Development, College of Creative Design and Entertainment Technology, Dhurakij Pundit University, Bangkok 10210, Thailand*
[2]*Department of Web Engineering and Mobile Application Development, College of Creative Design and Entertainment Technology, Dhurakij Pundit University, Bangkok 10210, Thailand*

([*]Corresponding author's e-mail: worasit.cha@dpu.ac.th)

Received: 23 September 2016, Revised: 18 April 2017, Accepted: 1 May 2017

Abstract

An exponentially increasing of knowledge in a knowledge management system is the main cause of the knowledge overload problem. A development of knowledge recommender service embedded in the knowledge management system becomes a challenging task. This paper proposed a hybrid approach by combining an item-based recommendation technique, also known as a collaborative filtering technique, with a tag-based recommendation technique, also known as a content-based filtering technique. To evaluate the performance of the proposed hybrid approach, a group of knowledge management system users were invited to be participants in this research study. As a criterion, the participants were asked to use the prototype of a knowledge management system embedded with the knowledge recommender service for 6 months. This would guarantee that each participant's interaction with knowledge items could be recorded. A confusion matrix was then used to compute an accuracy of the proposed hybrid approach. The result of the experiment revealed that the proposed hybrid approach outperformed the item-based approach and the tag-based approach. Hence, the proposed hybrid approach was a promising technique for a knowledge recommender service in the knowledge management system.

Keywords: Collaborative filtering, content-based filtering, item-based recommendation, tag-based recommendation, knowledge recommender service

Introduction

A knowledge management system helps its users to store, to retrieve and to disseminate their knowledge. In the era of information explosion, the knowledge management system becomes a necessary tool to store and disseminate important and interesting pieces of information among the users especially in an academic area. Education institutions exploit the usage of the knowledge management system to support their students and faculty members - the primary users of the knowledge management system. However, when tremendous collections of knowledge are stored in the knowledge management system, the users will encounter with the problem of knowledge overload. The knowledge retrieving service then becomes a major player to help the users to discover their desired knowledge. The smarter of the knowledge retrieving service is the more effective and efficient the knowledge management system become. However, with the effective knowledge retrieving service, the problem of knowledge overload can still be remained. It takes time for each user to find the interesting knowledge items in a long list of knowledge search results.

---

*Presented at 1st International Conference on Information Technology: October 27th - 28th, 2016

One way to overcome this problem is to develop an automatic knowledge dissemination mechanism or *a knowledge recommender service*. Because each user tends to have a limit set of interesting topics, it is possible that the knowledge recommender service takes users' interaction with knowledge items into consideration for predicting a list of recommendation items. Thus, the design and development of the knowledge recommender service embedded in the knowledge management system become a challenging task.

This paper is organized as follows: next section provides details on related work. The following section introduces a proposed hybrid recommendation mechanism. Then, the experimental setting and evaluation are explained. In addition, the results and discussion are revealed. Last but not least, the conclusion and future work are described.

Literature review

With a tremendous amount of information available, a recommender system plays a crucial role for filtering pieces of information that people may be interested in. Instead of asking for recommendations from other people, the recommender system provides suggestion on any items which are likely to be interesting for a user [1-3].

In 1992, Goldberg and his colleagues [4] introduced a recommender system, named *Tapestry*, and coined a term "Collaborative Filtering". The principle idea of Collaborative Filtering (CF) is that if 2 persons have the same behaviors, for instance, buying similar items, they will act on other items similarly [5]. The CF has been adopted and used to develop early generation of recommender systems, such as *GroupLens*, an online news recommender system [6]. Paul Resnick and his colleagues from MIT and University of Minnesota exploit the usage of user rating on news articles to provide news recommendation. Many commercial system, such as Amazon.com, use this technique because of easy implementation with highly effective [7,8]. The main drawback of the CF, however, is the cold-start problem. This problem occurs when there are a few interactions available for items. Thus, recommender systems cannot make any recommendation at the beginning [9]. Besides the CF, Content-Based Filtering is another technique for recommender systems. The principle idea of content-based filtering is that when each person interacts with any items, the content of those items will be recorded. A user profile of each person will then be created. A recommender mechanism in the content-based filtering technique will recommend any items based on a matched score calculation between the user profile and a content of each item. The content-based filtering, however, will only be appropriate if each item in the recommender system has textual content. The main difference between the collaborative filtering and the content-based filtering is that the collaborative filtering only uses user-item interactions to make predictions and recommendations, while the content-based filtering used the extracted feature of each user and each item for recommendation [10]. However, both techniques can be suffered when the system has a huge number of users and items and a few user-item interactions. This will lead to the problem of Sparsity [11]. To overcome the limitations of other techniques previously discussed, a hybrid recommender technique combines the collaborative filtering and the content-based filtering. It exploits the usage of item content and user-item interaction to create item recommendations [12].

As mentioned in the previous section, the users of the knowledge management system will encounter with the problem of knowledge overload, when tremendous collections of knowledge are stored. The knowledge retrieving service could derive benefit from applying the collaborative filtering or the content-based filtering to provide an automatic knowledge dissemination mechanism or a knowledge recommender service [13-19]. There were, however, few published research papers focused on a combination of the collaborative filtering and content-based filtering to develop the knowledge recommender service. Hence, this paper investigated how well the combination of the collaborative filtering, called item-based recommendation, and the content-based filtering, called tag-based recommendation, contributes to the task of the automated knowledge dissemination.

A proposed hybrid recommendation mechanism

Our proposed hybrid recommendation mechanism combines the item-based recommendation mechanism with the tag-based recommendation mechanism. To explain the proposed mechanism, the item-based recommendation mechanism and the tag-based recommendation mechanism are first described.

As illustrated in **Figure 1**, the item-based recommendation mechanism considers user-item interactions for calculating similarity measurements among users. The users, who interact with similar knowledge items, will be put together in the same group. The user-item interaction of all users in the same group can then be identified. The item-based recommendation mechanism will select knowledge items and disseminate the items to all the users in that group as illustrated in **Figure 2**.

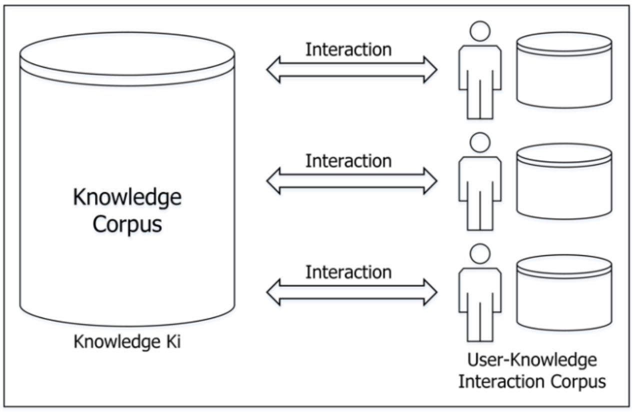

Figure 1 Concept of the item-based recommendation mechanism.

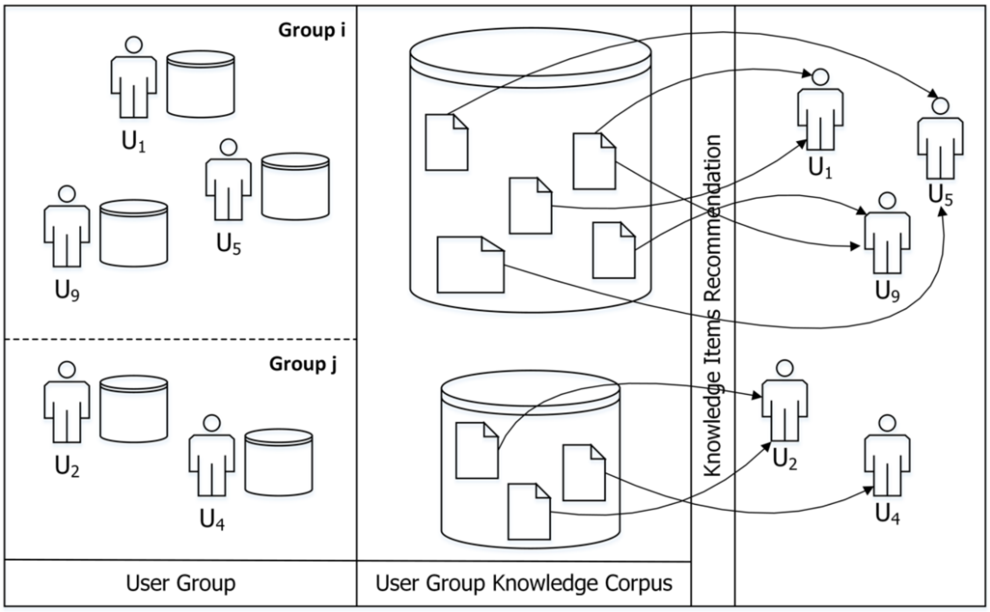

Figure 2 Example of knowledge items recommendation method.

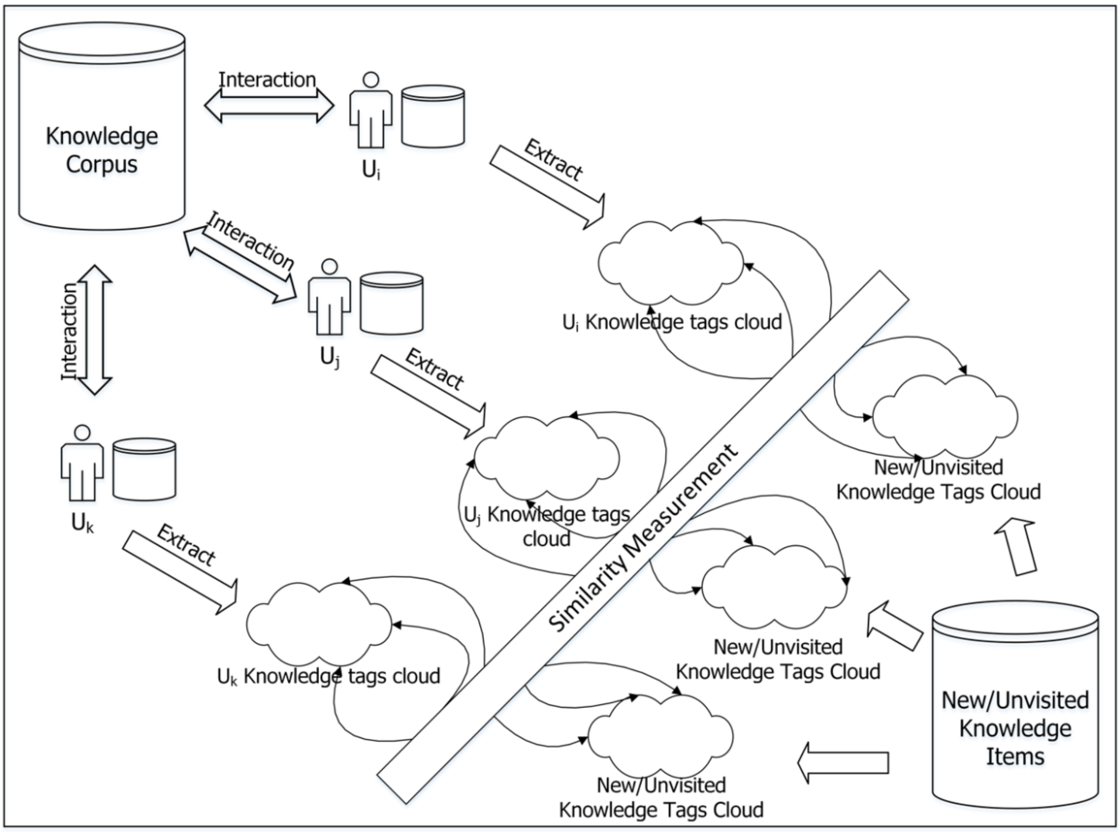

Figure 3 Concept of tag-based recommendation mechanism.

The tag-based recommendation mechanism, as illustrated in **Figure 3**, on the other hand, considers user-item interactions for extracting knowledge content, which is knowledge for this particular study. The extract knowledge tags will be put in a set of user's knowledge tags, which represents knowledge interest of each user. The tag-based recommendation mechanism will select new or unvisited knowledge items and disseminate the items by calculating similarity measurements between content of the set of user's knowledge tags and content of new or unvisited knowledge items.

To implement our proposed hybrid recommendation mechanism, there are 8 main components 1) set of users, 2) set of interactions with knowledge items, 3) set of users' knowledge tag, 4) set of new/unvisited knowledge items, 5) set of tags from new/unvisited knowledge items, 6) set of tags from all knowledge items in knowledge corpus, 7) similarity measurement, and 8) knowledge corpus.

Let N_u be the number of users and N_k be the number of knowledge items in a knowledge management system. Let U be a set of users and contains all users in the knowledge management system; $U = \{u_1, u_2, u_3, \ldots, u_n\}$, K be a set of knowledge items and contains all knowledge items in knowledge corpus; $K = \{k_1, k_2, k_3, \ldots, k_m\}$, and T be a set of knowledge tags and contains all knowledge tags associated with knowledge items; $T = \{T_1, T_2, T_3, \ldots, T_p\}$. Let M_{uk} be the $N_u \times N_k$ association matrix between users and knowledge items. $M_{uk}(u_x,k_y)$ will be equal to 1 when user u_x bookmarks a knowledge item k_y. Thus, each row, or UK_i, in M_{uk} represents user interaction with knowledge items. In addition, for each user u_x, let UTK_x be a set of user's knowledge tag derived from M_{uk}; $UTK_x = \{< u_x, t_p > \mid u_x \in U \wedge t_p \in T \wedge M_{uk}(u_x, k_y) = 1\}$ and $NTK_y(u_x)$ be a set of tags from new or unvisited knowledge items derived from M_{uk}; $NTK_y(u_x) = \{< k_y, t_p > | k_y \in K \wedge t_p \in T \wedge M_{uk}(u_x, k_y) = 0\}$.

A similarity measurement between each user is then performed. A group of users, who interact with similar knowledge items, can be identified if the similarity score, as illustrated in Eq. (1), is equal to or greater than a predefined threshold ($\alpha \geq 0.55$, for this particular study). Let $GroupU_x$ be a set of users who interact with similar knowledge items and $GroupKU_x$ be a set of knowledge items of every users in $GroupU_x$; $GroupKU_x = \{< u_i, k_j > | k_j \in K \wedge u_i \in GroupU_x \wedge M_{uk}(u_i, k_j) = 1\}$. Thus, $GroupKU_x$ can be used in the item-based recommendation mechanism as illustrated in **Figure 2**.

$$USim(UK_x, UK_y) = \frac{\sum_{j=1}^{t} UK_{xj} \cdot UK_{yj}}{\sqrt{\sum_{j=1}^{t} UK_{xj}^2 \cdot \sum_{j=1}^{t} UK_{yj}^2}} \quad (1)$$

Since UTK_x represents knowledge interest of user x and $NTK_y(u_x)$ represents content of new or unvisited knowledge, a similarity measurement between UTK_x and $NTK_y(u_x)$ can be calculated as illustrated in Eq. (2). Let KU_x be a set of new or unvisited knowledge items which $KSim(UTK_x, NTK_y(u_x))$ is equal to or greater than a predefined threshold ($\alpha \geq 0.55$, for this particular study). Hence, it can be used in the tag-based recommendation mechanism as illustrated in **Figure 3**.

$$KSim(UTK_x, NTK_y(u_x)) = \frac{\sum_{j=1}^{t} UTK_{xj} \cdot NTK_y(u_x)_j}{\sqrt{\sum_{j=1}^{t} UTK_{xj}^2 \cdot \sum_{j=1}^{t} NTK_y(u_x)_j^2}} \quad (2)$$

As mentioned previously, the proposed hybrid recommendation mechanism combines the item-based recommendation mechanism with the tag-based recommendation mechanism. The proposed hybrid approach takes the results from $GroupKU_x$ and KU_x into account when predicting the recommended list of knowledge items for each user.

Let $RecList1(u_x)$ be a set of recommended knowledge items for user u_i, which show up in both $GroupKU_x$ and KU_x; $RecList1(u_i) = \{< u_i, k_j > |(k_j \in GroupKU_x \wedge M_{uk}(u_i, k_j) = 0) \wedge (k_j \in KU_x)\}$ and $RecList2(u_x)$ be a set of recommended knowledge items for user u_i, which show up in either $GroupKU_x$ or KU_x; $RecList2(u_i) = \{< u_i, k_j > |(k_j \in GroupKU_x \wedge M_{uk}(u_i, k_j) = 0) \vee (k_j \in KU_x)\}$.

Experimental setting and evaluation

To evaluate the proposed hybrid recommendation mechanism, the data from a knowledge management system at Faculty of Information Technology, Dhurakij Pundit University, were crawled and loaded into our prototype knowledge management system embedded with knowledge recommender services. This prototype was designed and developed to make the task of recommendation mechanism evaluation easier.

Because the primary users of the knowledge management system are graduate students and faculty members, the knowledge items posted on the system are mainly IT and technical related knowledge items, for instance, programming techniques, new technology reviews, and research papers. However, some of knowledge items are for non-technical related, for example, news and announcements, curriculum information, and personal interested sharing items. The crawler randomly crawled the knowledge items from the knowledge management system. The final dataset loaded into the prototype system contained 5,359 knowledge items with 1,565 unique knowledge tags.

In this study, the results from the item-based recommendation mechanism and from the tag-based recommendation mechanism were compared with the results from the proposed hybrid recommendation mechanisms. There are 2 proposed hybrid recommendation approaches - *RecList1* and *RecList2*. In *RecList1*, a list of recommended knowledge items will be constructed by considering the knowledge items that both appeared in a set of knowledge items of every user in the same group ($GroupKU_x$) and a set of new or unvisited knowledge items(KU_x). On the other hand, a list of recommended knowledge items from *RecList2* will be constructed by considering the knowledge items that either appeared in a set of knowledge items of every user in the same group ($GroupKU_x$) or a set of new or unvisited knowledge items(KU_x).

The percentage of accuracy in knowledge items recommendation was an evaluation metric as illustrated in Eq. (3). **Table 1** shows a confusion matrix used for calculating the percentage of accuracy.

$$Percentage\ of\ Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \tag{3}$$

Table 1 A confusion matrix used for calculating the percentage of accuracy.

Actual	Predicted	
	Positive	Negative
Positive	True Positive (No. of Recommended User-Interested Knowledge Items)	False Negative (No. of Unrecommended User-Interested Knowledge Items)
Negative	False Positive (No. of Recommended User-Uninterested Knowledge Items)	True Negative (No. of Unrecommended User-Uninterested Knowledge Items)

Table 2 A table showing TP, TN, FP, FN values used for calculating the percentage of accuracy.

Period	Recommendation mechanism	All participants				% of accuracy
		TP	TN	FP	FN	
Month 1	Item-based Recommendation	97	209	554	640	20.40
	Tag-based Recommendation	216	242	623	419	30.53
	Hybrid Recommendation (RecList1)	54	104	316	1026	10.53
	Hybrid Recommendation (RecList2)	192	166	596	546	23.87
Month 2	Item-based Recommendation	347	336	361	456	45.53
	Tag-based Recommendation	469	366	415	250	55.67
	Hybrid Recommendation (RecList1)	339	425	211	525	50.93
	Hybrid Recommendation (RecList2)	516	483	242	259	66.60
Month 3	Item-based Recommendation	436	399	315	350	55.67
	Tag-based Recommendation	593	469	274	164	70.80
	Hybrid Recommendation (RecList1)	435	519	129	417	63.60
	Hybrid Recommendation (RecList2)	659	626	99	116	85.67
Month 4	Item-based Recommendation	424	381	326	369	53.67
	Tag-based Recommendation	615	479	256	150	72.93
	Hybrid Recommendation (RecList1)	426	514	136	424	62.67
	Hybrid Recommendation (RecList2)	638	602	121	139	82.67
Month 5	Item-based Recommendation	399	370	342	389	51.27
	Tag-based Recommendation	625	482	240	153	73.80
	Hybrid Recommendation (RecList1)	410	505	152	433	61.00
	Hybrid Recommendation (RecList2)	630	609	125	136	82.60
Month 6	Item-based Recommendation	360	351	365	424	47.40
	Tag-based Recommendation	620	475	252	153	73.00
	Hybrid Recommendation (RecList1)	399	495	162	444	59.60
	Hybrid Recommendation (RecList2)	625	593	132	150	81.20

Thirty subjects - consisted of 15 graduate students and fifteen alumni - were recruited to be participants in this research study. As a criterion, the participants were asked to use the prototype of a knowledge management system embedded with the knowledge recommender service for 6 months. This would guarantee that each participant's interaction with knowledge items could be recorded. During the 6 month period, each participant was asked to evaluate a list of recommended knowledge items at the end of each month. The list of recommended knowledge items - generated from the item-based recommendation mechanism, the tag-based recommendation mechanism, and the hybrid recommendation mechanism - were merged and all duplicated knowledge items were removed. Before evaluating the list, each participant was informed that the list would be displayed in a random order. The evaluation result provided by each participant were then associated with the original list produced by each recommendation mechanism. The percentage of accuracy for each recommendation mechanism was then calculated.

Results and discussion

To evaluation an effectiveness of each recommendation mechanism, the percentage of accuracy was examined. A higher accuracy indicates a more effectiveness of the recommendation mechanism. **Figure 4** provides a line chart of the percentage of accuracy for the item-based recommendation mechanism, the tag-based recommendation mechanism and the hybrid recommendation mechanism during the 6 month period. **Table 2** shows the TP, TN, FP, FN values used for calculating the percentage of accuracy for each recommendation mechanism.

Figure 4 showed that the proposed hybrid recommendation mechanisms outperformed the item-based approach and the tag-based approach. According to **Figure 4**, the performance of the proposed approach during the first month of experiment was lower than the performance of tag-based approach. When the interactions between users and items were not enough, the tag-based approach performed better because the set of users' knowledge tags, extracted from set of interactions with knowledge items, represented actual users' interest. On the other hand, recommending knowledge items using item-based approach suffered from the cold start problem and the false positive. It turned out that when the information in the set of interactions with knowledge items was not enough, it directly impacted on the performance of the similarity measurement between users resulting in assigning users to an inappropriate group. From the second to the sixth month of the experiment, the hybrid approach with RecList2 performed better than other approach. Exploiting the usage of both tag-based approach and item-based approach, not only built a users' profile for their interest items, but also took opportunities in exploring potential items of interest by considering other users' interaction. It is not surprising that the item-based approach and the hybrid approach with RecList1 has a lower percentage of accuracy comparing with the other 2 approaches. It is possible that users might interest different pieces of knowledge even though they have interacted with the same knowledge items.

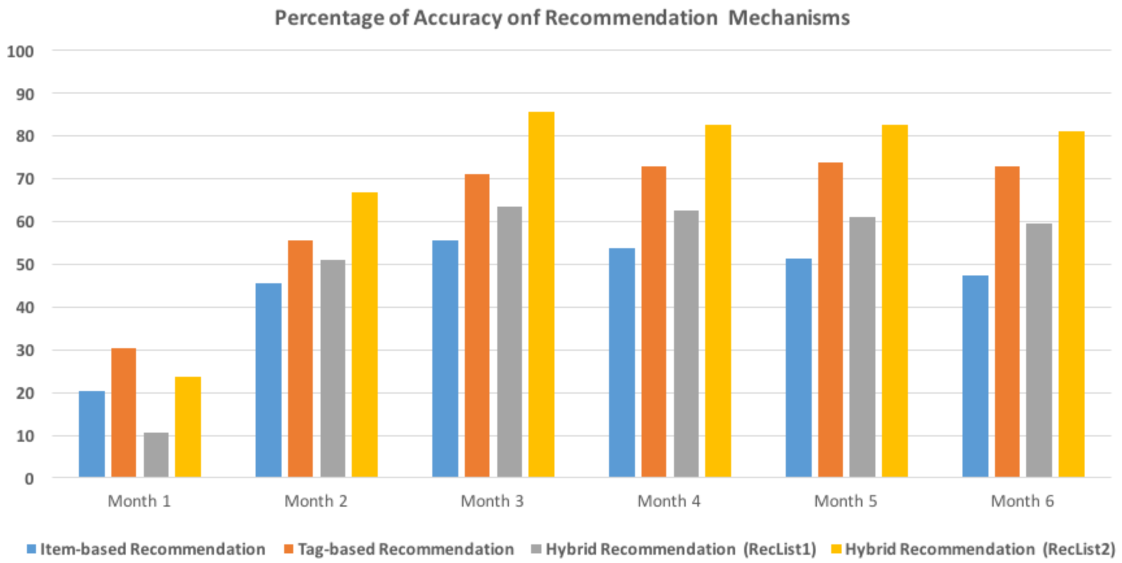

Figure 4 The percentage of accuracy of recommendation mechanisms.

From the result of the experiment, it can be concluded that both individual's item interactions and other people's item interactions should be considered when developing recommendation mechanism. The proposed hybrid method can be applied not only to improve a recommendation service for a knowledge management system but also to improve a recommendation service for systems that contain other content as well, for instance, a research paper recommender system.

Conclusion and future work

This paper investigated how well the combination of the collaborative filtering, called item-based recommendation, and the content-based filtering, called tag-based recommendation, augments to the task of automated knowledge dissemination services on a knowledge management system. The evaluation result suggests that the hybrid recommendation mechanisms outperformed the item-based recommendation mechanism - *a collaborative filtering technique* and the tag-based recommendation mechanism - *a content-based filtering technique*. Although the hybrid recommendation mechanism proposed in this paper relies on a simple combination between the item-based recommendation mechanism and the tag-based recommendation mechanism, the effectiveness of the proposed mechanism is remarkable. Considering a group of people interacting with the same items together with extracting content from individual's interaction with items does benefit a task of recommendation.

However, to improve a performance of the proposed mechanism, a further analysis on how to combine a content-based filtering technique with a collaborative filtering technique needs to be performed. It appears that people's interest may change as time goes by. Thus, a more robust technique to keep track of people's interest should be investigated. Last but not least, we would like to explore more on how to apply our proposed hybrid approach to build a recommender service on other domains such as research paper recommender service, news recommender service, and etc.

References

[1] F Ricci, L Rokach and B Shapira. *Introduction to Recommender System Handbook*. *In*: Recommender Systems Handbook. Springer, 2011, p. 1-35.
[2] R Bruke. *Hybrid Web Recommender System*. *In*: The Adaptive Web, Springer, 2007, p. 377-408.
[3] P Resnick and HR Varian. Recommender systems. *Commun. ACM* 1997; **40**, 56-8.
[4] D Goldberg, D Nichols, BM Oki and D Terry. Using collaborative filtering to weave an information tapestry. *Commun. ACM* 1992; **35**, 61-70.
[5] K Goldberg, T Roeder, D Gupta and C Perkins. Eigentaste: A constant time collaborative filtering algorithm. *Inform. Retrieval*. 2001; **4**, 133-51.
[6] P Resnick, N Iacovou, M Suchak, P Bergsrom and J Riedl. Grouplens: An open architecture for collaborative filtering of netnews. *In*: Proceedings of the ACM Conference on Computer Supported Cooperative Work. North Carolina, USA, 1994, p.175-86.
[7] G Linden, B Smith and J York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Comput*. 2003; **7**, 76-80.
[8] M Deshpande and G Karypis. Item-based top-N recommendation algorithms. *J. ACM Trans. Inform. Syst.* 2004; **22**, 143-77.
[9] FG Davoodi and O Fatemi. Tag based recommender system for social bookmarking sites. *In*: Proceedings of the ACM International Conference on Advances in Social Networks Analysis and Mining. Istanbul, Turkey, 2012, p. 934-40.
[10] L Si and R Jin. Flexible mixture model for collaborative filtering. *In*: Proceedings of the 20th International Conference on Machine Learning. Washington DC, USA, 2003, p. 704-11.
[11] S Jain, A Grover, PS Thakur and SK Choudhary. Trends, problems and solutions of recommender system. *In*: Proceedings of the International Conference on Computing, Communication and Automation. Noida, India, 2015, p. 955-8.
[12] R Prasad and VV Kumari. A categorical review of recommender systems. *Int. J. Distrib. Par. Syst*. 2012; **3**, 73-83.
[13] Z Huang, X Lu, H Duan and C Zhao. Collaboration-based medical knowledge recommendation. *Artif. Intell. Med*. 2012; **55**, 13-24.
[14] H Li, L Liu and C Lv. Knowledge recommendation services based on knowledge interest groups. *In*: Proceedings of the International Conference on Service Systems and Service Management. Troyes, France, 2006, p. 162-6.
[15] W Zhao, J Wang and G Lui. A knowledge recommendation algorithm based on content syndication. *In*: Proceedings of the 4th International Conference on Computer Sciences and Convergence Information Technology. Seoul, South Korea, 2009, p. 742-5.
[16] J Aryal, R Dutta and A Morshed. Development of an intelligent environmental knowledge recommendation system for sustainable water resource management using modis satellite imagery. *In*: Proceedings of the 2013 IEEE International Geoscience & Remote Sensing Symposium. Melbourne, VIC, Australia, 2013, p. 2204-7.
[17] A Vizcaino, J Portillo-Rodriguez, JP Soto, M Piattini and O Kusche. A recommendation algorithm for knowledge objects based on a trust model. *In*: Proceedings of the IEEE International Conference on Research Challenges in Information Science. Fez, Morocco, 2009, p. 93-102.
[18] K Liang, S Cai and Q Zhao. Context-based knowledge recommendation: A 3-D collaborative filtering approach. *In*: Proceedings of the 5th IEEE International Conference on Industrial Informatics. Vienna, Austria, 2007, p. 627-32.
[19] W Choochaiwattana. A comparison between item-based and tag-based recommendation on a knowledge management system: A preliminary investigation. *Inter. J. Inform. Educ. Tech*. 2015; **5**, 754-7.