Do recommender systems benefit users? a modeling approach

Chi Ho Yeung

Department of Science and Environmental Studies, The Hong Kong Institute of Education, 10 Lo Ping Road, Taipo, Hong Kong
E-mail: chyeung@ied.edu.hk

Received 16 July 2015
Accepted for publication 22 January 2016
Published 1 April 2016

Abstract. Recommender systems are present in many web applications to guide purchase choices. They increase sales and benefit sellers, but whether they benefit customers by providing relevant products remains less explored. While in many cases the recommended products are relevant to users, in other cases customers may be tempted to purchase the products only because they are recommended. Here we introduce a model to examine the benefit of recommender systems for users, and find that recommendations from the system can be equivalent to random draws if one always follows the recommendations and seldom purchases according to his or her own preference. Nevertheless, with sufficient information about user preferences, recommendations become accurate and an abrupt transition to this accurate regime is observed for some of the studied algorithms. On the other hand, we find that high estimated accuracy indicated by common accuracy metrics is not necessarily equivalent to high real accuracy in matching users with products. This disagreement between estimated and real accuracy serves as an alarm for operators and researchers who evaluate recommender systems merely with accuracy metrics. We tested our model with a real dataset and observed similar behaviors. Finally, a recommendation approach with improved accuracy is suggested. These results imply that recommender systems can benefit users, but the more frequently a user purchases the recommended products, the less relevant the recommended products are in matching user taste.

Keywords: critical phenomena of socio-economic systems, interacting agent models, socio-economic networks, communication, supply and information networks
Do recommender systems benefit users?

1. Introduction

Almost all popular websites employ recommender systems to match users with items [1–4]. For instance, news websites analyze the reading history of individuals and recommend news which matches their interests [5] and online social networks recommend new friends to individuals based on their existing friends [6]. Most commonly, online retailers analyze the purchase history of customers and recommend products to them to increase their own sales [7–9]. These examples show an increasingly crucial role of recommender systems in our daily life, influencing our various choices.

Due to their broad applications, great efforts have been devoted to study recommendation algorithms and to improve their accuracy [4]. Researchers in computer science, mathematics and management science employ various mathematical tools such as the Bayesian approach [11], matrix factorization [10] and latent models [12–14] to derive recommendation algorithms. On the other hand, user preferences are considered in knowledge-based recommender systems [15, 16], while the content of products are considered in content-based recommendation approaches [17, 18]. There are also studies to integrate different methods into a single recommender system, which are called hybrid recommendation approaches [19]. Recently, physicists and complex system scientists started to work in the area and incorporated physical processes such as mass diffusion and heat conduction in recommender systems [20]. The main goal of these studies is limited to recommendation accuracy, but the genuine benefits of recommender systems are less examined.

Although recommender systems have been shown to benefit retailers, whether the recommended products are relevant to customers is less explored [9, 21]. On one hand, many recommendation algorithms are based on product similarity and the recommended

doi:10.1088/1742-5468/2016/04/043401
products may be redundant since they are similar to the already purchased products [20]. On the other hand, instead of specific products which match individual needs, many recommender systems can only recommend popular but potentially irrelevant products [9, 22]. Nevertheless, users may be tempted to purchase the products due to recommendations, and in this case recommender systems benefit sellers but not customers.

In this paper, we introduce a simple model to examine the relevance between the recommended products and the preferences of users. Unlike empirical studies where the true preferences of users are unknown, each user in the model is characterized by a taste and the true recommendation accuracy can be measured. We found that recommendations can be either random or very accurate depending on the frequency with which the users select a product without recommendations. For some algorithms, an abrupt increase in accuracy is observed when this frequency exceeds a threshold. On the other hand, we found that a high accuracy indicated by common evaluation metrics does not necessarily imply a high real accuracy. We tested our model using the MovieLens dataset [23] and observed similar behaviors. Finally, a recommendation approach based on our findings was suggested which outperforms conventional approaches.

2. Model

Specifically, we consider a group of \( N \) users selecting products from a group of \( M \) items. Each user \( i \) and item \( \alpha \) is characterized by one of the \( G \) tastes or genres, denoted by \( g_i \) and \( g_\alpha \), respectively. For instance, in terms of movies, these tastes may correspond to science fiction, romantic comedies or thrillers. Unlike conventional models which represent user tastes and item genres by vectors with various components, we characterize users and items with individual taste groups such that the behavior of recommended systems would not be masked by the complication of taste vectors. This system where each individual user is characterized by a single taste is a simplified picture, and the case where users have multiple tastes are described in section 3.3.

The dynamics of the system proceeds as follows. Before the simulation starts, a taste or genre is randomly assigned to each individual user or item. The tastes are evenly distributed such that there is an equal number of \( N/G \) users and \( M/G \) items in each taste group. Each user \( i \) then randomly collects \( k_i \) items regardless of their genres before the start of the simulation. During the simulation, a user \( i \) is randomly drawn at each time step. A fraction \( f_{\text{sel}} \) of the time, user \( i \) randomly chooses an un-collected product which matches his/her own taste without using the recommender system. This is the conventional way to purchase a product and we call \( f_{\text{sel}} \) the frequency of deliberate selection. On the other hand, a fraction \( 1 - f_{\text{sel}} \) of the time, user \( i \) buys a product following the recommendations from the recommender system. In both cases, a product in his/her collection is randomly removed since all products are assumed to be consumable and can be brought and consumed more than once. The total number of products collected by user \( i \) remains constant at \( k_i \), which simplifies our model as network growth is not required and \( N \) and \( M \) remain constant. The above procedures are repeated a large number of times per user.

We remark that the recommender system has no direct knowledge of user taste and product genre, it can only infer user preferences through his/her purchase history. Since
$f_{sel}$ is the frequency a user makes purchases in the absence of recommender systems, on average at least $f_{sel}$ of the purchases of user $i$ must match his/her taste; $f_{sel}$ is thus proportional to the amount of available hints the recommender systems can exploit. We further define recommendation accuracy $A_{rec}$ to be the fraction of recommended products which match the taste of the user (i.e. in the same taste group with the user), and our goal is to examine $A_{rec}$ to reveal the benefit of recommender systems to users.

For simplicity, we employ the common item-based collaborative filtering (ICF) \cite{24} to be the recommendation algorithm in our model. ICF provides personalized recommendations to users by computing similarity between their purchased products with other products. We first denote the similarity between items $\alpha$ and $\beta$ at time $t$ to be $s_{\alpha \beta}(t)$. As shown by previous studies \cite{24}, the performance of the algorithm is strongly dependent on the definition of similarity. To show that our results are relevant to different recommendation algorithms, we will employ two definitions of similarity, namely the common neighbor (CN) similarity, given by

$$s_{\alpha \beta}^{(CN)}(t) = \sum_{i=0}^{N} a_{\alpha i}(t) a_{\beta i}(t),$$

and the cosine similarity \cite{24}, given by

$$s_{\alpha \beta}^{(\text{cosine})}(t) = \frac{1}{\sqrt{k_{\alpha} k_{\beta}}} \sum_{i=0}^{N} a_{\alpha i}(t) a_{\beta i}(t).$$

The adjacency variable $a_{\alpha i}(t) = 1$ if item $\alpha$ is collected by user $i$ at time $t$, and otherwise $a_{\alpha i}(t) = 0$. The recommendation score $r_{\alpha i}(t)$ of product $\alpha$ for user $i$ at time $t$ is given by

$$r_{\alpha i}(t) = \sum_{\beta=1}^{M} a_{\beta i}(t) s_{\alpha \beta}(t) = \sum_{\beta \in C_i(t)} s_{\alpha \beta}(t),$$

where $C_i(t)$ is the set of products collected by user $i$ at time $t$. Finally, the product with the highest score not yet collected by the user is recommended. Although only ICF is employed as the recommendation algorithm in the model, other recommendation algorithms such as those involving matrix factorization \cite{10} and latent variables \cite{12–14} can also be incorporated and studied via the model.

Our main goal is to employ the above model to examine the benefit of recommender systems on users. Based on the results, we will further suggest a simple method which is biased towards deliberately selected products to improve recommendation results. The simple method can be applied by real online retailers, given that the retailers can identify products deliberately selected by a user, for instance, by checking the purchases against the recommendation list a user receives during the purchase.

3. Results

3.1. Random versus accurate recommendations

To examine the benefit of recommender system to users, we first study the dependence of recommendation accuracy $A_{rec}$ on the frequency $f_{sel}$ of deliberate selection in
the model. The higher the value of $f_{sel}$, the more often the user chooses a product of a matching taste without recommendation, and the more information for the recommender system to exploit. If recommender systems work perfectly, $A_{rec} = 100\% = 1$ whenever $f_{sel} > 0$ as there exists non-zero information about user tastes in the dataset; on the other hand, if recommender systems do not work at all, recommendations are always random, and $A_{rec} = 1/G$ independent of $f_{sel}$.

As shown in figure 1, the recommendation accuracy falls between the two extreme cases. CN similarity is employed in figure 1(a), and $A_{rec} \approx 1/G$ which corresponds to the case of random recommendations when $f_{sel}$ is less than a threshold. This implies that at small $f_{sel}$, the performance of the employed recommendation algorithm is far from that of an ideal recommendation algorithm which is able to identify the smallest amount of genuine information to derive relevant recommendations. When $f_{sel}$ increases beyond a threshold, recommendation accuracy increases abruptly to $A_{rec} = 1$, which corresponds to the case of perfect recommendation. As shown in figure 1(b), cosine similarity is employed and a similar dependence of $A_{rec}$ on $f_{sel}$ is observed, though the transition between the two phases is more gentle. We remark that $A_{rec} = 1$ is an artifact of the model since each user and product is categorized by only one taste, and after users and products of the same taste formed an isolated bipartite cluster, only products within the cluster are recommended and lead to a persistent perfect accuracy. The same is not observed in section 3.3 when each user is characterized by two taste groups.

The accuracy $A_{rec}$ is also dependent on the number of taste groups $G$. Intuitively, the threshold value for perfect recommendation should decrease with $G$, since it seems to be easier to identify an item with the correct taste out of a smaller number of taste groups. However, simulated results in both figures 1(a) and (b) show that the threshold value increases when $G$ decreases. It is because users collect products of both relevant and irrelevant taste; when $G$ is small, the irrelevant products belong to a small number of taste groups, and there exists a strong connection between users and each irrelevant taste group, making it difficult for the recommender system to identify these false connections. In short, the more diverse and distinct the users and products, the fewer hints are required to provide correct recommendations.

Other than the number of taste groups, recommendation accuracy also depends on the number of items collected by each user. For simplicity, all users collect the same number of items, i.e. $k_i = k$ for $\forall i$. As shown in figures 2(a) and (b), perfect recommendation is more difficult to achieve for cases with larger $k$, where the stronger connection between users and irrelevant taste groups is again the reason. These results imply that when users collect a large number of products, false connections exist and may impact negatively on the recommender system. Indeed, true connections also increase with $k$, but $A_{rec}$ decreasing with increasing $k$ may show that false connections have a larger impact on decreasing recommendation accuracy compared to that of true connections to increase recommendation accuracy. This conjecture is consistent with figure 1, where $A_{rec}$ increases only when $f_{sel} > 1/G$, i.e. when the true connections are proportionately more than the false connections from other taste groups. Hence, instead of drawing recommendations based on all the available data, an algorithm which effectively eliminates the false connections may lead to a high recommendation accuracy.

The above results suggest that recommender systems may provide irrelevant recommendations when users do not provide sufficient information about their taste. On the
Do recommender systems benefit users?

Other hand, given sufficient hints, recommender systems utilize the information well in order to match users with products. The amount of hints required for accurate recommendation is different for different algorithms and systems.

3.2. Estimated accuracy versus real accuracy

In real systems, since the real preference of users is unknown, there is no way to measure the real recommendation accuracy. Various metrics are thus introduced to evaluate recommendation accuracy. Nevertheless, whether these metrics correctly measure

\[ A_{rec} = \frac{1}{N} \sum_{i=1}^{N} I \left( x_i = \text{predicted} \right) \]

\[ \text{Common Neighbor Similarity} \]

\[ \text{Cosine Similarity} \]

**Figure 1.** The accuracy $A_{rec}$ of the recommender system as a function of $f_{sel}$ for different number of taste groups $G$. The simulation results were obtained with $N = 2000$ users and $M = 100$ products. Each user collects $k = 7$ products and is updated $1 \times 10^5$ times. Each data point was averaged over 50 instances. The CN similarity in equation (1) and the cosine similarity in equation (2) were employed in (a) and (b) respectively.

**Figure 2.** The accuracy $A_{rec}$ of the recommender system as a function of $f_{sel}$ for different values of $k$, the number of products collected per user. The simulation results were obtained with $N = 2000$, $M = 100$ and $G = 10$. Each user was updated $1 \times 10^5$ times, and each data point was averaged over 50 instances. The CN similarity in equation (1) and the cosine similarity in equation (2) were employed in (a) and (b), respectively.
real accuracy is questionable. Since user taste and product genre are defined in our model, we can compare the accuracy measured by these metrics with real accuracy.

One common metric to evaluate recommendation accuracy is AUC, i.e. the area under the receiver operating curve (ROC). When recommendations are made for user i, AUC is computed as the probability that a correct product \( \alpha \) is ranked higher than an arbitrary product \( \gamma \), given by

\[
AUC_{\alpha \gamma} = \frac{n(r_\gamma < r_\alpha) + 0.5n(r_\gamma = r_\alpha)}{M - k_i}
\]

where \( n(r_\gamma < r_\alpha) \) is the number of products with score \( r_\gamma \) lower than that of the correct product; \( n(r_\gamma = r_\alpha) \) is the number of items with scores equal to that of the correct item. Based on the way to define correct predictions, we compute two AUC measures—(i) the conventional estimated AUC_{est}, obtained by dividing the dataset into a training set and a probe set; links in the probe set are removed and recommendations are considered to be correct if their existence are predicted by the algorithm; and (ii) the real AUC_{real} which quantifies the accuracy of the algorithm in matching products with users in the same taste group.

We remark that other common metrics in the computer science literature such as Precision and Recall\,\[^{25,26}\] can also be used to evaluate recommendation accuracy. Similar to AUC, two different versions of the above metrics can be defined depending on the way to define correct predictions, i.e. either to be the items in a probe set or the items with a matching taste. In this paper, we only study AUC since it is a popular metric and is less influenced by the number of correct products, which can differ greatly in the two definitions.

The dependence of AUC_{est} and AUC_{real} on \( f_{sel} \) is shown in figure 3. As we can see, AUC_{real} \( \approx 0.5 \) when \( f_{sel} \) is small since recommendations are random (see figure 1) such that the products of a matching taste are randomly ranked in the recommendation list. However, AUC_{est} is much higher and is not consistent with AUC_{real}. The reason for a large AUC_{est} at small \( f_{sel} \) is the frequent application of recommender systems, such that user purchases are strongly influenced by the algorithms regardless of their true preference. In this case, products which do not match their preference but are consistent with the algorithms are also collected by the users. This can be considered as a reinforcement by the recommender system in user choices, which favors the evaluation by AUC_{est} using a random probe set, and lead to a high AUC_{est} even if random recommendations are indeed provided.

When \( f_{sel} \) increases, AUC_{est} decreases since the user–product relations become less influenced by the recommender system. At the same time, AUC_{real} increases since more information about user tastes is present. We remark that although \( A_{rec} \approx 1/G \) when \( f_{sel} \) is smaller than the threshold (see figure 1(a)), the corresponding AUC_{real} is increasing in the same regime. Finally, AUC_{real} and AUC_{est} become consistent when \( f_{sel} \) further increases and the system achieves perfect recommendation.

The above results imply that the conventional evaluation of recommendation accuracy may not necessarily reflect the true accuracy. Indeed, AUC_{est} may overestimate the accuracy of the algorithm, especially in cases where users rely frequently on the recommender system and do not reveal their own taste by deliberately selecting products. This serves as an alarm for researchers and operators of recommender systems who
evaluate recommender systems merely with accuracy metrics. Alternative evaluations such as user satisfaction towards the recommended products are therefore necessary to supplement conventional accuracy metrics to quantify the benefit of recommender systems for users.

Nevertheless, we remark that unlike users in the model who always accept the recommended products, real users do not always follow recommendations provided to them. This may suppress the reinforcing impact of the recommender systems on user choices, and thus the difference between the true and the estimated accuracy.

3.3. Users with multiple tastes

Ordinary users usually have more than one interest; for instance, a user may be interested in both science fiction and action movies. To model this scenario, we assume that each user is characterized by two tastes, which we denote by taste 1 and taste 2. Similar to the previous case, \( f_{\text{sel}} \) of the time, the user selects a product in the absence of recommender systems; otherwise, the recommendation algorithm is applied. When a user selects a product, \( f_1 \) of the selected products are in taste 1 and the rest are in taste 2. To simplify the model, we only study cases with large \( f_{\text{sel}} \), with which perfect recommendation is achieved in the original single-taste system.

Since a fraction \( f_1 \) of the selected products of the user are in taste 1, the ratio \( f_1/(1-f_1) \) corresponds to his/her preference between the two tastes. If optimal recommendations are achieved, \( f_1 \) of the recommended products should be in taste 1 and \( 1-f_1 \) of them should be in taste 2. Nevertheless, as shown in figure 4, the fraction \( A_{\text{rec}}^{(1)} \) of the recommended products in taste 1 does not coincide with the optimal line \( A_{\text{rec}}^{(1)} = f_1 \). For instance, when \( f_1 \) is small, the recommendations are mainly in taste 2. It leads to a sub-optimal state which underrepresents the minority taste, i.e. taste 1 when \( f_1 < 0.5 \), among the recommended products. Similarly, taste 2 is underrepresented when \( f_1 > 0.5 \). As we can see in figure 4, the difference between \( A_{\text{rec}}^{(1)} \) and \( f_1 \) is larger when \( k \) is larger. This implies a larger difficulty for the recommender system in identifying a

Figure 3. The two different AUC measures, AUC\(_{\text{est}}\) and AUC\(_{\text{real}}\), as a function of \( f_{\text{sel}} \) obtained by ICF with CN similarity and cosine similarity (inset) in systems with \( N = 2000, M = 100, k = 3 \) and \( G = 10 \).
Do recommender systems benefit users?

secondary taste if the user–product connections are denser. The results by employing the CN similarity and the cosine similarity are almost identical.

On the other hand, one may expect a perfect recommendation regime at \( f_{\text{sel}}^{*} \), where \( f_{\text{sel}}^{*} \) denotes the threshold value in the corresponding single-taste scenarios, i.e. the smallest \( f_{\text{sel}} \) at which the system achieves perfect recommendation (see figure 1). For the system parameters employed in figure 4, \( f_{\text{sel}}^{*} \approx 0.73 \), but perfect recommendations in taste 1 are not achieved with \( f_{1} > f_{\text{sel}}^{*} \) (indicated by the dotted line in figure 4) due to the presence of taste 2. This again suggests that the presence of false connections, i.e. the taste 2 connections against achieving prefect recommendations in taste 1, may have a strong negative impact on the recommendation accuracy.

3.4. Tests with empirical datasets

To examine the emergence of the above phenomenon in real recommender systems, we incorporate our model with a real dataset obtained from MovieLens [23]. Since user taste and product genre are unknown in real systems, we only consider a movie to be a correct recommendation, which we call a correct movie, only if it was collected by the user in the original dataset and has received a rating of at least 3 or above from the user (in a scale from 1 to 5). Similar to our model, \( f_{\text{sel}} \) of the time, a user deliberately selects a correct movie and otherwise the recommendation algorithm is applied. To compute the recommendation accuracy, for all users who rated at least two movies with a score of 3 or above, we set their degree to be \( k_{i} - 1 \) such that an uncollected correct movie always exists. As in the previous simulations, a user randomly removes one of his/her collected movies when he/she obtains a new movie; the system is then repeatedly updated.

As shown in figure 5(a), the accuracy \( A_{\text{rec}} \) obtained by both similarity definitions starts at a low value and increases with \( f_{\text{sel}} \). Nevertheless, it does not show an abrupt jump to a high value similar to previous simulations, but a plateau at small \( f_{\text{sel}} \) and a
small jump at large $f_{sel}$ are observed in the case with cosine similarity. These results again suggest that sufficient hints about user taste are essential for the system to obtain accurate recommendations. When $f_{sel}$ approaches 1, $A_{rec}$ decreases since users have collected most of the correct movies through deliberate selection and it becomes more difficult for the recommender system to identify the fewer correct items among all the other items.

Next, we compare $A_{UC_{est}}$ and $A_{UC_{real}}$ obtained on the real dataset. While $A_{UC_{est}}$ is again computed by randomly dividing the current user–item relations into a training set and a probe set, $A_{UC_{real}}$ is defined as the accuracy of the algorithm in identifying a correct movie not found in the user’s current collection. As shown in figure 5(b), the dependence of $A_{UC_{est}}$ and $A_{UC_{real}}$ on $f_{sel}$ is similar to that observed from the previous simulations. When $f_{sel}$ is small, the conventional AUC metric overestimates the accuracy of the recommender system. Especially, $A_{UC_{est}}$ is highest when $A_{UC_{real}}$ is lowest, and $A_{UC_{est}} = A_{UC_{real}}$ only when $f_{sel} = 1$. This suggests that conventional metrics may again overestimate recommendation accuracy in real systems.

We remark that the general results in figures 5(a) and (b) are consistent with those in sections 3.1 and 3.2, i.e. (i) the increasing recommendation accuracy with increasing $f_{sel}$, and (ii) the disagreement between the estimated and real accuracy at small $f_{sel}$, but there are obvious differences between figures 1 and 5(a). The causes for the differences may include the simplified characterization of taste in the model as well as the absence of identification of deliberately selected products in the static dataset we studied. In real systems, it is possible to check the purchased product against the list of recommended products a user receives, so as to distinguish the deliberately selections from the recommended products. Such checking procedures would facilitate the implementation of an improved recommendation approach suggested in the following section. Nevertheless, we have no access to such data and can only assume products with a high rating are deliberately selected by the users, which constitutes a difference between the simulated model and the real data.
Based on the previous results, we slightly modify the ICF algorithm to improve the recommendation accuracy. The rationale is simple—since the products deliberately selected by users themselves usually match their own taste, we simply give a higher weight to these products during the computation of recommendation scores, by modifying the adjacency variable $a_{si}(t)$ as follows:

$$a_{si}(t) = \begin{cases} 
0 & \text{if } \alpha \notin C_i(t), \\
1 & \text{if } \alpha \in C_i(t) \text{ via recommendation,} \\
1 + b & \text{if } \alpha \in C_i(t) \text{ via selection,}
\end{cases}$$

(5)

where $C_i(t)$ is again the set of products collected by user $i$ at time $t$, and $b > 0$ is the bias on products collected via deliberate selection. The recommendation score of an item is then computed by the same formula (equation (3)). The recommendation accuracy obtained by the modified algorithm is compared to that of the original algorithm in figure 6. As we can see from figures 6(a) and (b), perfect recommendations are achieved at a smaller $f_{sel}$ when selected products are weighed more in the algorithm. The larger the value of the bias $b$, the smaller the value of $f_{sel}$ beyond which perfect
recommendations are achieved, or equivalently, the stronger the bias, the smaller the amount of genuine information required for the system to achieve accurate recommendations. Similar results are observed with the MovieLens datasets as shown in figures 6(c) and (d), but the behaviors with increasing $b$ are more nontrivial than those observed in the model, as the case with $b = 1$ may outperform the case with $b = 2$ for some values of $f_{sel}$. In general, these results imply that products deliberately chosen by users are essential information to improve recommendation accuracy, and the more the algorithm is biased towards the deliberately selected products, the less information is required for deriving accurate recommendations.

4. Conclusion

To reveal the benefit of recommender systems for users, we studied a simple model where users either choose their own products or follow recommendations from the system. Our results show that recommendations may be equivalent to random draws if users always follow recommendations and do not reveal their own taste by deliberately selecting products. On the other hand, if sufficient information about their taste is present, recommendation systems are able to achieve high accuracy in matching appropriate products to users. For some recommendation algorithms, the increase in accuracy is abrupt once the amount of available information exceeds a threshold. These results imply that recommender systems can benefit users, but relying too strongly on the system may render the system ineffective.

On the other hand, our study reveals the difficulties to obtain a realistic and accurate evaluation of recommendation accuracy. Since the preferences of real users are unknown, evaluation of recommendation algorithms usually involves removing a set of existing data and then estimates the accuracy of the algorithms by their success to retrieve the removed set. Our results show that such metrics do not necessarily reflect and may overestimate the true accuracy of the algorithm. This is because the choice of products collected by users was previously influenced by the recommendation algorithms; the presence of these products may not reflect their true preference and may favor evaluation by conventional accuracy metrics. Disagreement between the estimated and real accuracy was observed in simulations with both generated network and a real dataset. These results imply that a high recommendation accuracy indicated by conventional metrics may not necessarily imply a benefit for users. Alternative evaluations such as user satisfaction towards recommended products are necessary to supplement these metrics in order to quantify the effectiveness of the recommender systems.

Acknowledgment

I acknowledge support from the Internal Research Grant RG 71/2013-2014R and the Dean’s Research Fund 04115 ECR-5 2015 of the Hong Kong Institute of Education.
References

[1] Resnick P and Varian H R 1997 Commun. ACM 40 56–8
[2] Schafer J B, Konstan J and Riedl J 1999 Proc. of the 1st ACM Conf. on Electronic Commerce pp 158–66
[3] Ricci F, Rokach L, Shapira B and Kantor P B 2011 Recommender Systems Handbook (New York: Springer)
[4] Lu L, Medo M, Yeung C H, Zhang Y C, Zhang Z K and Zhou T 2012 Phys. Rep. 519 1–49
[5] Liu J, Pedersen E and Dolan P 2010 Proc. of the 15th Int. Conf. on Intelligent User Interfaces pp 31–40
[6] Chen J, Geyer W, Dugan C, Muller M and Guy I 2009 Proc. of the SIGCHI Conference on Human Factors in Computing Systems pp 201–10
[7] Brynjolfsson E, Hu Y and Smith M D 2003 Manage. Sci. 49 1580–96
[8] Chen P Y, Wu S Y and Yoon J 2004 Proc. of the 25th Int. Conf. on Information Systems p 58
[9] Fleder D and Hosanagar K 2009 Manage. Sci. 55 697–712
[10] Takács G, Pilászy I, Németh B and Tikk D 2007 Proc. KDD Cup Workshop at SIGKDD (San Jose, California) p 22–30
[11] Ungar L and Foster D 1998 Proc. of the Conf. of Automated Learning and Discovery
[12] Hofmann T 2004 Latent semantic models for collaborative filtering ACM Trans. Inf. Syst. 22 89–115
[13] Blei D L, Ng A Y and Jordan M I 2003 J. Mach. Learn. Res. 3 993–1022
[14] Griffiths T L and Steyvers M 2004 Proc. Natl Acad. USA 101 5228–35
[15] Burke R 2000 Encyclopedia Libr. Inf. Sci. 69 180–200
[16] Felfernig A and Burke R 2008 ACM Int. Conf. on Electronic Commerce Article 3, pp 1–10
[17] Mooney R J and Roy L 2000 Proc. of the 5th ACM Conf. on Digital Libraries pp 195–204
[18] Pazzani M J and Billsus D 2007 Content-based recommendation systems The Adaptive Web (Berlin: Springer) pp 325–41
[19] Adomavicius G and Tuzhilin A 2005 IEEE Trans. Knowl. Data Eng. 17 734–49
[20] Zhou T, Kuscsik Z, Liu J G, Medo M, Wakeling J R and Zhang Y C 2010 Proc. Natl Acad. Sci. USA 107 4511–5
[21] Liang T P, Lai H J and Ku Y C 2006 J. Manage. Inf. Sys. 23 45–70
[22] Zeng A, Yeung C H, Shang M S and Zhang Y C 2012 Europhys Lett. 97 18005
[23] MovieLens http://grouplens.org/datasets/movielens/ (retrieved May 2015)
[24] Sarwar B, Karypis G, Konstan J and Riedl J 2001 Proc. of the 10th Int. Conf. on World Wide Web pp 285–95
[25] Herlocker J L, Konstan J A, Terveen K, Riedl J T 2004 ACM Trans. Inf. Syst. 22 5–53
[26] Swets J A 1963 Science 141 245–50