Intent-Aware Diversification using Item-Based SubProfiles

Mesut Kaya
Insight Centre for Data Analytics
University College Cork, Ireland
mesut.kaya@insight-centre.org

Derek Bridge
Insight Centre for Data Analytics
University College Cork, Ireland
derek.bridge@insight-centre.org

ABSTRACT

In many approaches to recommendation diversification, a recommender scores items for relevance and then re-ranks them to balance relevance with diversity. In intent-aware diversification, diversity is formulated in terms of coverage of aspects, where aspects are either explicit such as movie genres or implicit such as the latent factors found during matrix factorization. Typically, the same set of aspects is used across all users. In this paper, we propose a form of personalized intent-aware diversification, which we call SPAD (SubProfile-Aware Diversification). The aspects we use in SPAD are subprofiles of the user's profile. They are not defined in terms of explicit or implicit features. We compare SPAD to other forms of intent-aware diversification. We present empirical results in support of SPAD.

KEYWORDS

Diversity; intent-aware; subprofiles.

ACM Reference format:
Mesut Kaya and Derek Bridge. 2017. Intent-Aware Diversification using Item-Based SubProfiles. In *Proceedings of RecSys 2017 Poster Proceedings, Como, Italy, August 27–31, 2017,* 2 pages.

1 INTRODUCTION

It has long been recognized that it is not enough for recommendations to be accurate or relevant. In many domains, recommendations must be novel to the user or serendipitous, and a set of recommendations must be diverse. Diversity is one response to uncertainty. A recommender cannot be certain of a user's short-term or longer-term interests, both because some user profiles are small and some, while they may not be so small, will contain preferences over different kinds of items. In the face of uncertainty, a diverse set of recommendations is more likely to contain one or more items that will satisfy the user.

In many approaches to recommendation diversification, a recommender scores items for relevance and then re-ranks them to balance relevance with diversity. In intent-aware diversification [3], the idea is that the re-ranked recommendations should cover the different tastes or interests revealed by the user's profile. The most common way to characterize a user's tastes is as a probability distribution over so-called aspects of the items in the user's profile. Aspects are usually either explicit features such as movie genres or implicit features such as the latent factors found during matrix factorization. Hence, typically, the same set of aspects is used across all users — only the probablilities vary across users.

Item aspects, such as genres, do not necessarily fully represent a user's tastes or interests and are not available in every recommendation domain. Hence, in this work, we propose a new intent-aware diversification framework based on user subprofiles, rather than item features. A subprofile is simply a subset of the items in a user's profile, each such subprofile representing one of the user's distinct tastes. We detect a user's subprofile by adapting DAMIB-COVER, a method designed for top-n recommendation to shared accounts [5]. Unlike the aspects used in earlier work, which are global across the set of users, subprofiles differ from user to user, making for a more personalized form of diversification. We refer to our new framework as SubProfile-Aware Diversification (SPAD).

2 RECOMMENDATION DIVERSITY

The dominant approach to diversification is greedy re-ranking, in which sets of recommendations RS for a user u are re-ranked by considering the marginal contribution that would be made by adding an item i to the result set RL. The marginal contribution is measured by an objective function $f_{obj}(i, RL)$ which is typically a linear combination of the item's relevance score $s(u, i)$ and the marginal contribution item i makes to the diversity of RL, $\mathrm{div}(i, RL)$, the trade-off between the two being controlled by a parameter λ ($0 \leq \lambda \leq 1$):

$$f_{obj}(i, RL) = (1 - \lambda)s(u, i) + \lambda\, \mathrm{div}(i, RL) \quad (1)$$

In early work, the diversity $\mathrm{div}(i, RL)$ is computed as the average (or sum) of the all-pairs intra-list distances (ILD) of the items in RL. The assumption in this early work is that a set of items that are dissimilar to each other is more likely to contain one or more items that satisfy the user's current needs or interests, but there is nothing in the operation of the system to explicitly ensure this. More recent approaches, going under the name *intent-aware diversification*, seek to select items that explicitly address different user interests.

Intent-aware diversification methods assume a set of aspects $\mathcal{A}$ which describe the items and for which user interests can be estimated. The aspects might be explicit: like genres such as comedy in a movie recommender. Alternatively, aspects might be implicit, e.g. corresponding to the latent factors found by a matrix factorization recommender system.

User u's interests can be formulated as a probability distribution $p(a|u)$ for aspects $a \in \mathcal{A}$. The probability of choosing an item i from the set of recommendations RS given an aspect a of user u is denoted by $p(i|u, a)$. In the Query Aspect Diversification framework (xQuAD) [2, 4] , diversification can be achieved by re-ranking a conventional recommender's recommendation set as Equation (1) but with $\mathrm{div}(i, RL) = \mathrm{nov}_{\mathrm{xQuAD}}(i, RL)$ defined as:

$$\mathrm{nov}_{\mathrm{xQuAD}}(i, RL) = \sum_{a \in A} p(a|u)p(i|u, a) \prod_{j \in RL} (1 - p(j|u, a)) \quad (2)$$

What characterizes the work on intent-aware diversification in recommender systems that we have described so far is the use of a global set of aspects. In our work, we infer the aspects from the user's profile, making them personalized: the aspects for one user need not be the same for another.

3 SUBPROFILE AWARE DIVERSITY

In this section, we explain our new approach to diversification in recommender systems, which we call SubProfile Aware Diversification (SPAD). It is a greedy re-ranking approach; it is intent-aware; but it is also personalized, based on identifiable subprofiles within the user's profile.

Let I be the set of all items. Subprofile detection works on positively-rated items in the user's profile. In the case of positive-only feedback, user u's profile, $I_u \subseteq I$, is the set of items she has interacted with (liked, clicked on, purchased, etc.). In the case of explicit ratings r_{ui} (e.g. 1-5 stars), then I_u must be defined in terms of items the user liked, which will usually involve thresholding the ratings, e.g. in our experiments, we use $I_u = \{i | r_{ui} \geq 4\}$. A user's subprofiles are subsets of I_u.

Our approach to detecting user subprofiles is based on a method for recommending to shared accounts, called DAMIB-COVER [5]. DAMIB-COVER identifies different tastes within the profile of a shared account (which it assumes come from the different users who share that account) and recommends items to satisfy each taste. We adapt DAMIB-COVER to take in the profile for a single-user account u and to extract the different subprofiles $\mathcal{S}_u$ that correspond to the different tastes of that user.

In the work on intent-aware diversification that we described earlier, the same set of aspects $\mathcal{A}$ was used for all users. In SPAD, aspects are user-specific: user u has set of aspects $\mathcal{A}_u$. And, in the earlier work, aspects were often based on explicit features $\mathcal{F}$, i.e. $\mathcal{A} = \mathcal{F}$. In SPAD, aspects are user subprofiles, i.e. $\mathcal{A}_u = \mathcal{S}_u$. Each subprofile $S \in \mathcal{S}_u$ contains a set of items from I_u. Different subprofiles can be of different lengths; the number of subprofiles can differ across users.

In SPAD, the set RS is greedily re-ranked using the objective function given as Equation (1) with $\text{div}(i, RL) = \text{nov}_{\text{xQuAD}}(i, RL)$ (Equation (2)). What differs is the computation of the probabilities used in Equation (2). Given that aspects are now subprofiles, we use $p(S|u)$ and $p(i|u, S)$ instead of $p(a|u)$ and $p(i|u, a)$ for $S \in \mathcal{S}_u$.

4 EXPERIMENTS

We compare SPAD to other re-ranking approaches on the MovieLens1M dataset with 5-fold cross validation. We show the results of taking recommendations made by matrix factorization (MF) and probabilistic latent semantic analysis (pLSA) algorithms and then re-ranking them using SPAD and other re-ranking approaches:

- MMR: Uses ILD with distance defined as the complement of Jaccard similarity on the item features [1].
- xQuAD: See Equation 2.
- RxQuAD: Relevance-based xQuAD that is based on maximizing relevance, rather than the probaility of choosing a single item [4].
- cplsa: Based on explicit aspects but the probabilities are learned by a constrained pLSA model [6].

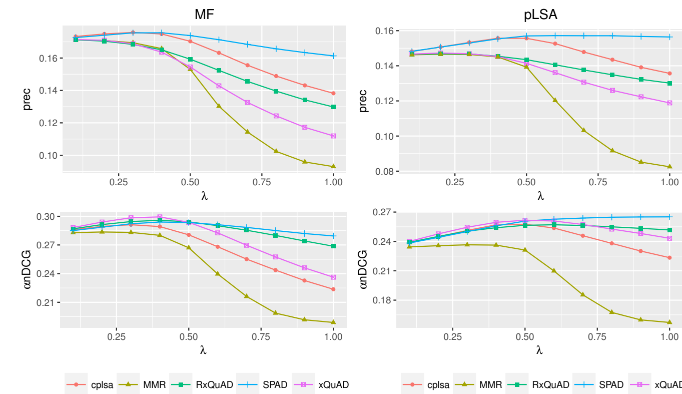

Figure 1: Results for MovieLens dataset.

In Figure 1, we plot precision (an accuracy metric) and α-nDCG (a diversity metric) for different values of λ, which controls the amount of diversification. Notice that α-nDCG measures diversity with respect to the explicit features $\mathcal{F}$ (the meta-data). It therefore may favour recommenders that re-rank using those features. Our new method, SPAD, makes no use of the features and so it is at a disadvantage in these experiments.

For both baseline algorithms (MF and pLSA), SPAD has the highest precision. For the pLSA baseline, SPAD also has the highest diversity; for the MF baseline, SPAD's diversity is competitive with the other re-ranking algorithms despite being at a disadvantage as mentioned earlier.

We plan to further explore the effectiveness of SPAD on other datasets, and with more baseline algorithms. We also plan to develop other subprofile detection methods instead of using DAMIB-COVER. We will also explore the interpretability of SPAD's recommendations in terms of subprofiles.

ACKNOWLEDGMENTS

This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289.

REFERENCES

[1] Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Procs. of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, 335–336.
[2] Rodrygo LT Santos, Craig Macdonald, and Iadh Ounis. 2010. Exploiting query reformulations for web search result diversification. In *Procs. of the 19th International Conference on World Wide Web.* 881–890.
[3] Saul Vargas, Pablo Castells, and David Vallet. 2011. Intent-oriented diversity in recommender systems. In *Procs. of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, 1211–1212.
[4] Saúl Vargas Sandoval. 2015. *Novelty and Diversity Evaluation and Enhancement in Recommender Systems.* Ph.D. Dissertation. Universidad Autónoma de Madrid, Spain.
[5] Koen Verstrepen and Bart Goethals. 2015. Top-N Recommendation for Shared Accounts. In *Procs. of the 9th ACM Conference on Recommender Systems.* ACM, 59–66.
[6] Jacek Wasilewski and Neil Hurley. 2016. Intent-Aware Diversification Using a Constrained PLSA. In *Procs. of the 10th ACM Conference on Recommender Systems.* ACM, 39–42.