A knowledge reuse framework for improving novelty and diversity in recommendations

Apurva Pathak[*]
Microsoft R&D. Pvt. Ltd., Hyderabad, India.
appatha@microsoft.com

Bidyut Kr. Patra[†]
VTT Technical Research Centre of Finland, Finland.
ext-bidyut.patra@vtt.fi

ABSTRACT
Recommender system (RS) is an important instrument in e-commerce, which provides personalized recommendations to individual user. Classical algorithms in recommender system mainly emphasize on recommendation accuracy in order to match individual user's past profile. However, recent study shows that novelty and diversity in recommendations are equally important factors from both user and business view points. In this paper, we introduce a knowledge reuse framework to increase novelty and diversity in the recommended items of individual users while compromising very little recommendation accuracy. The proposed framework uses features information which have already been extracted by an existing collaborative filtering. Experimental results with real datasets show that our approach outperfoms state-of-the-art solutions in providing novel and diverse recommended items to individual users and aggregate diversity gain achieved by our approach is on par with recently proposed rank based approach.

Categories and Subject Descriptors
H.3.3 [**Information Search and Retrieval**]: Information filtering.

General Terms

Keywords
Collaborative Filtering, Novelty, Diversity, Clustering, Knowledge Reuse Framework.

1. INTRODUCTION
Recommender system (RS) techniques have been successfully used to help people cope with information overload problem and they have been established as an integral part of e-business domain over last decades. The primary task of a recommender system is to provide personalized suggestions for products or items to an individual user filtering through large product or item space. Many recommender system algorithms have been developed in various applications such as e-commerce, digital library, electronic media, on-line advertising, etc. [2, 19, 22].

Collaborative filtering (CF) is the most successful and widely used recommendation system [2, 22]. In CF, item recommendations to a user are performed by analyzing past rating information of the system. Most of these approaches mainly emphasize on rating accuracy of the recommended items. As results, these traditional approaches recommend items highly similar to each other. After a certain point of time users may lose interest in using recommender system due to absence of novel and diverse items in their recommended lists. Recent study also supports this view and concludes that predictive rating accuracy is not adequate for providing most relevant and interesting items for a user [8, 17]. Scientists in RS community have identified another two essential utilities namely, *novelty* and *diversity* and started incorporating them into recommendation system [1, 5, 9, 24]. Novelty in recommended items provides an opportunity to a user to discover new (novel) and interesting items and recommending diverse range of items prevents any item from becoming obscure in a large item space. Novelty and diversity can only be achieved at the expense of recommendation accuracy as there is a trade-off between accuracy and diversity.

In this paper, we propose a novel framework, which incurs little accuracy loss in order to achieve huge gain in novelty and diversity in an existing collaborative filtering system. The proposed framework utilizes feature information which have already been extracted by existing matrix factorization based CF. We term the proposed framework as Knowledge Reuse Framework in CF (KRCF) as we do not access original rating data in order to improve novelty and diversity in recommendation. Having applied a MF based CF such as Regularized SVD [18] for predicting ratings of non-rated items of an active user, we employ clustering technique to the items which received predicted rating more than a predefined threshold. Finally, recommended list is generated by selecting items from clusters to provide maximum diversity in the list. Our framework is tested with real rating datasets. It is found that proposed KRCF outperfoms state-of-the-art approaches [1, 9] in providing novelty and diversity. Our contributions in this paper are summarized as follow.

[*]Corresponding Author
[†]Corresponding Author

- A knowledge reuse framework termed KRCF is proposed which can easily be deployed into an existing collaborative filtering (MF based CF) system to increase novelty and diversity in recommendation. Proposed KRFC incurs very little accuracy loss compared to the state-of-the art approaches [1, 9].
- In proposed KRCF, clustering technique is exploited and a list of N recommended items is generated for an active user by selecting items from different clusters.
- Experimental results with real rating datasets (MovieLens, Yahoo Music and Netflix) show effectiveness of our proposed framework.

It is worthy to mention that matrix factorization (MF) based CFs provide more accurate results compared to other variant (memory based [20]) of CFs [18, 14, 15]. Therefore, we choose MF based CF for incorporating novelty and diversity factors keeping the fact in mind that there exists trade-off between diversity and accuracy.

Rest of the paper is structured as follows. In section 2, we discuss background of the proposed framework and related work is described in section 3. In Section 4, we present our knowledge reuse framework. Experimental results of proposed framework are provided in Section 5. We conclude our paper in Section 6.

2. BACKGROUND OF THE PROPOSED FRAMEWORK

There are two main approaches for recommending items in CF category, *viz.* *neighborhood based CF* and *model based CF*.

Neighborhood based CF rely on a simple intuition that an item might be interesting to an active user if the item is appreciated by a set of similar users (neighbors) or she has appreciated similar items in the system [6]. Model-based CF algorithms learn a model from the training rating data using machine learning or other techniques [4, 2, 22]. Subsequently, the model is used for predictions. One advantage of the model-based approach is that it does not need to access whole rating data once model is built unlike neighborhood based CF.

Matrix factorization (MF) technique is a model based CF and has gained popularity in recent years because MF based approaches provide more accurate results than other variants of CF [14, 18, 7]. As we explained in previous section, we implemented Regularized Singular Value Decomposition (RSVD). In regularized SVD, rating ($\hat{r}_{ij}$) for an unrated item i of an active user u is predicted by computing inner product of two vectors P_u and Q_i as follows.

$$\hat{r}_{ui} = P_u^T Q_i \qquad (1)$$

where P_u and Q_i are K dimensional vectors of features for user u and item i, respectively. Values in user-feature vectors and item-feature vectors are estimated using gradient descent with regularization criteria such that the sum of squared error is minimized [18, 7]. Other matrix factorization techniques can be found in [15]. Having computed rating of all unrated items of an active user (*rating computation*), MF based approach recommends items to the active users using certain utility function which is based on predicted ratings (*recommendation*).

Proposed KRCF framework can be deployed into the MF based CF for performing the *recommendation* part for increasing novelty and diversity. For this purpose, we exploit clustering technique which is discussed briefly next.

Clustering method. Clustering methods are mainly divided into two categories based on the way they produce results *viz.,* partitional clustering and hierarchical clustering methods [10, 11]. Partitional clustering methods create a single clustering (flat clustering) whereas hierarchical methods create a sequence of nested clustering. The *k-means* clustering is one of most popular partitional clustering methods [10]. However, it suffers from few drawbacks [16]. On the other hand, hierachical clustering can produce satisfiable clustering results. Clustering methods like single-link, average-link, Ward's minimum variance method are few examples of hierarchical clustering methods. These methods differ mainly in distance measures between a pair of clusters [3]. For the sake of readability, we show how average-link and Ward's minimum methods compute distance between a pair of clusters C_1 and C_2.

- Ward's minimum variance: Distance between a pair of clusters is the amount of increase in squared Euclidean distance after merging the pair of clusters.

  $$dist(C_1, C_2) = \frac{|C_1|\ |C_2|}{|C_1| + |C_2|}\ ||\ m_{C_1} - m_{C_2}\ ||_2^2$$

  where m_{C_1} is centroid of cluster C_1.
- Average-link : Distance between a pair of clusters C_1 and C_2 is the average distance between pairs of points in $C_1 \times C_2$.

  $$dist(C_1, C_2) = \frac{1}{|C_1|.|C_2|} \sum_{x_i \in C_1} \sum_{x_j \in C_2} ||\ x_i - x_j\ ||$$

We use *k-cluster stopping condition* for selecting final result from the dendogram generated by hierachical method [13].

3. RELATED WORK

As mentioned in Section 1, recommendation accuracy is not enough for providing relevant and interesting items to individual users. In this section, we discuss approaches which focus on increasing novelty and diversity in recommendations briefly.

The problem of diversity are studied from two different perspectives , *i.e, user perspective* and *business perspective.* The user perspective or user view argue in support of increasing diversity in the recommended items (*individual diversity*) to help an individual user obtain more idiosyncratic items in her recommended list. Individual diversity is commonly measured by an average dissimilarity between all pairs of items in a set [21, 27] . An important concept related to individual diversity called *novelty/surprisal* has been coined in recommendation in recent years [26, 9]. It describes the ability of a recommender system to retrieve novel and unexpected items in a user's recommended list. Metric for novelty given in [26] can be used. It is the unexpectedness of an item relative to its popularity in the training set.

Smyth and McClave [21] observed that individual diversity in recommended items is an important factor and suggested a utility function termed as *quality* which combines similarity and individual diversity for optimizing accuracy-diversity trade-off. A greedy selection approach proposed in [21] is found to be performing well for incorporating individual diversity into a case based recommender system. However, proposed approach fails to provide novel relevant items as found in [9]. Ziegler et. al in [27] showed diversification in recommended items plays an important role in user satisfaction. They proposed a topic diversification algorithm which uses classification taxonomy of the products in order to compute *intra-list similarity*. The classification taxonomy utilizes content information of the product and the approach improves individual diversity of memory based (item and user) recommender systems. It is difficult to build taxonomy in many application domains. Zhang and Hurley propose a number of approaches for improving novelty and (individual) diversity in item based CF [24, 25, 9]. Zhang and Hurley in [25] use a clustering technique to partition active user's purchase history (P_u) into k numbers of clusters. Items in h clusters ($h < k$) which provide maximum diversity are chosen. Subsequently, item based CF (SUGGEST [12]) is modified in such a way that it recommends items by matching items from h clusters instead of items in P_u.

Hurley and Zhang propose more formal approach for (individual) diversity problem in [9]. They pose problem of diversity-accuracy trade-off as binary optimization problem with a tuning parameter for item based collaborative filtering. Recommended items of an active user are represented as a binary vector y of length M, which is determined by item based SUGGEST algorithm [12]. Objective function presented in this approach consists of two components, *diversity* and *matching*. The diversity of recommended items is represented as the quadratic function of y ($f_D(.)$) and matching part is a linear function of y ($f_m(.)$) as stated below.

$$f_D(y) = \alpha y^T D y,\ f_m(y) = \beta m_u^T y$$

where D is $M \times M$ dissimilarity matrix, m_u is matching vector, and α, β are normalization factors. These two terms are combined in linear and non linear ways to form two different objective functions. They offer several strategies to solve these objective functions. A greedy approach similar to the approach used in [21] is discussed. A relax version of binary (quadratic) optimization problem is solved by converting it as real-valued problem. Finally real-valued vector is quantized to obtain optimal binary vector $y*$. They show that both combinations (linear and non-linear) provide near-identical results.

In the proposed KRCF framework, we also use clustering approach for improving novelty and diversity. However, our approach differs in many ways- a) we use clustering approach on the predicted items, b) we donot access original rating dataset for computing similarity, c) simialrity function used in their approach may not be suitable if dataset is sparse and d) our approach can easily be adopted with state-of-the-art CF approaches.

Vargas and Castells in [23] present a generalize framework for the definition of novelty and diversity metrics. They argue that rank and relevance of recommended items are two important factors, which should be taken into account while computing novelty and diversity of recommendations. They generalize novelty and diversity metrics defined in [26, 9, 27] using the browsing model, which is based on three binary relations over set of users and the set of items.

On the other hand, business perspective (business view) inspects the impact of sale diversity by considering diversity of all users (*aggregate diversity*) in a system. A recommender system with high aggregate diversity can increases selling of obscure items and help business houses. Despite the potential importance of aggregate diversity in recommender system, it has received little attention from research community.

Recently, Adomavicius and Kwon [1] introduce ranking based algorithms to address the problem. They argue that high individual diversity may not necessarily imply high aggregate diversity. Aggregate diversity is measured as the total number of distinct items recommended across all users in a system. They suggest to combine standard ranking approach (predicted items arranged in decreasing order) with each of the six ranking approaches proposed in the paper. Proposed ranking approaches introduce two threshold T_R and T_H ($T_R > T_H$) on predicted rating of items. Subsequently, items are ranked based on their approaches if their predicted ratings are more than T_R and standard approach is followed if predicted ratings are less than T_R but higher than T_H. These six approaches are based on different statistical measures such *item popularity*, *item average rating*, *item rating variance*, *item likability*, *neighbors' rating variance* (for memory based CF) and *reverse predicted rating*. This increases aggregate diversity significantly incurring little accuracy loss. However, individual diversity and novelty in individual user's recommended items are ignored in their work.

Taking inspiration from their work [1], our research makes progress in this direction by introducing clustering based framework which takes care of all three factors (individual diversity, novelty and aggregate diversity) while trading very little accuracy loss.

4. PROPOSED FRAMEWORK

In this section, we explain proposed Knowledge Reuse Framework in CF (KRCF) in detail. As we describe in Section 2, proposed KRCF performs *recommendation* sub-task while it uses MF based approach for receiving predicted ratings of all users in a system.

Let I be the set of items and U be the set of users in CF system. Let $L_u \subset I$ be the list of predicted items received from a MF based CF for an arbitrary active user u. Let $\hat{R}(u, i)$ be the predicted rating for the active user u on an item i. We select a set of candidate items $C(u) \subseteq L_u$ in

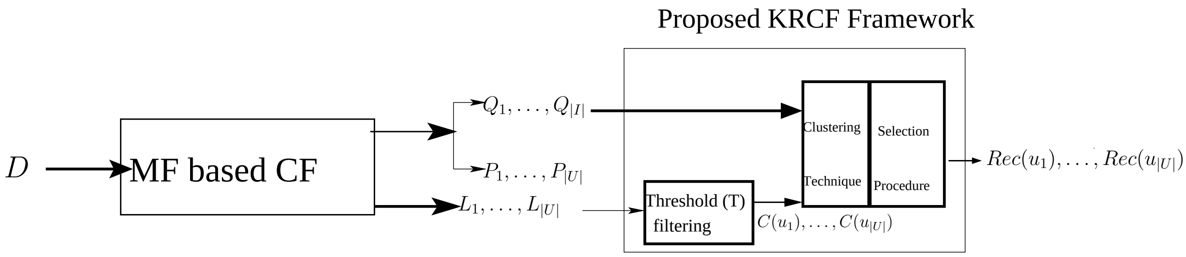


Figure 1: KRCF Framework is incorporated into MF based CF.

such a way that predicted rating of each item in $C(u)$ is above a user defined threshold T. Finally, N items from this candidate set $C(u)$ is recommended to the user u as stated below.

As we know similarity between patterns in a cluster is high while similarity (**diversity**) between patterns in different clusters is low (**high**), we use clustering technique to select Top N recommended items. We apply a clustering approach to obtain a clustering (partition) of items in $C(u)$, if there are more than N items in $C(u)$. As a clustering approach needs a similarity/dissimilarity metric over $C(u) \times C(u)$, we use Euclidean distance in this purpose. We reuse feature-vector of items extracted by MF based CF. We do not access original rating dataset to compute similarity between a pair of items unlike approaches in [9, 25]. Therefore, we term our framework as Knowledge Reuse Framework in CF (KRCF). Major steps of KRCF are shown in Figure 1.

We propose to use hierarchical agglomorative clustering such as Ward's minimum variance approach to obtain a clustering hierarchy of items in $C(u)$ and we apply *k-cluster stopping condition* to obtain k clusters $C_1, C_2, \ldots, C_k$ of items from the hierarchy. We suggest to use the value of $k = N$.

We chose one item from each cluster for recommendation. This ensures high individual diversity as the items in different clusters are dissimilar or diverse to each other. Since, there will be at least a cluster with more than one items, we need to devise a procedure for selecting an item from each cluster with more than one items. Taking inspiration from the work in [1], we select an item with the minimum predicted rating from each cluster. Let $L_{C(u)}$ be the items obtained after applying the selection procedure. More formally, we obtain $L_{C(u)}$ as follows.

$$L_{C(u)} = \bigcup_{i=1}^{k} \{ \; j \in C_i \quad | \;\; \hat{R}(u,j) \leq \hat{R}(u,l), l \in C_i \setminus \{j\} \quad \}.$$

Other selection procedures such as *picking item with median predicted rating, picking least populated items* are tested over the clustering output $\{C_1, C_2, \ldots, C_k\}$. However, *minimum predicted rating* selection is found to be outperforming others in providing diversity and novelty.

We apply other agglomorative clustering approaches such as average-link in similar fashion. Each of them provide different degree of diversity and novelty in recommendation. Partitional clustering approach k-means is also tested in our framework in similar line. This shows that any distance based clustering approach can be incorporated in proposed framework.

Finally, we combine standard ranking approach (predicted items arranged in decreasing order) and results obtained from one of the four clustering approaches discussed to recommend a list of items $Rec(u)$ as follow.

$$Rec(u) = \begin{cases} C(u) \text{ (Standard Ranking approach)} & \text{if } |C(u)| \leq N \\ L_{C(u)} \text{ (Clustering approach)} & \text{if } |C(u)| > N \end{cases}$$

As we know from clustering literature, these four traditional clustering approaches cannot identify and discard outlier (novel) points in a given dataset. They treat these outlier datapoints as valid datapoints and keep them in clusters. Frequently, these datapoints lie border of a cluster or appear as a singleton cluster. This characteristic of clustering approaches help increasing the novelty of items in the selected list $L_{C(u)}$. Experimental analysis supports our view and results show that clustering produces more than one singleton clusters which are unexpected items in $C(u)$.

One main advantage of our reuse framework is that it is immune to *data sparsity problem* as it does not access the original sparse data. Clustering technique plays an important role in providing high individual diversity and novelty in our framework. On the other hand, the *minimum predicted rating* selection procedure leads to increase high sale diversity (aggregate diversity) as items with less predicted values are generally long tail items (less popular items) [1].

The value of the parameter T controls the accuracy. A higher value of T means higher accuracy but lower diversity (aggregate diversity) and novelty and vice versa. The value of T can, thus, be varied to set a desired balance between accuracy, diversity and novelty. To ensure minimum accuracy loss, we suggest to use the value of $T > 4.0$ in a dataset with 1-5 rating scale such as MovieLens and Netflix datasets.

5. EXPERIMENTAL EVALUATION

To evaluate performance of proposed KRCF framework in providing diverse and novel items in recommendations, we implemented Regularized Singular Value Decomposition based CF (RSVD) and incorporated rank based approaches and KRCF approaches into RSVD. Memory based binary quadratic optimization (BQP) approach proposed by Hurley and Zhang in [9] is also implemented. In BQP, we use cosine similarity

Table 1: **Statistics of subsets**

Dataset	#User (| U |)	#Item (| I |)	#Rating (R)	κ $(\frac{R\times 100}{\|U\|\times\|I\|})$
Movie-Lens	3000	2000	$326,500$	5.44
Yahoo	1500	1000	$22,300$	1.48
Netflix	3000	2000	$427,200$	7.12

measure to compute simialrity between a pair of items. We compare our framework with rank based and BQP approach as these two approaches are found to be most popular (cited) in research community.

5.1 Data Preparation

The approaches described in earlier are tested with three real datasets, namely MovieLens [1], Netflix [2] and Yahoo Music[3]. We make training sets of various sparsity levels out of these datasets in such a way that each user should have enough number of relevant unrated items in the test set. For this purpose, we selected top 3000 users who rated maximum number of movies and subsequently, top 2000 movies which received ratings from those 3000 users are selected in MovieLens dataset. Similarly, we obtained other subsets. Sparsity level is parametarized by the *density index* (κ), which is the percentage of all possible ratings available in a dataset.The characteristics of all these subsets are summarized in Table 1.

5.2 Evaluation Metric

As discussed earlier, there is a trade-off between accuracy and diversity, therefore, we need to inspect accuracy loss suffered by each approach in the process of gaining diversity in recommendations. Let $Rec(u)$ be the set of top N recommended items for user u, recommendation accuracy of a RS can be written as

$$precision\text{-}in\text{-}top\text{-}N = \frac{\sum_{u\in U}|Rec(u)\cap Rev(u)|}{\sum_{u\in U}|Rec(u)|},$$

where $Rev(u)$ is the set of relevant items (ratings $\geq$ 4.0) of user u in the test set. As main objective of our work is increase diversity (aggregate, individual) and novelty in recommendations, we utilize frequently used metrics for computing them as follow.

Aggregate diversity (AD) [1] : Total number of distinct items recommended across all users in a RS. The AD of a recommender system is computed as $AD = \mid \bigcup_{u\in U} Rec(u) \mid$.

Individual diversity (ID) [24, 25, 9]: Individual diversity is also important factor for users' satisfactions. It is the average dissimilarity between each pair of recommended items to a user. The ID of an arbitrary recommender system is computed as

$$ID = \frac{1}{\mid U \mid}\sum_{u\in U}\frac{\sum\sum_{i,j\in Rec(u)} 1 - sim(i,j)}{N.(N-1)}$$

where $N = \mid Rec(u) \mid$, $sim(.)$ is similarity between a pair of items and it is computed using cosine similarity measure.

Another related concept Novelty describes unexpectedness of items in individual users' recommended list. These unexpectedness or surprisal is computed over all users in a system as follows [26, 23].

$Novelty = \frac{1}{|U|}\sum_{u\in U}\frac{\sum_{i\in Rec(u)}\log_2(\frac{|U|}{\#i})}{\mid Rec(u)\mid}$, where $\#i$ is the number of users rated item i in the training set.

5.3 Experiments and Results Analysis

We executed all variants of ranking and proposed KRCF approaches on MovieLens subset and results are reported in Table 2. We report results of *Reverse Predicted Rating Value* (RPR) based ranking approach only as this variant is found to be best among other variants in providing aggregate diversity in recommendations. We consider the top N recommendations for all experiments. However, value of N varies with datasets. For MovieLens, it is $N = 5$. The Reverse Predicted Rating (RPR) based ranking approach is very successful in providing aggregate ($AD = 151$) diversity of recommendation. However, it provide individual diversity (ID=0.753) less than the standard approach with accuracy loss 0.001 or 0.1% . Our proposed framework KRCF keeps balance of all these three factors, individual diversity, aggregate diversity and novelty. All variants (Ward's minimum variance, average-linkage and k-means) of KRCF provide same ID with that of the standard approach with accuracy loss 0.001 or 0.1%. All three variants of KRCF outperform standard approach in novelty measure (Novelty of standard CF=0.854, Novelty of KRCF in the range of 1.272-3.135) (Table 2).

Gain in aggregate diversity (AD) achieved by proposed KRCF is very close to RPR based ranking approach with same precision loss. The Ward's minimum variance and k-means based KRCF achieve AD gain of 4.25 (149 distinct movies out of 2000 movies in the datasets), whereas ranking based approach achieves AD gain of 4.31 (151) compared to the standard approach (Table 2). However, KRCF can provide more diverse and novel items to the individual users. The KRCF provides more novel items ($Novelty = 1.277$) in recommendations compared to ranking based approach with same accuracy loss. If system is allowed to suffer more accuracy loss, our approach can provide more and more diverse (ID) and novel items to individual users compared to ranking based approach. It can be observed from Table 2, proposed KRCF can provide more than 28.1% and 4.75% novelty and individual diversity, respectively, compared to ranking based approach if system is allowed to drop precision to 0.851 from 0.901 (5%). With equal amount of precision loss, ranking based approach can achieve AD gain of 14.74, whereas our approach achieve AD gain 14.37 over standard approach. Therefore, if business scarifies little (0.37) sale diversity (AD) compared to ranking approach , proposed KRCF can provide more novel and diverse items to indi-

[1] *http://www.grouplens.org*
[2] *http://www.netflixprize.com*
[3] *http://research.yahoo.com/academic_relations*

Table 2: Experimental results with MovieLens subset, $N = 5$.

Precision (Standard)	Precision Loss	Ranking Approach			KRCF (Ward's)			KRCF (Average-linkage)			KRCF (k-means)		
		AD	ID	Novelty	AD	ID	Novelty	AD	ID	Novelty	AD	ID	Novelty
0.901	-0.001	**151**	0.753	1.202	149	**0.759**	1.273	145	0.759	1.272	149	**0.759**	**1.277**
	-0.010	**252**	0.763	1.348	241	**0.782**	**1.578**	207	0.772	1.440	213	0.773	1.455
	-0.025	**362**	0.776	1.521	355	**0.806**	1.898	350	0.805	1.895	356	**0.806**	**1.909**
	-0.050	**516**	0.795	1.779	503	**0.833**	2.281	494	**0.833**	2.282	500	0.832	**2.285**
	-0.100	**1032**	0.853	2.595	1008	0.886	**3.135**	982	**0.887**	3.132	999	**0.887**	3.134
Standard	0.000	AD=35, ID=0.759, Novelty=0.854											

Table 3: Experimental results of BQP ($\theta = 0.5$) and KRCF on different subsets.

Metric	MovieLens ($N = 5$)		Yahoo Music ($N = 10$)		Netflix ($N = 5$)	
	BQP	KRCF	BQP	KRCF	BQP	KRCF
Precision Loss	-0.27	**-0.10**	-0.15	-0.10	-0.3	**-0.07**
AD	801	**1008**	686	**945**	743	**1015**
ID	0.784	**0.886**	0.700	**0.970**	0.845	**0.849**
Novelty	2.894	**3.135**	**6.058**	05.114	2.539	2.519

vidual users compared to ranking based approach. Similar trend is found with precision loss of 10% (Table 2).

We tested our KRCF with another popular approach binary quadratic optimization (BQP), which is built on memory based CF [9]. Detailed results are reported in Table 3. Proposed KRCF outperfoms BQP in all measures namely, precision loss, AD, ID and novelty. The BQP approach achieves AD of 40.05% suffering significant (27%) precision loss, whereas KRCF can achieves AD of 50.4% suffering 10% accuracy loss on MovieLens subset. It can be noted that we report results of Ward's minimum variance based KRCF as this is found to be best among other variants.

We conducted experiments with very sparse subset ($\kappa = 1.48$) of Yahoo Music. We incorporated hierarchical (Ward's variance) and partitional (k -means) clustering approaches in our KRCF framework and results are reported in Table 4. Experimental results show that KRCF outperforms ranking approaches in all metrics with suffering equal amount of precision loss (Table 4). It can be noted that standard approach can provide significant diversity (ID=0.943) to individual users. However, it fails to provide sale diversity adequately (AD=40) (Table 4). However, RPR based ranking approach and proposed KRCF increase ID, AD and novelty compared to standard approach with little (0.01) accuracy loss (*i.e.,* 0.8617 down to 0.8517). Experimental results show that both variants of proposed KRCF provide more sale diversity and individual diversity than ranking based approach. If system is allowed to suffer more accuracy loss, proposed KRCF gains diversity (AD, ID) and novelty significantly compared to ranking based approach. This clearly shows that clustering based approach can be an effective tool for providing diversity from business and user view points. If system precision can be degraded to 0.8117 from 0.8617, proposed KRCF has aggregate diversity (AD) gain of 16.35 (15.02 for k-means based KRCF), whereas, ranking based approach has AD gain of 11.62 compared to standard approach. With similar accuracy loss (-0.050), Ward's variance based KRCF, k-means based KRCF and ranking based approach gain 2.52, 2.42 and 2.12 times more novel items compared to standard approach, respectively. This results conclude that KRCF recommends more surprisal items to an individual user compared to ranking based approach consistently over various datasets.

Proposed KRCF (Ward's variance based) framework can recommend more than 94% (AD=945) of total items (musics) across all users with 10% precision loss, whereas the BQP can achieve AD of 68.6% with 15% accuracy loss (Table 3) on Yahoo Music subset. KRCF provides more than 38% diversity (ID) to individual users in their recommended items compared to BQP approach. However, BQP provides little more surprisal items compared to KRCF approach.

Experimental results with another popular real dataset Netflix are reported in Table 5. Standard approach perform poorly in providing AD, ID and novelty in recommendations. With accuracy loss 0.001 from 0.9215, ranking based approach and Ward's variance based, k-means based KRCF approaches gain 8.34 and 8.14, 8.06 times more AD respectively, compared to standard approach. However, our KRCF provides more diverse and novel items to individual users compared to ranking based approach. With accuracy loss 0.070 (7%), ranking based approach and KRCF approach can achieve aggregate diversity gains of 22.47 and 21.68 (21.59 for k-means based KRCF) times, respectively compared to standard approach. KRCF and ranking approaches have novelty gain of 2.94 (2.90 for k-means based KRCF) and 2.77 respectively, compared to standard approach.

Experimental results of BQP with Netflix subset is also shown in Table 3. Our proposed KRCF outperfoms BQP in providing diversity and novelty in recommendations. The BQP suffers 30% accuracy loss to achieve same level of individual diversity and novelty of that of the KRCF framework achieve suffering only 7% accuracy loss. The BQP can provide AD of 743 (out of total 2000 items) if user can tolerate 30% precision loss. However, KRCF can provide more sale diversity (AD=1015) suffering only 7% precision loss.

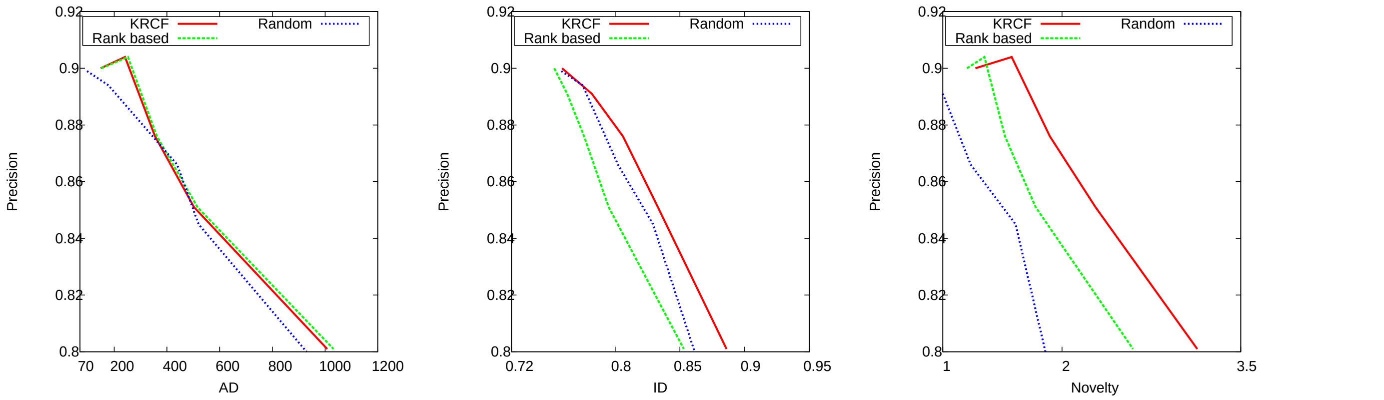
To have better understanding of the performance of KRCF and ranking based approaches, three plots are depicted in Figure 2 to show gain achieved by them in AD, ID and novelty. We report results of *random approach.* The random approach recommends N items randomly which have predicted values above a threshold T. Plot 2(a) shows that proposed KRCF produces AD, which is very close to the AD

Table 4: Experimental results with Yahoo Music subset, N=10

Precision of Standard Approach	Precision Loss	Ranking Approach AD	Ranking Approach ID	Ranking Approach Novelty	KRCF (Ward's) AD	KRCF (Ward's) ID	KRCF (Ward's) Novelty	KRCF (k-means) AD	KRCF (k-means) ID	KRCF (k-means) Novelty
0.8617	-0.010	**93**	0.966	4.74	**93**	**0.968**	**4.76**	**93**	0.967	4.72
	-0.025	323	0.967	5.12	**546**	**0.969**	**5.42**	516	0.967	5.32
	-0.050	465	0.968	5.15	**654**	**0.969**	**5.85**	601	0.968	5.63
	-0.100	750	0.969	4.98	**945**	**0.970**	**5.11**	899	**0.970**	5.00
Standard →	0.000	AD=40, ID=0.943, Novelty=2.32								

Table 5: Experimental results with Netflix subset, N=5

Precision of Standard Approach	Precision Loss	Ranking Approach AD	Ranking Approach ID	Ranking Approach Novelty	KRCF (Ward's) AD	KRCF (Ward's) ID	KRCF (Ward's) Novelty	KRCF (k-means) AD	KRCF (k-means) ID	KRCF (k-means) Novelty
	-0.001	**394**	0.773	1.743	383	**0.778**	1.745	379	0.777	**1.751**
	-0.010	**684**	0.803	2.026	655	**0.809**	2.105	650	0.807	**2.119**
0.9215	-0.070	**1067**	0.845	2.403	1019	**0.851**	**2.547**	1015	0.848	2.519
Standard→	0.000	AD=47, ID=0.72, Novelty=0.866								



(a) Precision Vs AD (b) Precision Vs ID (c) Precision Vs Novelty

Figure 2: Experimental results with MovieLens subset.

produced by ranking based approach. Both approaches outperform random approach. Gain in AD contributed by both approaches are significantly higher compared to standard approach. In the process of gaining AD, both approaches (KRCF and ranking based) suffer almost equal amount of precision loss.

Individual diversity helps a user receive more diverse items in her recommended list. Plot 2(b) shows that proposed KRCF provides more diverse items to individual users compared to rank based approach and random approach. However, our approach suffers less precision loss compared to both the approaches. It can be noted that random approach is found to be performing better than rank based approach.

Increasing novelty in recommendation provides more opportunity to individual user in receiving less popular and more novel items in recommended list. Proposed KRCF recommends more novel and surprisal items to individual users (Plot 2(c)). It can be mentioned that clustering technique in KRCF produces more than one singleton clusters, which contains non popular items. Proposed KRCF outperfoms random and ranking approaches in recommending novel items.

We conducted experiments to show the ability of KRCF and ranking based approaches in recommending non popular items to individual users. We analyzed recommended items of a set of random users (6 users) in each dataset. Items in each dataset are arranged in non increasing order of their popularity and we consider bottom 20% of the arranged items as *non popular* items. We count number of non popular items, which are recommended by KRCF and ranking based approach for the set of users. We report percentage of non popular recommended items for all three datasets in Figure 3. In MovieLens dataset, proposed KRCF recommends 30% non popular items to a user, whereas ranking based approach recommends 16% non popular items. Plot in Figure 3 shows that proposed KRCF recommends significantly more number of non popular items than that of the ranking based approach. It can be noted that we kept precision loss (0.05) same for both the approaches in each dataset. In Netflix dataset, ranking based approach recommends close to 20% non popular items. However, proposed KRCF recommends more than 26% non popular items to a user.

It can be observed that we do not report results of BQP approach in Figure 3 as BQP suffers huge accuracy loss to incorporate novelty and diversity in recommendations (Table 3). To capture the trade-off between novelty and precision in a single measure, we define a metric termed as *non popular item to precision loss (NPPL)* , which is the ratio of *average non popular recommended items* to precision loss incurred by an approach. The average non popular recommended item is the ratio of number of recommended items which are non popular items (bottom 20% of a list of items arranged in non increasing order of their popularity) to the total number of recommended items. We computed NPPL for all three approaches (Ranking, KRCF and BQP) in each dataset and results are shown in Figure 4. The plot clearly shows that proposed KRCF suffers least precision loss among the three approaches to provide highest number of non popular items in recommendations. Ranking based approach performs better than BQP in MovieLens and Netflix datasets, however, BQP outperfoms Ranking approach in Yahoo Music dataset. Proposed KRCF outperfoms both approaches (BQP and Ranking approaches) in NPPL measures in all three datasets.

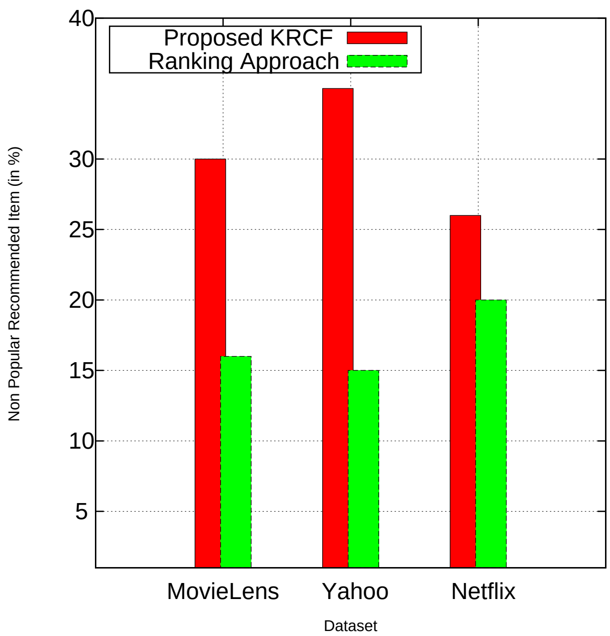


Figure 3: Non popular items recommended to a user by Ranking and proposed KRCF approaches in different datasets.

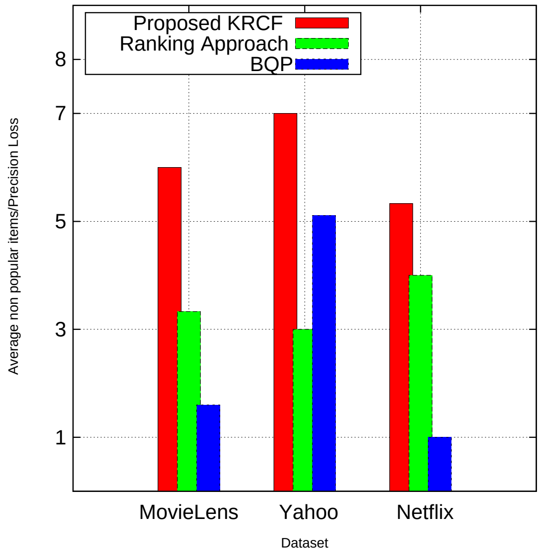


Figure 4: Average Non popular recommended item per Precision Loss in different datasets.

6. CONCLUSION

In this paper, we introduced a knowledge reuse framework, which exploits clustering technique for providing more diverse and novel items in recommendations. Experimental results with real rating datasets show that proposed framework is effective in maintaining balance among the various utility metrics namely, aggregate diversity, individual diversity, novelty and recommendation accuracy. Main advantage of the proposed framework is that it can easily be deployed into state-of-the-art matrix factorization based CF. The knowledge reuse framework can be explored for another collaborative filterings. This research can be extended in exploring biographical and social information along with the rating history for incorporating diversity in recommendation.

Acknowledgment

This work is carried out during the tenure of an ERCIM “Alain Bensoussan” Fellowship Programme. Bidyut Kr. Patra has received funding from the European Union Seventh Framework Programme ($FP7/2007-2013$) under Grant agreement 246016.

7. REFERENCES

[1] G. Adomavicius and Y. Kwon. Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering*, 24(5):896–911, 2012.

[2] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[3] P. Berkhin. A survey of clustering data mining techniques. In *Grouping Multidimensional Data - Recent Advances in Clustering*, pages 25–71. 2006.

[4] D. Billsus and M. J. Pazzani. Learning Collaborative

Information Filters. In *Proceedings of the Fifteenth International Conference on Machine Learning*, pages 46–54, 1998.

[5] O. Celma and P. Herrera. A new approach to evaluating novel recommendations. In *Proceedings of the 2008 ACM Conference on Recommender Systems*, pages 179–186, 2008.

[6] C. Desrosiers and G. Karypis. A Comprehensive Survey of Neighborhood-based Recommendation Methods. In *Recommender Systems Handbook*, pages 107–144. 2011.

[7] S. Funk. Netflix Update: Try this at Home. In *http://sifter.org/˜simon/journal/20061211.html*, 2006.

[8] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1):5–53, 2004.

[9] N. Hurley and M. Zhang. Novelty and diversity in top-n recommendation – analysis and evaluation. *ACM Trans. Internet Technol.*, 10(4):14:1–14:30, 2011.

[10] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recogn. Lett.*, 31(8):651–666, June 2010.

[11] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. *ACM Comput. Surv.*, 31(3):264–323, 1999.

[12] G. Karypis. Evaluation of item-based top-n recommendation algorithms. In *Proceedings of the Tenth International Conference on Information and Knowledge Management*, CIKM '01, pages 247–254, 2001.

[13] J. M. Kleinberg. An impossibility theorem for clustering. In *NIPS*, pages 446–453, 2002.

[14] Y. Koren. Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 426–434, 2008.

[15] Y. Koren. Collaborative filtering with temporal dynamics. *The Communication of ACM*, 53(4):89–97, 2010.

[16] C.-R. Lin and M.-S. Chen. Combining partitional and hierarchical algorithms for robust and efficient data clustering with cohesion self-merging. *Knowledge and Data Engineering, IEEE Transactions on*, 17(2):145–159, Feb 2005.

[17] S. M. McNee, J. Riedl, and J. A. Konstan. Being accurate is not enough: How accuracy metrics have hurt recommender systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, pages 1097–1101, 2006.

[18] A. Paterek. Improving regularized singular value decomposition for collaborative filtering. In *Proceeding of KDD Cup Workshop at 13th ACM Int. Conf. on Knowledge Discovery and Data Mining*, pages 39–42, 2007.

[19] P. Resnick and H. R. Varian. Recommender systems. *Communications of the ACM*, 40(3):56–58, 1997.

[20] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295, 2001.

[21] B. Smyth and P. McClave. Similarity vs. diversity. In *Proceedings of the 4th International Conference on Case-Based Reasoning: Case-Based Reasoning Research and Development*, ICCBR '01, pages 347–361, 2001.

[22] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009:4:2–4:2, 2009.

[23] S. Vargas and P. Castells. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, pages 109–116, 2011.

[24] M. Zhang and N. Hurley. Avoiding monotony: Improving the diversity of recommendation lists. In *Proceedings of the 2008 ACM Conference on Recommender Systems*, pages 123–130, 2008.

[25] M. Zhang and N. Hurley. Novel item recommendation by user profile partitioning. In *Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT '09. IEEE/WIC/ACM International Joint Conferences on*, volume 1, pages 508–515, 2009.

[26] T. Zhou, Z. Kuscsik, J. Liu, M. Medo, J. Wakeling, and Y. Zhang. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, 107(10):4511–4515, 2010.

[27] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, pages 22–32, 2005.