An Analysis of Users' Propensity toward Diversity in Recommendations

Tommaso Di Noia[1], Vito Claudio Ostuni[1], Jessica Rosati[1,2], 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
{firstname.lastname}@poliba.it

ABSTRACT

Providing very accurate recommendations to end users has been nowadays recognized to be just one of the main tasks a recommender systems must be able to perform. While predicting relevant suggestions, attention needs to be paid to their diversification in order to avoid monotony in recommendation. In this paper we focus on modeling users' inclination 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-rank the list of Top-N items predicted by a recommendation algorithm, in order to foster diversity in the final ranking. Experimental evaluation proves the effectiveness of the proposed approach.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Search and Retrieval

Keywords

Recommender System; Diversity; MMR; ILD; Movielens

1. INTRODUCTION

The main task of a recommendation engine is typically to predict the ratings of unknown items for each user 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 mainly devoted to maximizing recommendation accuracy. However, it has been recognized that improving only the predictive accuracy of recommendations is not enough to judge the effectiveness of a recommender system [4]. The most accurate recommendations for a user are often too similar to each other and attention has to be paid towards the goal of improving diversity in recommended items thus avoiding monotony. A number of works propose strategies to enhance the trade-off between accuracy and diversity, which however are not adaptive with respect to individual users' spontaneous needs [13, 12, 14]. Recently, the idea of exploiting the user profile in the diversification approach got ahead. In [11] the identification of diversity within the user profile is carried out through the extraction of user subprofiles to reflect the polyfacetic nature of user interests.

The main intuition behind our work is that some users may prefer diversification in suggestions while others may not. In addition, 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 interested in diverse suggestions related to other ones (e.g. *genre* in the movie domain). Following this idea, in this paper we investigate *individual* diversity[1] taking into account user diversity needs over item attributes. As in [11], we consider the principle that good recommendations should be not only diverse, but they should also realize the diversity within the user interests and behaviour, to provide recommendations that embody a better fitted and more comprehensive view of user preferences. We propose an *adaptive attribute-based* diversification approach able to customize the degree of recommendation diversity of the *Top-N* recommendation list taking into account the inclination to diversity of the user over different content-based item dimensions. Specifically, we employ Entropy as a measure of the degree of diversity in the user preferences and use it in conjunction with user profile dimensions for calibrating the degree of diversification of the list. We apply our approach to the movie domain, considering what leads a user to choose a movie in a huge collection of items. Reasonably the choice of a user can be led by several factors, such as *genre*, *actor*, *director* and *year of release*. However not all these factors have the same influence on different users: by way of example, a user can be interested in a movie director more than in its genre. Analogously, a user can decide to cling to a particular director and accept to watch several genres. As a consequence we need to measure such differences, that is to show the different tendencies of each user in diversifying her choices, dealing separately with the significant movie attributes mentioned above. Differently from [11], where the movie genre is the only attribute used to characterize subprofiles, we analyze user propensity to diversify, in different ways, the *genre*, the *actor*, *director* or *year of release*. The selection of these attributes follows the results pointed out

[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 of recommendations across all users [1].

RecSys'14, October 6–10, 2014, Foster City, Silicon Valley, CA, USA.

http://dx.doi.org/10.1145/2645710.2645774.

in [6] where the authors show a correlation between personality factors (such as openness and conscientiousness) and the degree of diversification in the user choices with respect to such attributes[2]. The main contributions of this paper are:

- a representation of user's propensity in diversifying her choices. By looking at the user's profile we model her inclination to diversify the items she chooses in relation to the attributes describing items in the collection.
- an adaptive attribute-based re-ranking approach that takes into account individual diversity while suggesting new items to the user. The ranking of items suggested to the end user is computed by considering her personal propensity to diversify the items she selects. The approach has been experimentally evaluated in the movie domain proving its effectiveness.

The remainder of the paper is structured as follows: after reporting an overview of the diversity property in Section 2, we propose the details of our recommendation approach in Section 3 and the experimental results in Section 4. Conclusion and future work close the paper.

2. DIVERSITY IN RECOMMENDATION

The activity of a recommender system can be divided into two phases: firstly there is the prediction of the ratings for unrated items and secondly the items can be re-ranked to maximize user's utility. According to [1], in order to improve the diversity of recommendations it is possible to deal only with the second phase. The individual diversity can be increased through several heuristics. In this work we consider greedy heuristics, which look for an intermediate solution and have been demonstrated to be efficient and effective. Particularly, we borrow the greedy strategy presented in [11]. 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 $\overline{\mathbf{P}} = \langle 1, ..., n \rangle$ be the recommendation list for user u generated using the predicted ratings and let us suppose we want to provide the user with the re-ranked list $\overline{\mathbf{S}}$ of recommendations, such that $\mathbf{S} \subset \mathbf{P}$ and whose length is $N \leq n$. The adopted greedy strategy can be explained through Algorithm 1.

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

1 $\overline{\mathbf{S}} = \langle\rangle$;
2 **while** $|\mathbf{S}| < N$ **do**
3 $\quad i^* = \underset{i \in \mathbf{P} \setminus \mathbf{S}}{\operatorname{argmax}}\ f_{obj}(i, \overline{\mathbf{S}})$;
4 $\quad \overline{\mathbf{S}} = \overline{\mathbf{S}} \circ i^*$;
5 $\quad \mathbf{P} = \mathbf{P} \setminus \{i^*\}$
6 **end**
7 **return S**.

Algorithm 1: The greedy strategy

At each step, we select the item which maximizes an objective function f_{obj} (line 3), which in turn can be defined to deal with the trade-off between accuracy and diversity, and we add it to the re-ranked list (line 4). In our work, we adopt the Maximal Marginal Relevance (MMR) [5], whose objective function is defined as:

$$f_{obj}(i, \overline{\mathbf{S}}) = \lambda \cdot r^*(u, i) - (1 - \lambda) \cdot \max_{j \in \mathbf{S}} sim(i, j) \qquad (1)$$

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

3. ADAPTIVE DIVERSIFICATION

In the following we detail the *adaptive attribute-based* diversification model proposed in this paper. We provide a way to compute similarity that considers the diversification attitude of each user with respect to different item attributes (i.e. *year*, *genre*, *director* and *actors* in the movie domain). For each attribute $a \in \mathbf{A}$, we measure the user propensity towards diversity through Shannon's entropy, which is a two-dimensional measure, meaning that it reflects both the categories of classification and the distribution of elements within those categories. Shannon's entropy for user u and attribute a with k values can be computed as:

$$\mathcal{H}_a(u) = -\sum_{i=1}^{k} p_i \cdot \log_k p_i, \qquad (2)$$

where p_i is the relative frequency of the i-th value (category) of a considering all the items (elements) belonging to the user profile (collection of the items rated by that 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 classification 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 defined as in Table 1.
The main modelling hypothesis behind this classification is the following. Users who have explored items with different characteristics in the past are willing to accept diverse recommendations. Given an attribute a, we interpret a high value of entropy as an attitude of the user to choose items with different values for a. Conversely, a low value of entropy is read as her willing to consider items similar for that attribute.
Of course, more groups can be defined thus identifying more than four quadrants. Nevertheless, we will show in our experimental section that already with such a coarse grained classification we obtain interesting results in terms of *precision* and *intra list diversity* (ILD) values.

The quadrants the user belongs to, potentially different for each item attribute, are used to define the similarity measure in Equation (1). Let us consider a user u and indicate with $\mathbf{A}$ the set of item attributes (for example in the movie domain $\mathbf{A} = \{year, genre, direction, starring\}$). We consider a function $q_u : \mathbf{A} \to \{1, 2, 3, 4\}$, which assigns, for each attribute, the quadrant to which user u belongs and then we define a quadrant weight $\omega_i \in [0, 1]$, with $i \in \{1, 2, 3, 4\}$. Taking into accounts such weights and set-

[2]In fact, in [6] the attribute *country* is considered too, but we prefer not considering it since in the dataset we use for our experiments *country* is specified for less than 10% of the items.

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

Table 1: Quadrants

ting $m = \max\{\omega_i \mid i = 1, 2, 3, 4\}$, 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:

$$sim(i, j) = \frac{\sum_{a \in \mathbf{A}} \omega_{q_u(a)} \cdot sim_a(i, j)}{m \cdot |\mathbf{A}|} \tag{3}$$

with $sim_a(i, j)$ being a similarity measure between i and j with respect to attribute a. The weights associated to user belonging quadrants influence the similarity score and hence the resulting objective function of MMR, with the consequence of varying the ILD of the recommendation list.

Specifically, the weights account for the user propensity in diversifying every single attribute, based on our modelling hypothesis. 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 $f_{obj}(i, \overline{\mathbf{S}})$ for the items i most similar to the ones already available in $\overline{\mathbf{S}}$. Those items are the ones whose number we want to reduce in $\overline{\mathbf{S}}$, in order to guarantee a high ILD value. Conversely, assigning low weights to the first and third quadrant (low weights for ω_1 and ω_3) results in a significant lowering of the original similarity score and hence in an increasing of the corresponding $f_{obj}(i, \overline{\mathbf{S}})$ values. This corresponds to foster items similar to the ones in the re-ranked list $\overline{\mathbf{S}}$.

4. EXPERIMENTS AND RESULTS

In order to test whether our proposal of using adaptive attribute-based recommendations is competitive, we carried out experiments on a subset of the well known `Movielens 1M`[3] dataset. The original dataset was enriched with further attribute information such as actors and directors extracted from `DBpedia`[4]. More details about this `DBpedia` enriched version of the dataset are available in [8]. We concentrated on users who gave at least fifty ratings: as a consequence, the final dataset contains 4297 users, 3689 items and 942590 ratings. We ordered the ratings by using the timestamps available in the dataset and built the training set by using, for each user, the first 60% of the ratings and the remaining 40% to build the test set. We compared our approach with two baselines: (i) recommendations generated considering the predicting ratings without any diversification strategy, denoted as *no-MMR*; (ii) recommendations computed by using Equation 1 without considering quadrants, wherein diversification is applied to all users, denoted as *MMR*. Hereafter we refer to our adaptive attribute-based diversification approach as *adaptiveMMR*. The main difference between *MMR* and *adaptiveMMR* is that the latter considers different values for each ω_i depending on the position, for each attribute, of the user in one of the four quadrants. We remark that the main insight behind *adaptiveMMR* is that the re-ranking is tailored to each user according to the idea that the diversification with respect to an attribute should be applied only for users who exhibited a propensity in diversifying such an attribute. Conversely, in *MMR*, diversification is applied indiscriminately to all users regardless of whether they are incline to diversifying their choices or not.

To compute values for *no-MMR* we adopted a generic user-based kNN Collaborative Filtering algorithm available in Apache Mahout[5] using Pearson correlation as similarity measure. Then, for each user we selected the top 200 recommendations generated by *no-MMR* to build the initial $\overline{\mathbf{P}}$ list used for performing the re-ranking as shown in Algorithm 1. In both *MMR* and *adaptiveMMR* the λ parameter was set to 0.5. As similarity measure for attribute a in (3), we decided to use the Jaccard index. Since the number of distinct attribute values was too large for the *year*, *actors* and *director* attributes, we divided movies in decades and performed a K-means clustering for actors and directors on the basis of their `DBpedia` categories. For both directors and actors we obtained 20 clusters. The number of values is 19 and 8 for *genre* and *year*, respectively.

For evaluating recommendation ranking accuracy and diversity we used the *TestItems* evaluation methodology presented in [3]. For measuring accuracy we used precision (P@k) and nDCG@k, while ILD@k was used for diversity. Furthermore we measured the balance between accuracy and diversity, as in [9], that is we standardized the metrics P@k and ILD@k in order to make the scales homogeneous and we computed the average. We indicate this combination metric as *avg(P,ILD)*. P@k is chosen instead of nDCG@k since they have a similar trend.

In the first part of the experiments we tested the validity of the modelling hypothesis behind our approach, that is users who have explored different items in the past are inclined to diversity. Specifically we compared the performances of the two baselines *no-MMR* and *MMR* on different groups of users considering those belonging to the same quadrant for at least three attributes as shown in Table 2. As expected, *MMR* dominates the *no-MMR* for quadrant 2 and 4 for both precision and ILD, demonstrating that users with high entropy benefit from diversification. In the other quadrants (1 and 3) there is a normal decrease of accuracy. Hence users with low entropy in their user profiles are not inclined to an uncontrolled diversification. This preliminary analysis supports the adoption of entropy as discriminant parameter for users' classification.

In order to test the effectiveness of *adaptiveMMR* we evaluated various combination of ω weights. Particularly, we conducted a grid search on ω, finding, as a first result, that our intuition of choosing small values for ω_1 and ω_3 and bigger ones for ω_2 and ω_4 is validated by accuracy and ILD results. Without such constraints, in fact, the accuracy values of *adaptiveMMR* get deeply worse. For lack of space we discuss here only three weights configurations $\langle \omega_1, \omega_2, \omega_3, \omega_4 \rangle$ (adopting the notation introduced in Section 2 for lists), A, B and C, respectively given by A $= \langle 0, 0, 0, 1 \rangle$, B $= \langle 0, 1, 0, 1 \rangle$, C $= \langle 0.1, 1, 0.1, 0.75 \rangle$. The values of list C have

[3]Available at `http://grouplens.org/datasets/movielens`
[4]`http://dbpedia.org`
[5]`http://mahout.apache.org`

	Quadrant 1 (1149 users)		*Quadrant 2* (469 users)		*Quadrant 3* (467 users)		*Quadrant 4* (1146 users)	
algorithm	**P@10**	**ILD@10**	**P@10**	**ILD@10**	**P@10**	**ILD@10**	**P@10**	**ILD@10**
no-MMR	**0.0455**	0.3890	0.0678	0.3663	**0.0904**	0.3961	0.1306	0.3544
MMR	0.0394	**0.4363**	**0.0706**	**0.4212**	0.0829	**0.4355**	**0.1325**	**0.4012**

Table 2: Accuracy and Diversity Results distributed among the different quadrants. *Quadrant 1* contains users belonging to Quadrant 1 for at least 3 attributes; analogously for the other quadrants (users who belong for each attribute to quadrant 2 are only 69, while those belonging always to quadrant 3 are 70).

algorithm	nDCG@10	P@10	ILD@10	avg(P,ILD)
no-MMR	0.0840	0.0842	0.3764	0.3019
MMR	0.0837^a	0.0827^a	$\mathbf{0.4236}^a$	0.5000
AdaptiveMMR-A	0.0851^{ab}	0.0849	0.3921^{ab}	0.6184
AdaptiveMMR-B	$\mathbf{0.0855}^b$	0.0850^{ab}	0.4049^{ab}	0.7592
AdaptiveMMR-C	0.0854^{ab}	$\mathbf{0.0852}^b$	0.4101^{ab}	**0.8561**

Table 3: Accuracy and Diversity Results on all users. The superscripts *a* and *b* indicate statistically significant differences (Wilcoxon signed rank with $p < 0.05$) with respect to the *no-MMR* and *MMR* algorithms, respectively.

been computed via grid search where we fixed ω_1 and ω_3 and varied ω_2 and ω_4 by considering a discretization step of 0.25. These configurations let us deal with emblematic situations and confirm our intuition on the importance of individual entropy values. In particular, configuration A acts on users who are in quadrant 4 for some attributes, configuration B on users belonging to quadrant 2 or quadrant 4, and configuration C imposes the constraints $\omega_1 < \omega_2$ and $\omega_3 < \omega_4$. The experimental results are shown in Table 3 and are relative to $k = 10$. With other values of k the results are analogous, so we omitted them for brevity. Consistently with the accuracy-diversity trade-off, the basic *MMR* approach improves the diversity by sacrificing the accuracy. *AdaptiveMMR*, compared to both *no-MMR* and *MMR*, gains the best balance between accuracy and diversity, represented by *avg(P,ILD)*, regardless of the aforementioned weights configurations. In terms of accuracy, *adaptiveMMR* out-performs *no-MMR* and *MMR*, while in terms of diversity, they perform better than *no-MMR*. Specifically, *adaptiveMMR-C* obtains the best results in terms of precision, diversity and their combination. Remarkably, the configuration C has an ILD value close to *MMR* but a significantly better accuracy values. The higher values of accuracy and diversity for *adaptiveMMR-B* and *adaptiveMMR-C* compared to *adaptiveMMR-A* suggest also that the diversification tendency, represented by entropy, should be considered even for users with a small profile length.

5. CONCLUSIONS AND FUTURE WORK

Computing effective recommendations calls for approaches taking into account not only accurate lists of results. Diversity in the final ranking plays an important role in the user satisfaction [4]. In this paper we focus on *individual* diversity. We firstly propose a novel approach to model users' propensity to accept diverse recommendations. The modeling hypothesis is that a user who selected many diverse items in the past is more willing to receive diverse recommendation. We then take into account the past diversification behaviour of each user and give the right weight to her attitude toward diversification with respect to each item attribute. This modelling is then exploited to re-rank the list of recommended items. An experimental evaluation in the movie domain on the `Movielens 1M` proved the effectiveness of the proposed approach. Experiments have been carried out considering the MMR criterion [5] and we are planning to test other criteria to diversify results in the final recommendation list such as IA-select [2] or xQuAD [10]. The usage of a more fine grained analysis on the user profile length, considering quantiles instead of medians will be also tested to evaluate how this affects final results. Moreover we are investigating how to apply our approach to estimate and detect peaks of interest as in [7].

Acknowledgements. The authors acknowledge partial support of PON01_00850 ASK-Health and PON02_00563_3470993 VINCENTE

6. REFERENCES

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

[2] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying search results. In *ACM WSDM '09*, pages 5–14, 2009.

[3] A. Bellogin, P. Castells, and I. Cantador. Precision-oriented evaluation of recommender systems: An algorithmic comparison. In *ACM RecSys '11*, pages 333–336, 2011.

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

[5] J. Carbonell and J. Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *ACM SIGIR '98*, pages 335–336, 1998.

[6] 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.

[7] H. Khrouf and R. Troncy. Hybrid event recommendation using linked data and user diversity. In *ACM RecSys '13*, pages 185–192, 2013.

[8] V. C. Ostuni, T. Di Noia, E. Di Sciascio, and R. Mirizzi. Top-n recommendations from implicit feedback leveraging linked open data. In *ACM RecSys '13*, pages 85–92, 2013.

[9] U. Panniello, A. Tuzhilin, and M. Gorgoglione. Comparing context-aware recommender systems in terms of accuracy and diversity. *User Modeling and User-Adapted Interaction*, 24(1-2):35–65, 2014.

[10] R.L.T. Santos, C. Macdonald, and I. Ounis. Exploiting query reformulations for web search result diversification. In *WWW '10*, pages 881–890, 2010.

[11] S. Vargas and P. Castells. Exploiting the diversity of user preferences for recommendation. In *OAIR '13*, pages 129–136, 2013.

[12] M. Zhang. Enhancing diversity in top-n recommendation. In *ACM RecSys '09*, pages 397–400, 2009.

[13] M. Zhang and N. Hurley. Avoiding monotony: Improving the diversity of recommendation lists. In *ACM RecSys '08*, pages 123–130, 2008.

[14] C. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *WWW '05*, pages 22–32, 2005.