Connecting with the Collective: Self-contained Reranking for Collaborative Recommendation

Yue Shi, Martha Larson, and Alan Hanjalic
Multimedia Information Retrieval Lab
Delft University of Technology
Delft, The Netherlands
{y.shi, m.a.larson, a.hanjalic}@tudelft.nl

ABSTRACT
Collaborative recommendation (CR) approaches have proven effective for the Top-N recommendation task. We introduce a novel approach, Rerank-CR, that further improves the Top-N results of an arbitrary CR algorithm using a post-processing step involving Bayesian reranking. The defining characteristic of Rerank-CR is that reranking is self contained, meaning that it requires no external resources, but rather makes use of information derivable from the original user-item matrix. Rerank-CR achieves top performance when used for incorporating collection-level information reflecting global tendencies as constraints into conventional CR, which we refer to as 'connecting with the collective'. Because information about the preferences of the collective is derived directly from the dataset, Rerank-CR has no need of an explicit model of rating styles within a certain community. Further, it is possible to adapt the domain of application (e.g., change to a different cultural setting) without explicit intervention. We evaluate Rerank-CR with experiments that demonstrate the ability of the basic Rerank-CR concept to improve an initial Top-N recommendation list and also the additional improvement achieved by 'multimodal' Rerank-CR, which integrates the collective modality. Additional experiments confirm that the performance of Rerank-CR is significant across different datasets.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *Information Filtering*

General Terms
Algorithms, Performance, Experimentation

Keywords
Recommender systems, recommendation, collaborative filtering, reranking, user-item matrix

1. INTRODUCTION
Collaborative recommendation (CR) techniques produce predictions by exploiting user ratings or usage/purchase statistics encoded in a user-item matrix. CR approaches achieve good performance on the Top-N recommendation task, i.e., the task of recommending new items that fit a user's taste and needs [1][3][6]. We propose an algorithm, Rerank-CR, that further improves the Top-N recommendations generated by a conventional CR system by exploiting additional information derived from the original user-item matrix that is otherwise overlooked. The key strength of Rerank-CR is its ability to create a balanced combination of collaborative information used by CR techniques and global, collection-level information reflecting collective tendencies. This strength makes the system particularly robust to domain variation. For example, it is known that users in different cultural settings display different rating behavior [10]. Instead of capturing this difference using an explicit model, Rerank-CR leverages patterns present in the user-item matrix, making it easy to change domains without explicit intervention or costly adaptation.

Reranking is a post-processing technique that operates on an initial results list. The goal of reranking is to reorder the list such that highly relevant items are encouraged to move upwards and less relevant items gravitate towards the bottom. In information retrieval, reranking is a valuable tool for integrating evidence from external information sources, e.g., [8]. However, reranking can also be used to exploit new perspectives, or 'views' on items that can be derived independently of external resources. Such *self-contained reranking* approaches are currently coming into their own in the area of multimedia retrieval, where multiple views correspond to *modalities* (e.g., speech, visual) and reranking improves performance by way of multimodal fusion [2][7][20].

Past attempts to improve CR via reranking [4][17] are not only limited in number, but also focus exclusively on using reranking to integrate external information sources. Rerank-CR is set apart from previous work by its use of self-contained reranking that draws only on information originally present in the user-item matrix. Specifically, Rerank-CR reorders the Top-N results list by leveraging indirect matches between recommended items and the user profile. An item is encouraged to climb in the list if similarity with high-ranked items suggests that it is a (indirect) match with the user profile. Conversely, an item falls in the list if similarity with low-rated items suggests that it is a mismatch.

Under the basic Rerank-CR approach, the similarity metric used to generate the initial results list is also employed as the reranking modality. The basic approach aims to exploit both direct and indirect similarities with the user profile. The true strength of Rerank-CR, however, lies not in the basic approach but in the 'multimodal' approach. Here, Rerank-CR exploits *collective* views of items – representations of items generated using collection-level information derived from the user-item matrix. Rerank-CR treats collective views as additional modalities, and for this reason the improvement that it achieves in recommendation performance compared to the basic (monomodal) Rerank-CR approach can be attributed to a mul-

CMM'10, October 29, 2010, Firenze, Italy.

timodal fusion effect. Specifically, we focus on popularity-based modalities, motivated by recent work analyzing NetFlix user rating patterns [19]. As with the basic approach, the multimodal approach exploits both direct and indirect similarities. Critically, however, it achieves performance improvement by *connecting with the collective*, i.e., incorporating collection-level information into conventional CR.

Connecting with the collective requires a careful balance between the influence of the personal preferences of the user as expressed by the user profile and the influence of collection-level tendencies. Reranking is particularly well suited to maintain this balance since the final output is constrained by the content of the initial results list, which is highly personalized by virtue of having been generated using the user's profile. This constraint prevents the incorporation of collection-level information from leading to over-generation of recommendations. Additionally, since reranking requires two steps, recommendation generation and then post-processing, the overall Rerank-CR framework is highly flexible, e.g., allowing, in real world application scenarios, the personalization step (the initial CR output) to scale-independently of the integration of collection-level information (connecting with the collective). Given these advantages it is surprising that little work has been devoted to the post-processing of CR results. With Rerank-CR we intend to contribute to research in this little-explored direction. In the reranking step, we choose to use Bayesian reranking as applied to multimodal video search in [20] since it provides a transparent balance between faithfulness to the original ranking and influence of additional modalities.

This paper assesses the viability of the Rerank-CR algorithm and thoroughly evaluates its ability to improve CR results by using self-contained reranking. Our experiments address the following research questions: (Q1) *Is there a potential for improving an initial Top-N recommendation list by relying on information internal to the user-item matrix?* (Q2) *How much extra improvement can be achieved by 'multimodal' Rerank-CR?* (Q3) *Is Rerank-CR effective for different application scenarios, i.e., different datasets?* Specific issues concerning the optimization of existing CR techniques and optimal formulation of additional modalities for reranking are touched on as necessary, but lie largely outside the paper's topical focus.

The remainder of the paper is structured as follows. In the next section, we summarize related work and position our approach with respect to it. Then, we present the Rerank-CR algorithm and validate it experimentally. The last section sums up the key aspects of the proposed approach and briefly addresses the possibilities for future work.

2. RELATED WORK

This section summarizes the existing approaches to collaborative recommendation and reranking and explains the main principles underlying the proposed Rerank-CR concept.

2.1 Collaborative Recommendation

Collaborative filtering makes use of social information, including usage patterns and user-contributed ratings, in order to predict user ratings [5][14][16] or to generate recommendation lists [3][9][11][15]. Our proposed Rerank-CR approach addresses the latter task, which is also referred to as collaborative Top-N recommendation (i.e., identifying the N most relevant items to be recommended to a user), or simply as collaborative recommendation (CR). CR approaches fall into two categories, namely memory-based and model-based CR [1], although sometimes a third combined category is also identified. Here, we discuss memory-based CR, since in this paper, Rerank-CR is used to rerank the recommendation list generated by a memory-based CR technique.

Memory-based techniques do not attempt to abstract away from the data, but rather make use of the entire user-item matrix. A distinction can be made between user-based and item-based approaches. User-based CR views the user for whom the recommendation is being made as a member of a neighborhood containing other users with similar interests. Contributions from the users in the neighborhood are aggregated to generate a recommendation or rating [5][14]. Item-based CR exploits co-occurrences between items in user profiles and recommends items based on similarities in ratings between items in which a user has already displayed an interest and new items [3][11][16]. Since the focus in this paper is exploiting collective-level information for reranking recommendation, rather than recommendation techniques, we choose the well-known item-based CR (Item-CR) as the recommendation technique used to test Rerank-CR. Note that a limitation is imposed on which items can be recommended by Item-CR. For an item to be recommended to a user, that user must have at least one other profile item that co-occurs with the recommended item in the profile of at least one other user in the collection.

2.2 Reranking

Reranking provides a convenient and flexible post-processing framework, which has been widely applied to improve the retrieval results lists. As mentioned above, a reranking approach may rely either on external information sources or on alternate views on the original collection. The latter category of approaches, which we refer to as self-contained reranking, has recently attracted increased attention in the multimedia retrieval community [2][7][20]. A typical video search-result reranking algorithm deploys visual information to rerank results generated by text-based search on speech transcripts. Naturally, some approaches exist that rerank using both alternate views and additional information, notably [13]. Reranking has also been successfully applied to CR tasks. However, as mentioned in Section 1, applications to CR are few [4][17] and differ from Rerank-CR because they use information sources that are external to the user-item matrix and not always available. The reranking approach that is most closely related to our own involves tag recommendation and, in particular, the use of collection-internal knowledge to rerank lists of recommended tags [18].

3. THE RERANK-CR APPROACH

Rerank-CR uses a framework for self-contained reranking to refine the initial recommendation list generated by an arbitrary recommendation technique. As previously mentioned, we chose a Bayesian reranking framework that provides a transparent balance between faithfulness to the original recommendation list and influence of additional information. Random walk reranking pursues similar objectives, as noted in [20], but since its performance is largely comparable, cf. [13], it is unlikely to yield dramatic additional insight and as such is excluded from further consideration in this context. As previously mentioned, Rerank-CR exploits indirect matching effects between the user profile and items in the initial recommendation list via item-to-item similarity. Fig. 1 illustrates the item-to-item similarities for items 4 and 6 from the depicted recommendation list. Thicker lines are used to represent greater similarity between items. The 'indirectness' is due to the fact that the similarity between item i and user profile is inferred via the similarity between this item and other items in the initial recommendation list. A key characteristic of the initial recommendation list is that different item pairs have different levels of item-to-item similarity. The similarity of a given item with items at the top of the initial list serves as evidence that that item should move upwards (as in the case of item 6) and similarity with items towards the bottom of the initial ranking list can be seen as evidence that that item should move downwards (as in the case of item 4).

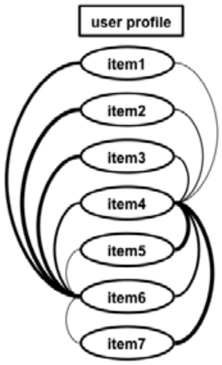

Fig 1: Initial recommendation list (ordered by decreasing similarity to user profile). Item-to-item similarity patterns are illustrated for items 4 and 6 (line thickness reflects similarity).

Why should items that fail to match the user profile when generating the initial recommendation list still be similar to top rated items and merit promotion up the list in the reranking step? One reason may be found in user profile deficiency. If the user profile is deficient in some respect, it may be the case that highly ranked items introduce a profile expansion effect because they provide a better representation of the underlying user preferences than the user profile itself. Another reason may lie in an inadequate similarity comparison. If the original modality provided a less than adequate measurement of the similarity of items with the user profile, the use of additional modalities, which compare items using different 'views,' has the effect of adding information to the system. For these reasons, we expect basic Rerank-CR to improve the initial ranked list and multimodal Rerank-CR to provide an even greater performance improvement.

3.1 Bayesian Reranking

The Bayesian reranking approach [20] takes as input an initial ranked list consisting of a well-ordered set of items represented by their rank scores, $r^{init}=\{r_1 \ldots r_n\}$, organized in descending or ascending order. It yields as output a reordering of the initial list, called the reranked list and denoted by r^{rrk}. The reranked list simultaneously maximizes consistency measured in terms of item similarity with respect to a specific modality X and faithfulness to the ordering of the initial ranked list. The Bayesian reranking approach is formally represented as:

$$r^{rrk} = \arg\max_{r} P(r \mid X, r^{init}) \tag{1}$$

The objective reranking r is taken to range over all possible reorderings of r^{init}. The formulation of Bayesian reranking is independent of the source of r^{init} and the identity of modality X. Expanding Eq. 1 using Bayes theorem leads to the following expression:

$$P\left(r \mid X, r^{init}\right) \propto P\left(r^{init} \mid r, X\right) \times P(r \mid X) \tag{2}$$

A conditional independence assumption can be imposed on X and r^{init}, given an objective reranked list, r, allowing Eq. 2 to be reformulated as:

$$P\left(r \mid X, r^{init}\right) \propto P\left(r^{init} \mid r, X\right) \times P(r \mid X) \approx P(r^{init} \mid r) \times P(r \mid X) \tag{3}$$

The term $P(r|X)$ is the *conditional prior* and imposes consistency of the reranked list with respect to modality X. The term $P(r^{init}|r)$ is the *likelihood* and ensures bias with respect to r^{init}. Both terms are now discussed in greater detail.

3.1.1 Conditional Prior

The conditional prior is formulated such that the probability of the results list rises to the degree that items similar with respect to modality X occupy similar rank positions. If we chose w_{pq} to denote the similarity between items p and q in r^{init} with respect to the modality X, then this prior can be defined as:

$$P(r \mid X) = \frac{1}{\beta_1}\exp\left(-\frac{1}{2}\sum_{p,q} w_{pq}\left(r_p - r_q\right)^2\right) \tag{4}$$

Here, r_p and r_q are the rank scores of elements p and q, and β_1 is a normalizing constant. As later will be discussed in detail, the choice of w_{pq} varies according to the reranking modality being deployed. For the purpose of implementing Rerank-CR the rank score is chosen to be the normalized rank. Normalized rank is defined as $(N-j+1)/N$, where N is the length of the initial ranked list and j is the rank position of the item (i.e., $j=1$ for the top hit).

3.1.2 Likelihood

The likelihood is formulated as:

$$P(r^{init} \mid r) = \frac{1}{\beta_2}\exp\left(-\alpha \times Dist(r, r^{init})\right) \tag{5}$$

Here, β_2 is a normalizing constant, and α is a scaling parameter, which also provides a tradeoff between the influence of the conditional prior and the likelihood discussed in further detail below. $Dist(r, r^{init})$ is the ranking distance between the initial recommendation list and a candidate reranked list. For the purpose of implementing Rerank-CR, we make use of a point-wise measure:

$$Dist(r, r^{init}) = \sum_{p=1}^{Z}\left(r_p - r_p^{init}\right)^2 \tag{6}$$

Here again, r_p and r_q are the rank scores, in our case, normalized ranks, and Z is the number of initially recommended items for a user.

3.1.3 Optimization

The conditional prior and the likelihood are combined according to Eq. 3 and with the objective to minimize the energy function $E(r)$ w.r.t. r, defined as:

$$E(r_1, r_2, \ldots, r_Z) = \frac{1}{2}\sum_{p=1}^{Z}\sum_{q=1}^{Z} w_{pq}(r_p - r_q)^2 + \alpha\sum_{p=1}^{Z}\left(r_p - r_p^{init}\right)^2 \tag{7}$$

The tradeoff parameter α weights the influence of the initial recommendation relative to the reranking modality. Note that as α approaches positive infinity, the reranked list approaches the initial list. Optimal performance can be achieved by tuning α with respect to specific use cases represented by different datasets or different reranking modalities. While various approaches can be applied to solve the minimization problem in Eq. 7, we use the iterative gradient descent approach in our implementation in order to be able to generate recommendations at reasonable computational cost.

3.2 Initial Recommendation

We test Rerank-CR using initial lists generated by Item-CR. Background for this approach was provided in Section 2 and details are given here.

Item-to-item similarity, denoted as sim_I, is the basic calculation contributing to Item-CR. We adopt the cosine similarity measure, which is commonly used for this purpose [3][11][16]. Assuming a user-item matrix encoding item scores for M users and K items, sim_I between two items p and q, can be expressed as:

$$sim_I(p,q) = \frac{\sum_{m=1}^{M} s_{mp} s_{mq}}{\sqrt{\sum_{m=1}^{M} s_{mp}^2}\sqrt{\sum_{m=1}^{M} s_{mq}^2}} \quad (8)$$

Here, s_{mp} denotes the score associated with item p in user m's profile. The score can be binary, taking on a value of 1 when the user views, buys, downloads or otherwise visits an item. Alternatively, it can be a rating that has been assigned by the user to the item and expresses the user's interest in the item. If a user has not rated an item, its score is 0. The sim_I we use in our experiments is calculated on the basis of ratings.

The conventional Item-CR approach [3][11][16] uses the sim_I from Eq. 8 to calculate a score for new items based on items already existing in the user profile. For a user m, the Item-CR score for a given new item x, is calculated by adding up the similarities over the K items in that user's profile:

$$Score_{Item-CR}(m,x) = \frac{\sum_{j=1}^{K} sim_I(x,j) \times s_{mj}}{\sum_{j=1}^{K} sim_I(x,j)} \quad (9)$$

$Score_{Item-CR}(m,x)$ is a prediction of the preference of user m for item x and can be used either as a predicted rating, or, in our case, to generate a ranked list of recommended items. Items the user has already rated are omitted from the recommendation list. We limit our implementation to only using the 10 most similar items for each new item, a restriction shown to have minimal impact on the results yet memory efficient [3]. Also following [3], we apply normalization of item-based similarity, which is known to improve the initial recommendation.

3.3 Reranking Modalities

Recall that basic (monomodal) Rerank-CR and multimodal Rerank-CR differ in their choices of reranking modality. This section presents the details of those modalities. Different modalities correspond to different choices of w_{pq} in Eq. 4.

3.3.1 Item-based Similarity.

Basic (monomodal)Rerank-CR uses the item-based similarity in Eq. (8), setting w_{pq} as:

$$w_{pq}^{I} = sim_I(p,q) \quad (10)$$

Since this is the similarity used to generate the initial recommendation result, it does not introduce a new modality.

3.3.2 Popularity-based Similarity.

Multimodal Rerank-CR makes use of an additional modality encoding collection-level information derived from the user-item matrix. This modality represents a collective view on each item. It is intended to capture collection-wide patterns of the sort that are 'invisible' to conventional collaborative recommendation approaches, which aggregate contributions from individual items or users based on pairwise similarity. In particular, we conjecture that the collection contains important information about the collective tendencies of users to vary along a spectrum of preference novelty. At one end are those users preferring hits (the most popular items) and at the other end are those users preferring niche items (the least popular items). Our assumption concerning this user pattern is supported in the findings of [19], which reports on an analysis of a large movie/user dataset from NetFlix. The user-level analysis of the NetFlix data turned up the observations that niche viewers and hit viewers are characterized by distinctive patterns of rating. We endow Rerank-CR with the capacity to cater separately for different categories of viewer along the hit-to-niche spectrum by adopting popularity-based similarities as additional modalities. Our first popularity-based modality, Pop_{Nr}, equates the popularity of an item with *Nr*(.), the raw number of users who have rated it (and thus, we assume, have viewed it).

Additionally, we investigate two other related popularity-based similarities, inspired by the observation that rating behavior also varies along the hit-to-niche spectrum. The modality Pop_{Ar}, makes use of the average rating, *Ar*(.), assigned to an item by all users in the collection. Naturally, we expect that the average rating of a movie carries information concerning its overall reception among viewers. However, it also contains implicit information arising from the tendency of 'hit watchers' to be undiscriminating and prone to assign high ratings and 'niche seekers' to be well-informed, critical and prone to assign low ratings. The tendency of occasional consumers to assign high ratings to popular products has been dubbed 'ill-informed good will' [21]. Exploratory analysis confirmed that the patterns in the datasets we choose for experimentation also reflect this trend. As an alternative to the average rating assigned to an item by all users, *Ar*(.), we also make use of the number of users who have assigned the item the highest possible rating, *NHr*(.) in our final popularity based modality Pop_{NHr}. These three popularity-based views of items can be calculated using information in the user-item matrix. We model them in a straightforward manner using an exponential function to encode them into appropriately-scaled alternatives for w_{pq}, the reranking modality:

$$w_{pq}^{Nr} = \exp\left(-\left\|Nr(p) - Nr(q)\right\|\right) \quad (11)$$

$$w_{pq}^{Ar} = \exp\left(-\left\|Ar(p) - Ar(q)\right\|\right) \quad (12)$$

$$w_{pq}^{NHr} = \exp\left(-\left\|NHr(p) - NHr(q)\right\|\right) \quad (13)$$

Other encodings are conceivable; however, these three provide us with range of possibilities necessary to explore the ability of Rerank-CR to make use of modalities derived from collection-level information encoded in the user-item matrix.

4. EXPERIMENTS AND EVALUATION

In this section, we provide details on our experimental setup and the results of the experiments carried out to evaluate the Rerank-CR approach. The experiments test different variants of Rerank-CR and involve using Eq. 7 to combine an initial ranking list, r^{init}, generated by Item-CR, with different reranking modalities.

4.1 Experimental Setup

4.1.1 Datasets

The experiments are performed on the Movielens [5] data set (ML) and a commonly used subset of the EachMovie data set (EM) [1]. The ML dataset consists of 943 users and 1682 items, where each user rated at least 20 items, while EM dataset consists of 2000 users and 1648 items, where each user rated at least 40 items. The sparseness of both datasets is around 94%.

4.1.2 Experimental Protocol

Adopting standard procedure for Top-N recommendation, we consider items that have been assigned high ratings by users to be relevant items. In the ML data set, where items are rated on a scale of 1-5, items with ratings of 4 or 5 are designated as relevant. In the EM data set with the rating scale 1-6 items are considered relevant if rated with 5 or 6. In each dataset, we randomly select 80% of the users as training users and 20% of the users as test users. Training users are used to estimate the proper tradeoff parameter α by means of 5-fold cross-validation. The optimal α is then used in generating recommendations for the test users. Testing involves applying 5-fold cross-validation to the set of test users. For each fold of testing, 20% of the relevant items are held out and the algorithm is required to generate a top-N list in order to predict these items. The remaining items are used to generate the top-N recommendation.

4.1.3 Evaluation Metrics

We report performance in terms of precision (P) and recall (R) measured for top-N lists of various lengths. P@N is defined as the proportion of the top-N items that are relevant and R@N is defined as proportion of relevant items included in the top-N. These metrics are commonly used for evaluation of top-N recommendation [6]. The scores reported here have been averaged across all folds of test users. Note that this P@N probably has a tendency to underestimate the performance of the algorithm as perceived by the user in a real-world application. Items in the top-N recommendation list are only counted as relevant if they were included in the held-out portion of the user profile. The number of items in the top-N list that would have been relevant had the user assigned them ratings (i.e., had they been included in the profile) is not taken into consideration. Arguably, due to this effect P@N and R@N are not the ideal metrics for comparing the performance of Item-CR. However, since our purpose is investigating the relative improvement achieved by Rerank-CR we can safely set this issue aside. P@N and R@N are adequate to evaluate the usefulness of Rerank-CR and we adopt them to maintain consistency with other Top-N recommendation research.

4.2 Tradeoff Parameter

Using 5-fold cross-validation in training users, the optimal tradeoff parameter α in Rerank-CR (cf. Eq. 7) is obtained for each Rerank-CR variant tested, i.e., for each combination of an initial recommendation and a reranking modality. Fig. 2 illustrates how the performance of Rerank-CR varies with α in the case of the Rerank-CR variant that reranks Item-CR using item-based similarity sim_I on ML dataset.

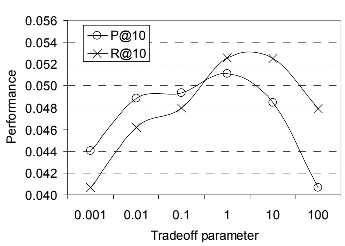

Fig. 2. Variation of Top-10 recommendation performance with α for the training users in the ML dataset, shown for an example Rerank-CR variant (reranking Item-CR using sim_I)

It can be seen that in this case the optimal α lies close to 1. Other Rerank-CR variants follow a pattern similar to that of this example. In each of the following experiments, the tradeoff parameter is set in the same manner, namely, the optimal α calculated using the training set is used with the test set to generate recommendation lists and measure the performance of Rerank-CR.

4.3 Item-CR + Item-based Similarity

Our first experiment investigates the performance of the basic (monomodal) Rerank-CR concept: the same similarity metric used by the CR approach that generates r^{init} is also used for the conditional prior. Specifically, we choose our memory-based approach, Item-CR, to generate r^{init} and item-based similarity, sim_I, as the reranking modality. Results are reported in the top two lines of Tables 1a. (ML dataset) and 1b. (EM dataset). Item-CR without reranking (i.e., Eq. 9) provides the baseline for comparison. Note that the performance levels achieved by the baseline are typical for Top-N CR, cf. [9][15]. Rerank-CR can be seen to outperform the baseline Item-CR approach across the board. The improvement ranges up to 20% in precision and 10% in recall on the ML dataset, and nearly 10% in precision and 5% in recall on the EM dataset. Most of the improvements are statistically significant with $p<0.05$, according to the Wilcoxon signed-rank significance test, as indicated by '^'. These results confirm the potential of information from the user-item matrix to improve Top-N recommendation (cf. research question Q1). The improvement can be attributed to the ability of sim_I to push relevant items upwards and less relevant items downwards by making use of similarities with initially highly ranked and less highly ranked items. Apparently, Rerank-CR is successful in exploiting indirect matching effects between user profiles and recommended items as discussed at the beginning of Section 3.

4.4 Item-CR + Popularity-based Similarity

Recall from Section 3 that we expect multimodal Rerank-CR, which incorporates information from collective views of items, will achieve yet larger gains. In order to investigate this hypothesis, we test the performance of Rerank-CR when it is used to combine Item-CR with the popularity-based conditional priors. Results are reported in the lower lines of Tables 1a. (ML dataset) and 1b. (EM dataset). It can be observed that using popularity-based reranking modalities result in a significant improvement, which ranges, in most cases, up to 50%–100% over the baseline and 20%–100% over the basic (monomodal) Rerank-CR approach Item-CR + sim_I. The results demonstrate that multimodal Rerank-CR yields substantial extra improvement (cf. Q2) and this improvement is stable across datasets (cf. Q3).

Table 1a. Rerank-CR performance: ML dataset for Item-CR

	P@3	P@5	P@10	R@3	R@5	R@10
Baseline: Item-CR	0.034	0.034	0.032	0.014	0.026	0.050
Item-CR+sim_I	0.038^	0.039	0.040^	0.014^	0.027^	0.055^
Item-CR+Pop_{Nr}	0.094^	0.094^	0.096^	0.029^	0.046^	0.096^
Item-CR+Pop_{Ar}	0.038	0.037	0.034	0.019	0.033	0.058
Item-CR+Pop_{NHr}	0.056^	0.055^	0.064^	0.021^	0.035^	0.084^

Table 1b. Rerank-CR performance: EM dataset for Item-CR

	P@3	P@5	P@10	R@3	R@5	R@10
Baseline: Item-CR	0.057	0.056	0.061	0.030	0.048	0.107
Item-CR+sim_I	0.064^	0.061^	0.065^	0.032^	0.051^	0.112^
Item-CR+Pop_{Nr}	0.080^	0.083^	0.089^	0.038^	0.064^	0.139^
Item-CR+Pop_{Ar}	0.122^	0.103^	0.089^	0.069^	0.097^	0.163^
Item-CR+Pop_{NHr}	0.090^	0.091^	0.095^	0.047^	0.080^	0.161^

The performance shows some variation with respect to the choice of the implementation of the popularity-based modality, suggesting that optimizing the encoding of the reranking modality might achieve further gains. The Pop_{Nr} modality emerges as the most reliable performer among the three popularity-based modalities, which also represents our most direct encoding of popularity (cf. Section 3.2).

5. SUMMARY

In this paper, we have introduced Rerank-CR, an approach that builds on established CR techniques and improves their recommendation performance by carrying out self-contained reranking that exclusively exploits information contained within the original user-item matrix. Rerank-CR successfully brings together two tried-and-true techniques, CR and reranking, to form an original combination capable of unlocking and successfully exploiting previously unused information in the user-item matrix. The key strength of Rerank-CR is its ability to achieve improvements in CR performance by exploiting collective views of items derived from collection-level information.

Experiments demonstrated the viability of improving performance with information internal to the user-item matrix (Q1), the added benefit of *connecting with the collective* via multimodal Rerank-CR (Q2) and finally the robustness of Rerank-CR across datasets (Q3).

Rerank-CR makes two central contributions. First, its potential impact is substantial due to the breadth of the domain of its applicability. Rerank-CR can be applied to the output of an arbitrary CR system, yielding an inexpensive improvement that is independent of the availability of external resources. The ability of Rerank-CR to integrate information from the collective without explicit interventions makes it a promising method for exploiting community-level information in a new domain (e.g., a cultural context with different rating behaviors), without the need for costly adaptation. Second, Rerank-CR succeeds in striking the balance necessary in order to successfully exploit collection-level information in combination with personalized recommendations based on user profiles. Although we focus on popularity-based modalities here, the formulation of Rerank-CR is sufficiently general to incorporate arbitrary additional modalities that are extractable from the collection.

6. ACKNOWLEDGEMENTS

The research leading to these results was carried out within the PetaMedia Network of Excellence and has received funding from the European Commission's 7th Framework Program under grant agreement n° 216444.

7. REFERENCES

[1] Adomavicius G., and Tuzhilin, A., 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. Knowledge and Data Engineering, 17, 6, 734-749.

[2] Cui, J., Wen, F. and Tang, X., 2008. Real time Google and live image search re-ranking. In Proc. ACM MM ′08, 729-732.

[3] Deshpande, M., and Karypis, G., 2004. Item-based top-N recommendation algorithms. ACM Trans. on Information Systems, 22, 1, 143-177.

[4] Herlocker, J. L. and Konstan, J. A., 2001. Content-Independent Task-Focused Recommendation. IEEE Internet Computing, 5, 6, 40-47.

[5] Herlocker, J., Konstan, J., Borchers, A., and Riedl, J., 1999. An algorithmic framework for performing collaborative filtering. In Proc. ACM SIGIR ′99, 230-237.

[6] Herlocker, J., Konstan, J., Terveen, L. G., and Riedl, J. 2004. Evaluating collaborative filtering recommender systems. ACM Trans. Information Systems, 22, 1, 5-53.

[7] Hsu, W. H., Kennedy, L. S., and Chang, S.-F., 2006. Video search reranking via information bottleneck principle. In Proc. ACM MM ′06, 35-44.

[8] Kamps, J., 2005. Improving Retrieval Effectiveness by Reranking. In Proc. ECIR ′05, 283-295.

[9] Kim, H.-N., Ji, A.-T., Kim, H.-J., and Jo, G.-S., 2007. Error-based collaborative filtering algorithm for top-n recommendation. In Proc. Joint International Conferences on Asia-Pacific Web Conference and Web-Age Information Management, 594-605.

[10] Koh, N. S., Hu, N., and Clemons, E. K. 2010. Do online reviews reflect a product's true perceived quality? - An investigation of online movie reviews across cultures. In *Proceedings of the 43rd IEEE Hawaii international Conference on System Sciences*, 1-10.

[11] Linden, G., Smith, B., and York, J., 2003. Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 7, 1, 76-80.

[12] Liu, T.-Y. 2009. Learning to rank for information retrieval. Foundations and Trends in Information Retrieval, 3, 3, 225-331.

[13] Liu, Y., Mei, T., and Hua, X., 2009. CrowdReranking: Exploring multiple search engines for visual search reranking. In Proc. ACM SIGIR ′09, 500-507.

[14] Resnick, P., Iacovou, N., Suchak, M., Bergstorm, P., and Riedl, J., 1994. Grouplens: An open architecture for collaborative filtering of netnews. In Proc. of ACM CSCW ′94, 175-186.

[15] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J., 2000. Application of dimensionality reduction in recommender system: A case study. In Proc. WebKDD Workshop.

[16] Sarwar, B., Karypis, G., Konstan, J., and Reidl, J., 2001. Item-based collaborative filtering recommendation algorithms. In Proc. WWW ′01, 285-295.

[17] Shih, Y. and Liu, D., 2008. Product recommendation approaches: Collaborative filtering via customer lifetime value and customer demands. Expert Syst. Appl., 35, 1-2, 350-360.

[18] Sigurbjörnsson, B., and van Zwol, R., 2008. Flickr tag recommendation based on collective knowledge. In Proc. ACM WWW ′08, 327-336.

[19] Tan, T.F., and Netessine, S., 2009. Is Tom Cruise threatened? Wharton Business School Working Notes Paper. University of Pennsylvania. Retrieved December 11, 2009. http://knowledge.wharton.upenn.edu/papers/1361.pdf

[20] Tian, X., Yang, L., Wang, J., Yang, Y., Wu, X., and Hua X.-S., 2008. Bayesian video search reranking. ACM MM ′08, 131-140.

[21] World of hits. In: Economist. Nov 26, 2009.