

Intent-Aware Diversification Using a Constrained PLSA

Jacek Wasilewski
Insight Centre for Data Analytics
University College Dublin
Belfield, Dublin, Ireland
jacek.wasilewski@insight-centre.org

Neil Hurley
Insight Centre for Data Analytics
University College Dublin
Belfield, Dublin, Ireland
neil.hurley@insight-centre.org

ABSTRACT

The intent-aware diversification framework was introduced initially in information retrieval and adopted to the context of recommender systems in the work of Vargas et al. The framework considers a set of aspects associated with items to be recommended. For instance, aspects may correspond to genres in movie recommendations. The framework depends on input *aspect model* consisting of item selection or relevance probabilities, given an aspect, and user intents, in the form of probabilities that the user is interested in each aspect. In this paper, we examine a number of input aspect models and evaluate the impact that different models have on the framework. In particular, we propose a constrained PLSA model that allows for interpretable output, in terms of known aspects, while achieving greater performance than the explicit co-occurrence counting method used in previous work. We evaluate the proposed models using a well-known MovieLens dataset for which item genres are available.

Keywords

intent-aware recommendation; diversity; evaluation

1. INTRODUCTION

User interactions as implicit or explicit evidence of user needs involve a great deal of uncertainty. Those interactions however do not represent all user interests and might lead to creation of filter bubbles. Predicting the user needs is an inherently difficult task as user interests are complex, dynamic, context-dependent or even contradictory. Diversity-oriented approaches have been used to address these issues, by widening the range of possible item types recommended to the user. Moreover, diverse recommendations enrich the user experience over time, helping to expand the user's horizon. From this viewpoint, diversity can be treated as a strategy to favour practical accuracy in matching true user needs in an uncertain environment.

Different frameworks have been developed to enhance and assess diversity. The most commonly used is a distance-

based intra-list diversity (ILD) where the average pairwise distance of the recommended items describes the recommendation diversity. Another notion of diversity, called the intent-aware diversity framework, has been adopted from the information retrieval (IR) field in [10]. The intent-aware framework in IR deals with ambiguous queries by attempting to ensure that all intents of the query are represented in the results. Adapted to a recommender system context, the framework attempts to ensure that all relevant interests of a user are represented in the recommendation list.

The intent-aware diversity framework assumes the existence of aspects through which user interests can be expressed and by which items are described. For instance, they may correspond to genres in the context of movie recommendation. A simple co-occurrence counting method has been applied to transform this explicit information into the aspect probabilities required by the intent-aware framework. The aspect probabilities are derived without direct optimisation of their utility. Latent aspects can be learned from the user-item interaction data, so that predictive performance of the aspect probabilities is maximised. The learned probabilities should better represent an underlying structure in the dataset. However, diversification over this hidden structure is not easily discernible by the user.

In this paper, we propose a constrained PLSA model that uses explicit aspects, but learns the aspect probabilities to directly optimise their predictive performance. We claim that, in doing so, we put the intent-aware framework on a stronger footing, since the aspect probabilities that drive the method, better model true user preference behaviour. Yet, we maintain the interpretability of the co-occurrence counting method, so that the framework diversifies across known, explicit aspects.

The rest of the paper is structured as follows. We briefly introduce and discuss the intent-aware diversification framework. In section 3 we discuss different strategies for estimating the probability components used by the framework and we introduce the constrained probabilistic aspect model. We empirically test and discuss the effectiveness of our intent model in section 4. We conclude with a summary in section 5.

2. INTENT-AWARE DIVERSIFICATION FRAMEWORK

The intent-aware diversification framework, proposed in the information retrieval field by Santos et al., [9], introduces the concept of explicit aspects in order to mitigate the uncertainty in queries by diversifying the ranked results. In IR,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

RecSys '16, September 15-19, 2016, Boston, MA, USA

© 2016 ACM. ISBN 978-1-4503-4035-9/16/09...\$15.00

DOI: <http://dx.doi.org/10.1145/2959100.2959177>

uncertainty arises due to ambiguous search queries that can be interpreted in different ways, depending on the context. If multiple aspects are related to a search query, without explicit information about the context of the query’s intent, a search engine tries to satisfy the query by serving as many different relevant aspects as possible, hoping that at least one will satisfy the query.

To apply the model into the recommender systems setting, Vargas et al. [10] introduced the notion of user aspect as an analogue of query intent. In recommender systems, users do not define their queries explicitly. Rather a system has to decide what a user would like to see based on previously recorded interactions of the user and the system. The system could assume that a user might be interested in any aspect that the interest in has been expressed before. So there is a similar ambiguity to the IR problem – for any given interaction, what aspect is the user currently interested in? If an incorrect assumption about the user’s intent is made, a recommender system may fail in generating recommendations that users find relevant.

Vargas’ formulation of the explicit Query Aspect Diversification framework (xQuAD) for recommender systems assumes a predefined set of $k > 0$ aspects, $\mathcal{A} = \{a_1, a_2, \dots, a_k\}$. The starting point for xQuAD is the computation of user’s interests in terms of a probability distribution $p(a|u)$, such that $\sum_a p(a|u) = 1$ and the computation of the probability of choosing an item, given a user and an aspect, $p(i|u, a)$. The xQuAD re-ranker uses these probabilities to re-rank a given recommendation list so that aspects are taken into account.

Starting with $\mathcal{S} = \emptyset$, and given the score $s(u, i)$ for item i produced by the baseline recommender system from which the original ranked list \mathcal{R}_u is obtained, xQuAD iteratively constructs the re-ranked list, \mathcal{S} , by greedily selecting at each iteration the item i that satisfies:

$$i^* = \arg \max_{i \in \mathcal{R}_u \setminus \mathcal{S}} (1 - \lambda)s(u, i) + \lambda \sum_{a \in \mathcal{A}} p(a|u)p(i|u, a) \prod_{j \in \mathcal{S}} (1 - p(j|u, a)),$$

and updating $\mathcal{S} \leftarrow \mathcal{S} \cup \{i^*\}$. The two terms in this expression can be summarised as, respectively, the item quality component and the item diversity component, and are mixed together using a tradeoff controlling parameter $0 \leq \lambda \leq 1$.

3. COMPONENTS ESTIMATION

The xQuAD formulation presented in the previous section depends on an *aspect model*, determined by the user and item aspect probabilities, given by:

$$p(i|u) = \sum_a p(a|u)p(i|u, a). \quad (1)$$

But, what exactly is an aspect? We consider two cases:

1. **Explicit Aspects:** Aspect information is directly available in the input data. Aspects are synonymous with item features, such as genres in the case of movies. In particular, from the input data it is possible to directly identify for each item i a set of aspects \mathcal{A}_i .
2. **Latent Aspects:** Aspect information is not explicitly available in the input data. Instead aspects are latent categories, which are learned from the user-item interaction data.

Diversification research most commonly considers explicit categories and these have the important advantage of being *interpretable* to the user. As the aspects are known, a user can clearly distinguish a recommendation list containing, say, movies of a single genre, from one which contains a mix of genres. Latent aspects do not afford this ability. It can be difficult to distinguish a highly diverse recommendation from a less diverse one, when the diversified aspects are latent.

Explicit Aspects. In [10], the xQuAD framework is developed and evaluated largely using explicit aspects, where the space of item features \mathcal{F} becomes the space of all possible aspects \mathcal{A} . Rather than learn the aspect probabilities to directly optimise for (1), they are heuristically estimated by counting user, item and aspects co-occurrences in the input data.

The estimates are as follows:

$$p(a|u) \sim \frac{|\{i \in \mathcal{I}_u : a \in \mathcal{A}_i\}|}{\sum_{a' \in \mathcal{A}} |\{i \in \mathcal{I}_u : a' \in \mathcal{A}_i\}|}, \quad p(i|u, a) \sim \frac{\mathbb{1}_{\mathcal{A}_i}(a)s(u, i)}{\sum_{j \in \mathcal{R}} \mathbb{1}_{\mathcal{A}_i}(a)s(u, j)}$$

where \mathcal{I}_u denotes the set of items in the profile of user u , $\mathbb{1}_{\mathcal{A}_i}(a) = 1$ if $a \in \mathcal{A}_i$ and 0 otherwise. $p(a|u)$ is the count of occurrences of aspect a over the sum of occurrences of all aspects in the profile of user u . $p(i|u, a)$ is proportional to the sum of scores of items in user profile that share the feature a . This model also assumes that $p(a|i) = 1/|\mathcal{A}_i|$ when $a \in \mathcal{A}_i$ and 0 otherwise. Each aspect connected to an item that a user likes, contributes equally to the estimation of the user’s intents. Consider, for example, a movie described by the genres *adventure*, *comedy* and *animation*. A user may have watched that movie more so because it is an *animation* movie than because it is a *comedy* or an *adventure*. But, the co-occurrence method gives equal weight to all three genres in the estimation of $p(a|u)$. We refer to this estimation method as **ExAs-Co0**, for explicit aspects co-occurrence estimation method.

Latent Aspects. The approach of using latent aspects has been pointed out previously in [10]. Latent feature extraction can be achieved using any dimensionality reduction methods, but (1) can be directly recognised as an instance of the probabilistic latent semantic analysis (**pLSA**) model [5, 6]. The **pLSA** method assumes that $p(i|u, a)$ is independent of u and learns $p(a|u)$ and $p(i|a)$ by a direct optimisation of (1) on training data. This can be achieved through the EM algorithm, as proposed in [6]. By applying **pLSA**, we are finding the best values for the aspect components, from the point-of-view of their predictive performance. The algorithm can be tuned by selecting the number of aspects. Despite its potential for selection of aspect components with a superior predictive performance, the hidden aspect model tends not to be preferred as a basis for diversification, because of the difficulty of interpreting its output.

Explicit Aspect Learning. The advantage of explicit aspects is their interpretability but they may be inaccurate when their predictive performance is considered. Latent aspects are chosen to directly optimise for predictive performance, but are much less interpretable. We propose a middle ground – we adopt an aspect model based on explicit aspects, but learn the components of the model by optimising for their predictive performance.

We propose a constrained pLSA model, which we'll refer to as c-pLSA, based on explicit aspects. Given a set of explicit aspects \mathcal{A} and their item correspondences \mathcal{A}_i , we impose the constraint that $p(i|a) = 0$ whenever $a \notin \mathcal{A}_i$. This additional constraint is easily incorporated into the EM algorithm for pLSA. The E-step of that algorithm requires the computation of

$$Q^*(a; u, i) = \frac{\hat{p}(i|a)\hat{p}(a|u)}{\sum_{a' \in \mathcal{A}} \hat{p}(i|a')\hat{p}(a'|u)}.$$

where $\hat{p}(\cdot)$ correspond to estimates from the previous iteration. The constrained algorithm simply initialises $\hat{p}(i|a) = 0 \forall a \notin \mathcal{A}_i$, so that $Q^*(a; u, i) = 0 \forall a \notin \mathcal{A}_i$. The optimal Q^* is used then in the M-step of the EM to derive:

$$\hat{p}(i|a) = \frac{\sum_{(u,i'): i'=i} Q^*(a; u, i)}{\sum_{(u,i)} Q^*(a; u, i)}$$

which maintains the condition.

A user has been selected that has 17 non-zero aspect probabilities, $p(a|u)$, estimated by ExAs-CoO. Top aspects with respect to $p(a|u)$ (probability values are given in parentheses) are: *comedy* (0.46), *adventure* (0.07), *action* (0.07), *sci-fi* (0.07), *romance* (0.06) and *drama* (0.05). The c-pLSA model sees the user differently, as *comedy* (0.86), *adventure* (0.10), *fantasy* (0.02) and *film-noir* (0.02). As the predictive performance of the c-pLSA is better, we believe the c-pLSA has learned the user aspect probabilities more accurately.

The proposed solution might be used not only to learn the real user interests but as well to find what drives people that they interacted with an item. For example, *Star Wars: Episode IV - A New Hope* is described as *action*, *adventure*, *fantasy* and *sci-fi*. With the c-pLSA we can derive weightings to express how people perceive this item in terms of those genres. In this case we would find out that people pick this item mostly because it is an *adventure* (0.5) than *sci-fi* (0.25), *fantasy* (0.24) or *action* (0.01).

4. EVALUATION

This section describes our experiments for evaluating the performance of aspect models.

4.1 Dataset

The MovieLens (ML-1m) dataset [4] contains the ratings information of users on movies. It holds 1,000,209 ratings from 6,040 users on 3,706 movies. Movies are described using at least one genre out of 18 available. Ratings are made on a 5-star scale, with whole-star ratings only. The dataset has been split into 2 subsets: the training set with 800,167 ratings and the test set of 200,042 ratings.

4.2 Experiment Setup

Four baseline algorithms have been used to generate recommendation candidates lists R_u of 100 items: user- (UB) and item-based (IB) neighbourhood [3], matrix factorisation (MF) [7] and probabilistic latent semantic analysis (pLSA) [6] recommenders.

Each candidate list has been re-ranked using xQuAD, based on the input aspect models. We denote xQuAD variations: xQuAD_E, xQuAD_C, xQuAD_P respectively for each of models: ExAs-CoO, c-pLSA and pLSA. The pLSA aspect model uses 18 latent aspects, c-pLSA uses 18 explicit aspects. Performance of xQuAD has been checked at different

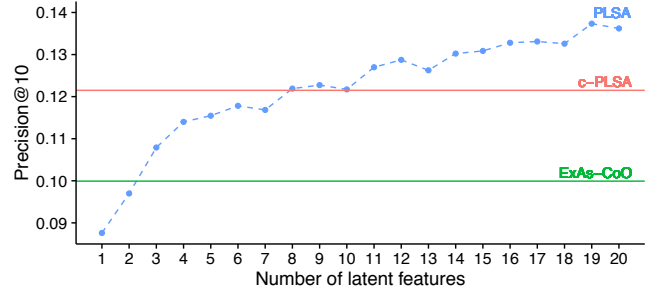


Figure 1: Predictive performance of aspect models.

values of λ , from 0 to 1 with the step of 0.1. Re-rankers produce final recommendations of top $N = 10$ items.

Experiments and the evaluation have been done using the implementations of the algorithms and metrics available in the RankSys framework¹.

4.3 Metrics

Among the many metrics that have been proposed for evaluating diversity, intent-aware metrics, such as ERR-IA [1], cannot be usefully applied to compare the performance of different aspect models as they require the aspect model as input to the metric. Nevertheless, in our results section, we note that on the ERR-IA metric, xQuAD_C outperforms xQuAD_E for MF and IB, even when the ExAs-CoO model is used as the input to the metric. We report feature-based metrics, such as α -nDCG ($\alpha = 0.5$) [2] and subtopic recall (S-recall) [11], to evaluate the diversity. The accuracy is assessed based on the precision (Prec), recall and nDCG [8]. Relevance-aware versions of the metrics are used, where 5 star rating is considered as relevant. Metrics are evaluated at $N = 10$.

4.4 Predictive Performance

It is worth noting that the item to aspect mapping is sparse – typically, each item has only a few aspects associated with it. For instance, in MovieLens, there are 18 movie genres and the average number of genres per movie is less than 2. Thus, for a given total number of aspects, the number of degrees of freedom in c-pLSA is much less than in standard pLSA and it is to be expected that c-pLSA fits the data less accurately than pLSA. Nevertheless, the structure of explicit aspects supports good recommendation. Since c-pLSA is optimised on predictive performance, it can be expected that in general c-pLSA outperforms ExAs-CoO in terms of predictive accuracy.

These observations are tested and in figure 1, the precision of recommendation using pLSA is plotted against the number of hidden aspects k , with the precision of ExAs-CoO and c-pLSA indicated on the plot. pLSA achieves the accuracy of c-pLSA when $k = 8$. The $s(u, i)$ scores of the best performing baseline, MF, have been used for the ExAs-CoO.

4.5 Results

Accuracy Performance. Figure 2 shows the precision performance when different input aspect models are used, for different values of the tradeoff parameter λ . The highest precision can be seen for $\lambda = 0.4$. The results confirm that

¹<http://ranksys.org>

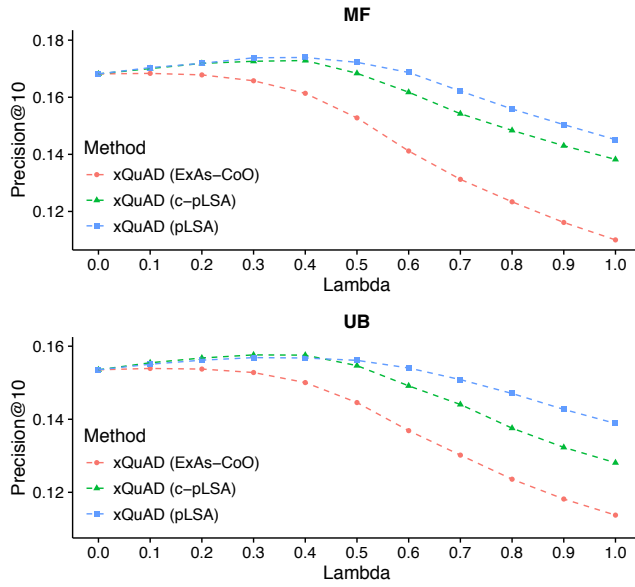


Figure 2: Intent-aware re-rankings for two recommendation algorithms in ML-1m dataset.

| | λ | Accuracy | | | Diversity | | |
|---------------------|-----------|--------------|--------------|--------------|--------------|----------------|--------------|
| | | Prec | nDCG | Recall | ERR-IA | α -nDCG | S-recall |
| MF | | | | | | | |
| +xQuAD _E | 0.4 | 0.168 | 0.274 | 0.270 | 0.168 | 0.278 | 0.157 |
| +xQuAD _C | 0.3 | 0.173 | 0.281 | 0.273 | 0.177 | 0.289 | 0.162 |
| +xQuAD _P | 0.3 | 0.174 | 0.283 | 0.276 | 0.176 | 0.288 | 0.162 |
| PLSA | | 0.143 | 0.226 | 0.222 | 0.143 | 0.234 | 0.139 |
| +xQuAD _E | 0.5 | 0.139 | 0.221 | 0.220 | 0.159 | 0.261 | 0.157 |
| +xQuAD _C | 0.5 | 0.152 | 0.243 | 0.237 | 0.157 | 0.254 | 0.149 |
| +xQuAD _P | 0.5 | 0.151 | 0.240 | 0.232 | 0.155 | 0.250 | 0.146 |
| UB | | | | | | | |
| +xQuAD _E | 0.3 | 0.154 | 0.252 | 0.243 | 0.167 | 0.274 | 0.157 |
| +xQuAD _C | 0.2 | 0.153 | 0.251 | 0.243 | 0.178 | 0.290 | 0.166 |
| +xQuAD _P | 0.3 | 0.157 | 0.259 | 0.248 | 0.173 | 0.281 | 0.159 |
| IB | | | | | | | |
| +xQuAD _E | 0.6 | 0.131 | 0.207 | 0.202 | 0.143 | 0.229 | 0.137 |
| +xQuAD _C | 0.5 | 0.128 | 0.203 | 0.198 | 0.153 | 0.244 | 0.149 |
| +xQuAD _P | 0.5 | 0.144 | 0.230 | 0.222 | 0.158 | 0.250 | 0.147 |
| PLSA | 0.6 | 0.148 | 0.235 | 0.227 | 0.157 | 0.253 | 0.148 |

Table 1: Accuracy and diversity performance results. For each baseline and re-ranker, the value of λ is presented for the best setting in terms of α -nDCG diversity metric. All differences with respect to the baselines are statistically significant (Wilcoxon $p < 0.001$) except when in italics.

the best precision is obtained when the pLSA model is used. For $\lambda \leq 0.4$, performance of c-pLSA is close to the pLSA. Both, c-pLSA and pLSA, outperform the ExAs-CoO model – the order from figure 1 in terms of performance is preserved.

Results on other baselines and other accuracy metrics are presented in table 1. It shows the generally superior accuracy performance of the pLSA however for the PLSA and UB baselines the c-pLSA model performs slightly better.

Diversity Performance. The impact of aspect models on diversification results is reported in table 1. The ExAs-CoO model generally offers the best diversity performance with an exception on α -nDCG on IB. The c-pLSA model performs similarly or slightly better than the pLSA. Although c-pLSA and pLSA models might not be as good as ExAs-CoO, they offer diversity capabilities that are better than the baseline algorithms. Also, their accuracy-diversity tradeoff is better than for ExAs-CoO model.

Neither pLSA nor c-pLSA perform the best on feature-based metrics – α -nDCG and S-recall. pLSA is unaware of explicit aspects and cannot directly optimise for those. In practise, the user aspect probabilities have fewer non-zero values when c-pLSA is used compared to when ExAs-CoO is used. Actually, for only 240 users out of 6,040, the number of non-zero aspect probabilities remain the same for ExAs-CoO and c-pLSA, and for almost 66% of users, the number of non-zero user aspect probabilities produced by c-pLSA is fewer by 4 or more than produced by ExAs-CoO. This is because ExAs-CoO gives equal weights to aspects of multi-aspect movies, while c-pLSA focuses only on those aspects that contribute to relevance. As a consequence, the re-ranker has more opportunity with ExAs-CoO to improve diversity metrics. c-pLSA results in a lower capability of diversifying over aspects, however we believe it better represents the user’s true intents.

5. CONCLUSION

Research presented here aimed to explore aspect models for the intent-aware diversification framework. A constrained PLSA model has been proposed, evaluated and compared with two other models. The proposed model mixes two other approaches used so far in the field. It maintains interpretability, providing better accuracy-diversity trade-off. Also, it models users and items aspects more accurately than the explicit aspect co-occurrence counting approach.

6. ACKNOWLEDGMENTS

This project has been funded by Science Foundation Ireland under Grant No. SFI/12/RC/2289.

7. REFERENCES

- [1] O. Chapelle, D. Metzler, Y. Zhang, and P. Grinspan. Expected reciprocal rank for graded relevance. *ACM CIKM '09 Conference Proceedings*, pages 621–630, 2009.
- [2] C. L. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon. Novelty and Diversity in Information Retrieval Evaluation. *ACM SIGIR'08 Conference Proceedings*, page 659, 2008.
- [3] C. Desrosiers and G. Karypis. A Comprehensive Survey of Neighborhood-based Recommendation Methods. *Recommender Systems Handbook*, 54:107–144, 2011.
- [4] F. M. Harper and J. a. Konstan. The MovieLens Datasets : History and Context. 5(4):1–19, 2015.
- [5] T. Hofmann. Unsupervised Learning by Probabilistic Latent Semantic Analysis. *Machine Learning*, 42(1-2):177–196, 2001.
- [6] T. Hofmann. Latent Semantic Models for Collaborative Filtering. *ACM Transactions on Information Systems*, 22(1):89–115, 2004.
- [7] Y. Hu, Y. Koren, and C. Volinsky. Collaborative Filtering for Implicit Feedback Datasets. 2008.
- [8] K. Järvelin and J. Kekäläinen. Cumulated Gain-Based Evaluation of IR Techniques. *ACM Transactions on Information Systems*, 20(4):422–446, 2002.
- [9] R. Santos, C. Macdonald, and I. Ounis. Exploiting Query Reformulations for Web Search Result Diversification. *WWW*, pages 881–890, 2010.
- [10] S. Vargas, P. Castells, and D. Vallet. Intent-oriented Diversity in Recommender Systems. *ACM SIGIR'11 Conference Proceedings*, pages 1211–1212, 2011.
- [11] C. X. Zhai, W. W. Cohen, and J. Lafferty. Beyond Independent Relevance: Methods and Evaluation Metrics for Subtopic Retrieval. 49(1):10–17, 2003.