Adaptive Diversity in Recommender Systems*

Tommaso Di Noia[1], Vito Claudio Ostuni[1], Jessica Rosati[1,2,3], 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
[3] University of Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, 20126 Milano
{firstname.lastname}@poliba.it

Abstract. The evaluation of a recommendation engine cannot rely only on the accuracy of provided recommendations. One should consider additional dimensions, such as diversity of provided suggestions, in order to guarantee heterogeneity in the recommendation list. In this paper we analyse users' propensity in selecting diverse items, by taking into account content-based item attributes. Individual propensity to diversification is used to re-rank the list of Top-N items predicted by a recommendation algorithm, with the aim of fostering diversity in the final ranking. We show experimental results that confirm the validity of our modelling approach.

1 Introduction

In the recommender systems field, most of the approaches have been devoted to maximizing recommendation accuracy. However, it has been recognized that improving only the predictive accuracy is not enough to judge the effectiveness of a recommender system [3], since 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 *individual* diversity, the degree of diversification in the recommendations provided to an individual user. A number of works propose strategies to enhance the trade-off between accuracy and diversity [9, 8, 10].

The main intuition behind our work is that some users may prefer diversification in suggestions while others may not and they could be inclined to diversify with respect to not all item attributes. We propose an *adaptive attribute-based* diversification approach able to customize the degree of individual diversity of the *Top-N* recommendation list, using the Entropy measure to represent the inclination to diversity of the user over different content-based item dimensions. We apply our approach to the movie domain, considering what reasonably leads a user to choose a movie in a huge collection of items, that is *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 decide to cling to a particular director and accept to watch several genres.

The main contributions of this paper are:

* An extended version of this paper has been published in [4].

– a representation of user's propensity in diversifying her choices.
– an adaptive attribute-based re-ranking approach based on the aforementioned representation.

2 Adaptive diversification

In the recommendation process, after the ratings prediction for unrated items, the maximization of user's utility and the improvement of individual diversity in the items list can be pursued through a re-ranking phase [1]. There are several heuristics which let to re-rank items in an efficient way, such as the MMR greedy strategy [7]. MMR iteratively selects the item which maximizes an objective function f_{obj}, which in turn can deal with the trade-off between accuracy and diversity and is defined as

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

where S is the previously re-ranked list, 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.

The diversification attitude of each user for each item attribute $a \in \mathbf{A}$ is measured through Shannon's entropy. For each attribute, users are clustered in four groups, referred to as *quadrants*, defined by the medians of the entropy and user profile length distributions across all users. 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. Table 1 provides a representation of quadrants. The main modelling hypothesis behind this classification is that 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.

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

Table 1. Quadrants

Quadrants 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 $q_u : \mathbf{A} \rightarrow \{1, 2, 3, 4\}$, which assigns, for each attribute, the quadrant to

	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.

which user u belongs to and then we define a quadrant weight $\omega_i \in [0,1]$, with $i \in \{1,2,3,4\}$. The overall similarity between items i and j in Equation (1), for user u, is tailored to the quadrants she belongs to and is defined as:

$$sim(i,j) = \frac{\sum_{a \in \mathbf{A}} \omega_{q_u(a)} \cdot sim_a(i,j)}{m \cdot |\mathbf{A}|} \tag{2}$$

with $m = \max\{\omega_i \mid i = 1,2,3,4\}$ and $sim_a(i,j)$ 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, eventually varying the diversity.

3 Experiments and Results

We carried out experiments on `Movielens 1M`[4] dataset, enriched with further attribute information (actors and directors) extracted from `DBpedia`[5], as in [5]. We concentrated on users who gave at least fifty ratings. The final dataset contains 4297 users, 3689 items and 942590 ratings. Training and test sets were built with a temporal 60-40% split. We compared our approach with two baselines: *no-MMR*, user-based kNN Collaborative Filtering algorithm with Pearson correlation; *MMR*, re-ranking with Equation 1 of the top 200 recommendations generated by *no-MMR* for each user. Our adaptive approach is denoted as *adaptiveMMR*. The λ parameter in Equation 1 was set to 0.5. As similarity measure for attribute a in (2), we used the Jaccard index. To reduce the number of distinct attribute values, we divided movies in decades and performed a K-means clustering for actors and directors on the basis of their `DBpedia` categories, obtaining 20 clusters. The number of values is 19 and 8 for *genre* and *year*, respectively.

We used the *TestItems* evaluation methodology presented in [2], with Precision (P@k) and nDCG@k for accuracy, ILD@k for diversity and *avg(P,ILD)* for the balance between accuracy and diversity, as in [6]. P@k is chosen instead of nDCG@k since they have a similar trend.

Firstly, we tested the validity of the hypothesis that users who have explored different items in the past are inclined to diversity. As shown in Table 2, *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.

Later, to test the effectiveness of *adaptiveMMR*, we conducted a grid search on ω, finding, as a first result, that our intuition of choosing small values for

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

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. 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.

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: 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 were computed via grid search fixing ω_1 and ω_3 and varying ω_2 and ω_4 with a step of 0.05. These configurations let us deal with emblematic situations: configuration A acts on users who are in quadrant 4 for some attributes and configuration B on users belonging to quadrant 2 or 4. Table 3 shows the results with $k = 10$. *AdaptiveMMR* gains the best balance between accuracy and diversity, represented by *avg(P,ILD)*. In terms of accuracy, *adaptiveMMR* out-performs *no-MMR* and *MMR*, especially *adaptiveMMR-C*. Remarkably, the configuration C has an ILD value close to *MMR* but a significantly better accuracy values. In conclusion, these results suggest that the diversification tendency, represented by entropy, should be considered even for users with a small profile length.

Acknowledgements The authors acknowledge partial support of VINCENTE (PON02_00563_3470993) and RES NOVAE (PON04a2_E)

References

1. G. Adomavicius and Y. Kwon. Improving aggregate recommendation diversity using ranking-based techniques. *IEEE TKDE*, 24(5):896–911, 2012.
2. A. Bellogin, P. Castells, and I. Cantador. Precision-oriented evaluation of recommender systems: An algorithmic comparison. In *ACM RecSys '11*, pages 333–336, 2011.
3. K. Bradley and B. Smyth. Improving Recommendation Diversity. In *Irish Conference in Artificial Intelligence and Cognitive Science*, pages 75–84, 2001.
4. T. Di Noia, V. C. Ostuni, J. Rosati, P. Tomeo, and E. Di Sciascio. An analysis of users' propensity toward diversity in recommendations. In *ACM RecSys '14*, pages 285–288, 2014.
5. 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.
6. 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.
7. S. Vargas and P. Castells. Exploiting the diversity of user preferences for recommendation. In *OAIR '13*, pages 129–136, 2013.
8. M. Zhang. Enhancing diversity in top-n recommendation. In *ACM RecSys '09*, pages 397–400, 2009.
9. M. Zhang and N. Hurley. Avoiding monotony: Improving the diversity of recommendation lists. In *ACM RecSys '08*, pages 123–130, 2008.
10. C. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *WWW '05*, pages 22–32, 2005.