Intent-Oriented Diversity in Recommender Systems

Saúl Vargas, Pablo Castells and David Vallet
Universidad Autónoma de Madrid
Escuela Politécnica Superior, Departamento de Ingeniería Informática

{saul.vargas,pablo.castells,david.vallet}@uam.es

ABSTRACT

Diversity as a relevant dimension of retrieval quality is receiving increasing attention in the Information Retrieval and Recommender Systems (RS) fields. The problem has nonetheless been approached under different views and formulations in IR and RS respectively, giving rise to different models, methodologies, and metrics, with little convergence between both fields. In this poster we explore the adaptation of diversity metrics, techniques, and principles from ad-hoc IR to the recommendation task, by introducing the notion of *user profile aspect* as an analogue of query intent. As a particular approach, user aspects are automatically extracted from latent item features. Empirical results support the proposed approach and provide further insights.

Categories and Subject Descriptors: H.3.3 [Information Search and Retrieval] *Information filtering, Retrieval models*

General Terms: Algorithms, Measurement, Performance, Experimentation, Theory

Keywords: Diversity, recommender systems, user profiles, diversity metrics, query intent, profile aspects

1. INTRODUCTION

Search result diversity is being actively researched in the IR field as a means to address query ambiguity and underspecification in ad-hoc IR [1],[2],[3]. Recommendation diversity is also an active research topic in the Recommender Systems (RS) area [5],[6]. In general terms, and most particularly in common practical scenarios, recommendation can be seen as an IR task. Interestingly, the diversity issue has been stated and addressed quite differently in the research on the topic so far in RS and ad-hoc IR respectively. In particular, diversity has been studied under a quite specific motivation and precise problem definition(s) in the IR community –building around the problem of uncertainty in user queries– along with formally grounded and well understood diversity metrics, with a theoretical depth and a drive towards standardization (backed by a specific TREC diversity task) which are not presently found or equally emphasized in the RS literature on the topic. It seems therefore natural to wonder whether, as far as it were possible to draw models and principles from one area to the other, research on RS diversity might benefit from the insights and ongoing progress in search diversity –and vice-versa.

In this poster we explore the adaptation of diversity models, metrics, and methods from ad-hoc IR into a RS setting. Specifically, we propose the notion of *user profile aspect* as an analogue of query intent, upon which we adapt the IR diversity techniques and methodology to a recommendation task. We consider two scenarios that differ in the available information for the construction of user aspects in the diversification method. We propose an approach for the extraction of user aspects based on latent factors when the only available information relates to the interaction between users and items.

2. RECOMMENDATION DIVERSITY VS. SEARCH DIVERSITY

Diversity in RS is generally motivated as a means to reduce redundancy under the assumption that recommending too similar items is less profitable for the user –and the vendor– than offering a more varied experience. The rationale for diversity is often stated in association with the notion of novelty and surprisal, upon the understanding that recommendation value is to a significant extent related to discovery in the user experience. Looking back for a connection to diversity in ad-hoc IR, one finds that the issues of ambiguity and underspecification are generally absent from the problem statement in the RS literature. This may seem natural as far as there is no query in the recommendation task to begin with. However, there is certainly a user information need, expressed in the form of a user profile (ratings or item access records). This implicit information need expression arguably involves far more ambiguity and incompleteness than an explicit user query, whereby the uncertainty-oriented motivation would certainly hold for RS diversity. So does the principle of diversification as a means to minimize the risk of underperformance extremes, which is also common in the IR literature [1].

Query ambiguity and underspecification are modeled in terms of query interpretations, categories, aspects, nuggets, subtopics, and similar elements in ad-hoc IR. An analogy can be drawn in the RS setting by considering an equivalent notion of *user profile aspect*. This is in fact a natural idea, since a single user's interests have many different sides and subareas (e.g. professional, politics, movies, travel, etc.). Different user preference aspects can be relevant or totally irrelevant at different times therefore, similar to query intent, there is uncertainty at recommendation time about what area of user interest should play in the given context.

If one is able to give a consistent approximation to user interest aspects in the context of a RS, the theories and metrics in search diversity could be adapted to the recommendation task. This would bring benefits such as a) a new perspective and rationale for diversity in RS in terms of theory and models, and b) new diversity metrics for RS, such as the intent-aware metrics [1] or α-nDCG [3]. Additionally, such metrics would bring in several important properties currently lacking in RS diversity studies: a) the introduction of metrics that take into account the order of items when measuring the overall recommendation diversity (i.e. top positions are more important); b) the consideration of diversity only in the presence of relevance; c) related to this, the assessment of accuracy and diversity altogether by a single metric; and d) a step towards a shared consensus on common metrics and methodologies.

3. ASPECT-BASED RECOMMENDATION DIVERSIFICATION AND METRICS

In the proposed approach, we take a space of item features (or attributes) as the basis to extract the user profile aspects, taking user profiles as the equivalent of search queries. The space of item features fulfils here a similar role to the notion of categories in [1]. Based on this, we apply a greedy diversification algorithm, where a baseline ranking R is diversified into a re-ranked list S by itera-

tively picking the item $i \in R - S$ which maximizes an objective function. We adapt two objective functions from search diversity, based on IA-Select [1] and MMR [2] respectively. In the IA-Select scheme, we define the objective function as:

$$\sum_{f \in \mathcal{F}} p(f|u)\hat{r}_{\mathrm{norm}}(u,i)p(f|i)\prod_{j \in S}\big(1 - p(f|j)\hat{r}_{\mathrm{norm}}(u,j)\big)$$

where u is the target user and $\mathcal{F}$ is the feature space. The item quality component represented by $\hat{r}_{\mathrm{norm}}(u,i)$ is a measure of user preference for items, given by the baseline RS being diversified, namely the scoring function returned by the system –typically a rating prediction– normalized to [0,1]. We estimate the distribution of features over user profiles and over items as:

$$p(f|u) \sim \frac{|\{i \in \mathbf{u}|f \in \mathbf{i}\}|}{\sum_{f' \in \mathcal{F}} |\{i \in \mathbf{u}|f' \in \mathbf{i}\}|} \qquad p(f|i) \sim \frac{\chi_{\mathbf{i}}(f)}{|\mathbf{i}|}$$

where $\mathbf{u}$ denotes the set of items in the profile of u, $\mathbf{i}$ denotes the set of feature values of item i, $\chi_{\mathbf{i}}(f) = 1$ if $f \in \mathbf{i}$, and 0 otherwise. In the MMR scheme, we define an objective function which combines a similarity and a diversity component, as follows:

$$(1 - \lambda)\,\hat{r}_{\mathrm{norm}}(u,i) + \lambda \underset{j \in S}{avg}\big(1 - sim(i,j)\big)$$

The similarity component is, again, the baseline retrieval function. The diversity component is defined by the complement of a similarity function in terms of the shared features between two items. In our experiments we choose to compute similarity as the cosine of the respective feature vectors.

As mentioned earlier, we consider two scenarios in our approach. One in which the item feature data is explicit and known to the diversification method (e.g. movie genre, director, etc., in movie recommendation), and one in which it is not, and only the user-item preference data (e.g. item ratings by users) is available. In the former case, the feature distributions estimation, and the similarity function in MMR, are based on the known feature information (using binary feature vectors in MMR). In the latter case, we base the diversification on latent item features extracted by a matrix factorization approach [4]. For IA-Select, the feature weight is binarized as "present" / "not present" upon a threshold, and for MMR we keep the weights in the feature vectors. The factorization approach thus works as a form of automatic user aspect extraction.

To evaluate the quality of diversified recommendations we adapt measures such as the intent-aware metrics [1] and α-nDCG [3], where explicit item features are the analogue of categories (or subtopics), and user profiles play the part of queries. That is, for instance, given a user u, the intent-aware nDCG of the recommendation to u is defined as $\mathrm{nDCG\text{-}IA} = \sum_{f \in \mathcal{F}} p(f|u)\mathrm{nDCG}(u|f)$ where, analogously to [1], $\mathrm{nDCG}(u|f)$ counts as relevant items only the ones that are relevant for u and have the feature f. In the evaluation of these metrics, the features $\mathcal{F}$ are always the explicit ones, regardless of what features – explicit or implicit– were used in the diversification method.

4. EXPERIMENTS

We have tested the behavior of the proposed approach on the Movie-Lens 100K dataset. We take as baselines two state of the art colaborative filtering algorithms: a common user-based nearest-neighbor (kNN) recommender, and a matrix factorization (MF) based algorithm [4]. As a simple feature space $\mathcal{F}$ for user interest aspects we take the set of movie genres. We test the two scenarios mentioned in the previous section, one in which the diversifier uses the known item genre data, and one in which it extracts latent user features as the space of user aspects in the diversification algorithm, using rating information only. The diversifiers re-rank the top 500 items returned by the baseline recommender for each user. We take the 80% training, 20% test data splits provided by the MovieLens distribution, with

5-fold cross-validation. For relevant judgments, we take as relevant (for each user) the items with a rating higher than 3 in the test set.

Table 1 shows the performance of the different configurations using three well known diversity metrics from ad-hoc IR, plus intra-list diversity (ILD) –based on the Jaccard similarity on genres–, a common metric used in RS diversity [6]. It can be seen that the proposed diversification methods work properly, consistently improving the non-diversified baselines (bottom row). The IA-Select approach performs overall significantly better than the MMR scheme on the three IR metrics. We believe this is because it builds upon a common formalization of diversity as do the metrics (after [1]). Somewhat surprisingly, diversification with latent features performs better than with explicit ones for IA-Select on kNN, and MMR on both. We attribute this to the fact that latent features provide a more dense representation of items, and also more significant in terms of explaining the differences in interests between users, and the similarity between items. On the ILD metric, MMR and IA-Select perform similarly, and explicit features work clearly better than latent. This is probably because ILD ignores relevance –with respect to which IA-Select and latent features seem to do better.

Table 1. Four diversity metrics (α = 0.5 in α-nDCG) on different diversification approaches: MMR (with λ=0.5) and IA-Select, combined with explicit and latent features, on two baseline RS, based on kNN and MF respectively. The best value of each column is underlined. All differences to baseline are statistically significant (p < 0.005, Wilcoxon), except values in parenthesis.

		α-nDCG@50		ERR-IA@50		nDCG-IA@50		ILD@50	
		kNN	MF	kNN	MF	kNN	MF	kNN	MF
IA-Sel	Explicit	0.1589	<u>0.1838</u>	0.0409	0.0516	0.0604	<u>0.0755</u>	<u>0.8659</u>	0.8734
IA-Sel	Latent	<u>0.1596</u>	0.1597	<u>0.0465</u>	0.0458	<u>0.0618</u>	0.0637	0.7951	0.7817
MMR	Explicit	0.1334	0.1652	(0.0367)	(0.0431)	0.0461	(0.0555)	0.8601	<u>0.8761</u>
MMR	Latent	0.1320	0.1742	0.0373	<u>0.0528</u>	0.0492	0.0705	0.7906	0.7740
Baseline RS		0.1213	0.1451	0.0352	0.0425	0.0440	0.0561	0.7787	0.7655

We have carried out additional experiments with further configurations, using different baseline recommender systems, and different metric cutoffs, the results from which also confirm our findings. The results were similarly positive with movie director as the explicit feature space. We plan to further explore the relation between the feature space and the effectiveness of diversification, under the intuition that the effectiveness should benefit from a higher dependency between features and user interests.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Agrawal, A., Gollapudi, S., Halverson, A., and Ieong, S. Diversifying search results. *WSDM 2009*. Barcelona, Spain, 2009, 5-14.

[2] Carbonell, J. G., and Goldstein, J. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. *SIGIR 1998*. Melbourne, Australia, 1998, 335-336.

[3] Clarke, C. L. A. et al. Novelty and diversity in information retrieval evaluation. *SIGIR 2008*. Singapore, July 2008, 659-666.

[4] Koren, Y., Bell, R. M, Volinsky, C. Matrix Factorization Techniques for Recommender Systems. *IEEE Computer* 42(8):30-37, 2009.

[5] Zhang, M. and Hurley, N. Avoiding Monotony: Improving the Diversity of Recommendation Lists. *RecSys 2008*. Lausanne, Switzerland, October 2008, 123-130.

[6] Ziegler, C-N., McNee, S. M., Konstan, J. A., Lausen, G. Improving recommendation lists through topic diversification. *WWW 2005*. Chiba, Japan, May 2005, 22-32.