

On the role of time and sessions in diversifying recommendation results (Extended Abstract)*

Vito W. Anelli, Vito Bellini, Tommaso Di Noia,
Wanda La Bruna, Paolo Tomeo, Eugenio Di Sciascio

Polytechnic University of Bari, Via Orabona, 4, 70125 Bari, Italy
{vitowalter.anelli,vito.bellini,tommaso.dinoia,
wanda.labruna,paolo.tomeo,eugenio.disciascio}@poliba.it

Abstract. Intent-aware approaches to diversification have been proposed in the last years to provide the user with a list of recommendations covering different aspects of her behavior. In this paper, we present two diversification methods taking into account temporal aspects of the user profile: in the first one, in order to emphasize the importance of more recent items, we adopt a temporal decay function while in the second one we take into account temporal sessions. The two proposed methods have been implemented as temporal variants of the well-known xQuAD framework and tested against the Netflix 100M dataset.

1 Introduction

Recommender systems are designed to meet the users' needs suggesting relevant items in a personalized fashion. As recommendations are usually presented in form of list or group, the user experience strongly depends on the overall quality of such recommendations and, the diversity among them has been identified as one of the most important quality factor [6]. The diversity issue has been originally addressed in the Information Retrieval field. As user queries are often ambiguous and their intent is not clear, proposing a set of answers covering different intents may increase the probability that users find at least one relevant document [9]. The concept of intent-aware diversification has then been applied to the Recommender Systems field [12] and extensively studied thus producing new algorithms and evaluation metrics. Here, user intents as defined in Information Retrieval have been mapped to user interests with reference to item characteristics. Very often, in the design of the model behind a recommendation engine, the user profile is considered as a static snapshot without taking into proper account its temporal dimension. Actually, the importance of analyzing temporal aspect for user modeling has been proved to affect the final recommendation results [3].

In this paper we investigate the effect on the trade-off between accuracy and diversity of a recommendation list when dealing with temporal aspects of the

*An extended version of this paper has been published in [1]

user profile. The intuition behind our idea is that temporal dynamics might allow to better understand the user interests with respect to the items characteristics and then provide a more accurate intent-aware diversification. We present two intent-based modeling methods that exploit the time dimension in a user profile. The first one analyses the frequency of interaction between the users and the items features using a temporal decay function in order to emphasize *persistence* and *recency* of an intent. The other method is based of a new session analysis technique of user ratings for intent modeling. Considering that a session is usually defined as a set of consecutive ratings with a very small gap of time among them, we provided a wide definition of session tailored for movie ratings. The experimental results on the Netflix dataset demonstrated that the analysis of temporal aspects in the user profile leads to better accuracy-diversity balance and intent-aware diversity compared to the original **xQuAD**. As an additional benefit, the aggregate diversity results improved too thus demonstrating to produce more personalized recommendations.

2 Intent-aware diversification for recommendations with temporal dynamics

Finding the most diverse results in a recommendation problem is a NP-hard problem and hence several heuristics have been proposed [5]. If we want to provide the top- N best items \mathbf{S} to the user u , the system originally computes a list of top- K elements \mathbf{R} (with $K > N$) with their associated ratings $r^*(u, i)$ and then re-ranks the list based on a given objective function. Most previous diversification approaches are based on a greedy selection strategy which selects the next most relevant item only if that item is diverse with respect to the items already selected. For the purpose of this work, we consider **xQuAD**, one of the most well-known intent-aware greedy heuristics. It maximizes the coverage of the inferred interests while minimizing their redundancy. **xQuAD** was originally proposed for search diversification in information retrieval by Santos et al. [9] and, More recently, it has been adapted for recommendation diversification [11], replacing query and relative aspects with user and items features, respectively. The expression of the **xQuAD** objective function is

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

with $\bar{\mathbf{S}}$ representing the set of the items belonging to \mathbf{R} not already in \mathbf{S} and $div(i, \bar{\mathbf{S}}, u)$ defined as

$$div(i, \bar{\mathbf{S}}, u) = \sum_f p(i|f, u) \cdot p(f|u) \cdot \prod_{j \in \bar{\mathbf{S}}} (1 - p(j|f, u)) \quad (2)$$

In (2) $p(i|f, u)$ 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 usually 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. In particular, this work focuses on the intent modeling in the **xQuAD** framework, namely the aforementioned $p(f|u)$ component in the Equation (2). We now propose two methods to exploit temporal analysis for intent modeling in diversification that we call **session-based** and **time-based intent modeling**. Both relies on the intuition that user intent can change during the interaction with the system and evaluating the importance of a feature merely computing its frequency in the user profile may not represent the current user interests.

Time-Based Intent Modeling. In order to valorize *persistence* and *recency* of an intent, we propose to analyze the frequency of interaction between the user u and the feature f and to weight each interaction by a temporal decay function. More formally, the following formula computes the interest of the user u with respect to the feature f :

$$p(f|u) = \frac{\sum_{i \in R(u)} cov(f, i) disc(u, i)}{\sum_{i \in R(u)} disc(u, i)} \quad (3)$$

where $R(u)$ indicates the set of rating provided by the user u ; $cov(f, i)$ is a binary function returning 1 if the item i is associated with the feature f , 0 otherwise; $disc(i, u)$ is a temporal decay function returning lower values for older ratings, and higher values for the most recent ones. Inspired by [4], as decay function we adopted the following exponential function

$$disc(u, i) = e^{-\beta \cdot |t_{u, last} - t_{u, i}|} \quad (4)$$

where $t_{u, last}$ indicates the timestamp of the last rating of the user u and $t_{u, i}$ the timestamp when user u rated i ; $\beta > 0$ controls the decay rate.

Session-Based Intent Modeling. In our setting, in order to identify user sessions we propose an EM clustering used to train two univariate Gaussian Mixture Models (with equivariance and variable variance). The number of clusters has been evaluated based on the Bayesian Information Criterion considering the fitted models with a number of clusters from 1 up to 300. In order to remove outliers from each session s , for each computed cluster we do not consider ratings falling outside the interval $[\mu_s - \sigma_s, \mu_s + \sigma_s]$, with μ_s and σ_s being respectively the mean and the standard deviation of ratings distribution for the session s . Once user sessions are determined, they can be used to analyze the user activities taking into account the temporal dynamics. In this work we present an approach to model the users intents over time, by considering three key properties: *importance*, *persistence* and *recency* of an intent among the user sessions. The following formula computes the interest of the user u with respect to the feature f :

$$p(f|u) = \frac{\sum_{s \in S(u)} \frac{\sum_{i \in I(s)} cov(f, i)}{|F(s)|} disc(s, u)}{\sum_{s \in S(u)} disc(s, u)} \quad (5)$$

Algorithm	Ndcg@10	ERR IA@10	Coverage@10	EPC@10
BPRMF+XQuAD	0.029264	0.01789	0.27540	0.02549
BPRMF_SB_XQuAD	0.03340	0.01510	0.30417	0.02737
BPRMF_TB_XQuAD	0.03433	0.01724	0.29820	0.02843
BPRSLIM+XQuAD	0.03072	0.01870	0.35799	0.02686
BPRSLIM_SB_XQuAD	0.03943	0.01736	0.40021	0.03240
BPRSLIM_TB_XQuAD	0.04026	0.01942	0.39183	0.03339

Table 1. Comparative results in terms of accuracy, individual diversity and aggregate diversity with $\lambda = 0.8$

where $S(u)$ indicates sessions computed for the user u ; $I(s)$ is the set of items in s ; $cov(f, i)$ is a binary function returning 1 if the item i is associated with the feature f , 0 otherwise; $F(s)$ represents the set of features associated with all the items in s ; $disc(s)$ is the temporal decay function adapted to handle the sessions instead of the items, considering a session as an item in Equation (4) where the session date is that of the last rated item in such session. As for the previous case, β value was set to $1/200$.

3 Experimental setting

In order to verify our research questions and evaluate our proposal, we used the popular movie datasets derived from the Netflix Prize Context [2] and we tested our approaches via a re-ranking of the BPRMF [8] and BPRSLIM [7] algorithms resulting recommendations. In our evaluation, the time-based and session-based intent modelings proposed in Section 2 are used as alternatives to the pure frequency based intent modeling in the original xQuAD. These two variations of xQuAD, are denoted as: TB_xQuAD, SB_xQuAD, where TB stands for time-based and SB for session-based.

In order to evaluate *accuracy*, we measured nDCG. As for *individual diversity*, namely the degree of dissimilarity among all items in the list provided to a user, was measured by ERR-IA as it has been shown [10] to be the metric targeted by xQuAD, while for *aggregate diversity* we computed the Catalog Coverage (percentage of items recommended at least to one user). An evaluation on the *novelty* of computed results has been done through EPC (Expected Popularity Complement) [10]. Our experiments show that the time-based variant of xQuAD performs better than the session-based one but for Catalog Coverage where we have better results for the session-based implementation of xQuAD. It is worth noticing that the base version of xQuAD outperforms its time-based variants up to a certain value of λ for both BPRMF and BPRSLIM. For the former this value lies between 0.8 and 0.9 while for the latter between 0.7 and 0.8. Hence, in case we are interested in higher values of diversity, time may play an important role. In Table 1 we see that with $\lambda = 0.8$ we obtain the best result in terms of trade-off among the various metrics we measured in our experiments.

References

1. V. W. Anelli, V. Bellini, T. Di Noia, W. La Bruna, P. Tomeo, and E. Di Sciascio. An analysis on time- and session-aware diversification in recommender systems. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, UMAP '17, pages 270–274. ACM, 2017.
2. J. Bennett, S. Lanning, and N. Netflix. The netflix prize. In *In KDD Cup and Workshop in conjunction with KDD*, 2007.
3. N. Koenigstein, G. Dror, and Y. Koren. Yahoo! music recommendations: Modeling music ratings with temporal dynamics and item taxonomy. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, pages 165–172, New York, NY, USA, 2011. ACM.
4. Y. Koren. Collaborative filtering with temporal dynamics. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 447–456, New York, NY, USA, 2009. ACM.
5. O. Küçüktunç, E. Saule, K. Kaya, and U. V. Çatalyürek. Diversifying citation recommendations. *ACM Transactions on Intelligent Systems and Technology*, 5(4):55:1–55:21, Dec. 2014.
6. S. M. McNee, J. Riedl, and J. 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.
7. X. Ning and G. Karypis. Slim: Sparse linear methods for top-n recommender systems. In *Data Mining (ICDM), 2011 IEEE 11th International Conference on*, pages 497–506. IEEE, 2011.
8. S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*, pages 452–461. AUAI Press, 2009.
9. R. Santos, C. Macdonald, and I. Ounis. Exploiting query reformulations for web search result diversification. In *WWW '10*, pages 881–890, 2010.
10. S. Vargas and P. Castells. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, pages 109–116, 2011.
11. S. Vargas and P. Castells. Exploiting the diversity of user preferences for recommendation. In *OAIR '13*, pages 129–136, 2013.
12. S. Vargas, P. Castells, and D. Vallet. Intent-oriented diversity in recommender systems. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 1211–1212, New York, NY, USA, 2011. ACM.