Adaptive Multi-attribute Diversity for Recommender Systems

Adaptive Multi-attribute Diversity for Recommender Systems

Tommaso Di Noia[1], Jessica Rosati[2,1], Paolo Tomeo[1*], Eugenio Di Sciascio[1]

[1] *Polytechnic University of Bari – Via Orabona, 4 – 70125 Bari, Italy*

[2] *University of Camerino – Piazza Cavour 19/f – 62032 Camerino (MC), Italy*

Abstract

Providing very accurate recommendations to end users has been nowadays recognized to be just one of the tasks an effective recommender system should accomplish. While predicting relevant suggestions, attention needs to be paid also to their diversification in order to avoid monotony in the returned list of recommendations. In this paper we focus on modeling user propensity toward selecting diverse items, where diversity is computed by means of content-based item attributes. We then exploit such modeling to present a novel approach to re-arrange the list of Top-N items predicted by a recommendation algorithm, with the aim of fostering diversity in the final ranking. An extensive experimental evaluation proves the effectiveness of the proposed approach as well as its ability to improve also novelty and catalog coverage values.

1. Introduction

Recommender systems have been proposed as essential tools in assisting users to face the "information overload" problem and they have been applied across several domains [8], such as music [26], TV programs [5], taxi suggestion [22], digital libraries [3], just to cite a few of them. The main task of a recommendation engine is suggesting unknown items in a personalized way and recommend the top N items by considering the highest predicted ratings. As a result, in the recommender systems field new algorithms and approaches have been proposed over the years mostly devoted to maximizing recommendation accuracy. However, more recently, the drawbacks of building recommendation engines focusing exclusively on accuracy maximization have been also widely explored and highlighted [1, 9, 30]. Simply put, the most accurate recommendations for a user are often too similar with each

*Corresponding Author.

E-mail addresses: tommaso.dinoia@poliba.it (T. Di Noia), jessica.rosati@unicam.it (J. Rosati), paolo.tomeo@poliba.it (P. Tomeo), eugenio.disciascio@poliba.it (E. Di Sciascio).

other (e.g., songs by the same artist), or *overspecialized*, thus causing user dissatisfaction and frustration [47]. The so called *portfolio effect* in recommender systems [10] has been widely recognized as a situation when very similar, almost identical, items appear in a recommendation list [38], correctly but bothering the user [49] (see Figure 1).

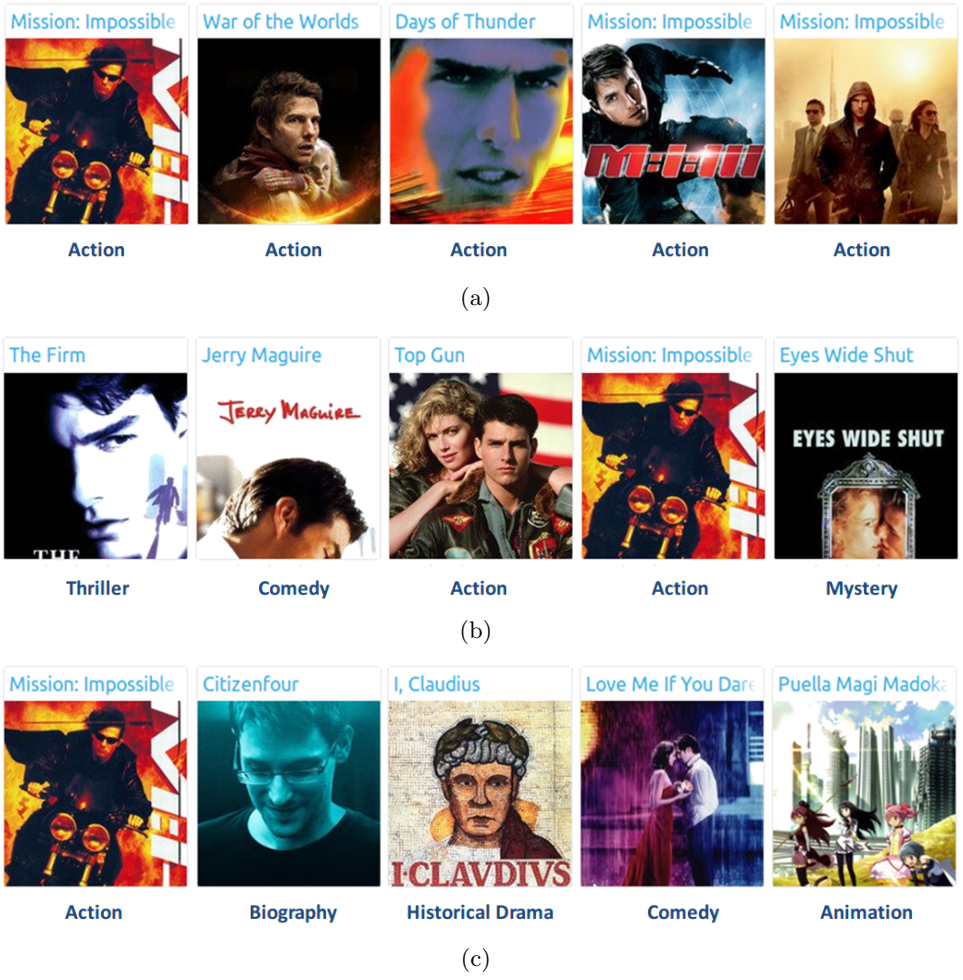

Figure 1: Example of three recommendation lists with different degree of diversity: (a) low diversity, all the movies have same actor (Tom Cruise) and genre (Action); (b) the actor is still the same but there are different genres; (c) higher diversity, in terms of both actor and genre. We see how the portfolio effect is more evident in (a) and (b).

The need to move beyond traditional accuracy metrics in the evaluation of a recommendation engine has been originally argued in [37] and several works have tackled the issue of diversification of recommendations as a way to increase user's utility [9, 6, 45, 40, 8], reaching the conclusion that a degree of diversity in the list can be increased at a cost of reducing system accuracy [12].

Since diversity is usually characterized as the dissimilarity degree between all the items in the recommendation list [27, 49, 47], one of the most important problems to address is the item-to-item dissimilarity evaluation. So far, diversity based on only one attribute (e.g. genre in movie and music domains, product category in e-commerce) [40] or collaborative filtering information (e.g. number of co-rating between items) [45] has been mainly considered in the literature. However, multi-attribute diversity is still underexplored. The main research questions we address in this paper, aiming at reducing the portfolio effect in a multi-attribute setting, are:

(i) *How to model different users' attitude with reference to diverse items in the recommendation list?*

(ii) *Does each user need diversity for every attribute?*

(iii) *What is the right level of diversity for each attribute?*

The main intuitions behind our work are that: (i) users could be inclined to diversifying only with respect to some specific item dimensions (e.g., item attributes as *director* and *year* in the movie domain) and not be interested in diverse suggestions related to other ones (e.g. *genre* in the movie domain); (ii) we can extract this information from the user's past interaction with the system. Following these ideas, we propose an *adaptive multi-attribute* diversification approach able to customize the degree of individual diversity[1] by taking into account the inclination of the user to diversifying over different content-based item dimensions. Specifically, we employ Entropy as a measure for the diversity degree while modeling user preferences and use it in conjunction with the user profile dimension for calibrating the degree of diversification of the list.

This paper considerably extends our previous work [15] where the notion of user quadrants defined in terms of attribute-based Entropy and profile dimension was originally introduced to foster the computation of diversified recommendation lists. The new contributions presented in this paper refer to different aspects of the overall approach. We introduce a new modeling of the user propensity towards diversity which is not based on an exclusive classification in four quadrants but allows the user to belong to all the quadrants to a certain degree (this is the main reason why we call this new modeling *fuzzy approach*). In fact, the classification of users in four quadrants originally proposed in [15] seemed a too strong hypothesis to be of practical use. We also compared how the two different modelings affect recommendation results in terms not just of diversity, but also in terms of accuracy and

[1] In this paper by *individual* diversity we mean the degree of diversification in the recommendations provided to an individual user, in contrast to *aggregate* diversity across all users [2]

novelty of recommendation as well as in terms of catalog coverage (a.k.a. aggregate diversity) [2]. We show that our approach to diversification on the one hand reduces the *portfolio effect* while remaining, on the other hand, effective compared to the other evaluation dimensions just mentioned. The two modelings have been tested against two recommendation datasets that refer to different domains. More specifically, the main contributions of this paper are:

- *Analysis of user needs in terms of individual diversity.* Other than the clustering of users in four disjoint quadrants originally introduced in [15], here we propose a more fine-grained analysis of users profiles introducing a fuzzy classification. For each attribute describing an item and according to the individual values of entropy and profile length, each user belongs to each quadrant with a certain degree.

- *Evaluation Methodology.* We propose an evaluation of our approach for individual diversity by considering also its performance in terms of accuracy, novelty and aggregate diversity. For the evaluation we tested both an implicit (`MMR` [11]) and an explicit (`xQuAD` [35]) method (see Section 2). The evaluation has been performed by considering Pareto optimal solutions.

- *Empirical Analysis.* We demonstrate the validity of our intuition via an extensive experimental evaluation on two datasets involving several baseline systems.

The remainder of the article is structured as follows. We provide a discussion of related works, followed by an overview on diversity in recommendation engines, in Section 2. The details of our adaptive multi-attribute diversification approach are shown in Section 3. Finally, after the presentation of the experimental set-up in Section 4, we evaluate the proposed strategies on two datasets related to the movie and book domains and we present the performance with different system settings (Section 5). A summary and an outlook on future research close the paper.

2. Related work

In the last few years, several approaches to the development of recommendation engines have been driven by the goal of improving not only the accuracy but also some form of utility associated to the recommended list. The attention to the concept of diversity and thus to the reduction of the *portfolio effect* arises from the need of increasing the utility associated to the returned list of items by avoiding monotony in recommendations. [17] shows that the diversity of recommendations has a positive influence on user's satisfaction, and is in turn a strong predictor of the user's final choice of the recommender system, at least for general-purpose movie recommendation.

Approaches to diversity are numerous in the literature. Given a set of items, most of previous proposals consider a definition of diversity based on content-based information (including the genre of a movie, the authors of a book, etc.) or on items feature space and try to maximize the sum of pairwise distances between elements in a set. [49] places itself in the former group, by defining the overall diversity of the recommended list through an intra-list-similarity metric using a taxonomy-based classification, in contrast for example to [47], which defines the reciprocal distance of a pair of items starting from the items feature space. Differently from existing diversity-promoting techniques based on pairwise comparisons, [33] acts on the feature space of the overall set of items, by adding an Entropy regularizer to the objective function which, in practice, is able to increase diversity. However, in offline evaluation settings, accuracy and diversity act in opposition with each other, since improving one of them usually leads to shrink the other. The concept of Pareto optimality could be used to face the trade-off of multi-objective problems [32, 34].

Recently, the idea of considering the user interests in the diversification approach in order to personalize the recommendation diversity received increasing interest. User modeling techniques have tried to characterize deeply the users-items interactions and to move beyond the network of users just based on the rating history, as in [28], where a trustworthy network made of users in which a user can rely on has been built. In [42] the identification of diversity within the user profile is carried out through the extraction of user sub-profiles to reflect the polyfacetic nature of user interests, where the definition of a sub-profile is done by analysing only the genre of a movie. The authors of [13] point out a causal relationship between personality factors (such as openness and conscientiousness) and the degree of diversification in the user choices with respect to genres, actors, directors, country or year of release of a movie. As a further validation, in [44], the same authors suggest a solution taking into account *personality* for generating more personalized diverse recommendations and consolidating their previous observations. To the best of our knowledge, [15] proposes the first attribute-based diversification approach, which is able to customize the degree of diversity of the recommendation list by taking into account the inclination to diversity of the user across different item attributes.

In addition to individual diversity, which is the main focus of this paper, aggregate diversity and novelty are recognized as essential objectives for user satisfaction [1, 2, 8] and should be considered for a complete evaluation of a recommendation engine. The necessity of improving aggregate diversity is particular important for online stores in the attempt to suggest a broader range of items including niche ones. [2] calculates rating prediction using existing filtering techniques and then re-ranks the list of candidate items thus pushing elements out of long tail. [43] falls in the second research line, by trying to improve the estimation process especially for rarely used items,

that is allowing a fair opportunity for most items to be recommended.

Novelty is defined differently in publications depending on the context and its purpose, e.g., the item novelty with respect to the user, which is related to the individual diversity, or the item novelty with respect to the total amount of recommended items, which is related to the aggregate diversity. The attempt to improve novelty runs often in parallel with the goal of increasing diversity, as in [21]. [23] proposes a regression model to predict the individual novelty preferences of a user analysing her recent past interactions. Its adaptive recommender also includes dynamic user's preferences for novelty.

2.1. Greedy algorithms for diversification

The activity of a recommender system can be divided into two phases: first there is the prediction of the ratings for unrated items and then the items can be re-ranked to maximize user's utility. According to [2], in order to improve the diversity (both individual and aggregate) of recommendations it is possible to deal only with the second phase. As finding the most diverse results set is NP-hard, several heuristics have been proposed [25]. Greedy heuristics, for example, select the next most relevant item only if that item is diverse with respect to the items already selected [25]. They have proven to be efficient and effective [16, 11, 35, 4].

Hereafter, we will use overlined bold capital letters to denote lists, e.g., $\overline{\mathbf{X}}$, and bold capital letters to represent the corresponding set of elements belonging to the list, e.g., $\mathbf{X}$. Let $\overline{\mathbf{R}} = \langle 1, ..., n \rangle$ be the recommendation list for user u generated using the predicted ratings and suppose we want to provide the user with the re-ranked list $\overline{\mathbf{S}}$ of recommendations, such that $\mathbf{S} \subset \mathbf{R}$ and whose length is $N \leq n$. The adopted greedy strategy can be explained through Algorithm 1. At each step, the algorithm selects the item which maximizes an objective function f_{obj} (line 3), which in turn can be defined to find a trade-off between accuracy and diversity, and add it to the re-ranked list (line 4). Thus, it requires $\mathcal{O}(N^2 n)$ computations of the function f_{obj}.

As for search results diversification, the diversity in a list of recommendations may be increased in an implicit or explicit manner [6]. The implicit diversification aims to increase the average distance between pairs of items in the recommendation list, while the explicit one tries to diversify the list by covering the user interests represented via categories or other information that can describe the items. Explicit diversification is also known as Intent-Aware. In fact, user intents in information retrieval correspond to user interests in recommender systems. Among state-of-the-art diversification algorithms, Maximal Marginal Relevance (`MMR`) [41] is an **implicit** approach, while Explicit Query Aspect Diversification (`xQuAD`) [42] represents an **explicit** strategy.

Data: The original list $\overline{\mathbf{R}}$, $N \leq n$
Result: The re-ranked list $\overline{\mathbf{S}}$

$\overline{\mathbf{S}} = \langle\rangle$;
while $|\,\mathbf{S}\,| < N$ **do**
    $i^* = \underset{i \in \mathbf{R}}{\text{argmax}}\ f_{obj}(i, \overline{\mathbf{S}}, u)$;
    $\overline{\mathbf{S}} = \overline{\mathbf{S}} \circ i^*$;
    $\mathbf{R} = \mathbf{R} \setminus \{i^*\}$
end
return $\overline{\mathbf{S}}$.

Algorithm 1: The greedy strategy. We remind that the overlined capitalized letters are used for lists and capitalized letters for the corresponding sets. The set cardinality is denoted with $|\cdot|$, the $\setminus$ symbol corresponds to set difference and the symbol $\circ$ is used for appending new elements to a list. $\langle\rangle$ indicates an empty list.

`MMR` **implicitly diversifies** a list considering a trade-off between the relevance of an item and its amount of new information provided with respect to previously selected items. More formally, the objective function of `MMR` is defined as:

$$f_{obj}(i, \overline{\mathbf{S}}, u) = \lambda \cdot r^*(u, i) + (1 - \lambda) \cdot \underset{j \in \mathbf{S}}{\text{avg}}\ (1 - sim(i, j)) \tag{1}$$

where r^* is a function for rating estimation, sim is a similarity measure on item pairs and the λ parameter lets to manage the accuracy-diversity balance.

Differently from `MMR`, `xQuAD` **is an explicit method** since it maximizes the coverage of the inferred interests while minimizing their redundancy. It was proposed for search diversification in information retrieval by Santos et al. [35], as a probabilistic framework to explicitly model an ambiguous query as a set of sub-queries that are supposed to cover the potential aspects of the initial query. More recently, it has been adapted for recommendation diversification by Vargas and Castells [42], replacing query and relative aspects with user and items features, respectively. The expression of the `xQuAD` objective function is

$$f_{obj}(i, \overline{\mathbf{S}}, u) = \lambda \cdot r^*(u, i) + (1 - \lambda) \cdot div(i, \overline{\mathbf{S}}, u) \tag{2}$$

with $div(i, \overline{\mathbf{S}}, u)$ defined as

$$div(i, \overline{\mathbf{S}}, u) = \sum_f p(i|f) \cdot p(f|u) \cdot \prod_{j \in \mathbf{S}} (1 - p(j|f)) \tag{3}$$

In (3) $p(i|f)$ represents the likelihood of item i being chosen given the feature f and is computed as a binary function that returns 1 if the item contains

f, 0 otherwise; $p(f|u)$ represents the interest of user u in the feature f and is computed as the relative frequency of the feature f on the items rated by user u. In other words, `xQuAD` fosters the idea of promoting items that are simultaneously highly related to at least one of the features of interest for the user and slightly related to the features of the items already recommended.

The computational complexity of both `MMR` and `xQuAD` is the same of Algorithm 1, since they do not change the algorithm but merely define the objective function f_{obj}.

3. Adaptive multi-attribute diversification

In this section we introduce and describe our proposal to model user attitude towards diversification in recommender systems. Figure 2 shows a possible representation of a recommendation engine that exploits our approach to mitigate the portfolio effect. To this aim, we adopt a re-ranking procedure [2] in an adaptive multi-attribute setting that acts on the recommendations lists provided by a generic recommendation algorithm. Re-ranking has been shown [2] to be effective in increasing diversity in results while not affecting the computational complexity of the overall recommendation procedure. Instead of relying on a multi-objective optimization function that tries to maximize both diversity and accuracy, the recommendation algorithm only takes care of accuracy and leaves to a simpler re-ranking procedure the task of increasing the diversity in the final recommendation list.

Before moving into a detailed description of the diversification procedure we briefly describe the different phases of the complete recommendation scenario.

- *Inputs.* The inputs of the system are: (i) the User-Item matrix where we have the rating history of each user; (ii) a structured description of the items belonging to the catalog. Such information can be extracted from external knowledge sources such as *Wikipedia*, *Google*, *last.fm*, *IMDb*, *MusicBrainz*, etc..

- *User modeling.* Based on the inputs, for each user the system computes a model of her propensity towards diversified recommendation. In our case, for each attribute describing the item, the system evaluates the quadrant the user belongs to (see Section 3.1 for more details).

- *Computation of the recommendation list.* The recommendation algorithm exploits the User-Item matrix and optionally the description of the items in the catalog to compute a list of recommended items. If we are interested in returning the *top-N* best items to the user, in this phase the recommendation engine computes the *top-M* best items, with $M > N$. It is noteworthy that we are not interested here in the specific recommendation algorithm as we only focus on the eventual

re-ranking phase. Indeed, in our experimental setting (see Section 5) we evaluated our diversification model against different state of the art algorithms (`BPRSLIM`, `BPRMF`, `WRMF`, `SoftMarginRankingMF`, `ItemKNN`).

- *Re-ranking.* Based on the user classification into quadrants, the system re-ranks the recommendation list previously computed.
- *Output.* The user is returned with the *top-N* items from the re-ranked list.

Note that this method needs a sufficient quantity of ratings for each user, since it relies on Entropy and profile length information. Therefore, it is not able to work properly for cold-start users, namely those users who have provided an exiguous number of ratings (usually less than 5) or even no rating at all. In such situations, additional information is required. For instance, personality information have been proved to be a good solution for facing the cold-start problem [18] and for adjusting diversity [44], although it is not always available or inferable from rating data.

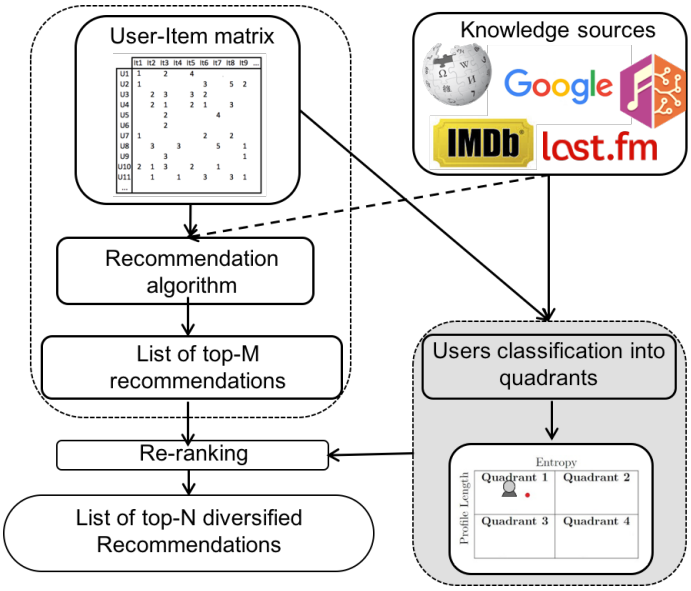

Figure 2: A schematic representation of the overall architecture.

In the following we detail how the *adaptive multi-attribute* diversification approach works. We start by introducing the notion of User Quadrants and then we move to their Fuzzy version. Subsequently, we show how the diversification approaches `MMR` and `xQuAD`, introduced in Section 2, may be adjusted to adaptive strategies under a multi-attribute setting. In other words, our intent is to modify the objective functions of `MMR` and `xQuAD` such that the diversification attitude of each user with respect to different

item attributes (i.e., *year*, *genre*, *director* and *actor* in the movie domain and *genre*, *author* and *subject* in the book domain) could stand out.

3.1. User Quadrants

In order to measure user's propensity to diversity on a specific attribute we used Shannon's Entropy which can be used as a measure of the information content associated with an attribute $A \in \mathcal{A}$ for each user u [29]. We compute Entropy with reference to each attribute $A \in \mathcal{A}$ to evaluate the degree of diversity with respect to u. Shannon's Entropy for user u and attribute A with $|dom(A)|$ values can be computed as:

$$\mathcal{H}_A(u) = -\sum_{k=1}^{|dom(A)|} p_k \cdot \log p_k \tag{4}$$

where p_k is the relative frequency of the k-th value of A considering all the items (elements) belonging to the user profile (collection of the items rated by the user).
Our model is adaptive in the way that it is based on the classification of users in four groups, referred to as *quadrants*, defined by considering as discriminating parameters the medians of the Entropy distribution and user profile length distribution across all users. A separate clustering is computed for each attribute describing the item. For example a user u is in the first quadrant for the *genre* attribute, if her Entropy $\mathcal{H}_{genre}(u)$ is less than the median of the Entropy computed across all users and she has a short user profile (her number of ratings is less than the median of users' ratings). The same user may belong to different quadrants in relation to different attributes. All the quadrants are represented in Figure 3.

The rationale behind our clustering hypothesis is that we can look at the previous interactions of the user with the system to infer whether she likes to enjoy items which result different with regards to some specific characteristics or not. If she uses to read books of the same subjects regardless of the author we may interpret this behavior as a clue that she is more willing to diversify with reference to authors while she is less willing with reference to subjects. It is noteworthy that such observation is more valid in the presence of longer interaction of the user with the system. We may imagine to have more information from a user who, during her whole interaction with the system, read dozens of books of the same genre from a high variety of authors rather than from a user who read only, say, five books of the same genre from five different authors. In the former case we have a stronger hint about the user attitude towards author diversification than in the latter one. Analogously we may say that the former user has a very low propensity towards genre diversification.

Given an attribute A, a high value of Entropy is then interpreted as an attitude of the user to choose items with different values for A. Conversely, a low value of Entropy is interpreted as her willingness to consider items similar with reference to that attribute. Furthermore, we are considering the user' profile length since we want to allow various values of Entropy to play a different role for users with a large or respectively short interaction with the system, making the Entropy computation potentially more meaningful if supported by a longer user experience.
The quadrants the user belongs to, potentially different for each item attribute, are used to rewrite $sim(i, j)$ in Equation (1) and $div(i, \overline{\mathbf{S}}, u)$ in Equation (2), as better explained in Sections 3.3 and 3.4 respectively. Given a user u and the set of item attributes $\mathcal{A}$, we then consider a function $q_u : \mathcal{A} \rightarrow \{1, 2, 3, 4\}$, which assigns, for each attribute, the quadrant to which user u belongs. Moreover, we introduce an absolute quadrant weight $\omega_k \in [0, 1]$, with $k \in \{1, 2, 3, 4\}$. Of course, more groups can be defined thus identifying more than four quadrants. However, we have already shown in [15] that even with such a coarse grained classification we are able to obtain interesting results in terms of *precision* and *intra list diversity* (ILD) values, and we will see how experiments described in Section 5 confirm this trend.

Profile Length \ Entropy		
	Quadrant 1 Low Entropy Small Profile	**Quadrant 2** High Entropy Small Profile
	Quadrant 3 Low Entropy Large Profile	**Quadrant 4** High Entropy Large Profile

Figure 3: Quadrants

3.2. Fuzzy Quadrants

Users hard clustering proposed in Section 3.1 and tested in our previous work [15] could seem too rigid because of a sharp discrimination into four quadrants. One way to overcome this inconvenience is to introduce a fuzzy users clustering (a.k.a. soft clustering) that permits a user to belong to more than one quadrant simultaneously with a different degree. In fact, we defined functions able to compute a membership degree for each quadrant. This setting can be regarded as the opposite extreme to the hard clustering in just four quadrants of the previous section, as it represents potentially infinite clusters to which a user may belong to. This allows us to get a comparison between the simplest version of clustering by median values and the fine-grained version represented by fuzzy clustering. In order to evaluate the membership grades to quadrants, we reproduced the quadrants subdivision in the unit square, normalizing in [0,1] the values of Entropy and profile

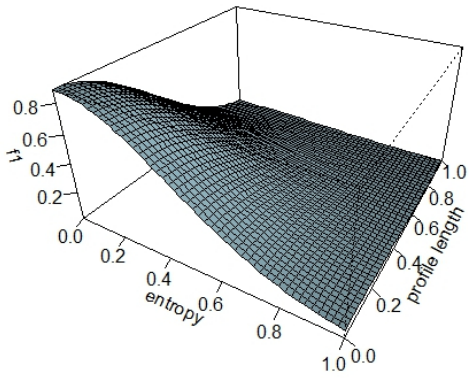

Figure 4: f_1 function

length and considering them as respectively the x and y coordinates. The x and y values then become the inputs for four bivariate Gaussian functions f_1, f_2, f_3, f_4 where $f_1 \sim \mathcal{N}((0,0), \sigma^2)$ (shown in Figure 4), $f_2 \sim \mathcal{N}((1,0), \sigma^2)$, $f_3 \sim \mathcal{N}((0,1), \sigma^2)$ and $f_4 \sim \mathcal{N}((1,1), \sigma^2)$. Whenever in the experiments we mention the fuzzy approach, we mean that we substituted the weights $\omega_{q_u(A)}$ introduced in Section 3.1 with a weighted sum

$$\omega_{q_u(A)} = \sum_{k=1}^{4} \omega_k \cdot f_k(x, y) \tag{5}$$

where x is the value of Entropy and y profile length for user u and ω_k are the absolute quadrants weights for attribute A. The choice of Gaussian functions was influenced by the need of having circular contour lines. The value of σ^2 is the same for the four functions and is chosen so that for point $(\frac{1}{2}, \frac{1}{2})$ each function assumes the maximum value of f divided by 4 ($\sigma^2 = 0.1803$).

3.3. Adaptive MMR

Here we are going to explain how the diversification algorithm `MMR` can be adjusted to incorporate the weights computed with User Quadrants or Fuzzy Quadrants settings. As we deal with a multi-attribute problem, sim has to consider similarities with respect to a set of attributes $\mathcal{A}$ and, for each attribute $A \in \mathcal{A}$, $sim_A(i, j)$ will hereafter denote the similarity between item i and item j with relation to A. The overall similarity between item i and item j in Equation (1), for the generic user u, becomes tailored to the quadrants she belongs to and is defined as:

$$sim(i, j) = \frac{\sum_{A \in \mathcal{A}} \omega_{q_u(A)} \cdot sim_A(i, j)}{m \cdot |\mathcal{A}|} \tag{6}$$

with $m = \max\{\omega_k \mid k = 1, 2, 3, 4\}$ and $sim_A(i, j)$ being a similarity measure between i and j with respect to attribute A. The weights associated

to quadrants the user belongs to influence the similarity score in Equation (6) and hence the resulting objective function of Equation (1). Specifically, based on our modeling hypothesis, the weights account for the user propensity in diversifying every single attribute. In fact, if a user is in the second or forth quadrant for a fixed attribute, then assigning a sufficiently big value to ω_2 and ω_4 corresponds to keeping a high value for the original similarity score and thus decreasing the overall value of $f_{obj}(i, \overline{\mathbf{S}}, u)$ for the items i most similar to the ones already available in $\overline{\mathbf{S}}$. These are the items we want to reduce in $\overline{\mathbf{S}}$, in order to guarantee a higher diversity value. Conversely, assigning low weights to the first and third quadrant (low values for ω_1 and ω_3) results in a significant lowering of the original similarity score and hence in an increase of the corresponding $f_{obj}(i, \overline{\mathbf{S}}, u)$ values. This corresponds to preferring items similar to the ones in the re-ranked list $\overline{\mathbf{S}}$.

3.4. Adaptive xQuAD

For the intent-aware diversification algorithm `xQuAD`, introduced in Section 2, we use the adaptation illustrated in [39] that allows to deal with the multi-attribute problem. Let $\mathcal{A}$ be the set of attributes and let us indicate with $A \in \mathcal{A}$ one of these attributes and with $f \in dom(A)$ the possible values or *features* of A. *div* in Equation (3) may be reformulated as follows

$$div(i, \overline{\mathbf{S}}, u) = \sum_{A \in \mathcal{A}} \frac{\sum_{f \in dom(A)} p(i|f) \cdot p(f|u) \cdot (1 - \text{avg}_{j \in \mathbf{S}}\, p(j|f))}{\sum_{f \in dom(A)} p(f|u)} \tag{7}$$

While `MMR` contains a simple similarity function where we can inject quadrants weights, `xQuAD` uses Equation (7) to compute the diversity across all the attributes via an explicit evaluation of the diversity between the features for each attribute. Therefore, we introduce weights in that formula changing the sum into a weighted sum. More formally, we rewrite $div(i, \overline{\mathbf{S}}, u)$ as

$$div(i, \overline{\mathbf{S}}, u) = \sum_{A \in \mathcal{A}} \omega_{q_u(A)} \cdot \frac{\sum_{f \in dom(A)} p(i|f) \cdot p(f|u) \cdot (1 - \text{avg}_{j \in \mathbf{S}}\, p(j|f))}{\sum_{f \in dom(A)} p(f|u)} \tag{8}$$

4. Experiments

4.1. Datasets

In order to test the effectiveness of our proposal for adaptive multi-attribute diversification, we carried out experiments on the well known `Movielens 1M`[2] dataset and on the `LibraryThing`[3] dataset.

[2] Available at `http://grouplens.org/datasets/movielens`

[3] Available at `http://www.librarything.com/services/`

`MovieLens 1M` dataset contains 1 million ratings from 6,040 users on 3,952 movies. The original dataset contains information about genres and year of release, and was enriched with side information such as actors and directors extracted from DBpedia[4]. More details about this enriched version of the dataset are available in [31]. Since not all movies have a corresponding resource in DBpedia, the final dataset contains 998,963 ratings from 6,040 users on 3,883 items. We built training and test sets by employing a 60%-40% temporal split for each user. Moreover, we used the `LibraryThing` dataset, which contains more than 2 million ratings from 7,279 users on 37,232 books. As in the dataset there are many duplicated ratings, when a user has rated more than once the same item, we selected her last rating. The unique ratings are 749,401. Also in this case, we enriched the dataset by mapping the books with BaseKB[5], the RDF version of Freebase[6] and then extracting three meaningful attributes: *genre*, *author* and *subject*. The subjects in Freebase represent the topic of the book, for instance Pilot experiment, Education, Culture of Italy, Martin Luther King and so on. The dump of the mapping is available online[7]. The final dataset contains 565,310 ratings from 7,278 users on 27,358 books. We built training and test sets by employing a 80%-20% hold-out split. The different ratio used for `LibraryThing` compared to `Movielens` (60%-40%) depends on its higher sparsity: holding 80% to build the user profile ensures a sufficient number of ratings to train the system.

	Movielens	**LibraryThing**
Number of users	6,040	7,278
Number of items	3,883	27,358
Number of ratings	998,963	565,310
Data sparsity	95.7%	99.7%
Avg users per item	275.57	20.66
Avg items per user	165.39	77.68

Table 1: **Statistics about the two datasets**

Since the number of distinct values was too large for *year*, *actors* and *director* attributes in `Movielens` and for all the attributes in LibraryThing, we convert years in the corresponding decades and performed a K-means clustering for other attributes on the basis of DBpedia categories[8] for `Movielens`

[4] `http://dbpedia.org`

[5] `http://basekb.com`

[6] `https://www.freebase.com`

[7] `http://sisinflab.poliba.it/semanticweb/lod/recsys/datasets/BaseKB2LibraryThing.zip`

[8] `http://purl.org/dc/terms/subject`

and Freebase classes[9] for `LibraryThing`. Table 2 and 3 report the number of attribute values and clusters. The number of clusters was decided according to the calculation of the within-cluster sum of squares (*withinss* measure from the R Stats Package, version 2.15.3), that is picking the value of K corresponding to an evident break in the distribution of the *withinss* measure against the number of extracted clusters.

	Num. Values	**Num. Clusters**
Genres	19	-
Decades	10	-
Actors	14736	20
Directors	3194	20

Table 2: **Statistics about `Movielens` attributes**

	Num. Values	**Num. Clusters**
Genres	270	30
Authors	12868	22
Subjects	2911	20

Table 3: **Statistics about `LibraryThing` attributes**

4.2. Recommendation Algorithms

Differently from [15], where the baseline was a generic user-based kNN Collaborative Filtering algorithm using Pearson correlation as similarity measure, here, for both datasets we adopt five different algorithms as baselines. We selected five state of the art algorithms available in MyMediaLite[10]: `BPRSLIM`, `BPRMF`, `WRMF`, `SoftMarginRankingMF` and `ItemKNN`. They were used to create a list of 200 recommendations to build the initial list $\overline{\mathbf{R}}$ used for performing the re-ranking step shown in Algorithm 1. With reference to Equation (1) and Equation (2) they represent $r^*(u, i)$. Jaccard index was used to compute $sim_A(i, j)$, as in [41, 45, 20], because each feature is represented by a binary value for each item: 1 if present, 0 otherwise[11].

4.3. Evaluation metrics

For evaluating the recommendation quality considering a wide range of evaluation metrics, we measured *Accuracy*, *Individual Diversity*, *Aggregate Diversity*, and *Novelty* in top-N recommendation task. In the experiments,

[9] `http://www.w3.org/1999/02/22-rdf-syntax-ns#type`

[10] `http://www.mymedialite.net`

[11] Cosine distance could be used to compute the distance between two items, but it is more appropriate in presence of weighted values [45].

the value of N is set to 10. Unless explicitly stated, each of the following metrics is computed with respect to a single user and then averaged across all users.

4.3.1. Accuracy

For accuracy, we used $Precision@N$, $Recall@N$ and $nDCG@N$. The first one represents the fraction of relevant items in the top-N recommendations. Let $rel(u, i)$ be a boolean function that represents the relevance of item i for the user u, with value 1 for relevant and 0 for non-relevant items, then $Precision@N$ is calculated as follows

$$Precision@N = \frac{\sum_{i=1}^{N} rel(u, i)}{N} \tag{9}$$

$Recall@N$ indicates the fraction of relevant items, in the user test set, occurring in the top-N list. Being $test(u)$ the set of relevant items in the test set for the user u, $Recall@N$ is defined as

$$Recall@N = \frac{\sum_{i=1}^{N} rel(u, i)}{\mid test(u) \mid} \tag{10}$$

Although precision and recall are good indicators to evaluate the accuracy of a recommendation engine, they are not rank-sensitive. On the other side, $nDCG@N$ takes into account the position of a relevant item in the recommendation list. More formally

$$nDCG@N = \frac{1}{\text{iDCG}} \cdot \sum_{i=1}^{N} \frac{2^{rel(u,i)} - 1}{log_2(1 + i)} \tag{11}$$

where iDCG is a normalization factor that sets nDCG@N value to 1 when an ideal ranking is returned [7].

4.3.2. Individual Diversity

The individual diversity of a recommendations set $\mathbf{R}$, whose size will be denoted as $|\mathbf{R}|$ and will match N in a top-N scenario, can be computed as the average dissimilarity of all pairs of items [21]:

$$\mathcal{ILD}(\mathbf{R}) = \frac{1}{|\mathbf{R}| \cdot (|\mathbf{R}| - 1)} \sum_{i \in \mathbf{R}} \sum_{j \in \mathbf{R}, j \neq i} div(i, j) \tag{12}$$

The distance function may correspond to the complement of some similarity measure in terms of the item features (content-based view) or their user interaction patterns (collaborative view) [41]. We used content-based ILD, that is we computed $div(i, j)$ as the complement of $sim(i, j) = \text{avg}_{A \in \mathcal{A}} sim_A$, where the similarity related to attribute A, sim_A, is given by Jaccard index computation.

Another diversity measure of a recommendation list is Subtopic Recall (S-Recall), proposed for evaluating subtopic retrieval in the information retrieval field, where documents may cover different subtopics of a query topic [46]. Adapted to recommendation task, S-Recall can evaluate the fraction of features covered in a recommendation list. More formally:

$$S{-}Recall(\mathbf{R}) = \text{avg}_{A \in \mathcal{A}} \frac{\left| \bigcup_{i=1}^{N} F_A(i) \right|}{|dom(A)|} \tag{13}$$

where $F_A(i)$ represents the set of features of attribute A in the i-th item in $\overline{\mathbf{R}}$. Intuitively, indicating the degree of subtopic coverage, S-Recall also represents the diversity of recommendation list. We used also the metric α-nDCG, that is the redundancy-aware variant of Normalized Discounted Cumulative Gain proposed in [14]. We adopt the adapted version for recommendation proposed in [36]:

$$\alpha{-}nDCG(\mathbf{R}, u) = \text{avg}_{A \in \mathcal{A}} \frac{1}{\alpha\text{-iDCG}} \cdot \sum_{i=1}^{|\mathbf{R}|} \frac{\sum_{f \in F_A(i)} (1-\alpha)^{cov(\mathbf{R}, f, i-1)}}{log_2(1+i)} \tag{14}$$

where $cov(\mathbf{R}, f, i-1)$ is the number of items ranked up to position $i-1$ containing the feature f. The α parameter is used to balance the emphasis between relevance and diversity. α-iDCG denotes the value of α-nDCG for the best "ideally" diversified list. Considering that the computation of the ideal value is NP-complete [14], we adopt a greedy approach: at each step we select solely the item with the highest value, regardless of the next steps.

4.3.3. Aggregate Diversity

Aggregate Diversity represents an important quality dimension for both business and user perspective, since improving the coverage of the items catalog and of the distribution of the items across the users may increase both the sales and the user satisfaction [43]. To evaluate Aggregate Diversity, we considered catalog coverage [19] (the percentage of items in the catalog recommended at least once), and Gini coefficient [2, 43] (for the distribution of recommended items). The latter is useful to analyse the concentration degree of top-N recommendations across all items and its scale is reversed, thereby forcing small values to represent low distributional equity and large values to represent higher equity.

$$coverage = \frac{| \bigcup_{u \in U} top\text{-}N(u) |}{| I |} \tag{15}$$

$$Gini\ coefficient = 2 \cdot \sum_{i \in I} \left(\frac{| I | + 1 - rank(i)}{| I | + 1} \right) \cdot \left(\frac{rec(i)}{|U|} \right) \tag{16}$$

In Equation (16) $rec(i)$ is the number of users to whom i has been recommended and $|U|$ is the number of users, while $rank(i)$ is the position of i if items were ordered according to the number of users they have been recommended to. The coverage metric needs to be considered together with a distribution metric like Gini coefficient, since the coverage gives an indication about the ability of a recommender to cover the items catalog, and the other one shows the ability to equally spread out the recommendations across all the items. Hence, only an improvement of both metrics indicates a real increasing of aggregate diversity, that in turn denotes a better personalization of recommendations [2].

4.3.4. Novelty

We evaluated the popularity-based novelty [41] which measures the unexpectedness of an object relative to its global popularity [48]. We used two popularity-based novelty metrics: Expected Popularity Complement (EPC) and the percentage of long-tail items among the recommendations across all users [2] (indicated with *total* in (18)) considering the 80 percent of less rated items in the training set as *Long-tail* items.

$$EPC = \frac{\sum_{i \in \mathbf{R}} (1 - pop(i))}{|\mathbf{R}|} \tag{17}$$

$$\%Long\text{-}tail = \frac{\sum_{i \in Long\text{-}tail} rec(i)}{total} \tag{18}$$

With $pop(i)$ in (17) we mean the number of users who rated item i, normalized by the maximum value over the items in the dataset.

5. Experimental Results

We conducted a comparative analysis of the adaptive methods we propose, the baselines without diversification introduced in Section 4 and the pure diversification algorithms (`MMR` and `xQuAD`). These latter consist of computing recommendations by using respectively Equation (1) and Equation (2) without considering the adaptive models. This implies that the diversification is applied indiscriminately to all users regardless of whether they are incline to diversifying their choices or not.

In the following we will indicate with $\texttt{MMR}_{quadr}$ the algorithm that carries out a hard users clustering and, given the list returned by the current baseline, performs re-ranking according to (1) with (6), as explained in Section 3.1. The $\texttt{MMR}_{fuzzy}$ model instead consists of a fuzzy clustering of users in four quadrants, as introduced in Section 3.2 and with quadrant weights as in Equation (5). Analogously, $\texttt{xQuAD}_{quadr}$ and $\texttt{xQuAD}_{fuzzy}$ represent the corresponding configurations for the diversification algorithm `xQuAD`.

It is common knowledge that building multi-objective recommender systems that suggest items that are simultaneously accurate and diversified may lead to a conflicting-objective problem, where the attempt to improve an objective further may result in worsening other competing objectives. We face the trade-off of multi-objective problems using the concept of Pareto optimality [34], according to which an individual (meant as the result of an algorithm in our case) dominates another if it performs better in at least one of the objectives considered. The Pareto Frontier is the set of all non-dominated individuals: none of them can get better without making at least one individual getting worse. We carried out the same type of comparative analysis based on Pareto frontier for `MMR` and `xQuAD`. In those analyses we vary the available parameters: only the value of λ can be modified for the diversification baselines, while λ and the quadrant weights w_1, w_2, w_3, w_4 are modified for *quadr* and *fuzzy* algorithms. The step size for variation was fixed in 0.05 for both λ and w_1, w_2, w_3, w_4. The results of this analysis are shown in Figures 5, 6, 7 and 8. However, a Pareto Frontier consists in potentially many individuals and in a realistic scenario the system designer would want to choose one or a few of them. In [34], an individual is chosen by means of a linear search on all of the individuals, selecting the one which maximizes a weighted mean on the objectives in the objective vector, where the weights in the weighted mean represent the priority given to each objective. For instance, if the objectives are accuracy and diversity, the objective vector [Accuracy = 0.7, Diversity = 0.3] allows the system to find the individual that strongly preserves the accuracy and slightly improves the diversity. In this work, in order to demonstrate the validity of the proposed adaptive diversification approach, we carried out a further comparison of the analysed algorithms selecting the most accurate individuals and those with the best mean between accuracy and diversity. Results are shown in Tables 4–11.

5.1. Comparative Results for MMR

The curves in Figures 5 and 6 show the relation between precision and other metrics, respectively for `Movielens` and `LibraryThing`, using the baseline `BPRSLIM` and the diversification algorithm `MMR`. Focusing on individual diversity, they point out that there is no particular difference in terms of ILD and α-nDCG, but there are improvements considering S-Recall. It means that using the adaptive models there is not an actual direct improvements on individual diversity, but the number of retrieved subtopics increases. Analysing aggregate diversity, the adaptive models improve both coverage and Gini coefficient, which indicates a real increment of aggregate diversity, as explained in Section 4.3. In particular, $\texttt{MMR}_{quadr}$ leads to a broader range of values compared to the diversification baseline and $\texttt{MMR}_{fuzzy}$. When considering the novelty dimension, $\texttt{MMR}_{quadr}$ leads to the best results, especially in terms of EPC, namely the popularity complement of the recommended items, while there is no relevant difference

between $\texttt{MMR}_{fuzzy}$ and MMR. The trends of the results are substantially similar across the other different baselines (BPRMF, WRMF, SoftMarginRankingMF, ItemKNN)[12].
Beyond the Pareto frontier, which is useful to point out the compromise between the involved objectives through many individuals, we analysed specific individuals as shown in Tables 4, 5, 6 and 7 [34]. We selected the best individuals for the algorithms involved in each comparison, according to two configurations. The first one considers as objective just accuracy, specifically Precision, while the second one corresponds to an unbiased balance between accuracy and diversity (ILD).

As already shown in [15], calibrating the diversity among different content-based attributes may lead to enhance diversity without penalizing accuracy. The same surprisingly good performance is observed in Tables 4 and 5 with the most accurate individuals for compared algorithms on Movielens and LibraryThing, respectively. Consistently with the accuracy-diversity trade-off, the basic MMR approach improves the diversity at the cost of the accuracy. The adaptive approaches we propose gain statistically significant improvements with respect to the baseline in terms of diversity, as we expected, but also accuracy, though non statistically significant. Furthermore the adaptive approaches significantly overcome MMR in terms of all the metrics, except for S-Recall on Movielens where $\texttt{MMR}_{fuzzy}$ obtains the same value of MMR. The individuals of Tables 6 and 7, related to the balance of accuracy and diversity on Movielens and LibraryThing respectively, improve the diversity with respect to the baseline, closely approaching the values reached by MMR, keeping high the accuracy values. A further observation corroborating our modeling hypothesis concerns the weights configurations $\langle \omega_1, \omega_2, \omega_3, \omega_4 \rangle$ for the best individuals found. It is useful to recall that, as explained in Section 3, we introduced quadrants weights in the attempt to provide recommendations with a diversity degree reflecting users propensity towards diversification. As a general trend, ω_4 gains the highest values while ω_1 and ω_3 lowest values, which basically means that users with higher entropy and longer profile will receive more diverse recommendations with respect to the other users. Interestingly, ω_2 shows a discordant behaviour between Movielens and LibraryThing using $\texttt{MMR}_{fuzzy}$. This could be a clue saying that for small profiles the propensity towards diversification is domain dependent and needs more investigations. As we will see in the next section, the same behaviour holds for xQuAD.

[12]For the sake of conciseness we do not report the plots here and made them available at http://sisinflab.poliba.it/recommender-systems/adaptive-multi-attribute-diversity.html.

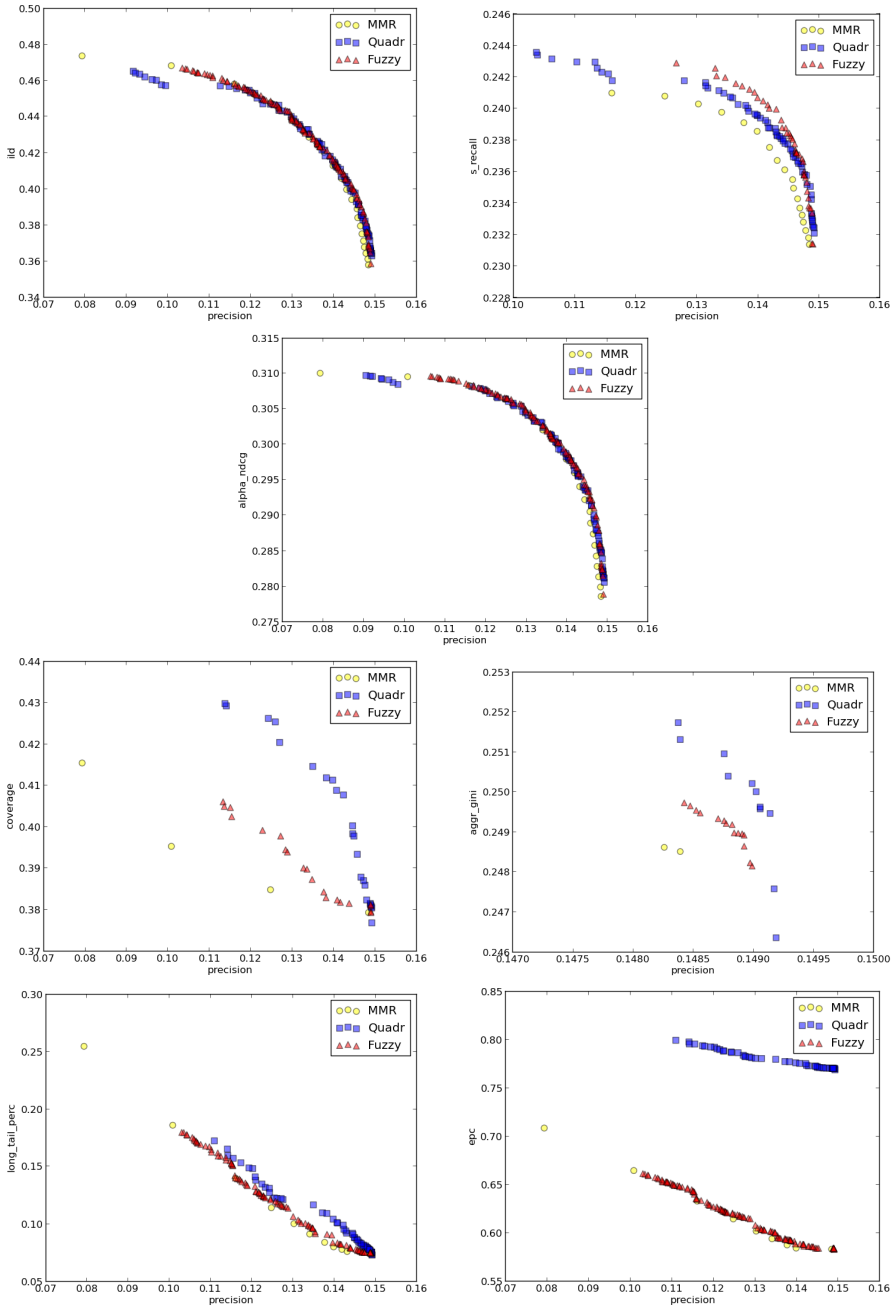

Figure 5: Pareto Frontiers for `Movielens` Dataset, using *BPRSLIM* and `MMR`

5.2. Comparative Results for xQuAD

In this section we investigate the results of the proposed adaptive methods used with the diversification baseline `xQuAD`. Figures 7 and 8 show the curves between precision and different other metrics, respectively for

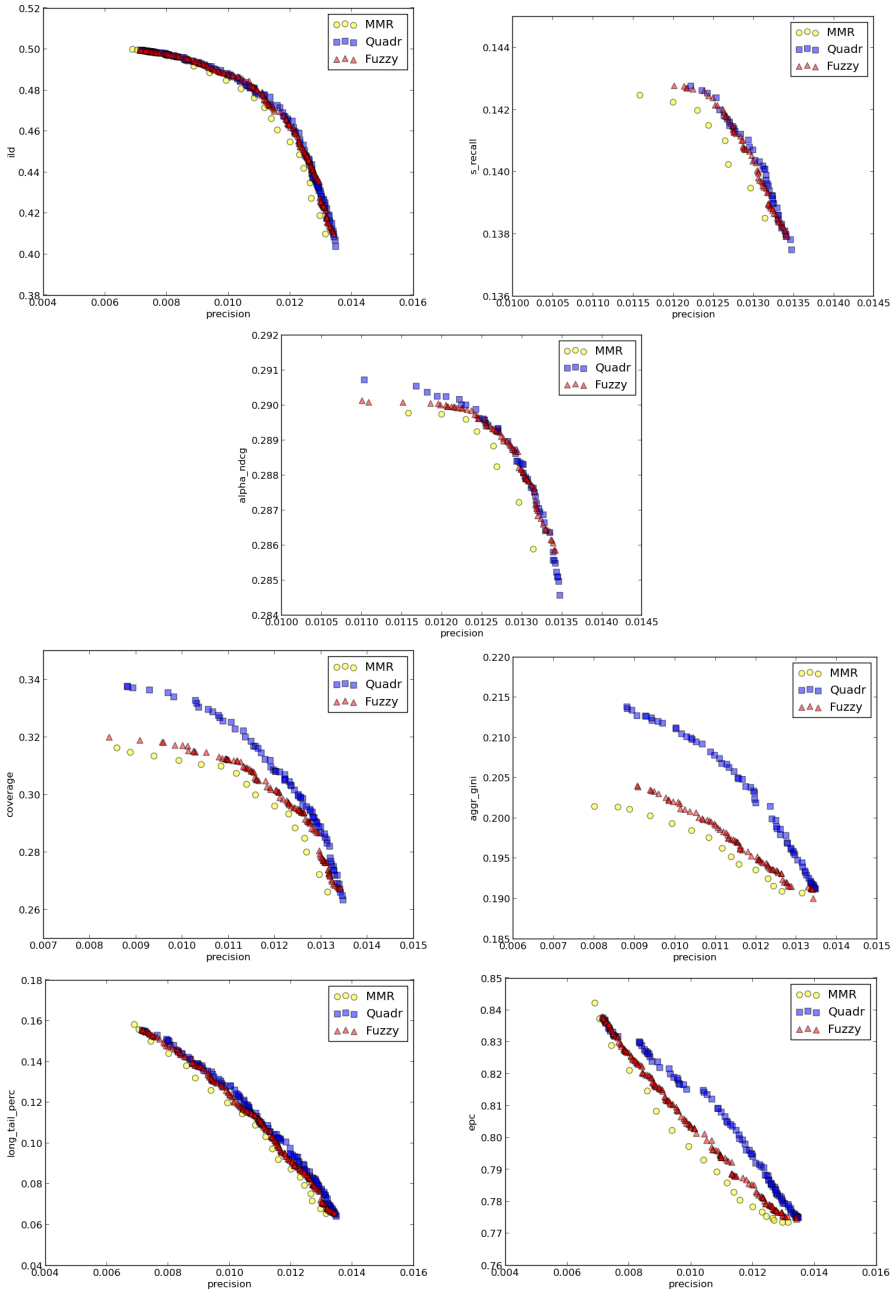

Figure 6: Pareto Frontiers for `LibraryThing` Dataset, using *BPRSLIM* and `MMR`

`Movielens` and `LibraryThing`, using the baseline `BPRSLIM` and the diversification algorithm `xQuAD`. Using the `Movielens` dataset, the adaptive models $\texttt{xQuAD}_{quadr}$ and $\texttt{xQuAD}_{fuzzy}$ lead to improvements in terms of ILD and α-nDCG, and reductions in terms of S-Recall. With regard to aggregate

	weights	λ	**Precision**	**Recall**	**nDCG**	**ILD**	**S-Recall**	**α-nDCG**
BS			0.1488	0.0692	0.1634	0.3551	0.2310	0.2773
MMR		0.95	0.1484^{a}	$0.0686^{\mathbf{a}}$	0.1630^{a}	$0.3579^{\mathbf{a}}$	$0.2314^{\mathbf{a}}$	$0.2786^{\mathbf{a}}$
QUADR	$\langle$0.0, 0.0, 0.2, 0.8$\rangle$	0.55	0.1492^{b}	$0.0690^{\mathbf{b}}$	0.1637	$0.3629^{\mathbf{ab}}$	$0.2321^{\mathbf{ab}}$	$0.2806^{\mathbf{ab}}$
FUZZY	$\langle$0.0, 0.0, 0.1, 0.9$\rangle$	0.6	$0.1490^{\mathbf{b}}$	$0.0689^{\mathbf{b}}$	$0.1636^{\mathbf{a}}$	$0.3585^{\mathbf{ab}}$	$0.2314^{\mathbf{a}}$	$0.2789^{\mathbf{ab}}$

Table 4: Most accurate individuals from Pareto Frontiers for `Movielens` Dataset, using *BPRSLIM* and `MMR`. The superscripts a and b indicate statistically significant differences (Wilcoxon signed rank with $p < 0.05$) with respect to the baseline and `MMR` algorithms, respectively. Bold superscripts indicate stronger statistically significant differences (Wilcoxon signed rank with $p < 0.001$)

	weights	λ	**Precision**	**Recall**	**nDCG**	**ILD**	**S-Recall**	**α-nDCG**
BS			0.0132	0.0146	0.0180	0.3993	0.1375	0.2836
MMR		0.95	0.0131	0.0145	0.0179^{a}	$0.4099^{\mathbf{a}}$	$0.1385^{\mathbf{a}}$	$0.2859^{\mathbf{a}}$
QUADR	$\langle$0.1, 0.3, 0.0, 0.6$\rangle$	95	0.0135	0.0149	$0.0184^{\mathbf{b}}$	$0.4039^{\mathbf{ab}}$	$0.1375^{\mathbf{ab}}$	$0.2846^{\mathbf{ab}}$
FUZZY	$\langle$0.0, 0.9, 0.0, 0.1$\rangle$	85	0.0134	0.0147	0.0183^{a}	$0.4100^{\mathbf{ab}}$	$0.1379^{\mathbf{ab}}$	$0.2858^{\mathbf{ab}}$

Table 5: Most accurate individuals extracted from Pareto Frontiers for `LibraryThing` Dataset, using *BPRSLIM* and `MMR`

	weights	λ	**Precision**	**Recall**	**nDCG**	**ILD**	**S-Recall**	**α-nDCG**
BS			0.1488	0.0692	0.1634	0.3551	0.2310	0.2773
MMR		0.3	$0.1377^{\mathbf{a}}$	$0.0569^{\mathbf{a}}$	$0.1509^{\mathbf{a}}$	$0.4203^{\mathbf{a}}$	$0.2391^{\mathbf{a}}$	$0.2999^{\mathbf{a}}$
QUADR	$\langle$0.1, 0.1, 0.2, 0.6$\rangle$	0.15	$0.1417^{\mathbf{ab}}$	$0.0616^{\mathbf{ab}}$	$0.1554^{\mathbf{ab}}$	$0.4109^{\mathbf{ab}}$	$0.2377^{\mathbf{ab}}$	$0.2970^{\mathbf{ab}}$
FUZZY	$\langle$0.0, 0.1, 0.3, 0.6$\rangle$	0.1	$0.1405^{\mathbf{ab}}$	$0.0610^{\mathbf{ab}}$	$0.1541^{\mathbf{ab}}$	$0.4151^{\mathbf{ab}}$	$0.2395^{\mathbf{a}b}$	$0.2986^{\mathbf{ab}}$

Table 6: Individuals with best mean between Precision and ILD from Pareto Frontiers for `Movielens` Dataset, using *BPRSLIM* and `MMR`

	weights	λ	**Precision**	**Recall**	**nDCG**	**ILD**	**S-Recall**	**α-nDCG**
BS			0.0132	0.0146	0.0180	0.3993	0.1375	0.2836
MMR		0.7	$0.0123^{\mathbf{a}}$	$0.0133^{\mathbf{a}}$	$0.0168^{\mathbf{a}}$	$0.4486^{\mathbf{a}}$	$0.1420^{\mathbf{a}}$	$0.2896^{\mathbf{a}}$
QUADR	$\langle$0.1, 0.1, 0.3, 0.5$\rangle$	0.4	$0.0123^{\mathbf{a}}$	$0.0134^{\mathbf{a}}$	$0.0168^{\mathbf{ab}}$	$0.4591^{\mathbf{ab}}$	$0.1425^{\mathbf{ab}}$	$0.2898^{\mathbf{a}}$
FUZZY	$\langle$0.1, 0.6, 0.1, 0.2$\rangle$	0.75	$0.0129^{\mathbf{ab}}$	$0.0140^{\mathbf{ab}}$	$0.0176^{\mathbf{ab}}$	$0.4355^{\mathbf{ab}}$	$0.1407^{\mathbf{ab}}$	$0.2887^{\mathbf{ab}}$

Table 7: Individuals with best mean between Precision and ILD extracted from Pareto Frontiers for `LibraryThing` Dataset, using *BPRSLIM* and `MMR`

diversity, $\texttt{xQuAD}_{quadr}$ is able to improve coverage and Gini coefficient, while $\texttt{xQuAD}_{fuzzy}$ improves the Gini coefficient but not the coverage reached by `xQuAD`. Analysing novelty of recommendations, $\texttt{xQuAD}_{quadr}$ gives the highest values of EPC with small loss of Precision and best balance between Precision and EPC. It is noteworthy that the same trend on EPC occurs using $\texttt{MMR}_{quadr}$ on `Movielens`, as we may see in Figure 5.

Considering the `LibraryThing` dataset, there are relevant differences with `MMR`. $\texttt{xQuAD}_{quadr}$ and $\texttt{xQuAD}_{fuzzy}$ overcome `xQuAD` only in terms of S-Recall, while there is no evident difference in terms of ILD. α-nDCG shows a critical situation: $\texttt{xQuAD}_{fuzzy}$ gives the worst results while $\texttt{xQuAD}_{quadr}$ is able to increase the α-nDCG with non significant losses of precision. Moreover, $\texttt{xQuAD}_{quadr}$ and $\texttt{xQuAD}_{fuzzy}$ overcome `xQuAD` in terms of both coverage and Gini coefficient, therefore giving a real improvement of aggregate diversity. In particular, $\texttt{xQuAD}_{quadr}$ gives the highest values and the best compromise between accuracy and aggregate diversity. Also analyzing novelty of recommendations, $\texttt{xQuAD}_{quadr}$ and $\texttt{xQuAD}_{fuzzy}$ overcome `xQuAD`, giving better bal-

ance between precision and %Long-Tail and between precision and EPC[13].

Just as for MMR, we show in Tables 8, 9, 10 and 11 the best individuals of compared algorithms according to the configurations of objectives described above. The situation is analogous to the one depicted for MMR, since the main outcome is that our adaptive multi-attribute approaches xQuAD$_{quadr}$ and xQuAD$_{fuzzy}$ are able to improve the diversity without accuracy loss, while the pure xQuAD increases the diversity penalizing the accuracy, as expected. The statistically significance test validate the results even further, especially for diversity measures. The same considerations made for MMR on weights ω_4, ω_1 and ω_3 are still effective, and an analogous discordant behaviour for ω_2 can be observed too. In fact, for xQuAD$_{quadr}$ on LibraryThing in both configurations of objectives and xQuAD$_{fuzzy}$ on Movielens just in the first configuration, the value ω_2 is even higher than ω_4, while is almost zero in the other cases.

	weights	λ	**Precision**	**Recall**	**nDCG**	**ILD**	**S-Recall**	**α-nDCG**
BS			0.1488	0.0692	0.1634	0.3551	0.2310	0.2773
XQUAD		0.95	0.1479^{a}	0.0676^{a}	0.1621^{a}	$0.3633^{\mathbf{a}}$	$0.2339^{\mathbf{a}}$	$0.2815^{\mathbf{a}}$
QUADR	0.0,0.2,0.1,0.7	0.8	0.1494^{b}	0.0692	0.1638	$0.3631^{\mathbf{ab}}$	$0.2330^{\mathbf{ab}}$	$0.2806^{\mathbf{ab}}$
FUZZY	0.1,0.5,0.1,0.3	0.95	$0.1489^{\mathbf{b}}$	0.0688^{b}	0.1634^{b}	$0.3575^{\mathbf{ab}}$	$0.2315^{\mathbf{ab}}$	$0.2784^{\mathbf{ab}}$

Table 8: Most accurate individuals from Pareto Frontiers for Movielens Dataset, using *BPRSLIM* and xQuAD The superscripts a and b indicate statistically significant differences (Wilcoxon signed rank with $p < 0.05$) with respect to the baseline and xQuAD algorithms, respectively. Bold superscripts indicate stronger statistically significant differences (Wilcoxon signed rank with $p < 0.001$)

	weights	λ	**Precision**	**Recall**	**nDCG**	**ILD**	**S-Recall**	**α-nDCG**
BS			0.0132	0.0146	0.0180	0.3993	0.1375	0.2836
XQUAD		0.95	0.0131	0.0145	0.0179^{a}	$0.4099^{\mathbf{a}}$	$0.1385^{\mathbf{a}}$	$0.2859^{\mathbf{a}}$
QUADR	0.1,0.8,0.0,0.1	0.90	0.0135	0.0152	0.0184	$0.4123^{\mathbf{ab}}$	$0.1404^{\mathbf{ab}}$	$0.2867^{\mathbf{ab}}$
FUZZY	0.3,0.2,0.0,0.5	0.90	0.0134	0.0152	0.0183^{a}	$0.4165^{\mathbf{ab}}$	$0.1410^{\mathbf{ab}}$	$0.2864^{\mathbf{a}b}$

Table 9: Most accurate individuals extracted from Pareto Frontiers for LibraryThing Dataset, using *BPRSLIM* and xQuAD

	weights	λ	**Precision**	**Recall**	**nDCG**	**ILD**	**S-Recall**	**α-nDCG**
BS			0.1488	0.0692	0.1634	0.3551	0.2310	0.2773
XQUAD		0.8	0.1433^{a}	$0.0620^{\mathbf{a}}$	$0.1566^{\mathbf{a}}$	$0.3859^{\mathbf{a}}$	$0.2405^{\mathbf{a}}$	$0.2907^{\mathbf{a}}$
QUADR	0.1,0.1,0.1,0.7,	0.35	$0.1401^{\mathbf{ab}}$	$0.0578^{\mathbf{ab}}$	$0.1528^{\mathbf{ab}}$	$0.4143^{\mathbf{ab}}$	$0.2395^{\mathbf{ab}}$	$0.2966^{\mathbf{ab}}$
FUZZY	0.1,0.2,0.0,0.7,	0.35	$0.1401^{\mathbf{ab}}$	$0.0581^{\mathbf{ab}}$	$0.1528^{\mathbf{ab}}$	$0.4142^{\mathbf{ab}}$	$0.2401^{\mathbf{ab}}$	$0.2968^{\mathbf{ab}}$

Table 10: Individuals with best mean between Precision and ILD from Pareto Frontiers for Movielens Dataset, using *BPRSLIM* and xQuAD

	weights	λ	**Precision**	**Recall**	**nDCG**	**ILD**	**S-Recall**	**α-nDCG**
BS			0.0132	0.0146	0.0180	0.3993	0.1375	0.2836
XQUAD		0.7	$0.0123^{\mathbf{a}}$	$0.0133^{\mathbf{a}}$	$0.0168^{\mathbf{a}}$	$0.4486^{\mathbf{a}}$	$0.1420^{\mathbf{a}}$	$0.2896^{\mathbf{a}}$
QUADR	0.1,0.6,0.1,0.2,	0.9	$0.0134^{\mathbf{b}}$	$0.0151^{\mathbf{b}}$	$0.0183^{a\mathbf{b}}$	$0.4165^{\mathbf{ab}}$	$0.1410^{\mathbf{ab}}$	$0.2864^{\mathbf{ab}}$
FUZZY	0.1,0.1,0.0,0.8,	0.9	$0.0134^{\mathbf{b}}$	$0.0152^{\mathbf{b}}$	$0.0183^{a\mathbf{b}}$	$0.4165^{\mathbf{ab}}$	$0.1410^{\mathbf{ab}}$	$0.2864^{\mathbf{ab}}$

Table 11: Individuals with best mean between Precision and ILD extracted from Pareto Frontiers for LibraryThing Dataset, using *BPRSLIM* and xQuAD

[13]According to the Figures available at http://sisinflab.poliba.it/recommender-systems/adaptive-multi-attribute-diversity.html, the aforementioned trends on the results are generally confirmed using the adaptive diversification models upon other recommendation algorithms (BPRMF, WRMF, SoftMarginRankingMF, ItemKNN).

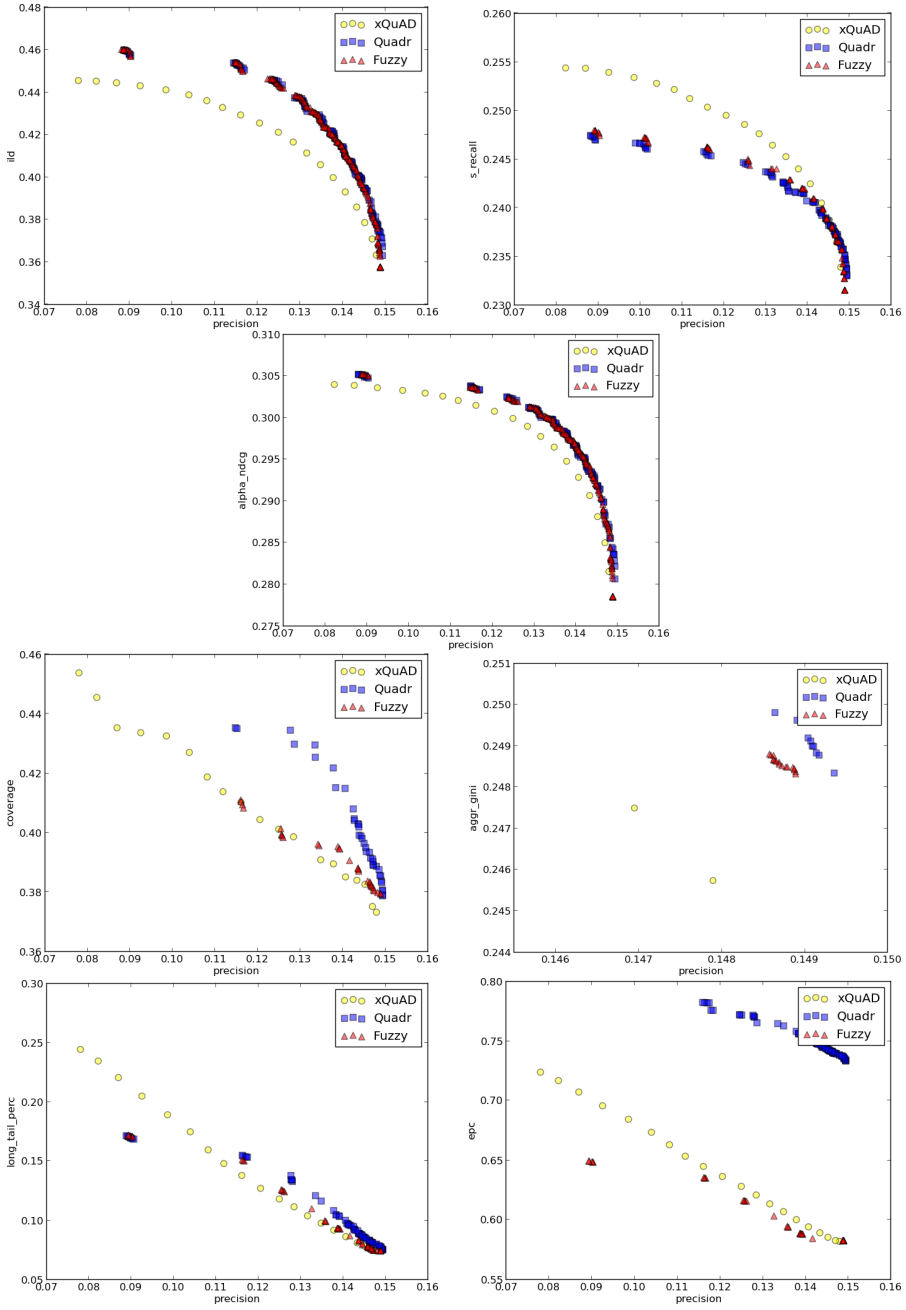

Figure 7: Pareto Frontiers for `Movielens` Dataset, using *BPRSLIM* and `xQuAD`

5.3. Results discussion

Summing up, previous results show that our proposed adaptive diversifications model is able to foster the recommendations quality in a multi-objective scenario. More specifically, considering the individual diversity,

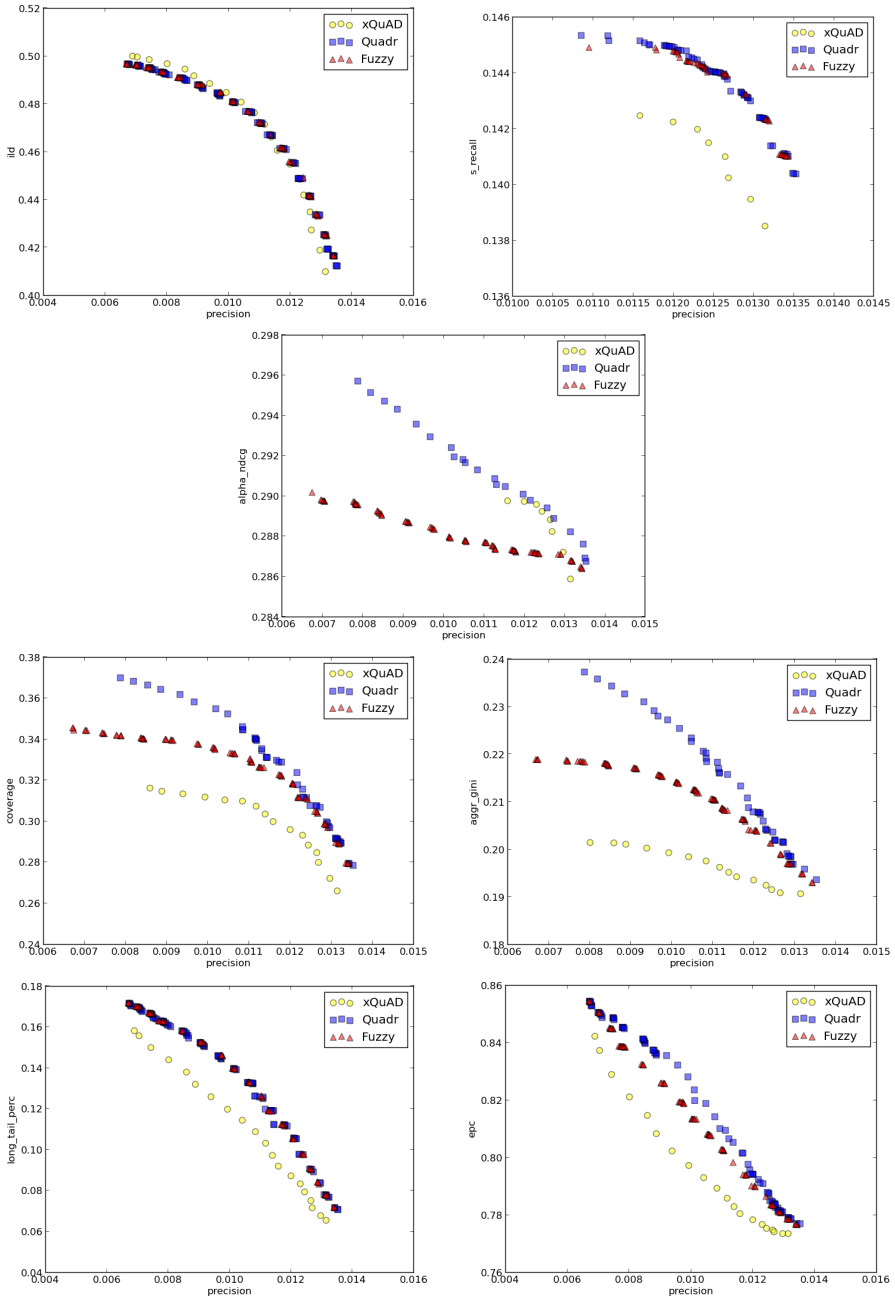

Figure 8: Pareto Frontiers for `LibraryThing` Dataset, using *BPRSLIM* and `xQuAD`

`MMR` benefits from the adaptive model in terms of S-Recall on both the datasets, while there is no significant difference in terms of ILD and α-nDCG. It is worth to note that the basic `MMR` gives the best results in terms ILD and S-Recall among different diversification algorithms, as demonstrated

in [40] and here the results show that it is possible to further improve those metrics with the $\texttt{MMR}_{quadr}$ and $\texttt{MMR}_{fuzzy}$. Moreover, it always obtains improvements considering novelty and aggregate diversity, especially using $\texttt{MMR}_{quadr}$. On the other hand, the adaptive models applied with `xQuAD` show different behaviours on the two datasets, especially considering the individual diversity. They are able to improve the ILD and α-nDCG results on `LibraryThing`, but not S-Recall, while on `Movielens` they improve S-Recall and only $\texttt{xQuAD}_{quadr}$ gives better results in terms of α-nDCG. As for `MMR`, also `xQuAD` obtains better results in terms of novelty and aggregate diversity, specially using $\texttt{xQuAD}_{quadr}$. In other words, the results suggest that generally using an adaptive model may improve all the balances between accuracy and the other quality dimensions, or at least improve some of them and do not make the other worse. As an additional consideration, the values for quadrant weights proposed in Tables from 4 to 11 give worth and effectiveness to our idea of using profile size and entropy to cluster users into groups and approaching their predilection to diversity through belonging groups.

The main difference between the hard clustering and the fuzzy one is that the former assumes that a user can belong to only a quadrant for each attribute, while the second lets a user belong to different quadrants simultaneously with different degree for the same attribute. As a consequence, the hard clustering is straighter, while the fuzzy version tends to distribute more equally the quadrants weights since each quadrants gives a more or less significant contribution. Although the hard version is a simple clustering of users by means of median values, the results point out a positive impact of a clear division of users on most of the evaluation metrics. The hard clustering is able to beat both the fuzzy one and the diversification baseline in terms of Aggregate Diversity (coverage and Gini coefficient) and also in terms of novelty. This is more evident when considering EPC. On the other hand, fuzzy clustering remains very close to the diversification baseline. The reason of this outcome could be found in the aforementioned difference: the hard clustering is more straight than the fuzzy one, therefore it is much more selective during the re-raking phase.

6. Conclusion

Computing effective recommendations calls for approaches which are able to provide not just accurate lists of results. Modern recommendation engines need to go beyond accuracy and consider, while computing a recommendation list, also other dimensions such as diversity in the recommendation list to reduce the *portfolio effect*, catalog coverage to maximize the number of items in the catalog recommended to the users and novelty of results to mitigate the popularity bias thus suggesting also items in the long tail. In particular, it has been shown [17] that reducing the portfolio effect by increasing diversity in the recommendation list plays an important role

on user satisfaction. The task is not trivial especially when we deal with a multi-attribute personalized diversification results. In the recent years, the importance of adapting the recommendation diversity to user's needs with respect to different attributes has strongly emerged, although research on multi-attribute diversity is still in its early stage.

In order to fill this gap, in this work we introduced an adaptive multi-attribute diversification method according to the hypothesis that a user who selected many diverse items in the past could be more willing to receive diverse recommendations. With reference to the research questions pointed out in Section 1, as an answer to question (i), we proposed to model the user profile by taking into account her attitude to enjoy (or not) items which result diverse with regard to different attributes and eventually adopt this modeling to foster diversity in the list returned by a recommendation engine. Our modeling has been exploited to re-rank the list of items produced by whatever recommender system to reduce the portfolio effect. In order to evaluate the effectiveness of our hypothesis we tested two different versions of our profile modeling (we called them hard and fuzzy), built upon two different state-of-the-art diversification methods - `MMR` and `xQuAD`- in the movie domain on `Movielens 1M` dataset and in the book domain on `LibraryThing` dataset. As for the evaluation of the recommendation quality we considered a wide range of metrics to measure four important quality dimensions: Accuracy, Individual Diversity, Aggregate Diversity, and Novelty in top-N recommendation task. As an answer to question (ii), the experimental results confirmed our intuition on the need of tailoring diversity degree to actual user's interests in a personalized way, pointing out the inadequate performances of non adaptive diversification baselines. Finally, our approach can be considered as a step forward to solve the challenge posed by question (iii). In fact, the construction of a content-based user profile in terms of diversity allowed not only to customize the degree of individual diversity in the recommendation list but led to better recommendation quality in a multi-objective scenario. In particular, our adaptive model overcame the traditional accuracy-diversity trade-off issue, improving different quality objectives, without affecting the others. Hence, the results let us draw the conclusion that diversification methods tailored to actual user's needs produce better recommendations from a broad user utility perspective.

The outcomes presented in this paper pave the way to further investigations and research directions in the design and evaluation of multi-attribute diversity approaches. From the point of view of attributes selection, other domain independent side information may be taken into account such as popularity or even latent dimensions. A further related aspect to be considered is that of time-aware selection of attributes and corresponding values. Interesting results to estimate and detect peaks of interest have already been presented in [24] while in [23] the idea to model individual needs is put forward with respect to the novelty property, with emphasis on user's dynamic

behaviour and time dependency. Additional investigation to understand the approach that should be used for users with a small profile and a high value of entropy needs also to be done together with the role of individual diversity in cold-start situations. Reasonably, a hybrid system like the one used in [3], able to switch between different approaches depending on the actual needs, could be conveniently applied to match the demand of both cold and expert users.

Acknowledgements. The authors acknowledge partial support of project PON03 PE_00136_1 Digital Services Ecosystem: DSE. Jessica Rosati also acknowledges support of I.B.M. Ph.D. fellowship 2015.

References

[1] Panagiotis Adamopoulos and Alexander Tuzhilin. On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology*, 5(4):54:1–54:32, December 2014.

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

[3] Álvaro Tejeda-Lorente, Carlos Porcel, Eduardo Peis, Rosa Sanz, and Enrique Herrera-Viedma. A quality based recommender system to disseminate information in a university digital library. *Information Sciences*, 261:52 – 69, 2014.

[4] Azin Ashkan, Branislav Kveton, Shlomo Berkovsky, and Zheng Wen. Optimal greedy diversity for recommendation. In Qiang Yang and Michael Wooldridge, editors, *IJCAI*, pages 1742–1748. AAAI Press, 2015.

[5] Ana Belén Barragáns-Martínez, Enrique Costa-Montenegro, Juan C. Burguillo, Marta Rey-López, Fernando A. Mikic-Fonte, and Ana Peleteiro. A hybrid content-based and item-based collaborative filtering approach to recommend tv programs enhanced with singular value decomposition. *Information Sciences*, 180(22):4290 – 4311, 2010.

[6] Fabiano Belém, Rodrygo Santos, Jussara Almeida, and Marcos Gonçalves. Topic diversity in tag recommendation. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, pages 141–148. ACM, 2013.

[7] Alejandro Bellogín, Iván Cantador, and Pablo Castells. A comparative study of heterogeneous item recommendations in social systems. *Information Sciences*, 221:142–169, February 2013.

[8] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. Recommender systems survey. *Knowledge-Based Systems*, 46:109 – 132, 2013.

[9] K. Bradley and B. Smyth. Improving Recommendation Diversity. In *Proceedings of the Irish Conference in Artificial Intelligence and Cognitive Science*, pages 75–84, 2001.

[10] Robin Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370, November 2002.

[11] Jaime Carbonell and Jade Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '98, pages 335–336. ACM, 1998.

[12] Pablo Castells, Neil J. Hurley, and Saul Vargas. Novelty and diversity in recommender systems. In *Recommender Systems Handbook*, pages 881–918. Springer US, Boston, MA, 2015.

[13] L. Chen, W. Wu, and L. He. How personality influences users' needs for recommendation diversity? In *Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 829–834, 2013.

[14] C. L.A. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon. Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 659–666. ACM, 2008.

[15] T. Di Noia, V. C. Ostuni, J. Rosati, P. Tomeo, and E. Di Sciascio. An analysis of users' propensity toward diversity in recommendations. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, pages 285–288. ACM, 2014.

[16] Marina Drosou and Evaggelia Pitoura. Comparing diversity heuristics. Technical report, Technical Report 2009-05. Computer Science Department, University of Ioannina, 2009.

[17] Michael D. Ekstrand, F. Maxwell Harper, Martijn C. Willemsen, and Joseph A. Konstan. User perception of differences in recommender algorithms. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, pages 161–168. ACM, 2014.

[18] Ignacio Fernández-Tobías, Matthias Braunhofer, Mehdi Elahi, Francesco Ricci, and Iván Cantador. Alleviating the new user problem

in collaborative filtering by exploiting personality information. *User Model. User-Adapt. Interact.*, 26(2-3):221–255, 2016.

[19] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach. Beyond accuracy: Evaluating recommender systems by coverage and serendipity. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys '10, pages 257–260. ACM, 2010.

[20] Sreenivas Gollapudi and Aneesh Sharma. An axiomatic approach for result diversification. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 381–390. ACM, 2009.

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

[22] Ren-Hung Hwang, Yu-Ling Hsueh, and Yu-Ting Chen. An effective taxi recommender system based on a spatio-temporal factor analysis model. *Information Sciences*, 314(C):28–40, September 2015.

[23] Komal Kapoor, Vikas Kumar, Loren Terveen, Joseph A. Konstan, and Paul Schrater. "I Like to Explore Sometimes": Adapting to dynamic user novelty preferences. In *Proceedings of the 9th ACM Conference on Recommender Systems*, RecSys '15. ACM, 2015.

[24] H. Khrouf and R. Troncy. Hybrid event recommendation using linked data and user diversity. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, pages 185–192. ACM, 2013.

[25] Onur Küçüktunç, Erik Saule, Kamer Kaya, and Ümit V. Çatalyürek. Diversifying citation recommendations. *ACM Transactions on Intelligent Systems and Technology*, 5(4):55:1–55:21, December 2014.

[26] Seok Kee Lee, Yoon Ho Cho, and Soung Hie Kim. Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations. *Information Sciences*, 180(11):2142 – 2155, 2010.

[27] Andrii Maksai, Florent Garcin, and Boi Faltings. Predicting online performance of news recommender systems through richer evaluation metrics. In *Proceedings of the 9th ACM Conference on Recommender Systems*, RecSys '15, pages 179–186. ACM, 2015.

[28] C. Martinez-Cruz, C. Porcel, J. Bernabé-Moreno, and E. Herrera-Viedma. A model to represent users trust in recommender systems using ontologies and fuzzy linguistic modeling. *Information Sciences*, 311:102 – 118, 2015.

[29] D. G. Mcdonald and J. Dimmick. The conceptualization and measurement of diversity. *Communication Research*, 30(1):60–79, 2003.

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

[31] V. C. Ostuni, T. Di Noia, E. Di Sciascio, and R. Mirizzi. Top-n recommendations from implicit feedback leveraging linked open data. In *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys '13, pages 85–92. ACM, 2013.

[32] Umberto Panniello, Alexander Tuzhilin, and Michele Gorgoglione. Comparing context-aware recommender systems in terms of accuracy and diversity. *User Modeling and User-Adapted Interaction*, 24(1-2):35–65, February 2014.

[33] Lijing Qin and Xiaoyan Zhu. Promoting diversity in recommendation by entropy regularizer. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI '13, pages 2698–2704. AAAI Press, 2013.

[34] Marco Tulio Ribeiro, Nivio Ziviani, Edleno Silva De Moura, Itamar Hata, Anisio Lacerda, and Adriano Veloso. Multiobjective pareto-efficient approaches for recommender systems. *ACM Transactions on Intelligent Systems and Technology*, 5(4):53:1–53:20, December 2014.

[35] R.L.T. Santos, C. Macdonald, and I. Ounis. Exploiting query reformulations for web search result diversification. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 881–890, 2010.

[36] Y. Shi, X. Zhao, J. Wang, M. Larson, and A. Hanjalic. Adaptive diversification of recommendation results via latent factor portfolio. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 175–184, 2012.

[37] Barry Smyth and Paul McClave. Similarity vs. diversity. In *Proceedings of the 4th International Conference on Case-Based Reasoning: Case-Based Reasoning Research and Development*, ICCBR '01, pages 347–361, London, UK, UK, 2001. Springer-Verlag.

[38] Nava Tintarev and Judith Masthoff. Similarity for news recommender systems. In *In Proceedings of the AH06 Workshop on Recommender Systems and Intelligent User Interfaces*, 2006.

[39] Paolo Tomeo, Tommaso Di Noia, Marco de Gemmis, Pasquale Lops, Giovanni Semeraro, and Eugenio Di Sciascio. Exploiting regression trees as user models for intent-aware multi-attribute diversity. In *Proceedings of the 2nd Workshop on New Trends on Content-Based Recommender Systems co-located with 9th ACM Conference on Recommender Systems (RecSys 2015)*, pages 2–9, 2015.

[40] Saúl Vargas, Linas Baltrunas, Alexandros Karatzoglou, and Pablo Castells. Coverage, redundancy and size-awareness in genre diversity for recommender systems. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, pages 209–216. ACM, 2014.

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

[42] Saúl Vargas and Pablo Castells. Exploiting the diversity of user preferences for recommendation. In *Proceedings of the 10th Conference on Open Research Areas in Information Retrieval*, OAIR '13, pages 129–136, Paris, France, 2013.

[43] Saúl Vargas and Pablo Castells. Improving sales diversity by recommending users to items. In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, pages 145–152. ACM, 2014.

[44] Wen Wu, Li Chen, and Liang He. Using personality to adjust diversity in recommender systems. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, HT '13, pages 225–229. ACM, 2013.

[45] Cong Yu, Laks Lakshmanan, and Sihem Amer-Yahia. It takes variety to make a world: Diversification in recommender systems. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, EDBT '09, pages 368–378. ACM, 2009.

[46] Cheng Xiang Zhai, William W. Cohen, and John Lafferty. Beyond independent relevance: Methods and evaluation metrics for subtopic retrieval. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '03, pages 10–17. ACM, 2003.

[47] Mi Zhang and Neil Hurley. Avoiding monotony: Improving the diversity of recommendation lists. In *Proceedings of the 2008 ACM Conference on Recommender Systems*, RecSys '08, pages 123–130. ACM, 2008.

[48] 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*, 107:4511–4515, 2010.

[49] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, pages 22–32. ACM, 2005.