Non-personalized Movie Recommendation by Maximum k-Coverage

Nícollas Silva[1], Diego Carvalho[2], Adriano C. M. Pereira[1], Fernando Mourão[3], Leonardo Rocha[2]

[1] Universidade Federal de Minas Gerais, Brazil
ncsilvaa@dcc.ufmg.br adrianoc@dcc.ufmg.br
[2] Universidade Federal de São João del-Rei, Brazil
lcrocha@ufsj.edu.br dcarvalho@ufsj.edu.br
[3] Seek AI Labs
fernando.mourao@catho.com

Abstract. Turning first-time users into returning ones is a major task for the success of e-commerce systems. However, providing effective recommendations for these users remains as an open challenge for the area due to the absence of consumption information. In this context, non-personalized RSs emerge as the main approach adopted in real scenarios. Such approaches are based on the premise that the consumption is generally biased towards items that arouse interest in the majority of a population. Despite being valid for mass consumption, by adopting this premise RSs fail to help users interested in items different from the common taste. In this context, this work proposes a new RSs based on Maximum k-Coverage strategy to merge popular and non-popular items in order to retain different profiles of first-time users. The premise of this approach is that maximizing diversity, while maintaining the relevance of the recommended items, satisfies the preferences of different user profiles. Indeed, the results show a mean gain of 13.5% w.r.t. utility and diversity, when compared to traditional strategies based on Popularity, Best Rated and Recent Items. In addition, our results indicate that the proposed strategy is able to smooth the popularity bias in recommendations, satisfying at least 97% of different users.

Categories and Subject Descriptors: H.3.3 [**Retrieval tasks and goals**]: Recommender Systems

Keywords: Recommender Systems, Maximum Coverage, Ramp-up Problem, Diversify

1. INTRODUCTION

Recommendation Systems (RSs) are among the most important tools used to help users in the decision-making process [1]. These tools concern to satisfy properly the needs of existing users in several real domains. However, under the business perspective, a relevant scenario has been neglected by RSs: the adhesion of first-time users. Retaining these users in the system requires addressing issues such as: *Considering a product catalog, which items have the highest potential to turn a given first-time user into a returning one?*; and *How to recommend available items in order to retain the maximum number of first-time users?*. This problem is known in the literature in two ways: (1) Cold-Start problem; and (2) Ramp-up problem. The Cold-Start problem is related to generating recommendations for new users, whose consumption history is small and little relevant [2, 3]. On the other hand, the Ramp-up problem is even more complicated, since it is related to first-time users, for whom there is still no information in the system [4, 5]. In the absence of consumption histories, traditional strategies of Collaborative Filtering (CF) and Content-Based (CB) RSs are not able to provide personalized recommendation models and, consequently, they cannot issue good recommendations. Hence, the so-called non-personalized RSs emerge as the main recommendation strategies for this scenario.

Non-personalized RSs are strategies that exploit metadata derived from the items and/or consumption in a domain, such as items popularity, best-rated items or consumer recurrence [6]. Such approaches are based on the premise that the consumption is generally biased towards one or more of those dimensions, allowing achieving high efficiency through simple and unsupervised prediction strategies. Although this assumption is valid for the "mass consumption", in which users consume items that emerge as a common interest of the majority, we cannot assume it for "niche consumers", who are interested in non-popular items. Therefore, straightforward recommendation of popular, recent or best-rated products for "niche consumers" may not represent an appropriate strategy. Properly identifying items that satisfy this user profile, however, is crucial for various scenarios, since these users may represent more than half of the profit on e-commerce systems, composing what is known as the long tail consumption [7].

In this study, we introduce a novel non-personalized recommendation strategy that best suits the needs of both mass and niche consumers. The proposed strategy is a practical application of the well-know problem of Maximum Coverage on recommendation scenarios [8]. The Maximum Coverage strategy aims to identify a subset of unrated items in a specific domain that can be potentially relevant to the largest number of different users. The premise, in this case, is that it is possible to recommend items that satisfy the largest number of users, bringing also more diversity when compared to strategies that just recommend the popular, recent or best rated items. Furthermore, the chances of a specific user to find at least one item that suits his/her personal needs increases since the strategy aims to cover a wide range of distinct preferences. It is important to note that recommending items for mass consumption users is considered an easy task, since we have available a large volume of information about them. On the other hand, recommending items to satisfy niche consumers remains as a challenge and has implications for several scenarios.

In order to assess the practical relevance of the recommendations generated by the Maximum coverage strategy, we evaluate these recommendations considering four quality dimensions: accuracy, precision, recall and diversity. Further, we consider two data collections from *MovieLens*, related to movie recommendation, and we compare our approach against three non-personalized strategies: (1) popularity; (2) best-rated; (3) recent items and (4) random popularity. Indeed, the results corroborate the main premise assumed in this article. Applying the Maximum Coverage strategy tends to bring more diversity to domain users, and it is able to satisfy at least once 97% of the users in each data collection. Recommendations generated by Maximum Coverage present, in average, 13% of items that are not considered popular, whereas the other strategies recommended only popular items. Considering the trade-off between accuracy and diversity, our approach presents gains up to 13.5% over the baseline strategies. Focusing the analysis on the top-5, 10 and 20 recommendation tasks, which correspond to classical tasks in real scenarios, our approach achieved gains up to 23.8%, 18.5% and 14.3%, respectively. Besides these results, recommendations based on Maximum Coverage are comparable to those provided by the other strategies w.r.t. classical accuracy metrics.

Contributions. We present a new strategy focused on non-personalized recommendations based on Maximum Coverage that aims to retain mass and niche users. We also proposed and adopted a new evaluation methodology that considers the accuracy and diversity, reflecting the desirable real scenarios. We emphasize that the contributions of this study are particularly relevant for scenarios on which the adopted RSs are focused in recommend non-popular items that aim to reach users with different preferences. Finally, it is important to mention that we did not find in the literature studies that address the issues raised on the applicability of Maximum Coverage problem for the ramp-up problem.

The remainder of this article is organized as follows. In Section 2, we describe the main concepts hereby exploited, as well as some related work. Then, Section 3 presents the Maximum k-Coverage algorithm, adapted for the ramp-up problem. In Section 4, we present the experimental design used to evaluate the Maximum Coverage strategy. Later, in Section 5 we present the results related to the evaluation of the proposed strategy. Finally, we present the conclusions in Section 6.

2. RELATED WORK

Turning first-time users into returning ones is a major problem for the success of e-commerce systems [9–11]. The literature in Recommender Systems defines this problem as: (1) Cold-Start problem; and (2) Ramp-up problem. The Ramp-up problem is commonly deemed as a variation of the Cold-Start problem [3]. Despite being closely related, both problems should be addressed differently. Whereas the Cold-Start problem deals with users with small consumption histories (i.e., inactive or new users), in the Ramp-up there is no consumption information about the users (i.e., first-time users). For e-commerce systems, any information is better than none and, for this reason, Ramp-up is a major challenge for Recommender Systems. The challenge raises due to the absence of information about the consumption history of users, impairing or precluding the consolidation of effective models. We identified three main categories of RSs designed to deal with the ramp-up problem: (1) Interactive RSs; (2) Hybrid RSs; (3) Non-personalized RSs.

A straightforward strategy to overcome this problem is to adopt the so-called Interactive RSs [12,13], wherein the system acquires further information about the users by applying small questionnaires before recommending the first items [4]. In this direction, *Zhou et. al* [14] modified a matrix factorization method according to the responses provided by the users to a questionnaire. Such as search engines, the main drawback of this strategy is that the quality of the recommendations depends on the information provided by the users. Since users may struggle to properly define their needs in some domains, adopting Interactive RSs may become a challenge. The second strategy comprises using external information (e.g., social, demographic or personal information) about users to build preliminary profile models through hybrid strategies, such as *CF-content based*, *CF-demographic based* and *CF-social based* [3,15, 16]. However, this approach may be not valid for several e-commerce scenarios, where users are interested on buying/consuming items without providing any personal, social or demographic information.

The third category of strategies includes non-personalized RSs, which recommend items based on global information derived from the available data or user behavior in the system [3]. Hence, the issued recommendations are the same for all users [17]. Simplicity, generalization capability, domain independence and good performance are characteristics that make non-personalized RSs the main strategy to address the ramp-up problem in practical scenarios. Most of these RSs exploits one of three pieces of information: item popularity, best rated items or release date. Popularity-based RSs consist of a simple and intuitive strategy that always recommends the most popular items of a dataset, regardless the target user. Item popularity is estimated by the number of distinct users who have consumed each item in the past. This approach has proven to be effective in several domains [3]. Non-personalized RSs based on the best rated items generate a ranking of items ordered decreasingly by the mean rating received by each item in the system. The assumption of this approach is that the best evaluated items tend to interest many users [3]. Finally, RSs based on release date present to users the items more recently added to the system. Despite being a simple strategy, this non-personalized recommendation strategy is very popular in real scenarios, since users normally appreciate recent items or novelties [2].

In general, existing non-personalized RSs rely on the premise that consumption is strongly skewed towards one of these dimensions in scenarios wherein RSs are applicable, allowing high predictive efficacy. However, such premise may be valid only for "mass consumption", in which users consume items that arouse interest in the majority of a population. The same cannot be said about "niche consumption", in which users are interested in items different from the common taste. Thus, recommending items belonging to mass consumption to potential niche consumers would not represent a proper strategy to retain the latter in the system. Given the relevance of niche users in real scenarios, this work proposes a novel non-personalized solution for the ramp-up problem able to retail both mass and niche consumers. The proposed strategy applies the Maximum Coverage in order to identify k distinct items that maximize the coverage of users who have consumed at least one of these items.

The use of Maximum Coverage in the recommendation task is not totally novel [18,19]. In [18], the authors proposed a method based on Maximum Coverage to maximize the chances of a customer to make purchases. In turn, [19] models items as a complete similarity graph and applied the Maximum Coverage on it. The goal is to identify the k items unknown by each user u similar to the maximum number of items relevant to u. In this approach, the authors consider aims to enhance simultaneously diversity and utility. The present work differs from both ones due to two reasons: (1) we focus on the ramp-up problem, considering the main goal of retaining first-time users in the system while diversifying recommendations; and (2) we maximize the user coverage rather than the item one.

3. MAXIMUM USER COVERAGE IN RECOMMENDATION

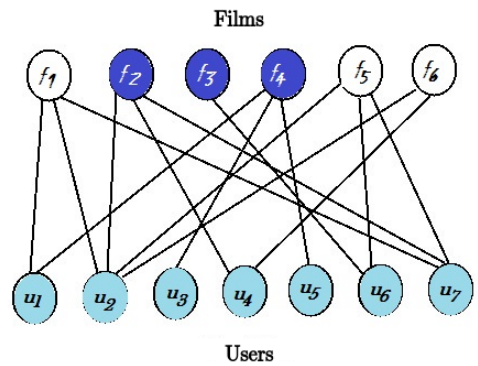
The proposed strategy aims to issue recommendations that suit the needs or interests of both mass and niche consumers. In this sense, let us formalize this and let $U = \{u_1, ..., u_m\}$, be the users that previously have consumed items and let $F = \{S_1, ..., S_n\}$ denote the items in the items selection. If we associate the users that have purchased item S_i , we can view S_i as a set of those users. The family F represents the items, where u_j is in S_i if and only if user u_j has consumed item S_i. Our objective now is to establish the k sets $F^* = \{S_{i1}, ..., S_{ik}\}$ that together contain as many different users u_j as possible. This is a well-known problem from Combinatorial Optimization denoted the Maximum k-Coverage problem and is formally defined as follows:

Maximum k-Coverage
Instance: A universe of elements $U = \{u_1, ..., u_m\}$, an integer value k and a family of sets $F = \{S_1, ..., S_n\}$, where each set S_i is a subset of U.
Objective: Find a subfamily $F^* \subseteq F$ such that $|F^*| \leq k$ and the number of covered elements $|\bigcup_{S \in F^*} S|$ is maximised, i.e. using up to k sets, cover as many elements as possible.

In general, we can say that the goal of Maximum Coverage is to find a subset of F^* of items, such that $|F^*| \leq k$, that maximizes the number of different users reached. In other words, we want to find the k movies that interested the most different users and recommend them to end users. Note that this problem is not interested in the consumption past of a specific user, but rather of all users of the data collection. Figure 1(a) depicts the movie context and can be replicated for various scenarios in the literature. In this context, users evaluate movies using a scale from 1 to 5. The value "0" indicates that the user does not evaluate the movie. Therefore, we can model the set of users and items presented in a system as a bipartite graph $G = (\{U, I\}, E)$, where U represents the users, I the items and E denotes the relationship between items and users (i.e., which user have consumed each item). The computational representation of this model can use any graph structure, such as adjacent matrix and incidence list. The result of this modeling is illustrated in Figure 1(b).

$$\begin{array}{cc} & \textbf{Films} \\ \textbf{Users} & \begin{bmatrix} 3 & 0 & 0 & 3 & 0 & 0 \\ 2 & 2 & 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 & 2 & 0 \\ 2 & 1 & 0 & 0 & 2 & 0 \end{bmatrix} \end{array}$$

(a)



(b)

Fig. 1. An illustrative example of the recommendation scenario, where Maximum k-Coverage can be applied.

Algorithm 1 GREEDY-MAX-COVERAGE(U,k,F)

1: $R \leftarrow U$
2: $F^* \leftarrow \emptyset$
3: **for** i from 1 to k **do**
4: $\quad S \leftarrow max_{S \in (F \setminus F^*)} |S \cap R|$
5: $\quad F^* \leftarrow F^* \cup S$
6: $\quad R \leftarrow R \setminus \{S\}$
7: $\quad$ **if** $|R| = 0$ **then**
8: $\quad\quad$ **break**
9: $\quad$ **end if**
10: **end for**
11: **return** F^*

The Maximum Coverage problem is a variant of the well-studied problem of Vertex Cover [20]. Unfortunately, these problems belong to the NP-complete class and it is unknown an optimal solution to solve them in polynomial time. However, a simple greedy heuristic that, at each iteration, finds the item that maximizes the number of users approximates by 63% the maximum number of users that can be covered, as demonstrated in [21]. Algorithm 1 presents this heuristic such as implemented for the recommendation task. This algorithm is implemented with time complexity $O(kmn)$, where k is the number of recommended items, m the number of users and n the number of items.

A preliminary assessment on the proposed strategy may lead us to assume the wrong conclusion that it simply retrieves the most popular items, since the goal is to find the items that coverage the largest number of users. However, observing the greedy algorithm presented in Algorithm 1, it is possible to realize that the chosen items are the ones that maximize the number of users at each iteration. In other words, the covered users are excluded from the set R, which corresponds to the analyzed users (line 6), so that the S item that has not been previously selected ($S \in F \setminus F^*$) is selected and has the largest intersection with uncovered users ($|S \cap R|$), as shown in line 4. Consequently, as the number of iterations approximates to k, the selected items become less popular. Despite not being consumed by many users, these items are the ones that maximize the coverage of users. In the example presented in Figure 1(b), the proposed algorithm first selects item f_2, which is the most popular one. Since users u_2, u_4, u_7 are covered by f_2, the algorithm removes them. At the next iteration, item f_4 is selected and users u_1, u_3, u_5 are removed. Finally, the algorithm selects item f_3, covering user u_6. Notice that f_3 is not popular, because only user u_6 has consumed it.

4. EXPERIMENTAL PROJECT

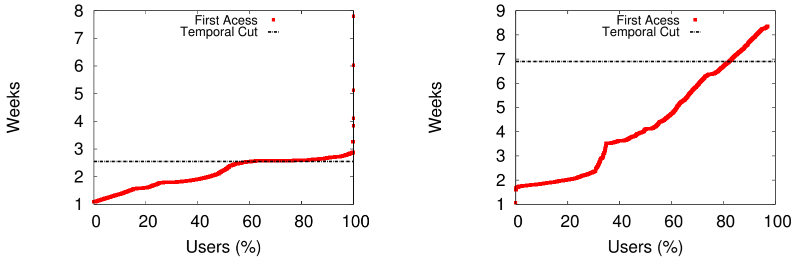
This section presents the experimental project designed to evaluate the applicability of Maximum Coverage strategy to the ramp-up problem. First, we present the process to select first-time users in MovieLens datasets. Then, we discuss the quality requirements of RSs, emphasizing the need to get useful and diversified recommendations. We measure utility of recommendations using classical metrics, such as accuracy, precision and recall. Moreover, we adopt the framework proposed by Vargas et. al [22] to measure diversity. Finally, we present our validation methodology, which aims to characterize and evaluate the generated recommendations by each strategy, considering the trade-off between utility and diversity.

4.1 First-time users data consolidation

Given the relevance of these scenarios, we chose MovieLens 1M and MovieLens 10M datasets. Both collections were compiled by GroupLens and contains, respectively, 1 million and 10 millions explicit ratings assigned by users to movies from various categories. Users rated each movie in a range from 1

Datasets	Users	Items	Sparsity
MovieLens 1M	6,040	3,682	95.82%
MovieLens 10M	69,878	10,677	98.66%

Table I. Datasets.



(a) ML-1M (b) ML-10M

Fig. 2. Temporal data split of each data set.

to 5 stars. In each data set, there are at least twenty ratings assigned by each user to her/his movies of interest. This approach results in a highly sparse data sets, as shown in Table I.

For this work, the set of first-time users was defined considering a temporal analysis of user's consume in the data set. Basically, we perform a temporal split in the data in a specific week, which was able to establish that around 15% to 20% of users would exist in the system after this data separation. This set of users is considered as first-time users and all information of item consumption of them were excluded to represent the ramp-up problem. Figures 2(a) e (b) show the temporal split for scenarios of ML-1M e ML-10M, respectively, considering the first user access in the system. With these data transformations, we have 1,277 and 10,633 first-time users in ML-1M e ML-10M datasets, respectively, which account for 20.25% e 15.21% of total users of each data set.

4.2 Baselines Selected

Considering the effectiveness of non-personalized strategies to mitigate the Ramp-up Problem, this work has selected some state-of-the-art strategies such as *baselines*. They are:

(1) **Popularity:** aims to recommend the k most popular items in the domain, on the assumption that items that interest a large number of users cover the different preferences. Basically, the popularity of an i item is estimated by the number of distinct users who consumed the i.

(2) **TopRated:** is the strategy to recommend the k best evaluated domain items. Basically, the average grade of each item is calculated considering the total number of users and the k items with the highest grades are recommended. The assumption of this approach is that the items most valued tend to interest more distinct users.

(3) **Recent Items:** consists of recommending the k last items consumed by users in the domain. The assumption of this approach is that users tend to be interested in items that are fashionable and are currently consumed.

(4) **Random Popularity:** aims to recommend k random items within the group of items that are rated as popular. The popular items group is defined as the items present at the head of the popularity distribution, which are a percentage of items in the domain [23]. This strategy is not used in practice, but is used to compare whether proposed strategies are effective in selecting potentially relevant items, or whether only a random selection would be sufficient.

4.3 Quality Requirements

The main works related to evaluation of RSs [3, 24] describe three quality requirements: utility, novelty and diversity. The concept of **utility**, which is considered the main goal of a RSs, refers to the capacity of RSs to identify and present to users items that better correspond to their preferences [25]. **Novelty** refers to how distinct the recommended items are in relation to the previous items consumed by a user [26]. Some works consider that a RSs is valuable if the system is able to offer new items or information to users [27, 28]. **Diversity** is related to how different the recommended items are between each other [22]. Although there exists a variety of domains where RSs are applied, the recommendations are, in general, not well diversified [14, 29].

In this work, the RSs utility is calculated using the classical metrics of accuracy, precision and recall. Accuracy consists in simply enumerating how many ranked items were really consumed by users, considering the test data set. Precision represents the probability of an item to be relevant to a user, which can be defined by the ratio of relevant recommend items (N_{rs}) by the total number of recommended items (N_s), as specified by Equation 1.

$$Precision = \frac{N_{rs}}{N_s} \tag{1}$$

On the other hand, recall is defined by the ratio of the number of relevant recommended items by the total number of relevant items in the test set (N_r). Basically, recall represents the item relevance probability to be selected, as specified by Equation 2. In this work, we consider an item as relevant for a user if the user rating assigned to this item is greater or equal to the user's average rating.

$$Recall = \frac{N_{rs}}{N_r} \tag{2}$$

The *novelty* of a information generally refers to how different this information is when compared to everything that had previously been observed by a particular user, or by a community as a whole [30]. For the ramp-up problem, all generated recommendations present maximum level of novelty, since all items are considered new to a user that has not consumed any items before. On the other hand, *diversity* metric generally applies to a set of items, and relates to how different items are when compared with each other [30]. Basically, it is calculated as the complement of Pearson's similarity of the recommended items R, as shown by the equation 3. This metric comes from the framework proposed by [22]. This metric represents the diversity of items based on the expected average distance from an item to an item list (ILD).

$$div(R|u) = ILD = \frac{2}{|R|(|R|-1)} \sum_{i_k \in R, l<k} d(i_k, i_l) \tag{3}$$

Finally, we propose and use a new metric in this work that aims to measure the harmonic mean between accuracy and diversity. This metric has the objective to consider the trade-off between diversity and accuracy, just to evaluate a recommender systems based on both criteria. In order to do it, we normalize the accuracy values considering its highest value according to parameter k. High values indicate that the recommender is able to present a useful and diverse set of items to the users. The results of this metric emphasizes the need for recommender to achieve both accuracy and diversity.

$$F - measure = 2 \times \frac{div(R|u) \times accuracy(u)}{div(R|u) + accuracy(u)} \tag{4}$$

4.4 Evaluation Methodology

Under the premise that in the real world, there are several user profiles and that RSs should be able to satisfy them, we intend to distinguish and characterize the proposed strategies, considering the trade-off between utility and diversity. Thus, our evaluation methodology consists first of distinguishing the recommendations generated in order to verify if the items recommended by Maximum Coverage differ from traditional strategies. In order to perform this analysis, we calculate the intersection between the sets generated by the top-100 recommendations, comparing all four strategies evaluated.

Then, we vary the amount of items recommended by each technique, selecting 5, 10, 20, 50 and 100 first items ranked by each strategy. For each of these sets, we evaluate the recommendations based on utility, diversity and the trade-off between utility and diversity. This process is detailed as follows:

Step-1: Evaluate the utility of each implemented strategy in order to determine whether the recommended items are potentially relevant to users. Calculate the accuracy, precision and recall of each recommender and calculate these metrics average.

Step-2: Evaluate the *diversity* for each strategy. For each obtained ranking, calculate the average value for each recommendation.

Step-3: Evaluate the trade-off between utility and *diversity* for each strategy. For each obtained ranking, calculate the *F-measure* average value for each recommendation.

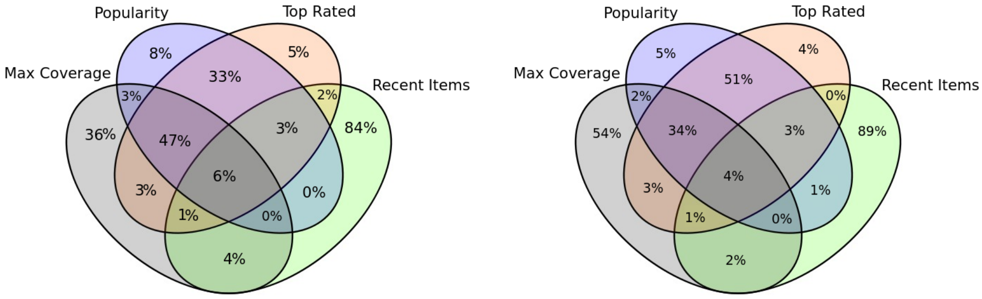
After, we evaluate whether the proposed strategy is able to mitigate the bias of popularity, which makes the most popular items to be more recommended. This analysis allows us to say if the proposed strategy is able to merge popular and non-popular items into the recommendations, thus reaching the niche consumers. Finally, we evaluate the application of proposed strategy in possible realistic scenarios, where only 5, 10 or 20 items are recommended for each user. The objective of this analysis is to verify the gains obtained by the Maximum Coverage strategy in relation to these scenarios. In addition, we evaluated the execution time required for each strategy to generate 5, 10 and 20 items to recommend. With this analysis, we hope to consolidate the practical efficiency of each non-personalized strategy evaluated.

5. EXPERIMENTAL ANALYSIS

In this section, we discuss the main results of applying the proposed evaluation methodology to first-time users selected in MovieLens 1M and 10M datasets. Initially, we performed a comparison and analysis of the different recommendation strategies, considering the set of items recommended. For this, we propose to evaluate the intersection of the items recommended by each strategy using a Venn diagram. Next, we evaluate the utility and diversity of each strategy to evaluate the performance of our Maximum Coverage strategy. Our objective is to evaluate the quality of the recommendations of each of the strategies considering the quality dimensions. Finally, we perform specific analyses of the strategies, considering a real scenario in which a maximum of twenty items are recommended. In this step, we specifically evaluated the recommendation of up to twenty items, such as a real e-commerce system such as *Amazon* or *Netflix*, which has a maximum of 20 items at a time for users.

5.1 Similarity Analysis

First, we analyze the intersection of the recommendation lists of each strategy, as shown in the Venn diagrams of Figure 3 for each scenario. We can note that only 6% and 4%, for ML-1M and ML-10M, respectively, of the recommended items are present in all recommendation lists. This low intersection value is justified when we observe that 84% and 89%, for the respective datasets, of the recommended items are unique to the Recent Items strategy. In fact, when we look at the intersection between Popularity, Best Rated and Maximum Coverage, we note that 47% and 34% of items for ML-1M and



(a) ML-1M (b) ML-10M

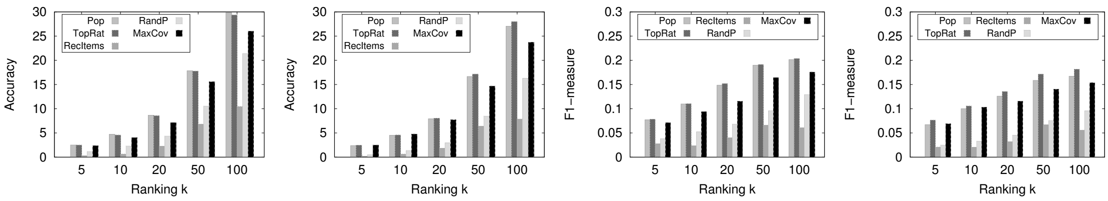
Fig. 3. Venn diagrams to represent the intersection rate of the recommendation lists issued by distinct non-personalized RSs, considering the $k = 100$.

ML-10M are recommended by all of them. We can also note that there is an intersection of 33% and 51% of the items for Popularity and Best Rated strategies. Finally, it is important to highlight that the Maximum Coverage strategy presents 36% and 54% of distinct items from the baselines.

Briefly, these results show that: (1) the Popularity strategy has similarities to strategy to recommend Best Rated items, which can be confirmed by the high intersection value between them; (2) the Recent Items strategy presents items totally different from those recommended by other strategies; (3) the Maximum Coverage strategy is able to present popular items due to its intersection with Popularity strategy and also Best Rated items. Moreover, it has a good diversity, since it presents a significant amount of distinct items provided only by itself. This fact indicates that the Maximum Coverage may introduce novelty and diversity to recommendations.

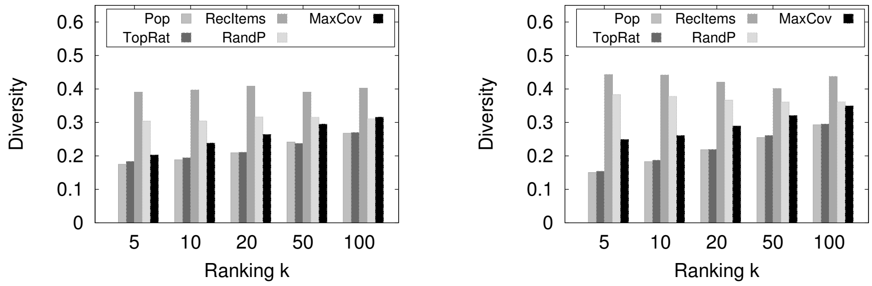
5.2 Utility vs. Diversity

Considering the utility related to the evaluated RSs, Figure 4(a) and 4(b) presents the mean accuracy of each RSs, considering distinct top-k lists. As expected, Popularity presents the highest accuracy, since popular items occur in a larger number of distinct user lists. Given the high level of intersection between Popularity and Best Rated items, we also observe high accuracy rates related to the latter strategy. Despite recommending items different from those issued by the Popularity strategy, Maximum Coverage exhibits accuracy rates comparable to the latter, mainly for smaller ranking k lists. Evaluating the F1 metric, which performs a harmonic mean of the precision and recall values, we can note that the good performance of the recommendation metrics based on Popularity and Best Rated items. These strategies, as shown in Figure 4 (c) and (d), are able to return relevant items to users, since the studied scenarios are related to mass consumption. On the other hand, it may be noted that the Recent Items and Random Popularity strategies do not have relevant items for users.



(a) Accuracy ML-1M (b) Accuracy ML-10M (c) F1 to ML-1M (d) F1 to ML-10M

Fig. 4. Results related to metrics Accuracy and F1 achieved by each strategy regarding to the items recommended for the first-time users. In fact, Popularity and Best Rated items strategies present the highest accuracy and F1 for both datasets, since the consumption of these users has a bias for the mass consumption.



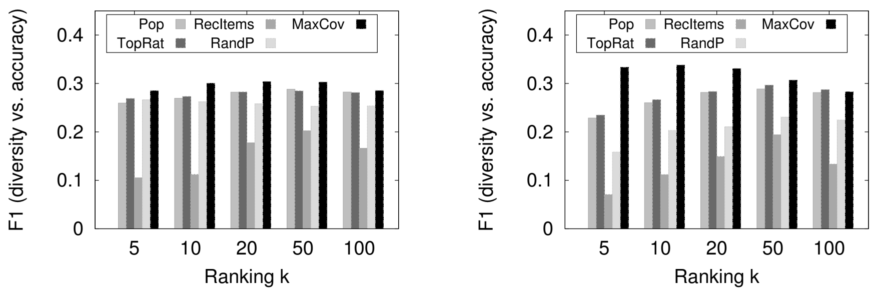
(a) Diversity - ML-1M

(b) Diversity - ML-10M

Fig. 5. Average of Diversity values related to each non-personalized strategy. Recent Items strategy presents the highest values of diversity related to its items recommended, followed by Maximum Coverage strategy. As expected, the Popularity and Best Rated do not present high values of diversity.

Next, we evaluate the diversity metric for non-personalized strategies. Figures 5(a) and 5(b) show the average values obtained for this metric, providing a global analysis compared to all list of recommended items. We can observe that the Random Popularity and Recent Items strategy present more diverse items than the other ones. However, these results, when evaluated in conjunction with the usefulness of the items. The Maximum Coverage strategy brings more diverse items than the other strategies. This result is related to the approaches used by each strategies that aim to select random (Random Popularity) or even items that interest to different users (Maximum k-Coverage).

The desirable trade-off between accuracy and diversity by a recommender system can be calculated through the F-measure metric, which performs a harmonic mean of diversity and accuracy, as shown in Figure 6(a) and 6(b). Table II compare trough Area Under Curve (AUC) of the generated F-measure rankings. This trade-off shows that the strategy of recommending the Recent Items is not useful to end users, making the recommendation inefficient. The Popularity and Best Rated strategies, which present high accuracy, are not able to diversify the set of presented items, not satisfying users of niche consumers. On the other hand, the Maximum Coverage strategy is statistically better than the baselines, with a p-value of less than 0.001 using the Wilcoxon test for non-normal distributions. In general, the Maximum Coverage strategy presents an average gain of 13.5% in relation to the strategy of recommending the Best Rated items, which is the strongest baseline in these analyses. These results show that, even though it is not a strategy widely used in the literature, this Maximum Coverage strategy presents potentially relevant results for non-personalized recommendations and may be applicable in real and practical scenarios.



(a) F-measure - ML-1M

(b) F-measure - ML-10M

Fig. 6. F-measure on accuracy and diversity for non-personalized RSs. These results indicate that the Maximum Coverage strategy is superior the traditional ones.

	k-value	Popularity	TopRated	Recent Items	MaxCoverage	Statics Gain	
ML-1M	$k = 5$	0.1708	0.1821	0.0243	0.2543	8%	▲
	$k = 10$	0.2034	0.2110	0.0511	0.2697	9.25%	▲
	$k = 20$	0.2255	0.2270	0.0997	0.2693	10%	●
	$k = 50$	0.2316	0.2409	0.1627	0.2574	9%	●
	$k = 100$	0.2257	0.2343	0.1186	0.2375	4%	●
ML-10M	$k = 5$	0.2089	0.2161	0.0582	0.2335	39.7%	▲
	$k = 10$	0.2302	0.2329	0.0714	0.2544	27.8%	▲
	$k = 20$	0.2476	0.2461	0.1478	0.2709	18.6%	▲
	$k = 50$	0.2533	0.2494	0.1756	0.2718	6.8%	▲
	$k = 100$	0.2447	0.2439	0.1435	0.2537	1.35%	▲

Table II. AUC related to the rankings achieve by the metric F-measure based on all k items recommended. The symbol ▲ denotes significant positive gains, ● non significant gains and ▼ significant negative losses from the F-measure distribution. In fact, Maximum Coverage strategy presents the best rankings providing better utility and diversity between the items recommended.

5.3 Application on Real Scenarios

In a real scenario of recommendation, such as *Amazon* or *Netflix*, the RSs are interested in presenting a list or ranking of 5, 10 or 20 items of potential interest of a set of users or consumers. Therefore, we present in Figure 7 the results of our proposed recommendation strategy for top-5, top-10 and top-20 rankings. From these analyses, it is possible to note that Maximum Coverage strategy presents a similar utility performance when compared to Popularity and Best Rated strategies (see Figures 7(a), 7(b), 7(e), and 7(f)), showing a better accuracy than these baselines in ML-10M (Figure 7(e)) for top-5 and top-10. In terms of diversity of items, we can observe that there is a diversity of items, even on a small list of 5-20 items, as can be confirmed by Figures 7(c) and 7(g). Moreover, based on the F-measure on accuracy and diversity (see Figures 7(d) and 7(h)), we can see a high level of diversity and a significant accuracy, providing a performance that overcomes the other strategies. These superior results are also demonstrated by the AUC of F-measure metric for accuracy and diversity. Observing the Table II, the first three columns of each dataset, it is possible to note high levels of gains. For recommendations of 5, 10 or 20 items, our strategy presents gains of 8%, 9.25% e 10% for ML-1M, and 39.7%, 27.8% and 18.6% for ML-10M.

Despite the cubic complexity of the greedy algorithm used, the execution time required to recommend these items is scalable to the real scenarios applied. Table III shows the execution time of the algorithms used, in seconds, showing that Maximum Coverage is feasible in practice. Thus, all these results demonstrate the Maximum Coverage is a complementary option to be adopted behind the traditional methods, being effective in real scenarios.

	k-value	Popularity	TopRated	Recent Items	MaxCoverage
ML-1M	$k = 5$	0.0296	0.1542	0.1565	1.8573
	$k = 10$	0.0302	0.1529	0.1576	2.3515
	$k = 20$	0.0314	0.1581	0.1618	3.2807
ML-10M	$k = 5$	0.4791	2.1652	1.8166	114.5131
	$k = 10$	0.4795	2.1397	1.8047	133.8829
	$k = 20$	0.4911	2.1604	1.8162	157.8802

Table III. Runtime of the algorithms used (in seconds).

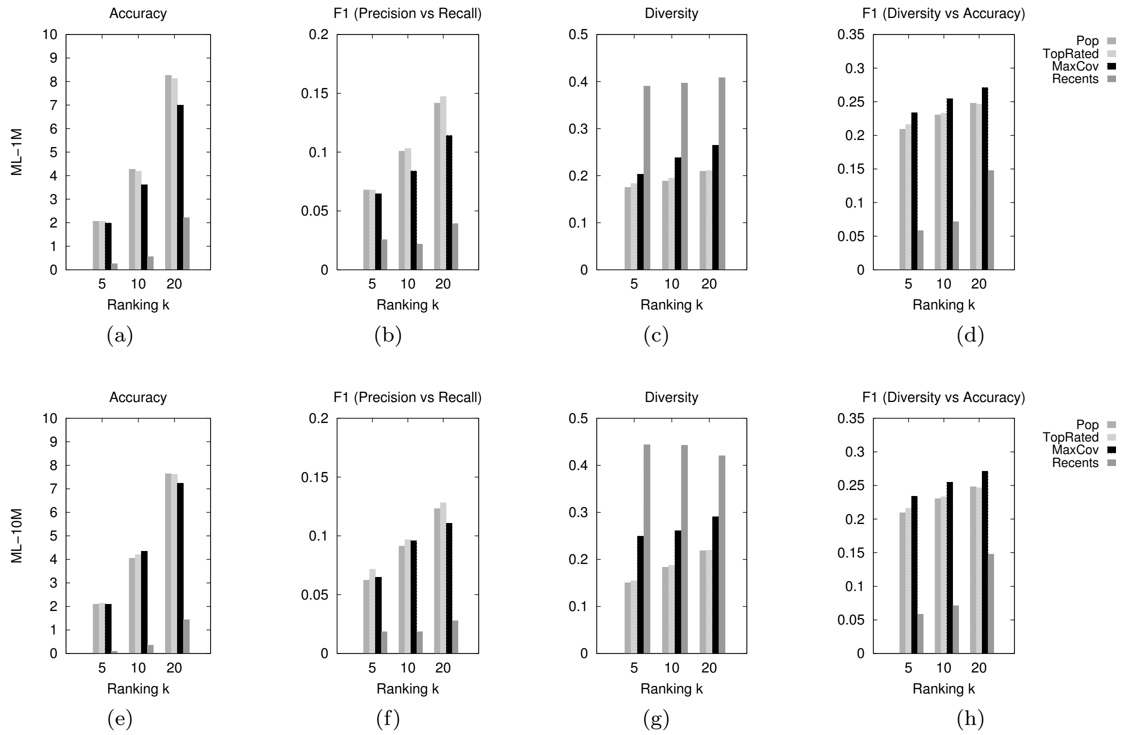


Fig. 7. Results of performance metrics for ML-1M and ML-10M data sets, considering the classical recommendation task for real applications, which recommends only 5, 10 or 20 items. We can verify a good performance for Maximum Coverage strategy compared to the other strategies or baselines, with a good level of accuracy and better diversity level.

5.4 Recommending non-Popular Items

The impact of applying the Maximum Coverage strategy is evident in scenarios whose user consumption is strongly biased towards popularity (i.e., mass consumption). By relating the popularity of items and users' consumption history in the studied collections, as shown in Figures 8 (a) and 8 (b), we observe an effect similar to the so-called long-tail distribution [1] caused by mass consumption. In this context, mass consumption-related recommendations do not represent a good strategy for acquiring niche consumers, as these RSs would be trapped to recommend items from the *head* of this distribution (i.e., widely consumed items).

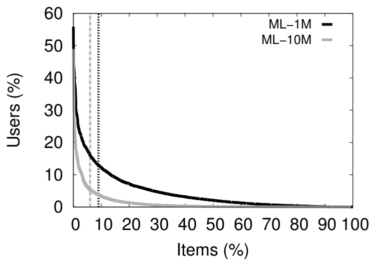


Fig. 8. Popularity items distribution in ML-1M and ML-10M datasets. The vertical lines represent the division between head and tail preformed in elbow point, for each evaluated scenario.

[1]A power distribution on the consumption of users who tend to consume a few popular items.

When assessing whether items recommended by Maximum Coverage belong to the *head* or *tail* of the popularity distribution, we note that such a strategy is indeed valid for niche consumers. By calculating the second derivative of the popularity distribution (i.e., the elbow point), we found that 14% and 12% of the recommended items appear in the distribution tail for ML-1M and ML-10M, respectively. Hence, recommendations generated by Maximum coverage tend to mitigate the problem of long-tail, bringing benefits to various real applications and easing the task of satisfying the niche consumers.

In fact, analyzing how many distinct users each strategy is able to satisfy minimally (i.e. accuracy superior or equal to 1), we can note that Maximum Coverage strategy achieves 97% of the users for dataset ML-10M. In turn, the remaining traditional strategies have achieved just 94% of the users in the same dataset. Despite it corresponds a small percentage difference (i.e. 3%), it represents 318 users, which is a considerable number. This result corroborates to our hypothesis that recommendations merging popular and non-popular items, which focus on different user preferences, are effective and applicable in practice.

6. CONCLUSION & FUTURE WORK

In this work we present a new approach for the user ramp-up problem in recommendation research area. Basically, we analyze the hypothesis that present diversified recommendations, which are able to combine popular items (i.e., related to mass consumption) and non popular items (i.e., related to niche consumption) may satisfy different user profiles. In this sense, we propose to model the domain as a bipartite graph and, then, apply the Maximum k-Coverage strategy to return the k items that interest the higher number of distinct users. Under these assumptions, we evaluate this strategy using actual datasets from MovieLens (i.e., ML-1M e ML-10M) and consolidate the results based on an evaluation methodology that is able to consider the desirable trade-off between utility and diversity in Recommender Systems (RSs). We compare our proposal with traditional strategies for ramp-up problem: Popularity, Best Rated and Recent Items.

First, we analyze the intersection of the recommendation lists of each strategy and we note that Maximum Coverage strategy presents up to 36% of distinct items from the all other strategies, indicating that the Maximum Coverage may introduce novelty and diversity to recommendations. Next, we evaluate the utility of the recommendation list of each strategy using the metrics Accuracy, Precision and Recall. We observe that the traditional strategies Popularity and Best Rated are more effective, since the consumption of the users has a bias for the mass consumption. However, evaluating the diversity of the recommendations, we conclude that Recent Items and Maximum Coverage strategies are better than the other ones. Adopting an evaluation that balances utility (i.e., Accuracy) and diversity, we observe that: (1) Despite Recent Items strategy presents high diversity, it is not useful to end users; (2) The Popularity and Best Rated strategies, which present high accuracy, are not able to diversify the set of presenting items, not satisfying users of niche consumers; and (3) The Maximum Coverage strategy presents the best results related to the trade-off between utility and diversity, satisfying 97% of the users. In sum, our approach, based on a greedy algorithm for Maximum Coverage, presents gains up to 13.5% when compared with traditional strategies, with a cubic complexity, corresponding to a practical and scalable strategy for scenarios with millions of items and users.

As future work, in order to consolidate even more our results, we intend to evaluate these strategies adopting new datasets on which it is possible to distinguish different user profiles present in real scenarios. Moreover, our goal is to propose a new approach, similar to the one present in this article, for the cold-start problem, on which the users are participatory and perform few evaluations of items.

ACKNOWLEDGMENTS

This work was partially funded by the Brazilian National Institute of Science and Technology for the Web (grant no. 573871/2008-6), MASWeb (grant FAPEMIG/PRONEX APQ-01400-14), CAPES, CNPq, Finep, and Fapemig.

REFERENCES

Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, no. 8, pp. 30–37, 2009.

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*, vol. 17, no. 6, pp. 734–749, 2005.

J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, pp. 109–132, 2013.

A. M. Rashid, I. Albert, D. Cosley, S. K. Lam, S. M. McNee, J. A. Konstan, and J. Riedl, "Getting to know you: learning new user preferences in recommender systems," in *Proceedings of the 7th international conference on Intelligent user interfaces*, pp. 127–134, ACM, 2002.

R. Martin, M. Walid, W. Robert, and Z. Thomas, *Recommendation Systems in Software Engineering.* Springer, 2014.

A. Iskold, "The art, science and business of recommendation engines," *Retrieved April*, vol. 5, p. 2012, 2007.

C. Anderson, *The long tail: Why the future of business is selling less of more.* Hyperion, 2006.

R. G. Michael and S. J. David, "Computers and intractability: a guide to the theory of np-completeness," *WH Free. Co., San Fr*, 1979.

J. D. O. Hinz and D.-K. J. Eckert, "The impact of search and recommendation systems on sales in electronic commerce," *Business & Information Systems Engineering*, vol. 2, no. 2, pp. 67–77, 2010.

J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 230–237, ACM, 1999.

M. Batet, A. Moreno, D. Sanchez, D. Isern, and A. Valls, "Turist@: Agent-based personalised recommendation of tourist activities," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7319–7329, 2012.

T. Mahmood and F. Ricci, "Learning and adaptivity in interactive recommender systems," in *Proceedings of the ninth international conference on Electronic commerce*, pp. 75–84, ACM, 2007.

C. He, D. Parra, and K. Verbert, "Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities," *Expert Systems with Applications*, vol. 56, pp. 9–27, 2016.

T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling, and Y.-C. Zhang, "Solving the apparent diversity-accuracy dilemma of recommender systems," *Proceedings of the National Academy of Sciences*, vol. 107, no. 10, pp. 4511–4515, 2010.

L. Safoury and A. Salah, "Exploiting user demographic attributes for solving cold-start problem in recommender system," *Lecture Notes on Software Engineering*, vol. 1, no. 3, p. 303, 2013.

S. Loh, F. Lorenzi, R. Granada, D. Lichtnow, L. K. Wives, and J. P. M. de Oliveira, "Identifying similar users by their scientific publications to reduce cold start in recommender systems.," in *WEBIST*, vol. 9, pp. 593–600, 2009.

J. B. Schafer, J. Konstan, and J. Riedl, "Recommender systems in e-commerce," in *Proceedings of the 1st ACM conference on Electronic commerce*, pp. 158–166, ACM, 1999.

M. Hammar, R. Karlsson, and B. J. Nilsson, "Using maximum coverage to optimize recommendation systems in e-commerce," in *Proceedings of the 7th ACM conference on Recommender systems*, pp. 265–272, ACM, 2013.

S. A. Puthiya Parambath, N. Usunier, and Y. Grandvalet, "A coverage-based approach to recommendation diversity on similarity graph," in *Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 15–22, ACM, 2016.

N. Alon, B. Awerbuch, and Y. Azar, "The online set cover problem," in *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pp. 100–105, ACM, 2003.

V. Chvatal, "A greedy heuristic for the set-covering problem," *Mathematics of operations research*, vol. 4, no. 3, pp. 233–235, 1979.

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*, pp. 109–116, ACM, 2011.

C. Anderson, "Long tail, the, revised and updated edition: Why the future of business is selling less of more," 2008.

F. H. d. J. Mourão, "A hybrid recommendation method that combines forgotten items and non-content attributes," 2014.

J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 5–53, 2004.

L. McGinty and B. Smyth, “On the role of diversity in conversational recommender systems,” in *International Conference on Case-Based Reasoning*, pp. 276–290, Springer, 2003.

B. M. Sarwar, *Sparsity, scalability, and distribution in recommender systems.* University of Minnesota, 2001.

N. Kawamae, “Serendipitous recommendations via innovators,” in *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pp. 218–225, ACM, 2010.

D. McSherry, “Diversity-conscious retrieval,” in *European Conference on Case-Based Reasoning*, pp. 219–233, Springer, 2002.

F. Ricci, L. Rokach, and B. Shapira, *Introduction to recommender systems handbook.* Springer, 2011.