Injecting Semantic Diversity in Top-N Recommender Systems Using Determinantal Point Processes and Curated Lists

Surya Kallumadi
Kansas State University
USA
surya@ksu.edu

Gabriel Necoechea
Kansas State University
USA
gnecoechea@ksu.edu

ABSTRACT

Top-N Recommender Systems usually suffer from intra-list diversity as they are tailored for relevance and predicted rating accuracy. This problem is magnified in the case of cold start setting - resulting in users being restricted to popular set of items and can result in a *"rich getting richer eco-system"*. As a result, in recent years, more attention is being paid to improving the diversity of recommender system results. List creation has become a popular way for users to express preferences over items on online platforms such as imdb.com and goodreads.com. These user curated lists tend to contain a coherent semantic representation of the domain the list of items belong to. List curation can be seen as a way to capture fine grained topic-specific item-lists by users. Understanding and modeling user preferences expressed in these curated lists can help with diverse set of applications such as recommendations, user modeling, session understanding etc.

In this paper, we propose an approach to improve the diversity of results generated by Top-N recommender systems, by using Determinantal Point Processes (DPPs) over user curated lists in the movie domain and incorporating them to rerank the Top-N recommender systems. For this work, we use the user curated lists in the imdb.com domain. We evaluate our approach over the Movielens 1-Million dataset and compare the results with other baseline approaches. Our early results show that incorporating semantic similarity expressed in user lists as a diversity proxy results in a more diverse set of recommendations.

KEYWORDS

Recommender Systems, Diversity, User Curated Lists, Diversity metrics

ACM Reference Format:
Surya Kallumadi and Gabriel Necoechea. 2018. Injecting Semantic Diversity in Top-N Recommender Systems Using Determinantal Point Processes and Curated Lists. In *UMAP'18 Adjunct: 26th Conference on User Modeling, Adaptation and Personalization Adjunct, July 8–11, 2018, Singapore, Singapore.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3213586.3226223

UMAP'18 Adjunct, July 8–11, 2018, Singapore, Singapore

ACM ISBN 978-1-4503-5784-5/18/07…$15.00
https://doi.org/10.1145/3213586.3226223

1 INTRODUCTION

Modern Recommender systems use interaction data between users and items - either in the form of explicit or implicit feedback - to predict user preference of unobserved items. These approaches tend to use relevance metrics such as RMSE and ranking metrics to predict a user's proclivity towards items, but focusing solely on these relevance metrics leads to recommending highly similar homogeneous set of items that exhibit low diversity [2]. The drawback of such an approach is that the user is constrained to a low entropy result set resulting in lower user satisfaction. This will also result in reduced coverage of the item set and lower exploration opportunities for the user to discover novel and serendipitous items [10]. Improving diversity of recommender systems has become an important research topic in order to increase the discoverability of items. In [7] the authors note that lack of diversification of results lead to *"filter bubbles"* and over time recommender systems expose users to a narrowing set of items. Vargas et. al. use various latent user preferences over items to improve quality of recommendations[8]. The authors propose identifying user sub profiles by creating subsets of user interests from the set of user preferences over items. In this work, first the user profiles are partitioned into pre-defined categories over the items and then use these partitioned user sub-profiles to generate partition specific recommendations. The recommendations from various partitions are then aggregated to generate a diverse set of recommendations.

User-curated lists span a wide range of domains and usually contain items that users view to be of a coherent topic. These lists of items can be videos belonging to a particular topic on Youtube, movies on the Internet Movie Database (IMDb), books on GoodReads domain, lists of users and accounts on twitter, playlists on music platforms such as Spotify and wish lists on e-commerce domains such as Amazon. On domains such as Spotify, most user-item preference activity happens predominantly through *"list activity"*. On IMDb, a movie domain and GoodReads, a books domain the items in lists that are curated by users typically share some common attribute such as genre, tag, director, actor, author etc. These lists capture semantically meaningful items at various granularities across various dimensions of user interests. For example in the list titled "TOP WAR MOVIES"[1] on IMDb, the user lists a set of "war" movies. A user list titled "Great Old Movies (pre-1960)"[2] deals with pre-1960 movies. In these examples, the items in the lists co-exist in different contexts such as Genre and Temporal similarity, respectively. While a 'tag' such as "western" exists on imdb.com,

[1] https://www.imdb.com/list/ls026329851/
[2] https://www.imdb.com/list/ls000000580/

a user identifies "Modern western" movies in the user list "Modern Westerns:The Ultimate List."[3] On goodreads.com, within lists tagged with the tag "love,"[4] we can see a wide variety of contexts within which items can be categorized semantically. In this paper we propose to leverage the semantic context that exists within lists to provide a more diverse set of recommendations. In [7] Nguyen et. al. use information encoded in user-generated tags to measure the content diversity of items recommended by Recommender systems. The user-generated tags in this instance capture a context association of a user to an item.

Determinantal Point Processes (DPPs)[5] have been succesfully used in machine learning tasks such as information retrieval, diverse subset sampling, and document summarization. Determinantal Point Processes have the ability to model the balance between quality and diversity of sets as they model *repulsion*. In this exploratory work, we propose to improve the diversity of Top-N recommender systems, by using DPPs over user created lists. We crawled the imdb.com[5] domain to generate a lists dataset, the semantic similarity measure from these user-generated lists is used by the DPP to model the diversity of the items. We use an average dissimilarity metric to measure the diversity of the resulting re-ranked list. Our early results on the MovieLens 1-million ratings dataset[1] show that incorporating semantic similarity expressed in user lists as a diversity proxy results in a more diverse set of recommendations. The contributions in this work are the following:

- We propose leveraging user-curated lists for improving the intra list diversity of Top-N recommender systems using Determinantal Point Processes(DPPs). We empirically show that the intra-list diversity score of Top-N systems can be improved by DPP re-ranking.
- We argue for using more diversity metrics apart from the "popularity-biased" co-occurrence similarity that is popular in the recommender systems literature. We observe that to improve our understanding of this field, we need objective metrics that would inform us of the utility of various diversity metrics.

2 BACKGROUND

2.1 DPPs

A Determinantal Point Process (DPP) is useful in models and applications where repulsive effects or diversity are important. For example, creating a model with the assumption of a uniform spatial distribution (of e.g. particles [6], cluster centers [9], etc.) may be unwarranted, and using a DPP may be a better choice. In recommender systems and information retrieval settings, it may be desirable to return a more diverse collection of items, and a DPP can be used to incorporate this preference for diversity.

Given a set Y of cardinality N, a DPP can be thought of as a probability distribution on the subsets of Y, where the probabilities are proportional to the determinant of some matrix. Such distributions are important because the determinant of a matrix captures the volume of the parallelepiped spanned by its columns, which provides a useful measure of diversity amongst the column vectors.

[3] https://www.imdb.com/list/ls055895628/

[4] https://www.goodreads.com/list/tag/love

[5] https://www.imdb.com/

In principle, one can encode relevant information into an N-by-N matrix, each row and column being indexed by an item in Y. The determinants of interest are those corresponding to the principal minors of this matrix. Different approaches to DPPs exist, depending on the conditions the N-by-N matrix satisfies. In the *L-ensemble* approach, our N-by-N matrix L is positive semidefinite, and the probability that our randomly selected subset $R \subseteq Y$ is exactly the subset $A \subseteq Y$ is given by the equation

$$P_L(A = R) = \frac{\det(L_A)}{\det(L+I)},$$

where I is the $N \times N$ identity matrix and L_A is the principal minor of L whose rows and columns are indexed by the items in A.

The L-ensemble approach is useful when we have a feature space representation for the items in Y. In this case, L is just a Gram matrix associated to these items.

For sufficiently large N, it is infeasible to compute the determinants of the all the principal minors and select a subset of maximal determinant. For this reason, applications rely on sampling algorithms that probabilistically constructs a subset with the probability of constructing A being approximately $P_L(A)$.

2.2 k-DPPs

In general, sampling from a DPP involves two random variables: the random subset itself and the size of the subset. For some applications, the desired number of items to select, say k, is already known. (For example, a recommender system may recommend k = 5 items for purchase.) In this case, a *k-DPP* is used. Technical details, including the contents of this section, can be found in [4] and [5] by Kulesza and Taskar.

Given an L-ensemble DPP P_L, we may recalculate probabilities to only take into consideration k-item subsets:

$$P_L^k(A = R) = \frac{\det(L_A)}{\sum_{|A'|=k} \det(L_{A'})}. \tag{1}$$

Since

$$\sum_{A' \subseteq Y} \det(L_{A'}) = \det(L + I) = \det(L + I) \sum_{A' \subseteq Y} P_L(A'),$$

the denominator in (1) can be rewritten as

$$\sum_{|A'|=k} \det(L_{A'}) = \det(L + I) \sum_{|A'|=k} P_L(A') \tag{2}$$

while the numerator in (1) can be rewritten as

$$\det(L_A) = \det(L + I) P_L(A). \tag{3}$$

Using results in the theory of symmetric functions, we derive from (2) that

$$\det(L + I) \sum_{|A'|=k} P_L(A') = e_k(\lambda_1, ..., \lambda_N), \tag{4}$$

where e_i is the ith elementary symmetric polynomial and the $\{\lambda_j\}$ are eigenvalues of L.

Using (3) and (4), we have that the atomic probability can be written as

$$P_L^k(A = R) = \frac{\det(L+I) P_L(A)}{e_k(\lambda_1, \ldots, \lambda_i)}.$$

The practical significance of this form is that the elementary symmetric polynomials may be computed in polynomial time, whereas the sum $\sum_{|A'|=k}$ has exponentially many terms, presenting a priori time complexity challenges. Furthermore, it can be shown that the marginal probability is given by

$$P(i \in R) = \frac{\lambda_i e_{k-1}(\lambda_1, \ldots, \lambda_{i-1})}{e_k(\lambda_1, \ldots, \lambda_i)},$$

yielding a sampling algorithm described by Kulesza and Taskar in [4].

Input: k Eigenvector/Eigenvalue Pairs $\{(v_n, \lambda_n)\}$
Output: A sampled subset R
for $n \leftarrow 1$ **to** N **do**
  if $u \sim U[0,1] < \lambda_n \frac{e_{k-1}(\lambda_1, \ldots, \lambda_{n-1})}{e_k(\lambda_1, \ldots, \lambda_n)}$ **then**
    $J \leftarrow J \cup \{n\}$
    $k \leftarrow k - 1$
    if $k = 0$ **then**
      break
    end
  end
end
$V \leftarrow \{v_n\}_{n \in J}$
$R \leftarrow \emptyset$ **while** $|V| > 0$ **do**
  $R \leftarrow R \cup \{i\}$ with probability $\frac{\sum_{v \in V} (v^T e_i)^2}{|V|}$
  $V \leftarrow V_\perp$, orthonormal basis of the subspace of V orthogonal to e_i
end

Algorithm 1: Procedure for Sampling from k-DPP

The sampling process from the k-DPP is described in Algorithm 1. It's input is an eigendecomposition of matrix L, i.e. a collection of eigenvectors of L along with their associated eigenvalues. It outputs a set containing k items, and the probability of outputting any particular k-element set is given by the k-DPP. To use DPP model in the top-N recommendation task, we should construct the similarity matrix. In the context of recommender systems, we might think of L as a similarity matrix between all pairs of items. In Section 3, we provide more details on obtaining L from user-curated lists.

3 DATA SET, EXPERIMENTS AND RESULTS

The Movielens 1-million ratings dataset contains ~1 million ratings of 3,076 movies provided by 6,040 users.

We obtained the IMDb user lists by performing targeted crawls for lists on the imdb.com domain. We performed various crawls on the imdb.com domain over a period of 4 weeks from 11/24/2017 to 12/24/2017. The crawl was able to capture historical interactions of 74134 users, these interactions were of the form "list activity" and "rating activity". A user performs list activity when they create a list and the lists has at least 1 item in the lists. These ~74k users generated 352543 user-curated lists. These lists contained 13 million interactions between users and items. Of these 350k lists, we pruned generic lists that contain mostly popular items(as these generic lists are associated with no semantic coherence). This resulted in a pruned set of 155496 lists. From these ~155k lists we removed items that were not part of the movielens dataset.

3.1 Experiments and Results

We divide the Movielens dataset into training and test sets, where we randomly select 1 item from the user's item set to be part of the test set and the rest of the items are part of the training set. For each user, using the user-item interaction in the training set as the user profile, we identify the top 20 similar items of each item and the aggregate of all the top 20 similar items of every item in the profile set forms the final candidate set. We then rank the items in the candidate set and items are evaluated for N= 10 and 20. The baseline algorithm aggregates the item-item similarity to generate a top-N list.

The semantic item-item similarity $SemSim_{ij}$ utilized by k-DPP is calculated based on the co-occurrence of the item pairs across the item lists. We can think of $SemSim$ as a Gram matrix by viewing each item as a vector of 0s and 1s, with the m-th coordinate 1 if the item shows up in list m and 0 otherwise. Under this representation, $SemSim_{ij}$ is just the dot product of the vectors for items i and j, and $SemSim$ is a Gram matrix. As a Gram matrix, $SemSim$ is positive semidefinite and therefore able to be used in our k-DPP.

3.1.1 *Evaluation metric.*

We use the average dissimilarity metric to measure the diversity of the resulting set. The similarity of 2 items i, j is given by Sim_{ij} which is the co-occurrence similarity of the items in the Movielens datsets as defined in [3].

$$\text{Diversity(AverageDissimilarity)} = \underset{i,j \in R_u, i \neq j}{\text{mean}} (1 - \mathbf{Sim}_{ij})$$

Table 1: Average dissimilarity measure of Diversity

Approach	diversity-Top 10	diversity-Top 20
Top-100 + Random	0.45	0.47
Top-100	0.72	0.74
Top-100 + k-DPP	**0.79**	**0.86**

3.1.2 *Results.*

Table 1 contains the empirical results for the baseline approaches compared to semantic similarity aware diverse recommendations. Top-100 + Random randomizes the top 100 items of the candidate set, Top-100 just uses the top 100 ranked items as it is, and Top-100 + k-DPP obtains the k-DPP sampling from the top 100 ranked items. As shown in the Table 1, applying k-DPP to the Top-100 will improve the diversity of the Top-N recommender systems.

4 A DISCUSSION ON DIVERSITY METRICS AND THEIR UTILITY

One popular way of measuring diversity of a recommendation list is to look at rating vectors associated with the items and calculate similarity of items based on these rating vectors. One major problem with this approach is that if two movies have similar user rating vectors, that should lead to the conclusion that these are similarly liked, not the conclusion that their content is similar [7]. By this approach, most popular movies would have a high similarity even though from a content and taste perspective they could be highly diverse. Thus, even for prevalent metrics we do not have a consensus on the utility of these metrics. To further complicate things, one can consider a wide array of diversity measures based on content and attributes associated with entities when large amounts of side information is available. We need a structured approach to identify the utility of diversity metrics and such a framework would help us take informed decisions based on a wide range of diversity metrics.

5 CONCLUSIONS AND FUTURE WORK

This paper shows early results for improving the diversity of top-N recommendation algorithms using DPPs over user-curated lists. User-curated lists capture context, and they tend to have properties opposite of diversity, we can leverage this contextual similarity defined by users to improve diversity of recommender systems. We use this intuition and use determinantal point processes on similarities learnt over user list to improve diversity of a top-k list. Our results show that user lists contain vital semantic and contextual information, and can be utilized to improve the diversity of Recommender systems. Diversity is of high significance under certain conditions such as cold start, when user intent is exploratory. Identifying these conditions to improve the discoverability of items can have a significant effect on the user experience. For this reason, we need more nuanced diversity metrics apart from metrics such as Pairwise Mutual information (PMI) and average dissimilarity. Future work will investigate better metrics for measuring diversity in Top-N recommender systems. Furthermore, the sampling process from k-DPP is not efficient and to make this practical, we need to improve the sampling efficiency of k-DPPs. This is another direction for future investigation.

REFERENCES

[1] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. *ACM Trans. Interact. Intell. Syst.* 5, 4, Article 19 (Dec. 2015), 19 pages. https://doi.org/10.1145/2827872

[2] Neil Hurley and Mi Zhang. 2011. Novelty and Diversity in Top-N Recommendation – Analysis and Evaluation. *ACM Trans. Internet Technol.* 10, 4, Article 14 (March 2011), 30 pages. https://doi.org/10.1145/1944339.1944341

[3] George Karypis. 2001. Evaluation of Item-Based Top-N Recommendation Algorithms. In *Proceedings of the Tenth International Conference on Information and Knowledge Management (CIKM '01)*. ACM, New York, NY, USA, 247–254. https://doi.org/10.1145/502585.502627

[4] Alex Kulesza and Ben Taskar. 2011. k-DPPs: Fixed-size Determinantal Point Processes. In *Proceedings of the 28th International Conference on International Conference on Machine Learning (ICML'11)*. Omnipress, USA, 1193–1200. http://dl.acm.org/citation.cfm?id=3104482.3104632

[5] Alex Kulesza and Ben Taskar. 2012. *Determinantal Point Processes for Machine Learning*. Now Publishers Inc., Hanover, MA, USA.

[6] Odile Macchi. 1975. The coincidence approach to stochastic point processes. 7 (03 1975), 83–122.

[7] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. 2014. Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14)*. ACM, New York, NY, USA, 677–686. https://doi.org/10.1145/2566486.2568012

[8] Saúl Vargas and Pablo Castells. 2013. Exploiting the Diversity of User Preferences for Recommendation. In *Proceedings of the 10th Conference on Open Research Areas in Information Retrieval (OAIR '13)*. LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, Paris, France, France, 129–136. http://dl.acm.org/citation.cfm?id=2491748.2491776

[9] Yanxun Xu, Peter Muller, and Donatello Telesca. 2016. Bayesian inference for latent biologic structure with determinantal point processes (DPP). *Biometrics* (2016).

[10] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving Recommendation Lists Through Topic Diversification. In *Proceedings of the 14th International Conference on World Wide Web (WWW '05)*. ACM, New York, NY, USA, 22–32. https://doi.org/10.1145/1060745.1060754