A Comparison between Item-Based and Tag-Based Recommendation on a Knowledge Management System: A Preliminary Investigation

Worasit Choochaiwattana

Abstract—**Knowledge is one of the most important critical factor for the success of organizations. As huge collections of knowledge are stored in a knowledge management system, hence the development of an automated knowledge dissemination mechanism becomes a challenging task. To improve performance of such a system, a "knowledge recommendation service" should be embedded. This paper compared the effectiveness of two recommendation techniques namely, an item-based recommendation (a collaborative filtering technique) and a tag-based recommendation (a content-based filtering technique). In order to evaluate the performance of the two proposed recommendations, a group of knowledge management system users were ask to be participants in this research study. The percentage of accuracy in knowledge items recommendations was used as an evaluation metric. The result of the experiment revealed that the tag-based recommendation had an accuracy of 86.67%, which is higher than that of item-based recommendation. Thus, the tag-based recommendation showed a promising approach to the task of automated knowledge dissemination in a knowledge management system.**

Index Terms—**Collaborative filtering, content based filtering, item-based recommendation, tag-based recommendation, knowledge management system.**

I. MOTIVATION

Knowledge is an asset and one of the most important critical factors for the success of a modern organization. A knowledge management system has become a critical part of modern business management in both profit and non-profit organizations. While business organizations exploits the usage of various information systems to gain competitive advantages, the knowledge management system helps members of the organization store, retrieve and disseminate their knowledge and also helps the business organization maintain its core knowledge. Thus, it becomes a center of learning about organization core knowledge for all staffs in the organization.

One of the main services embedded in any knowledge management systems is a knowledge retrieval service. The more effective this service provides, the more competent a business organization staff becomes. However, users of knowledge management system may spend a lot of time for finding their desired knowledge.

As huge collections of knowledge are stored in the

knowledge management system, a development of an automatic knowledge dissemination mechanism becomes a challenging task.

II. RELATED WORK

With a huge amount of information available, recommendation system (also known as *recommender systems*) are now play an important role for selecting and recommending pieces of information that people may be interested in. Instead of asking for recommendations from friends, relatives and colleagues about things to buy, movies to watch, music to listen and so forth, the recommendation systems provide suggestions on any items to be of interest and valuable to a user [1]-[3].

In 1992, a term "Collaborative Filtering" was coined by the developers of Tapestry [4]. Collaborative Filtering (CF) has been one of several techniques used in recommendation systems. The fundamental assumption of CF is that if user X and user Y have similar behaviors (e.g. rating, buying, watching, or listening similar items), and they will act on other items similarly [5].

Early generation of recommendation systems adopting the collaborative filtering technique, such as *GroupLens*, an online news recommender system [6], exploited the usage of user rating on news articles to provide recommendation based on the calculated similarity values. This technique is called *Memory-based CF Methods*. Many commercial systems, such as http://www.amazon.com, used this technique because it is easy to implement and highly effective [7], [8]. When data are sparse, however; the similarity value will be difficult to calculate. Therefore, the effectiveness of the recommendation performance can be degraded.

Besides collaborative filtering, another technique for the recommendation systems is content-based filtering. In content-based filtering, the system analyzes the textual content and recommends items to be of interest by finding fidelity in the content that a user interacts with. The main difference between collaborative filtering and content-based filtering is that the collaborative filtering only uses the user-item interaction (e.g. rating, buying, watching, listening and so forth) to make predictions and recommendations, while the content-based filtering uses the extracted feature of users and items for recommendation [9].

As mentioned in the previous section, the users of knowledge management systems may spend a lot of time for finding their desired knowledge. Thus, the knowledge management system could benefit from applying

collaborative filtering or content-based filtering to provide automated knowledge dissemination mechanism [10]-[16].

There were, however, few published research papers focused on comparing effectiveness of collaborative filtering and content-based filtering to augment the automatic knowledge dissemination services on the knowledge management system.

Hence, this paper investigated how well the following two recommendation mechanisms namely, item-based recommendation — a collaborative filtering technique, and tag-based recommendation — a content-based filtering technique, can contribute to the task of automated knowledge dissemination.

III. PROPOSED RECOMMENDATION TECHNIQUES

A. Item-Based Recommendation Mechanism

On a knowledge management system, users normally read, search, and share knowledge items based on their interest. When they find interesting knowledge items, they will bookmark them for future references.
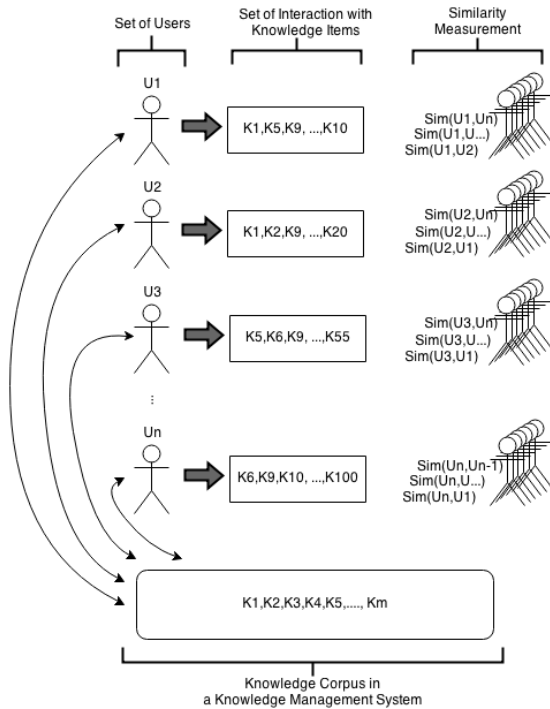

Fig. 1. Concept of item-based recommendation mechanism.

The item-based recommendation mechanism as illustrated in Fig. 1 looks at the pattern of other users in the system and tries to recommend knowledge items based on the matches between a user's knowledge interest and that of others. To implement the item-based recommendation mechanism, there are four main components—*set of users, set of interaction with knowledge items, similarity measurement and knowledge corpus.*

Assuming that there are N_u users and N_k knowledge items in a knowledge management system. Let U be a set of users and contains all users in the knowledge management system; $U = \{u_1, u_2, u_3, ..., u_n\}$, and K be a set of knowledge items and contains all knowledge items in knowledge corpus; $K = \{k_1, k_2, k_3, ..., k_m\}$. 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 knowledge item k_y. Thus, each row, or UK_i, in M_{uk} represents user interaction with knowledge items.

To recommend knowledge items to each user in the knowledge management system, cosine similarity scores between each user will be calculated as illustrated in (1).

$$Sim(A, B) = \frac{\sum_{j=1}^{t} A_j \cdot B_j}{\sqrt{\sum_{j=1}^{t} A_j^2 \cdot \sum_{j=1}^{t} B_j^2}} \qquad (1)$$

After all cosine similarity scores between each user are calculated, a group of similar users, who interact with similar knowledge items, can be identified. For this particular study, the users will be in the same group if the similarity score is greater or equal to a predefined threshold ($\alpha = 0.65$). Then, the bookmarked knowledge items of users in the same group will be put together. The item-based recommendation mechanism will then recommend knowledge items to all the users in that group as shown in Fig. 2.
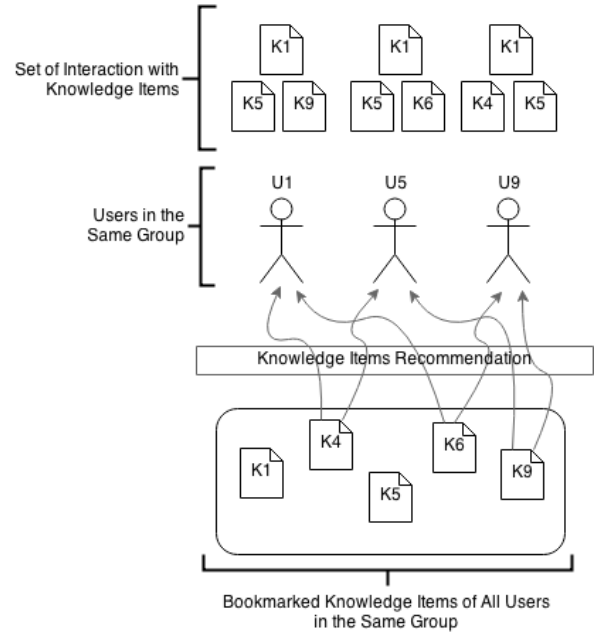

Fig. 2. Example of knowledge items recommendation.

B. Tag-Based Recommendation Mechanism

The tag-based recommendation mechanism as illustrated in Fig. 3, on the other hand, exploits the usage of knowledge tags to recommend knowledge items. To implement the tag-based recommendation mechanism, there are five main components—*set of users, set of interaction with knowledge items, set of users' knowledge tag, set of new/unvisited knowledge items, and set of tags from new/unvisited knowledge items.*

Assume that there is N_u users and N_k knowledge items in a knowledge management system. Let U be a set of users and contains all users in the knowledge management system; $U = \{u_1, u_2, u_3, ..., u_n\}$, K be a set of knowledge items and contains all knowledge items in knowledge corpus; $K = \{k_1, k_2, k_3, ..., k_m\}$, and T be a set of knowledge tags and contains all knowledge tags associated with knowledge items; $T = \{t_1, t_2, t_3, ..., t_p\}$. 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 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 that derived from M_{uk}; $UTK_x = \{< u_x, t_p > | u_x \in U \wedge t_p \in T \wedge M_{uk}(u_x, k_y) = 1\}$ and $NTK_y(u_x)$ be a set of tags from new or unvisited knowledge items that derive from M_{uk}; $NTK_y(u_x) = \{< k_y, t_p > | k_y \in K \wedge t_p \in T \wedge M_{uk}(u_x, k_y) = 0\}$
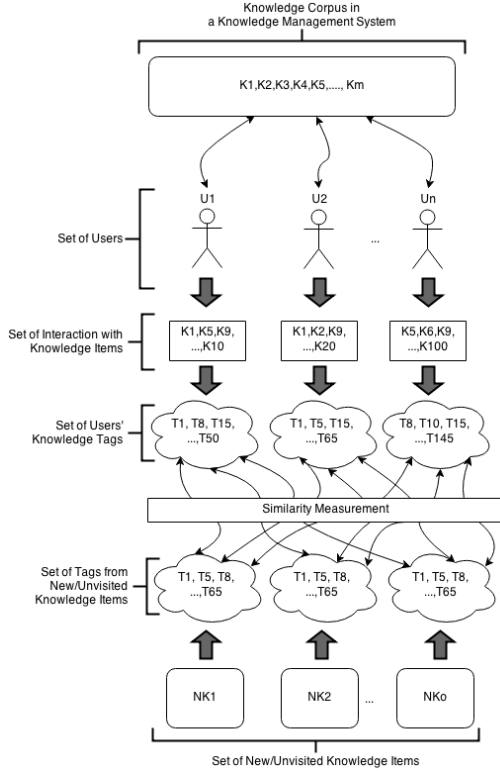

Fig. 3. Concept of tag-based recommendation mechanism.

To recommend knowledge items to each user in the knowledge management system, cosine similarity scores between UTK_x, representing knowledge interest of user x, and $NTK_y(u_x)$, representing content of new or unvisited knowledge items of user x, will be calculated as illustrated in (1).

For this particular study, when the similarity score is greater or equal to a predefined threshold ($\alpha = 0.65$), the tag-based recommendation mechanism will then recommend the knowledge items to that user.

IV. EVALUATION

A. Data Collection

Data from a knowledge management system at Faculty of Information Technology, Dhurakij Pundit University, were used to evaluate two proposed recommendation mechanism. To make it easier to evaluate, the data were crawled between June and December 2013. The data then were loaded into a prototype of knowledge management system embedded with the two proposed recommendation mechanisms.

B. Evaluation Metric

To evaluate the performance of the two proposed recommendation mechanism, a group of knowledge management system users were ask to be participants in this research study. The percentage of accuracy in knowledge items recommendations was used as an evaluation metric illustrated in the (2). Table I shows the confusion matrix used to calculate the percentage of accuracy.

TABLE I: CONFUSION MATRIX

Actual	Predicted	
	Positive	Negative
Positive	TruePositive	FalseNegative
Negative	FalsePositive	TrueNegative

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \qquad (2)$$

where

TP = a number of recommended user-interested knowledge items

TN = a number of not recommended user-uninterested knowledge items

FP = a number of recommended user-uninterested knowledge items

FN = a number of not recommended user-interested knowledge items

C. Experimental Setting

Five faculty members and staff, ten M.Sc. students and five alumni were asked to be participants in this research study. However, as a criterion, participants need to be familiar with the knowledge management system. They were asked to use the prototype system at least three months before the evaluation period. This would guarantee that each participant's interaction with knowledge items could be recorded and used.

When the two recommendation mechanisms finished creating a list of recommended knowledge items, the lists from the two different recommendation mechanisms were merged and all duplicated knowledge items were removed. The percentage of accuracy in knowledge items recommendation for each mechanism was then calculated.

V. RESULT AND DISCUSSION

The main objective of the experiment was to compare the effectiveness of collaborative filtering and content-based filtering in augmenting the automatic knowledge dissemination services in a knowledge management system. The two proposed recommendation mechanisms were item-based recommendation — a collaborative filtering technique, and tag-based recommendation — a content-based filtering technique.

To assess performance, the percentage of accuracy for each mechanism was examined. A higher accuracy indicates a more effective mechanisms. Table II provides a summary of the percentage of accuracy for the proposed mechanisms.

TABLE II: EVALUATION RESULTS OF THE PROPOSED MECHANISMS

	Item-Based Recommendation	Tag-Based Recommendation
Accuracy	78.33%	86.67%

The result suggests that the tag-based recommendation mechanism provides a higher accuracy in knowledge item recommendation task and outperforms the item-based recommendation mechanism. Although the performance of

the tag-based recommendation is at 86.67% accuracy, the causes of error were from inappropriate value of a predefined threshold setting. The knowledge items associated with more knowledge tags tend to have a lower similarity score. In addition, the users who have a broad interest on a variety of topics tend to have more associated knowledge tags. This results in lowering of the similarity score when comparing their knowledge tags to a set of knowledge tags of a new or unvisited knowledge item.

After performing further investigation, the reason that the tag-based recommendation mechanism outperforms the item-based recommendation mechanism is a set of knowledge tags can represent a user's knowledge interest in more details. Using only a set of interaction with knowledge items of a user can roughly represent the user's knowledge interest. It is not guaranteed that users who interact with the same set of knowledge items will always have a similar knowledge interest.

VI. CONCLUSION AND FUTURE WORK

This research paper focused on comparing effectiveness of collaborative filtering and content-based filtering to augment the automatic knowledge dissemination services on the knowledge management system.

The result of experiment showed that the tag-based recommendation (a content-based filtering technique) does benefit a knowledge item recommendation task. However, to develop an effective and efficient knowledge dissemination services, more detailed investigation on the user's knowledge interest representation as well as further analysis on the setting of the predefined threshold needs to be performed.

REFERENCES

[1] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender system handbook," *Recommender Systems Handbook*, Springer, 2011, ch. 1, pp. 1-35.
[2] R. Bruke, "Hybrid web recommender systems," *The Adaptive Web*, Springer, 2007, pp. 377-408.
[3] P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56-58, 1997.
[4] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of ACM*, vol. 35, no. 12, pp. 61-70, 1992.
[5] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, "Eigentaste: a constant time collaborative filtering algorithm," *Information Retrieval*, vol. 4, no. 2, pp. 437-476, 2001.
[6] P. Resnick, N. Iacovou *et al.*, "Grouplens: an open architecture for collaborative filtering of netnews," in *Proc. the ACM Conference on Computer Supported Cooperative Work*, New York, NY, USA, 1994, pp. 175-186.
[7] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," *IEEE Internet Computing*, vol. 7, no. 1, pp. 76–80, 2003.
[8] M. Deshpande and G. Karypis, "Item-based top-N recommendation algorithms," *ACM Transaction on Information Systems*, vol. 22, pp.143–177, 2004.
[9] L. Si and R. Jin, "Flexible mixture model for collaborative filtering," in *Proc. the 20th International Conference on Machine Learning (ICML '03)*, Washington, DC, USA, August 2003, vol. 2, pp. 704–711.
[10] Z. Huang, X. Lu, H. Duan, and C. Zhao, "Collaboration-based medical knowledge recommendation," *Artificial Intelligence in Medicine*, vol. 55, no. 1, pp. 13–24, 2012.
[11] L. Si and R. Jin, "Flexible mixture model for collaborative filtering," in *Proc. the 20th International Conference on Machine Learning (ICML '03)*, vol. 2, Washington, DC, USA, August 2003, pp. 704–711.
[12] H. Li, L. Liu, and C. LV, "Knowledge recommendation services based on knowledge interest groups," in *Proc. the International Conference on Service Systems and Service Management*, France, October 2006, pp. 162–166.
[13] W. Zhao, J. Wang, and G. Lui, "A knowledge recommendation algorithm based on content syndication," in *Proc. the 4th International Conference on Computer Sciences and Convergence Information Technology*, Seoul, Korea, November 2009, pp. 742–745.
[14] 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 *Proc. the 2013 IEEE International Geoscience & Remote Sensing Symposium*, Melbourne, Australia, July 2013, pp. 2204–2207.
[15] A. Vizcaino, J. Portillo-Rodriguez, J. P. Soto, M. Piattini, and O. Kusche, "A recommendation algorithm for knowledge objects based on a trust model," in *Proc. the IEEE International Conference on Research Challenges in Information Science*, Morocco, April 2009, pp. 93–102.
[16] K. Liang, S. Cai, and Q. Zhao, "Context-based knowledge recommendation: A 3-D collaborative filtering appraoch," in *Proc. the 5th IEEE International Conference on Industrial Information*, Vienna, Austria, 2007, pp. 627–632.

Worasit Choochaiwattana was born in 1975. He earned a B.B.A. (marketing) in 1996 from Chulalongkorn University, Thailand, a M.Sc. (technology of information system management) in 2000 from Mahidol University, Thailand, a M.Sc. (computer science) in 2003 from Chulalongkorn University, Thailand, and a Ph.D. (information science) in 2008 from University of Pittsburgh, Pennsylvania, United States.

Currently, he is an assistant professor at Faculty of Information Technology, Dhurakij Pundit University, Thailand. His research interests include search engines, adaptive web and intelligent information systems.