

A Coverage-Based Approach to Recommendation Diversity On Similarity Graph

Shameem A Puthiya Parambath

Nicolas Usunier *

Yves Grandvalet

Sorbonne Universités, Université de technologie de Compiègne
CNRS, Heudiasyc UMR 7253, Compiègne
firstname.lastname@utc.fr

ABSTRACT

We consider the problem of generating diverse, personalized recommendations such that a small set of recommended items covers a broad range of the user's interests. We represent items in a similarity graph, and we formulate the relevance/diversity trade-off as finding a small set of unrated items that best covers a subset of items positively rated by the user. In contrast to previous approaches, our method does not rely on an explicit trade-off between a relevance objective and a diversity objective, as the estimations of relevance and diversity are implicit in the coverage criterion. We show on several benchmark datasets that our approach compares favorably to the state-of-the-art diversification methods according to various relevance and diversity measures.

Keywords

Diversity; Recommender Systems

1. INTRODUCTION

Most information retrieval systems are designed with the assumption that the relevance of the answers to the query are independent of each other. However, in many real world applications like web search and recommender systems, the usefulness of results depend on each other. The inherent limitations of the retrieval system to represent and capture the complex and time-varying user requirements often results in imprecise responses [18]. The above two factors demand a system to respond with diverse relevant results. For web search, diversification is usually preferred to achieve low query abandonment and to compensate for the ambiguity in the user query [8]. Similarly, in recommender systems, for example movies, a user may have eclectic movie tastes, and diversification is a means to cover movies from all relevant types.

*Now at Facebook AI Research, Paris, France.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

RecSys '16, September 15 - 19, 2016, Boston, MA, USA

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-4035-9/16/09...\$15.00

DOI: <http://dx.doi.org/10.1145/2959100.2959149>

The most usual way of inducing diversity is to perform a two-step approach, in which a ranked list of top- k results is first retrieved, and a re-ranking algorithm is then run on the list such that diverse results appear at the top ranking positions. The re-ranking algorithms optimize an objective function that explicitly trades-off a relevance term and a diversity term. There are many variants of this scheme and we broadly refer to them as re-ranking algorithms for diversification.

We consider in this paper the diversity/relevance trade-off in the context of item-based recommendation, where personalized recommendations propose items that are similar to items that are known to be of interest to the user. Item-based recommender systems have proved effective in practice; they are also particularly relevant in online recommendation settings since recommendations can be generated on-the-fly after any update to the user history.

We view items as nodes in a similarity graph, and we define the coverage of a set of items by another set of items from the similarities between pairs of nodes. The objective of recommendation is then to generate a set of unrated items that covers the set of items that were positively rated by the user. In this approach, diversity is obtained by defining the coverage as a submodular function: there is little gain in improving the coverage of a rated item that is already covered, whereas there can be a large gain in covering a new positively rated item.

As for other diversification algorithms, the submodularity of the objective function provides approximation guarantees to the greedy algorithm iteratively building the set of recommended items. We also experiment with a slightly more involved inference algorithm based on a convex relaxation of the problem, but with limited success. We conclude that the greedy algorithm gives satisfactory results in practice.

In contrast to existing approaches that rely on two separate objectives for relevance and diversity, coverage accounts for both relevance and diversity: relevance is captured through the set to be covered, defined by positively rated items, and diversity through the preference towards lightly covering many items instead of covering heavily a few items. We compare our approach to existing baselines for the diversity/relevance trade-off in an item-based collaborative filtering setup on benchmark datasets, and show that our algorithm compares favorably in terms of various relevance and diversity metrics.

The remainder of this paper is structured as follows. We give a brief overview of the state-of-the-art re-ranking algo-

gorithms for diversification in Section 2. Section 3 describes our framework, the optimization problem, and its relationship with other well-known problems. We carry out large scale experiments on benchmark data and report the results in Section 4. We conclude the paper in Section 5.

2. RELATED WORK

Earlier work on diversification is mostly confined to two strongly inter-connected domains: document retrieval and recommender systems, where many re-ranking algorithms have been suggested in different settings. In general, the re-ranking algorithms can be broadly classified into two categories, (i) based on random walk [25, 27, 21, 9] and (ii) based on the maximum set cover algorithm in offline settings [4, 8, 1, 16, 23, 2] and online settings [15].

In web search settings, re-ranking algorithms operate on the list of items returned by a given query; they re-order this list such that top-ranked results are diverse. One can use random walk based algorithms like PageRank to give an estimation of the importance of the nodes in a graph. Then the nodes are re-ordered based on this score to induce diversity. Zhang et al. [25] iteratively select diverse items by updating the PageRank scores of the unselected items w.r.t. the last selected item. Tong et al. [21] and He et al. [9] use greedy algorithm for set cover on top of PageRank based algorithm to select diverse items by explicitly trading-off relevance and diversity.

In the set cover based algorithms, a submodular objective function which is a linear combination of two components, one for the relevance part and the other for diversity part, is maximized to select the top- k items with maximal relevance-diversity trade-off. The relevance component measures the similarity between the user query and an item, and the diversification component measures the dissimilarity between the item and the set of already selected items. The idea was first proposed in the seminal work of Carbonell and Goldstein [4]. Gollapudi and Sharma [8] study the theoretical properties of the diversification objective and propose many possible choices of objective functions. Agrawal et al. [1] and Chapelle et al. [5] propose algorithms based on retrieving pages covering many intents of the user query. Radlinski et al. [15] suggest algorithms based on multi-armed bandits to learn diverse lists from click data in online settings. Their objective function aims to maximizing the probability of finding a relevant document in the top- k list. Few earlier works focused on performance metrics for quantifying diversity. Agrawal et al. [1] propose intent-aware versions of commonly used performance metrics in IR, whereas Chapelle et al. [5] inspects the theoretical properties of the metrics.

Recommender system is another area where diverse results are demanded. Here there is no query associated to the task, and the relevance of an item is conveyed through a graded rating value. The recommendation list diversification problem has been studied from different points of view. Vargas and Castells [22] propose a unified view of the state-of-the-art metrics used in recommender system diversification. Vargas et al. [23] discuss a diversity metric based on probabilistic models for genre coverage, and propose a re-ranking algorithm to diversify the recommendation list obtained using a baseline recommender system. Ashkan et al. [2] maximize a modular objective function with submodular constraints and demonstrate the effectiveness of the

approach on benchmark datasets. A re-ranking based approach by penalizing the system wide popularity bias with individual user specific popularity is proposed by Oh et al. [13]. Hurley [10] proposes a diversification method which does not require a ranked list beforehand, by weighting the pairwise rank difference with the dissimilarity score. However, the theoretical properties of the objective function do not give a clear insight about the relevance-diversity trade-off. A similar approach, proposed by Su et al. [20], creates a user profile based on relevant and irrelevant items, and optimizes an objective function defined over the proper subsets of the profile. The algorithm focuses on intra-list diversity whereas we focus on inter-list diversity. Wu et al. [24] propose a framework for diversification similar to the one of Carbonell and Goldstein [4].

Our approach differs from the previous work in that we propose a single criterion to account for both diversity and relevance. The trade-off is dealt with by the exact definition of coverage we use, which we describe in the next section.

3. UTILITY-WEIGHTED COVERAGE FOR RELEVANT DIVERSE SETS

We consider that we have n objects $\mathcal{X} = \{x_1, \dots, x_n\}$, together with a similarity matrix $\mathbf{W} = (W_{ij})_{i,j=1..n}$. To be concrete, in a personalized recommender system, each object corresponds to an item (e.g. movie/book), and the similarity between node i and j could be given as the cosine similarity between the columns i and j of the user \times item rating matrix. We do not assume the similarity to be symmetric, but we will restrict ourselves to the case where $W_{ij} \geq 0$.

We can view $(\mathcal{X}, \mathbf{W})$ as a weighted graph, where W_{ij} , which weights the edge between nodes x_i and x_j , should be interpreted as how much item x_i is *similar* to item x_j . Our goal is to return diverse relevant items, and we formalize it as a property of the returned solution set \mathcal{S} , based on the coverage of the subset of nodes in the graph \mathcal{X} representing already rated items \mathcal{R} . In this respect, our approach is similar to the set-cover based re-ranking algorithms [4, 1, 16].

3.1 Coverage of a Node

From now on, to simplify the notation, we identify the set of nodes \mathcal{X} with $\{1, \dots, n\}$. For a subset \mathcal{S} of \mathcal{X} , given a node $i \notin \mathcal{S}$, we define the coverage score of i by \mathcal{S} as:

$$\text{cov}(i, \mathcal{S}) = f\left(\sum_{j \in \mathcal{S}} f^{-1}(W_{ij})\right). \quad (1)$$

where $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is an invertible function that will be chosen non-decreasing and concave, so as to ensure that $\mathcal{S} \rightarrow \text{cov}(i, \mathcal{S})$ is non-decreasing with respect to inclusion and submodular. We call the function f in (1) the *saturation function*, because its main usage is to make the coverage of a single node of the graph saturate as we enlarge \mathcal{S} .

3.2 Utility-Weighted Coverage of a Set of Nodes

We now extend the definition of the coverage of a node to a set of nodes through the utilities attached to the nodes $\{v_1, \dots, v_n\}$. Utility is a degree of liking for an item given by a user, such as the rating given to this item. Here, we assume that $v_j \geq 0$. Given a set of m items $\mathcal{R} = \{\kappa_1, \dots, \kappa_m\} \subset \mathcal{X}$ with their corresponding observed utility values $\Upsilon = \{v_1, \dots, v_m\}$, we define the *profile* of the

Algorithm 1 Greedy Algorithm

Input: set of items \mathcal{X} , profile \mathcal{P} , similarity matrix \mathbf{W} , # of recommendations k
 $\mathcal{X} = \mathcal{X} \setminus \mathcal{R}, \mathcal{S} = \emptyset$
repeat
 $i^* = \operatorname{argmax}_{i \in \mathcal{X}} \operatorname{cov}(\mathcal{P}, \mathcal{S} \cup \{i\})$
 $\mathcal{S} = \mathcal{S} \cup \{i^*\}$
 $\mathcal{X} = \mathcal{X} \setminus \{i^*\}$
until $|\mathcal{S}| = k$
Output: set of diverse items \mathcal{S}

user as $\mathcal{P} = \{(\kappa_j, v_j)\}_{j=1, \dots, m}$. In the recommendation example, the profile can be the pair of sets (item, rating) given by a user.

Now, given a profile \mathcal{P} and a set of items \mathcal{S} such that $\mathcal{S} \subset \mathcal{X} \setminus \mathcal{R}$, the coverage of profile \mathcal{P} by \mathcal{S} is defined as:

$$\operatorname{cov}(\mathcal{P}, \mathcal{S}) = \sum_{(\kappa, v) \in \mathcal{P}} v \operatorname{cov}(\kappa, \mathcal{S}) . \quad (2)$$

Here, we use a slight abuse of notation for cov , which can take as first argument either a profile or an item, but we assume that the context is clear considering the use of calligraphic notation for sets. From now on, we use the terminology of *point-wise* coverage for (1) and *profile* coverage for (2).

3.3 Optimal Utility-Diversity Trade-Off

Given a fixed *saturation function*, a set \mathcal{S}^* realizes the optimal utility-diversity trade-off if it solves:

$$\max_{\substack{\mathcal{S} \subset \mathcal{X} \setminus \mathcal{R} \\ |\mathcal{S}| \leq k}} \operatorname{cov}(\mathcal{P}, \mathcal{S}) . \quad (3)$$

Since f in (1) is concave, the objective function of problem (3) is submodular. As a result, the greedy approximation algorithm for the maximum set coverage problem, outlined in Algorithm 1, can be used to approximately solve (3), with the approximation guarantees stated in Theorem 1.

THEOREM 1 (NEMHAUSER ET AL. 12). *For a submodular function $\operatorname{cov}(\mathcal{P}, \mathcal{S})$, let \mathcal{S}^* be the optimizer of (3) and $\hat{\mathcal{S}}$ be the set returned by the greedy Algorithm 1, then*
 $\operatorname{cov}(\mathcal{P}, \hat{\mathcal{S}}) \geq (1 - (1 - \frac{1}{k})^k) \operatorname{cov}(\mathcal{P}, \mathcal{S}^*) \geq (1 - \frac{1}{e}) \operatorname{cov}(\mathcal{P}, \mathcal{S}^*)$.

3.4 Convex Relaxation for Inference

Instead of solving the Problem (3) approximately using the greedy Algorithm 1, another approach is to solve exactly a convex relaxation of Problem (3). As f is concave, the constrained problem:

$$\begin{aligned} \max_{\alpha \in \mathbb{R}^n} \quad & \sum_{(\kappa, v) \in \mathcal{P}} v f \left(\sum_{j=1}^n f^{-1}(W_{\kappa j}) \alpha_j \right) \\ \text{such that} \quad & \forall j \in \mathcal{R}, \alpha_j = 0 \\ & \forall j \in \mathcal{X} \setminus \mathcal{R}, 0 \leq \alpha_j \leq 1 \\ & \sum_{j=1}^n \alpha_j = k , \end{aligned} \quad (4)$$

has a concave objective with simple linear constraints, which can be solved efficiently using the Frank-Wolfe algorithm [7] given in Algorithm 2. The inequality constraints are expected to have a sparsifying effect, leading to a vast majority

Algorithm 2 Algorithm Based on Frank-Wolfe

Input: set of items \mathcal{X} , profile \mathcal{P} , similarity matrix \mathbf{W} , # of recommendations k
 $i = 1, \operatorname{tol} = 10^{-6}, \operatorname{maxiter} = 10^6$
 $\mathcal{S} = \mathcal{X} \setminus \mathcal{R}, \alpha^i = \mathbf{0}$
repeat
 $\mathbf{L}^i = -\nabla_{\alpha} \left(\sum_{(\kappa, v) \in \mathcal{P}} v f \left(\sum_{j=1}^n f^{-1}(W_{\kappa j}) \alpha_j \right) \right) \Big|_{\alpha^i}$
 $\alpha^{i+1} = \operatorname{argmin}_{\alpha \in \mathbb{R}^n} \alpha^T \mathbf{L}^i$ (subject to the constraints in (4))
 $\delta = \alpha^{i+1} - \alpha^i$
 $\alpha^{i+1} = \alpha^i + \frac{2}{2+i} \delta$
 $i = i + 1$
until $\operatorname{abs}(\delta) \leq \operatorname{tol}$ **or** $i \geq \operatorname{maxiter}$
Output: indexes of k largest elements in α

of the α_j to be zero. In any case, a solution of cardinality k can always be returned by taking the k items with largest values of α_j .

3.5 Special Cases

The non-linearities of the *saturation function*, if any, are the critical features that will allow us to make the trade-off between cumulated utility and diversity in profile coverage.

3.5.1 0/1 Saturation Function and Covering Problems

We first consider the limiting case where $f(t) = \lim_{\epsilon \rightarrow 0} t^\epsilon$. For the sake of clarity, we assume that $v = 1$ for every $(k, v) \in \mathcal{P}$, and that items are embedded in a metric space.

Covering \mathcal{R} with balls:

Let us assume that W_{ij} is 1 if i lies in the ball of radius ρ centered on j and 0 otherwise, for some fixed radius $\rho > 0$. Then, $\operatorname{cov}(\mathcal{P}, \mathcal{S})$ counts the number of items of \mathcal{R} that are covered by the balls of radius ρ centered on the items of \mathcal{S} . Maximizing $\operatorname{cov}(\mathcal{P}, \mathcal{S})$ with a cardinality constraint on \mathcal{S} corresponds to finding a maximum subset of \mathcal{R} that is covered with k balls of radius ρ centered in points of $\mathcal{X} \setminus \mathcal{R}$. Problem (3) is then a maximal coverage problem.

k -nearest neighbors and clusters:

If \mathbf{W} is the adjacency matrix of a k -nearest neighbor graph, then $\operatorname{cov}(\mathcal{P}, \mathcal{S})$ counts the number of items in \mathcal{R} that are in the k -nearest neighbors of items in \mathcal{S} . Likewise, assume that the items are clustered and that the similarity W_{ij} is 1 if i and j belong to the same cluster and 0 otherwise. Then, $\operatorname{cov}(\mathcal{P}, \mathcal{S})$ is the number of items in \mathcal{S} that are in the same cluster as at least one item in \mathcal{R} .

3.5.2 Linear Saturation Function

Here we consider the case where the *saturation function* is a linear. If f is, say the identity function, then maximizing coverage boils down to choosing the set of items $j \in \mathcal{S}$ such that $\sum_{(\kappa, v) \in \mathcal{P}} v W_{\kappa j}$ is maximal. This corresponds to one way of performing item-based collaborative filtering [17].

4. EXPERIMENTS

In this section, we evaluate our approach in movie recommender settings. We compare our results against the state-of-the-art re-ranking algorithms.

4.1 Baselines

We chose two baselines: Maximal Marginal Relevance (MMR) [4] and Max-Sum Diversification (MSD) [3].

4.1.1 Maximal Marginal Relevance

Given the set of already selected items \mathcal{S} (initialized to the empty set), the MMR algorithm selects an item i^* such that

$$i^* \in \operatorname{argmax}_{i \in \mathcal{X} \setminus (\mathcal{R} \cup \mathcal{S})} \lambda \operatorname{sim}_1(u, i) - (1 - \lambda) \max_{j \in \mathcal{S}} \operatorname{sim}_2(i, j) ,$$

where sim_1 and sim_2 are similarities, and u is related to the user profile. In our settings, this reads:

$$i^* \in \operatorname{argmax}_{i \in \mathcal{X} \setminus (\mathcal{R} \cup \mathcal{S})} \lambda \sum_{(\kappa, v) \in \mathcal{P}} v W_{\kappa i} - (1 - \lambda) \max_{j \in \mathcal{S}} W_{ij} .$$

As the trade-off parameter $\lambda \in [0, 1]$ is decreased, more emphasis is put on the diversity of the resulting set. MMR can be interpreted as a greedy scheme for maximizing a non-monotone submodular objective function, for which the approximation guarantees of Theorem 1 do not apply [11].

4.1.2 Max-Sum Diversification

The Max-Sum diversification (MSD) algorithm is based on the facility dispersion problem, where one aims to find a subset of optimal locations such that the distance between the selected locations is maximized. Like in MMR, the objective function comprises of two terms, a modular relevance term and a supermodular sum of distance diversity term. Formally, MSD returns the set \mathcal{S} of cardinality k , such that:¹

$$\operatorname{argmax}_{\mathcal{S} \subseteq \mathcal{X} \setminus \mathcal{R}} \lambda g(\mathcal{S}) + (1 - \lambda) \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S} - \{i\}} \operatorname{dist}(i, j) \text{ s.t. } |\mathcal{S}| \leq k$$

where λ is the trade-off parameter, $g(\mathcal{S})$ is the utility function and $\operatorname{dist}(i, j)$ is the distance function between item i and j . The problem is NP-Hard, but efficient greedy algorithm with provable approximation guarantees exist [3]. In our settings, the utility $g(\mathcal{S})$ is defined exactly as the sim_1 in MMR and $\operatorname{dist}(i, j)$ as $1 - W_{ij}$.

4.2 Performance Metrics

Our experiments aim to assess the diversity of the recommended set. Even though there has been some work on unifying the performance metrics for diversity, there is no clear consensus, especially in recommendation tasks. So we measure several features of the different solutions in the movie recommender settings, involving relevance, coverage, and popularity bias [14].

Genre Coverage – A diversifying algorithm should produce results that cover different relevant interest groups. One way to measure the coverage of the user interests is to count the number of relevant genres recommended to the user. We define the *Genre Coverage* for the set \mathcal{U} of users as the average ratio of relevant genres recommended to each user.

$$\frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{\left| \bigcup_{i \in \mathcal{S}_u} \operatorname{genres}(i) \cap \bigcup_{i \in \mathcal{R}_u^+} \operatorname{genres}(i) \right|}{\left| \bigcup_{i \in \mathcal{R}_u^+} \operatorname{genres}(i) \right|} ,$$

¹The original formulation of Borodin et al. [3] is slightly different but equivalent.

where, for user u , \mathcal{R}_u^+ is the set of relevant rated movies, \mathcal{S}_u is the set of recommended items, and $\operatorname{genres}(i)$ returns the genres associated with item i .

Intra-List Distance (ILD) – It measures the diversity of the set of recommended items by the mean distance between all pairs of items in this set. In our experiments, we used the Hamming distance between genre vectors:

$$\frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{1}{k(k-1)} \sum_{(i,j) \in \mathcal{S}_u} |\operatorname{genres}(i) - \operatorname{genres}(j)| .$$

Catalog Coverage – It is the fraction of the relevant items that are recommended at least once, across all users. Higher values of catalog coverage indicate that the algorithm counterbalances the popularity bias by covering a large portion of the overall set of items. Formally, it is defined as:

$$\frac{\left| \bigcup_{u \in \mathcal{U}} \mathcal{S}_u^+ \right|}{|\mathcal{X}|} .$$

Popularity Stratified Recall@k – This metric is suggested by Steck [19] to measure the ability of a recommender system to compensate for the popularity bias. As argued earlier, a diversity inducing recommendation system should cover diverse interests and may in turn cover items from the tail of the item-popularity distribution. *Popularity Stratified Recall@k* is defined as:

$$\frac{\sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{S}_u^+} \left(\frac{1}{N_i^+} \right)^\beta}{\sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{T}_u} \left(\frac{1}{N_i^+} \right)^\beta} ,$$

where \mathcal{S}_u^+ is the set of recommended items that are known to be relevant for user u (among the k recommended items), \mathcal{T}_u is the set of items in the test set that are known to be truly relevant for user u , N_i^+ is the number of relevant ratings for item i in the test set and β is a hyperparameter which adjusts for the popularity bias. Higher values of *Popularity Stratified Recall@k* indicate that more relevant movies from the tail distribution are recommended. In our experiments, we used $\beta = 0.5$ and k was set to 5, 10, 20 or 50.

Intra-List Distance (ILD) – Proposed by [26], it measures the diversity of the set of recommended items by the mean distance between all pairs of items in this set. In our experiments, we used the Hamming distance:

$$\frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{1}{k(k-1)} \sum_{(i,j) \in \mathcal{S}_u} |\operatorname{genres}(i) - \operatorname{genres}(j)| .$$

Discounted Cumulative Gain (DCG) – It is a commonly used metric in ranking problems. It measures the relevance of a ranked list by the sum of the graded relevance discounted by the rank of the item. In our experiments, we used:

$$\frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{S}_u} \frac{2^{r_i} - 1}{\log(i+1)} ,$$

where r_i is the *graded* relevance score of the i th item. In our experiments, the i th item is either the i th item entering \mathcal{S}_u for greedy algorithm, or the one with i th largest α_i in the convex relaxation formulation (4).

Precision@k – It is the fraction of relevant items in the recommended list of k items.

$$\frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{1}{k} |\mathcal{S}_u^+ \cap \mathcal{T}_u| .$$

Genre Coverage and *ILD* measure diversity, *Catalog Coverage* and *Stratified Recall@k* mix diversity and relevance, and *DCG* and *Precision@k* measure the relevance. Higher values of the aforementioned metrics indicate a better recommendation list.

4.3 Experimental Protocol

We used two benchmark datasets (i) MovieLens and (ii) Yahoo! Movies to evaluate the proposed algorithm. Following Cremonesi et al. [6], we carried out holdout validation by splitting the data randomly into training and test set such that 3% of the original data goes into testing and remaining goes into training. To reduce the variability in the result, split is carried out five times and the reported results are the average over the five splits. The rating values and the corresponding movies in the training set is used to create the profile \mathcal{P} and the unrated movies in the training set is used as \mathcal{S} . For the purpose of evaluation, whenever necessary, we discretized rating scores to binary values such that rating scores of 4 and 5 are deemed as relevant and otherwise irrelevant.

We estimated the unobserved rating values for MMR and MSD in the training set using item-based collaborative filtering [17] (matrix-factorization based methods gave inferior results). We used the the observed movie ratings to create the user profiles \mathcal{P} and the similarity matrix \mathbf{W} , which is computed by a cosine similarity. For evaluation purposes, we used the original observed rating values or whenever applicable, their binarized version, in the test set. We used $f(t) = t^\gamma$ with $\gamma = \{0.1, 0.5, 0.8\}$ and the limiting case where $\gamma \rightarrow 0$, which corresponds to the ℓ_∞ -norm, for the *saturation function* in the submodular (SUB) setting, and $f(t) = t$ for the modular (MOD) setting.

4.4 Results

4.4.1 MovieLens

MovieLens² 1M dataset contains ratings from 6040 users on 3706 movies (excluding movies with no rating values). Each movie is associated with a set of genres, among 18 distinct categories. Performance of different algorithms on the MovieLens dataset for $k = 10$ is given in Table 1 (all values in %). The relevance-diversity values for different values of k is given in Figure 1.

4.4.2 Yahoo! Movies

Yahoo! Movies³ dataset contains separate training and test set, but we used only the training set due to the unavailability of genre information. The training data contain 211,231 rating values for 7,642 users and 11,915 movies. We removed the movies with missing genres, being left with 187,435 ratings spanning 7,636 users and 8,647 movies. Yahoo! Movies span a total of 25 distinct genres. Table 2 contains the result for Yahoo! Movies for $k = 10$ (all values in %) and Figure 2 contains the relevance-diversity values for different values of k .

²<http://grouplens.org/datasets/movielens/>

³<https://webscope.sandbox.yahoo.com/>

4.4.3 Discussion

It should be noted that the modular version of our algorithm is equivalent to MMR and MSD with $\lambda = 1$. Surprisingly, MMR does not exhibit any trade-off between relevance and diversity metrics as λ is varied. As the λ value is increased from 0.1 to 1, values corresponding to DCG, Precision@k, Catalog Coverage and Genre Coverage remain the same or decrease only marginally. In effect, MMR does not recommend very relevant and less diverse movies by weighting the relevance term highly. The same trend can be noted for Yahoo! Movies as well. On the other hand, MMR compensates for popularity bias by recommending less popular movies covering a larger spectrum of the set as the diversity term is weighted high, thus increasing Stratified Recall@k and Catalog Coverage. But for MSD, on MovieLens, as the λ value is increased, the recommendation list becomes more relevant but nothing can be inferred about diversity. But on Yahoo! Movies, trade-off between relevance (DCG) and diversity (Genre Coverage) is clearer. As the λ value is increased, recommended list becomes more relevant and less diverse. However, there is no clear indication that MSD compensates for popularity bias, even though it recommends movies spanning large spectrum of movies from the set. The modular (equivalent to item-based collaborative filtering) version perform as good as the MMR and MSD versions.

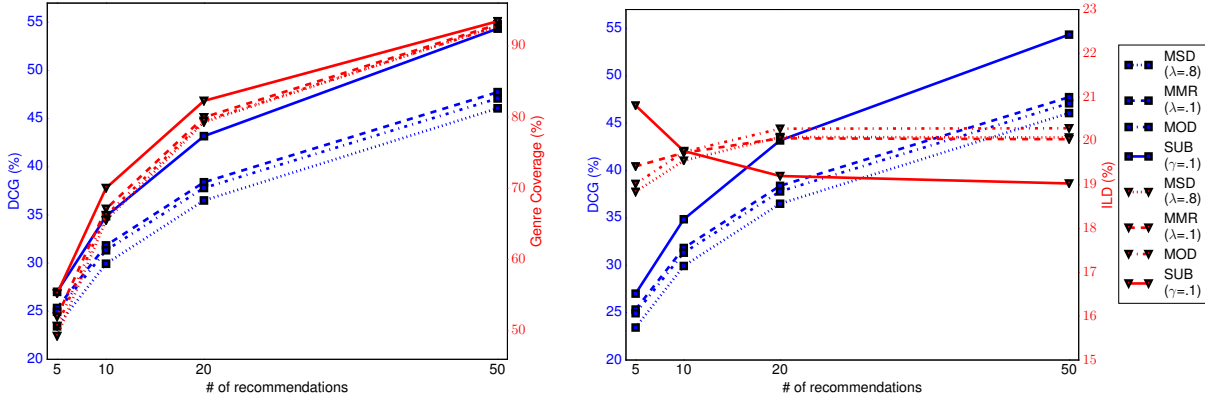
For our algorithm, as the γ value is increased, both the relevance and the diversity values increase and the increase is more significant. The best in-class relevance-diversity values are obtained for $\gamma = 0.1$. For MovieLens, we see a 3% increase for both DCG and Genre Coverage metrics compared to the second best algorithm. For Yahoo! Movies, $\gamma = 0.1$, gives the best diversity value (Genre Coverage) for a marginally smaller value of relevance (DCG). On MovieLens, our proposed algorithm returns greater number of movies from the tail distribution which are collectively distinct, i.e. covering large spectrum of movies, as indicated by the larger values of Stratified Recall@k and Catalog Coverage. But this effect is not very evident on Yahoo! Movies. As the γ value approaches zero the quality of the recommendation list deteriorates. Convex relaxation based algorithm performance is on par with other algorithms, but we found it computationally more expensive. The performance of relevance-diversity metrics for varying recommendation size is given in Figure 1 and 2. On MovieLens, submodular (solid blue line) algorithm returns greater number of relevant movies (square markers) which are diverse (red solid line with triangle markers on the left plot), whereas MMR and MSD returns diverse movies which are less relevant (red non-solid lines with triangle markers on the left plot). It can also be noted that, as the k value increases, our algorithm recommends the most diverse, but relevant movies (higher DCG & Genre Coverage) whereas MMR and MSD recommend the most diverse, but irrelevant movies (higher ILD, lower Genre Coverage & DCG; see right plot in Figure 1). However, in Yahoo! Movies, even though the diversity metrics (Genre Coverage & ILD) are marginally higher for the proposed algorithm, the relevance values are close to MMR and MSD algorithms.

4.4.4 Effect on Eclectic Users

The problem of diverse recommendation is more critical for users with eclectic interests. Here we study the effect of diversification on eclectic users by sampling an arbitrary

Table 1: Experimental Results on MovieLens (k=10)

	MMR $\lambda=0.1$	MMR $\lambda=0.5$	MMR $\lambda=0.8$	MSD $\lambda=0.1$	MSD $\lambda=0.5$	MSD $\lambda=0.8$	MOD	SUB $\gamma=0.8$	SUB $\gamma=0.5$	SUB $\gamma=0.1$	SUB $\gamma \rightarrow 0$	SUB conv $\gamma=0.5$
Genre Coverage	67.12	66.34	66.26	66.84	65.35	65.50	66.23	66.31	66.56	70.04	63.64	66.27
ILD	19.73	19.75	19.75	19.22	19.44	19.55	19.75	19.78	19.88	19.77	18.78	19.70
Catalog Coverage	6.55	6.20	6.19	7.92	6.37	6.51	6.18	6.26	6.57	10.49	5.37	6.76
Stratified Recall@k	7.78	7.57	7.56	8.00	7.36	7.60	7.56	7.63	7.90	10.27	6.23	7.89
DCG	31.83	31.38	31.35	25.09	25.33	29.93	31.33	31.51	32.10	34.86	24.42	31.95
Precision@k	4.06	3.98	3.98	4.02	3.86	3.97	3.98	4.01	4.10	4.60	3.19	4.08


Figure 1: Relevance-Diversity values for the MovieLens according to k

number of eclectic users from the MovieLens and Yahoo! Movies. We define eclectic users based on their affinity towards many diverse items, as measured by their number of positive ratings and the mean similarity between rated items, based on W_{ij} , below some threshold. We selected 209 users from MovieLens by setting the minimum number of relevant ratings to 100 and the mean similarity between already rated movies below 0.2, and 109 users from Yahoo! Movies by setting the minimum number of relevant ratings to 50 and the mean similarity between already rated movies below 0.1. The experimental results are given in Tables 3 and 4 for $k = 10$. The proposed algorithm significantly improves the DCG and Genre Coverage values compared to the second best diversification algorithm. In case of MovieLens, DCG and Genre Coverage values improve by 5% and 6% respectively whereas in case of Yahoo! Movies for the same relevance value we get 5% improvement. The relevance and diversity metric values for different recommendation size for eclectic users is given in Figures 3 and 4. As we can see from the figures, the blue solid line with square markers and red solid line with triangle markers dominate the DCG-Genre Coverage graph for varying sizes of k . On eclectic user set, as the recommendation size grows, MSD and MMR algorithm return movies which are diverse with respect to each other (higher ILD values) but less relevant to the users (smaller DCG and Genre Coverage) compared to the proposed algorithm.

5. CONCLUSION

We presented a new criterion that captures both relevance and diversity for recommendation. The criterion can be approximately optimized with an efficient greedy algorithm; the algorithm can be applied in any recommendation scenario where we have access to similarities between items

and a subset of items which are known to be of interest to the user. Experiments on benchmark datasets showed that the algorithm performs well, both in terms of relevance and diversity compared to a strong baseline.

6. ACKNOWLEDGMENT

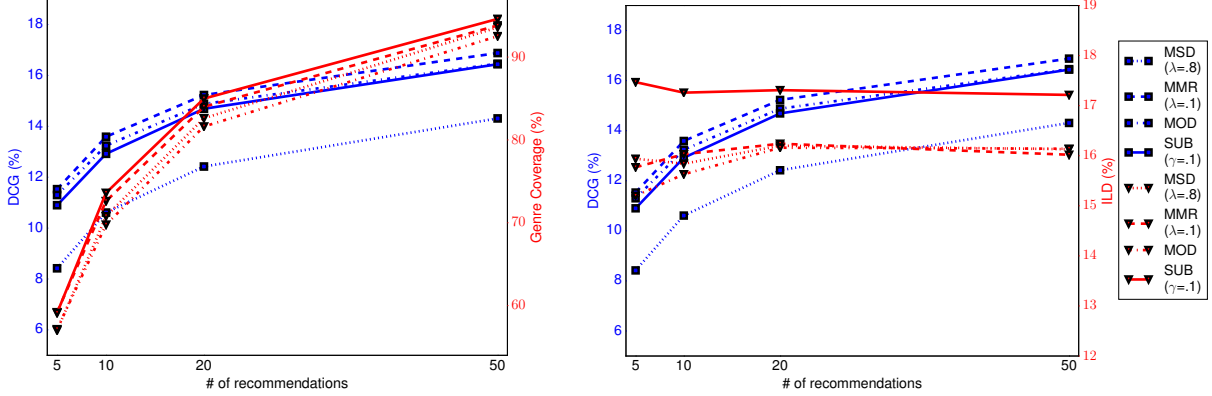
This work was carried out and funded in the framework of the Labex MS2T. It was supported by the Picardy Region and the French Government, through the program “Investments for the future” managed by the National Agency for Research (Reference ANR-11-IDEX-0004-02). We thank our anonymous reviewers for their valuable suggestions and comments.

References

- [1] Rakesh Agrawal, Sreenivas Gollapudi, Alan Halverson, and Samuel Ieong. Diversifying search results. In *WSDM*. ACM, 2009.
- [2] Azin Ashkan, Branislav Kveton, Shlomo Berkovsky, and Zheng Wen. Optimal greedy diversity for recommendation. In *IJCAI*, 2015.
- [3] Allan Borodin, Hyun Chul Lee, and Yuli Ye. Max-sum diversification, monotone submodular functions and dynamic updates. In *PODS*, pages 155–166. ACM, 2012.
- [4] Jaime Carbonell and Jade Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *SIGIR*, pages 335–336. ACM, 1998.
- [5] Olivier Chapelle, Shihao Ji, Ciya Liao, Emre Velipasaoglu, Larry Lai, and Su-Lin Wu. Intent-based diversification of web search results: metrics and algorithms. *Information Retrieval*, 14(6):572–592, 2011.

Table 2: Experimental Results on Yahoo! Movies (k=10)

	MMR $\lambda=0.1$	MMR $\lambda=0.5$	MMR $\lambda=0.8$	MSD $\lambda=0.1$	MSD $\lambda=0.5$	MSD $\lambda=0.8$	MOD	SUB $\gamma=0.8$	SUB $\gamma=0.5$	SUB $\gamma=0.1$	SUB $\gamma \rightarrow 0$	SUB conv $\gamma=0.5$
Genre Coverage	72.61	70.42	69.96	72.80	71.67	70.78	69.77	70.44	71.75	73.61	66.59	69.58
ILD	16.03	15.75	15.67	14.99	15.93	15.84	15.63	15.72	15.82	15.52	14.87	14.84
Catalog Coverage	2.96	2.61	2.55	3.15	2.78	2.63	2.54	2.64	2.87	3.03	1.91	1.71
Stratified Recall@k	16.44	15.08	14.83	15.36	15.51	14.85	14.75	15.33	16.48	16.00	11.36	14.45
DCG	13.59	13.37	13.27	9.64	10.30	10.61	13.24	13.47	13.88	12.92	9.75	13.02
Precision@k	1.62	1.55	1.53	1.40	1.55	1.51	1.53	1.56	1.63	1.50	1.21	1.53

**Figure 2: Relevance-Diversity values for the Yahoo! Movies according to k** **Table 3: Experimental Results on MovieLens (Eclectic Users $k = 10$)**

	MMR $\lambda=0.1$	MMR $\lambda=0.5$	MMR $\lambda=0.8$	MSD $\lambda=0.1$	MSD $\lambda=0.5$	MSD $\lambda=0.8$	MOD	SUB $\gamma=0.8$	SUB $\gamma=0.5$	SUB $\gamma=0.1$	SUB $\gamma \rightarrow 0$	SUB conv $\gamma=0.5$
Genre Coverage	59.54	59.41	59.39	58.02	59.03	59.40	59.39	59.46	59.51	64.54	59.87	59.01
ILD	20.69	20.69	20.69	20.23	20.45	20.69	20.69	20.72	20.82	21.03	20.55	20.34
Catalog Coverage	3.12	3.12	3.12	2.97	3.19	3.12	3.12	3.16	3.29	4.13	2.61	3.36
Stratified Recall	4.62	4.62	4.62	4.32	4.67	4.62	4.62	4.67	4.87	6.33	4.06	4.89
DCG	87.41	87.45	87.45	67.43	86.91	89.23	87.44	87.97	90.02	95.32	71.05	90.22
Precision@k	11.74	11.76	11.76	11.01	11.77	11.76	11.76	11.86	12.21	13.21	9.64	12.13

Table 4: Experimental Results on Yahoo! Movies (Eclectic Users $k=10$)

	MMR $\lambda=0.1$	MMR $\lambda=0.5$	MMR $\lambda=0.8$	MSD $\lambda=0.1$	MSD $\lambda=0.5$	MSD $\lambda=0.8$	MOD	SUB $\gamma=0.8$	SUB $\gamma=0.5$	SUB $\gamma=0.1$	SUB $\gamma \rightarrow 0$	SUB conv $\gamma=0.5$
Genre Coverage	55.88	54.12	54.03	56.66	56.46	55.74	53.91	55.47	57.22	60.03	45.93	56.32
ILD	14.57	14.03	13.98	14.58	14.43	14.32	13.96	14.26	14.68	14.00	11.78	14.49
Catalog Coverage	0.28	0.25	0.25	0.31	0.26	0.25	0.24	0.27	0.29	0.27	0.19	0.28
Stratified Recall@k	4.46	3.97	3.88	5.75	4.30	4.00	3.82	4.38	5.17	5.62	3.47	4.90
DCG	19.92	18.40	18.15	18.13	15.87	18.89	17.95	19.34	21.20	19.54	12.53	20.15
Precision@k	2.59	2.39	2.35	2.99	2.42	2.44	2.33	2.53	2.88	2.59	1.63	2.66

- [6] Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. Performance of recommender algorithms on top-n recommendation tasks. In *RecSys*, pages 39–46. ACM, 2010.
- [7] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110, 1956.
- [8] Sreenivas Gollapudi and Aneesh Sharma. An axiomatic approach for result diversification. In *WWW*, pages 381–390. ACM, 2009.
- [9] Jingrui He, Hanghang Tong, Qiaozhu Mei, and Boleslaw Szymanski. Gender: A generic diversified ranking algorithm. In *NIPS*, pages 1142–1150, 2012.
- [10] Neil J Hurley. Personalised ranking with diversity. In *RecSys*, pages 379–382. ACM, 2013.
- [11] Hui Lin and Jeff Bilmes. A class of submodular functions for document summarization. In *ACL*, pages 510–520, 2011.
- [12] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 14(1), 1978.
- [13] Jinoh Oh, Sun Park, Hwanjo Yu, Min Song, and Seung-Taek Park. Novel recommendation based on personal popularity tendency. In *ICDM*, pages 507–516. IEEE, 2011.

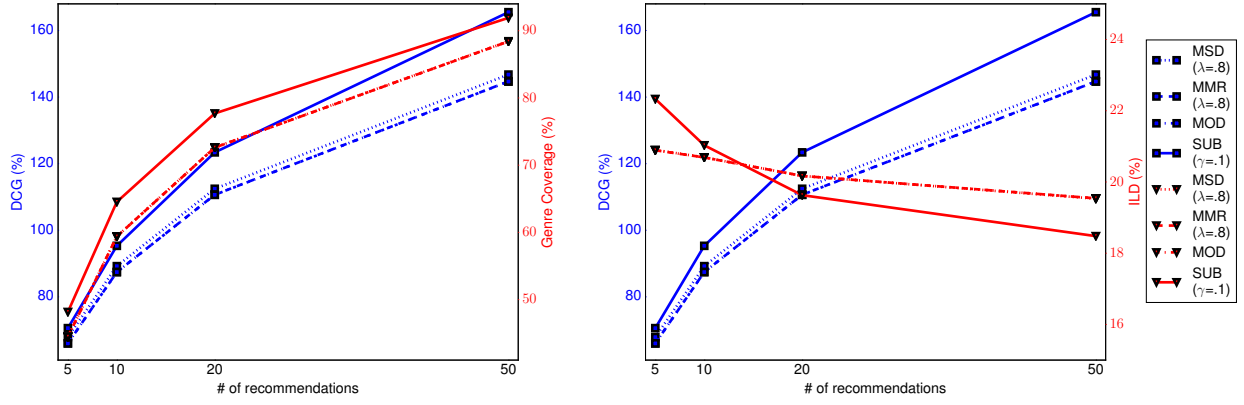


Figure 3: Relevance-Diversity values for the MovieLens (Eclectic Users) according to k

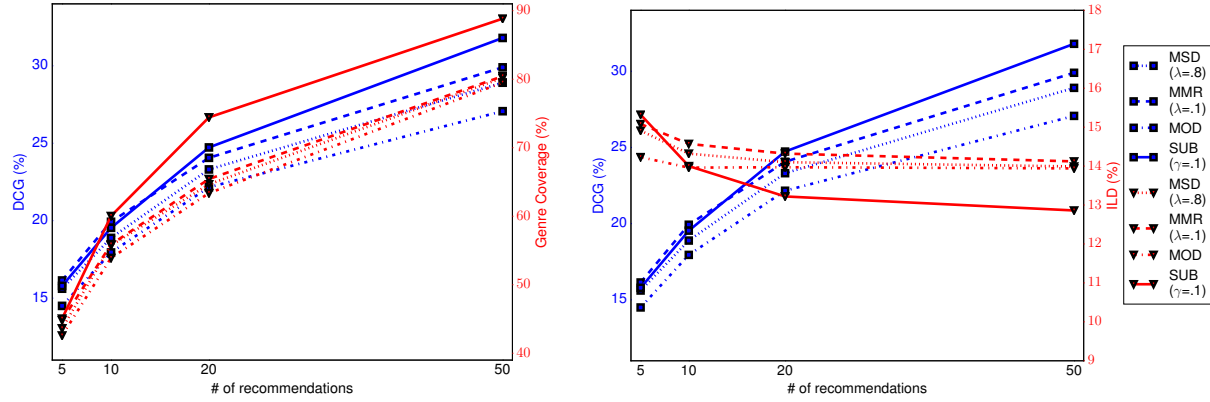


Figure 4: Relevance-Diversity values for the Yahoo! Movies (Eclectic Users) according to k

- [14] Bruno Pradel, Nicolas Usunier, and Patrick Gallinari. Ranking with non-random missing ratings: influence of popularity and positivity on evaluation metrics. In *RecSys*, pages 147–154. ACM, 2012.
- [15] Filip Radlinski, Robert Kleinberg, and Thorsten Joachims. Learning diverse rankings with multi-armed bandits. In *ICML*, pages 784–791. ACM, 2008.
- [16] Rodrygo LT Santos, Craig Macdonald, and Iadh Ounis. Exploiting query reformulations for web search result diversification. In *WWW*. ACM, 2010.
- [17] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *WWW*. ACM, 2001.
- [18] Karen Spärck-Jones, Stephen E Robertson, and Mark Sanderson. Ambiguous requests: implications for retrieval tests, systems and theories. In *SIGIR Forum*. ACM, 2007.
- [19] Harald Steck. Item popularity and recommendation accuracy. In *RecSys*, pages 125–132. ACM, 2011.
- [20] Ruilong Su, Li’Ang Yin, Kailong Chen, and Yong Yu. Set-oriented personalized ranking for diversified top-n recommendation. In *RecSys*, pages 415–418. ACM, 2013.
- [21] Hanghang Tong, Jingrui He, Zhen Wen, Ravi Konuru, and Ching-Yung Lin. Diversified ranking on large graphs: an optimization viewpoint. In *SIGKDD*, pages 1028–1036. ACM, 2011.
- [22] S. Vargas and P. Castells. Rank and relevance in novelty and diversity metrics for recommender systems. In *RecSys*, pages 109–116. ACM, 2011.
- [23] Saúl Vargas, Linas Baltrunas, Alexandros Karatzoglou, and Pablo Castells. Coverage, redundancy and size-awareness in genre diversity for recommender systems. In *RecSys*, pages 209–216. ACM, 2014.
- [24] Le Wu, Qi Liu, Enhong Chen, Nicholas Jing Yuan, Guangming Guo, and Xing Xie. Relevance meets coverage: A unified framework to generate diversified recommendations. *ACM Trans. Intell. Syst. Technol.*, 7(3), 2016.
- [25] Benyu Zhang, Hua Li, Yi Liu, Lei Ji, Wensi Xi, Weiguo Fan, Zheng Chen, and Wei-Ying Ma. Improving web search results using affinity graph. In *SIGIR*, pages 504–511. ACM, 2005.
- [26] Mi Zhang and Neil Hurley. Avoiding monotony: improving the diversity of recommendation lists. In *RecSys*, pages 123–130. ACM, 2008.
- [27] Xiaojin Zhu, Andrew B Goldberg, Jurgen Van Gael, and David Andrzejewski. Improving diversity in ranking using absorbing random walks. In *HLT-NAACL*, pages 97–104, 2007.