Title	Personalised Ranking with Diversity
Authors(s)	Hurley, Neil J.
Publication date	2013-10-16
Publication information	Hurley, Neil J. "Personalised Ranking with Diversity." ACM, 2013.
Conference details	RecSys '13: 7th ACM conference on Recommender systems, Hong Kong, China, 12-16 October 2013
Publisher	ACM
Item record/more information	http://hdl.handle.net/10197/10880
Publisher's statement	© ACM, 2013. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in PUBLICATION, {VOL#, ISS#, (DATE)} http://doi.acm.org/10.1145/2507157.2507226

Personalised Ranking with Diversity

Neil Hurley
UCD INSIGHT
University College Dublin
Ireland
neil.hurley@ucd.ie

ABSTRACT

In this paper we discuss a method to incorporate diversity into a personalised ranking objective, in the context of ranking-based recommendation using implicit feedback. The goal is to provide a ranking of items that respects user preferences while also tending to rank diverse items closely together. A prediction formula is learned as the product of user and item feature vectors, in order to minimise the mean squared error objective used previously in the `RankALS` and `RankSGD` methods, but modified to weight the difference in ratings between two items by the dissimilarity of those items. We report on preliminary experiments with this modified objective, in which the minimisation is carried out using stochastic gradient descent. We show that rankings based on the output of the minimisation succeed in producing recommendation lists with greater diversity, with just a small loss in relevance of the recommendation, as measured by the error rate.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

Keywords

Diversity; Implicit Ratings

1. INTRODUCTION

A large amount of research on recommender system algorithms has been driven by the task of predicting ratings for unrated items, given a database of items that have been explicitly rated by users in the past. Matrix factorisation models have proven very successful in minimising the root mean squared prediction error and fast algorithms have been developed to fit these models. However, there is growing interest within the recommender system community in producing recommendation algorithms that optimise for different quality measures in other contexts. One context of par-

ticular interest is that of producing personalised *rankings* of items given a database of implicit feedback [2, 4, 5, 3, 6]. For instance, in a music recommendation scenario, implicit positive feedback is gathered in terms of the number of times that a track or an artist has been listened to by a user. In this case, the absence of a rating can be interpreted as negative feedback that should be incorporated into the recommendation algorithm. Of particular interest to the work presented here are the methods of [3, 6], in which a matrix factorisation model is learned by the minimisation of a ranking objective.

Ranking based on relevance only can result in recommendation sets that have low diversity – many of the highly relevant items can be very similar to each other. This has led to interest in diversifying recommendation lists while maintaining high relevance [10, 8, 7]. This paper focuses on this problem. Differently to past work that has focused on diversifying nearest-neighbour algorithms [8, 9] or on selecting diverse subsets of items from a set of relevant items [10], here we examine whether a diversification criterion can be incorporated into a ranking-based objective. Using a matrix factorisation model, user- and item-feature vectors are learned by minimising this objective to produce predicted ratings, that when used to rank the items in the database, result in recommendation sets that are highly diverse, while remaining highly relevant. We present preliminary results that evaluate our *ranking with diversity* model on the Netflix dataset.

2. RANKING WITH IMPLICIT FEEDBACK

Adopting the notation of [6], let U and I be the number of users and items, respectively, in the database and write $\mathcal{U}$ and $\mathcal{I}$ for the sets of users and items. We use $u \in \mathcal{U}$ to index users and $i, j \in \mathcal{I}$ to index items. Implicit ratings are denoted as r_{ui} and the matrix of ratings is denoted as $R \in \mathbb{R}^{U \times I}$. Predictions are denoted as $\hat{r}_{ui}$. $\mathcal{T}$ denotes the set of (u, i) indices of R where positive feedback is provided. If $(u, i) \notin \mathcal{T}$ then an implicit negative feedback rating of $r_{ui} = 0$ is assumed. $\mathcal{U}_i = \{u | (u, i) \in \mathcal{T}\}$ is the set of users who have given positive feedback on item i.

Our ranking with diversity strategy is based on the objective function proposed originally in [3] and further developed in [6]:

$$f_R(\Theta) = \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}} c_{ui} \sum_{j \in \mathcal{I}} s_j \left[(\hat{r}_{ui} - \hat{r}_{uj}) - (r_{ui} - r_{uj}) \right]^2 \quad (1)$$

where Θ are the parameters of the prediction model and c_{ui} and s_j are parameters of the objective function. The key is

to find a good ranking by learning the rating difference of two items for a given user u. The parameters c_{ui} select user-item pairs with positive feedback and are taken as $c_{ui} = 1$ if $(u, i) \in \mathcal{T}$ and $c_{ui} = 0$ otherwise. The parameters s_j represent the importance of item j in the objective function. The original KDD 2011 challenge [1] for which the objective was designed, required that items that receive positive feedback from a given user should be distinguished from generally popular items. For this task, an importance weighting of $s_j \propto |\mathcal{U}_j|$ is appropriate.

A matrix factorisation formulation is chosen for the prediction model such that $\hat{r}_{ui} = \mathbf{p}_u^T \mathbf{q}_i$, where $\mathbf{p}$ is an F-dimensional user feature vector associated with user u, $\mathbf{q}$ is an F-dimensional item feature vector associated with item i and F is the number of features. Hence $\Theta = (\mathbf{p}_u, \mathbf{q}_i)$ The computational challenge of minimising $f_R(\Theta)$ is that the objective contains $T \times I$ terms. Originally in [3] a stochastic gradient descent algorithm (SGD), in which the items j of the third summation in (1) are sampled at random, was proposed. This is referred to as the `RankSGD` algorithm in [6], where an efficient alternating least squares (ALS) algorithm, `RankALS`, that avoids sampling is proposed. Our focus in not on such computational issues, but rather on exploring whether the ranking objective can be modified to take account of item diversity. In the experiments presented later in this paper, we use `RankSGD`, although `RankALS` could also be applied.

3. RANKING WITH DIVERSITY

Assume that we have an $I \times I$ diversity or distance matrix $\mathrm{D} = \{d_{ij}\} \in \mathbb{R}^{I \times I}$ expressing how dissimilar item i is to item j. This matrix may be obtained from any source; it may, for example, represent a content-based dissimilarity of the items. We assume that the dissimilarity is derived from a positive definite, normalised similarity matrix, $\mathrm{S} = \{s_{ij}\}$, with $s_{ii} = 1$, using $d_{ij} = \sqrt{1 - s_{ij}}$.

We wish to form an I-dimensional ranking vector $\mathbf{q}$ based on this dissimilarity matrix. Let π_i denote the item whose rank is i where items are ranked according to $\mathbf{q}$ i.e. $q_{\pi_i} \geq q_{\pi_j}$. Consider an objective that attempts to choose $\mathbf{q}$ so that the squared difference of the difference between two components i and j of the ranking vector and the dissimilarity d_{ij} is minimised:

$$f_1(\mathbf{q}) = \sum_i \sum_j (|q_i - q_j| - d_{ij})^2$$

This objective may be minimised by gradient descent, using:

$$\begin{aligned}
\nabla_q(f_1) &= \sum_j (|q_i - q_j| - d_{ij})\mathrm{sgn}(q_i - q_j) - \\
&\quad \sum_j (|q_j - q_i| - d_{ji})\mathrm{sgn}(q_j - q_i) \\
&= 2I(q_i - \bar{q}) - 2\sum_j \mathrm{sgn}(q_i - q_j)d_{ij} \quad \text{(D symmetric)} \\
&= 2I(q_i - \bar{q}) - 2(\sum_{j|q_i > q_j} d_{ij} + \sum_{j|q_i < q_j} d_{ij})
\end{aligned}$$

where $\bar{q}$ is the mean of $\mathbf{q}$. We conduct an experiment in which we form a similarity matrix $\mathrm{S} = \mathrm{XX}^T$ for a random vector X with $I = 100$. We derive the dissimilarity matrix D and minimise $f_1(\mathbf{q})$. The resulting $\mathbf{q}$ is used to sort the items and a set of $N = 10$ items is selected in order, starting
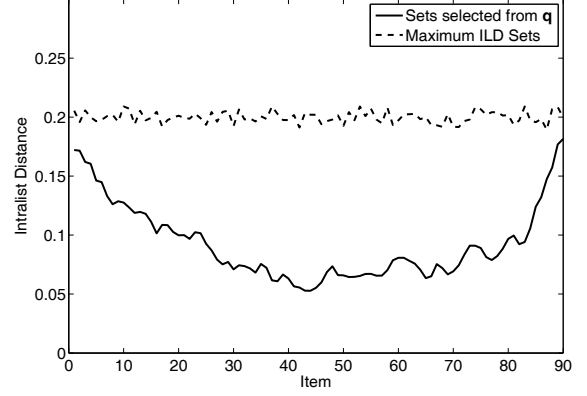

Figure 1: The ILD of a list of $N = 10$ items, consisting of items $\{\pi_i, \ldots, \pi_{i+N-1}\}$, plotted against the item π_i, where ILD $= \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} d_{\pi_i \pi_j}$.

from each item in turn. The intra-list distance (ILD) [10, 8] of the resulting sets is plotted in Figure 1 and compared against sets selected by starting with the same item but greedily maximising the ILD. We note that choosing items from either end of the ordering produced by $\mathbf{q}$ results in sets with high ILD.

3.1 Diversifying Relevance

Of course, if diversity were the only criterion of interest, then we could simply choose sets that maximise the ILD. However, we also want to maximise the relevance of the resulting set. Given ranking vector $\mathbf{r}$, chosen on the basis of relevance, we can ask if we can find a new ranking vector $\mathbf{q}$ which respects the given ranking in $\mathbf{r}$ but otherwise ranks items according to their diversity. To do so, consider the objective

$$f_2(\mathbf{q}) = \sum_i \sum_j ((q_i - q_j) - d_{ij}(r_i - r_j))^2 .$$

Again, computing the gradient, we obtain (assuming symmetric D)

$$\begin{aligned}
\nabla_q(f_2) &= 4\sum_j (q_i - q_j) - 2\sum_j (d_{ji} + d_{ij})(r_i - r_j) \quad (2) \\
&= 4I(q_i - \bar{q}) - 4\sum_j (r_i - r_j)d_{ij}
\end{aligned}$$

Setting the gradient to zero, we obtain

$$(q_i - \bar{q}) = \frac{1}{I}\sum_j (r_i - r_j)d_{ij} = r_i\bar{d}_i - \frac{1}{I}\sum_j d_{ij}r_j \quad (3)$$

Using the same dissimilarity matrix as the previous experiment with $I = 100$, we now construct a random relevance vector r with each component assigned a number between 1 and 5, with uniform probability. In Figure 2 we show the ILD and the average relevance of sets of $N = 10$ items chosen in order from the ranking derived from $\mathbf{q}$ as computed by (3). We see that, while the ordering of $\mathbf{q}$ respects the order of $\mathbf{r}$ (the average relevance is monotonically decreasing), it also results in sets of high diversity as measured by the ILD.
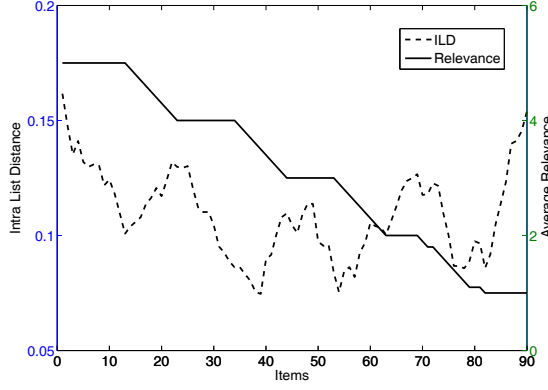
Figure 2: The vector q keeps the same rank order as r when the ratings differ but sets chosen in the rank order of q also have high ILD.

```
Inputs: Learning Rate:  η  Num.of features:  F
1. forall (u,i) ∈ 𝒯 do
2.      draw an item j not rated by u, with
            probability proportional to popularity
3.      r̂_ui ← p_u^T q_i
4.      r̂_uj ← p_u^T q_j
5.      e ← s_j[(r̂_ui − r̂_uj) − d_ij(r_ui − r_uj)]
6.      c = p_u
7.      p_u = p_u − η e (q_i − q_j)
8.      q_i = q_i − η e c
9.      q_j = q_j + η e c
10. end
```

Figure 3: The `RankSGD` algorithm used to minimize the ranking with diversity objective.

3.2 Optimising for Relevance and Diversity

In the previous section, we assumed that the relevance vector was pre-computed and then modified it to take account of diversity. Now we ask if it is possible to optimise for relevance and diversity simultaneously. To do so, we propose the following modification to the ranking objective function (1), to obtain a *ranking with diversity* objective for implicit feedback systems:

$$f_{RD}(\Theta) = \sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{I}} c_{ui} \sum_{j \in \mathcal{I}} s_j \left[(\hat{r}_{ui} - \hat{r}_{uj}) - d_{ij}(r_{ui} - r_{uj}) \right]^2$$

The `RankALS` or `RankSGD` algorithms of [6] and [3] respectively, can be used to minimise this objective to obtain rating estimates $\hat{r}_{ui} = p_u^T q$, that are used to rank items to be recommended to each user u. In our experiments, we have employed the `RankSGD` algorithm with pseudocode shown in Figure 3.

4. EVALUATION

4.1 Datasets

We carry out preliminary evaluation of the ranking with diversity objective on the Netflix dataset. Although this is
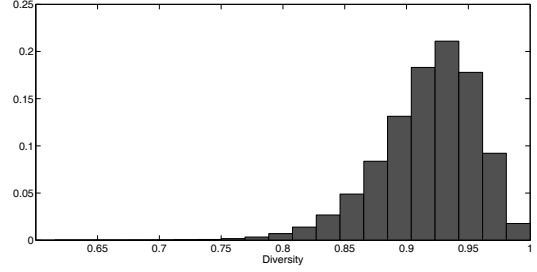

Figure 4: The Distribution of ILD values for $N = 10$ randomly selected items from Netflix

not an implicit rating dataset, it has previously been used for evaluation of `RankALS` and `RankSGD` in [6], so it is possible to compare the results we obtain with these previous results. The Netflix training set consists of 100,480,507 ratings ranging from 1 to 5 from 480,189 users on 17,770 items. Following [6], we defined the implicit rating matrix by assigning 1 to user-items pairs with rating values of 5 and 0 to all other user-item pairs. We trained using the Netflix training set with the probe set removed (containing 22,783,659 implicit ratings) and we tested on the probe set (containing 384,573 implicit ratings).

4.2 Performance Measures

To evaluate the performance of the ranking in terms of relevance, we use the `errorrate` metric as proposed in Track 2 of the 2011 KDD cup [1]. To apply this metric, for each user u, the test set consists of two sets of items $\mathcal{I}_u^+$ and $\mathcal{I}_u^-$ containing items for which u has given positive and negative feedback. $\mathcal{I}_u^-$ is created by drawing $|\mathcal{I}_u^+|$ items randomly with probability proportional to $|\mathcal{U}_i|$. The items in $\mathcal{I}_u^+ \cup \mathcal{I}_u^-$ are ranked using the predicted rating and the errorrate is the proportion of the top $|\mathcal{I}_u^+|$ ratings that are not from $\mathcal{I}_u^+$. A lower errorrate implies better performance. To evaluate the diversity of the recommended sets, we use the ILD of the top N items, obtained by ranking all $\mathcal{I} - \mathcal{I}_u^+$ items using the predicted ratings. We evaluate using $N = 10$.

4.3 Pairwise Item Distances

In the presented evaluation, we use a distance matrix D based on item profile distance. In particular, we obtain the similarity matrix S as the normalised dot product of item profiles:

$$S = (RE^{-\frac{1}{2}})^T (RE^{-\frac{1}{2}})$$

where E is a diagonal matrix with diagonal elements $e_i = \sum_u r_{u,i}^2$. Because of the high sparsity of the Netflix dataset, this leads to generally low similarity values (and consequently very high pairwise distances). To obtain a greater spread of diversity values, the similarity values are filtered through the sigmoidal function $\tanh(\alpha s)$, where we take $\alpha = 20$. The resulting distribution of ILD values of sets of $N = 10$ items selected uniformly at random from the 17,770 items in Netflix is shown in Figure 4.

4.4 Results

In [6] it is stated that an item importance weighting proportional to the item popularity is appropriate when the errorrate measure is being optimised. Hence, we take $s_j =$
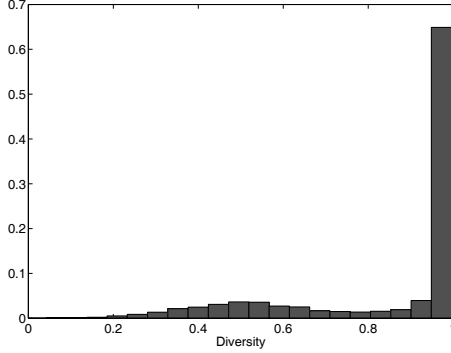
Figure 5: The Distribution of ILD values for $N = 10$ items selected by ranking with diversity $f_{RD}(\Theta)$
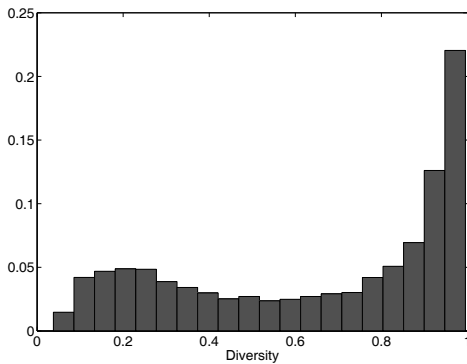

Figure 6: The Distribution of ILD values for $N = 10$ randomly selected by ranking with $f_R(\Theta)$

$|\mathcal{U}_j|/\max|\mathcal{U}_j|$. We compare the `RandSGD` algorithm run to convergence with $\eta = 0.008$ to minimise the ranking objective $f_R(\Theta)$ with the same algorithm used to minimise the ranking with diversity objective $f_{RD}(\Theta)$. The results are summarised in Table 1. The distribution of ILD values obtained from the ranking with diversity objective is shown in Figure 5 and the distribution obtained from optimising $f_R(\Theta)$ is shown in Figure 6. It is clear that ranking with diversity obtains much more diverse recommendation sets at some cost to performance in terms of relevance. As the raw ILD values are dependent on the distance matrix, it is informative to look at the results in terms of percentiles of the distributions. For instance, the 80th percentile of the distribution of ILD values of randomly selected sets, corresponds to the 79th percentile of the ILD values obtained when only relevance is considered. However, it is only the 36th percentile of the distribution of ILD values obtained by ranking with diversity. Ranking with diversity is highly biased towards highly diverse recommendations.

Table 1: Performance on Netflix Dataset

Obj	Errorrate	Mean ILD
$f_R(\Theta)$	0.25	0.64
$f_{RD}(\Theta)$	0.29	0.84

5. CONCLUSION

Preliminary results with the ranking with diversity objective are encouraging and further evaluations on other datasets will follow. Other formulations for incorporating diversity within ranking objectives will be considered in future work.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] G. Dror, N. Koenigstein, Y. Koren, and M. Weimer. The yahoo! music dataset and kdd-cup '11. *Journal of Machine Learning Research - Proceedings Track*, 18:8–18, 2012.

[2] Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, ICDM '08, pages 263–272, Washington, DC, USA, 2008. IEEE Computer Society.

[3] M. Jahrer and A. Töscher. Collaborative filtering ensemble for ranking. *Journal of Machine Learning Research - Proceedings Track*, 18:153–167, 2012.

[4] I. Pilászy, D. Zibriczky, and D. Tikk. Fast als-based matrix factorization for explicit and implicit feedback datasets. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 71–78, New York, NY, USA, 2010. ACM.

[5] 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*, UAI '09, pages 452–461, Arlington, Virginia, United States, 2009. AUAI Press.

[6] G. Takács and D. Tikk. Alternating least squares for personalized ranking. In *Proceedings of the sixth ACM conference on Recommender systems*, RecSys '12, pages 83–90, New York, NY, USA, 2012. ACM.

[7] 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, New York, NY, USA, 2011. ACM.

[8] M. Zhang and N. Hurley. Avoiding monotony: Improving the diversity of recommendation lists. In *Proceedings of 2nd ACM International Conference on Recommender Systems*, 2008.

[9] M. Zhang and N. Hurley. Niche product retrieval in top-n recommendation. In *2010 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2010, Toronto, Canada, August 31 - September 3, 2010, Main Conference Proceedings*, pages 74–81. IEEE, 2010.

[10] C. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web*, pages 22–32, 2005.