Visualizing Ratings in Recommender System Datasets

Diego Monti[a], Giuseppe Rizzo[b] and Maurizio Morisio[a]

[a]Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy
[b]LINKS Foundation, Via Pier Carlo Boggio 61, 10138 Turin, Italy

Introduction

Being able to correctly interpreting the results obtained during an offline evaluation of different recommender systems is of paramount importance for understanding the quality of the suggested items [1]. However, this task is particularly difficult as it requires knowing several details regarding the evaluation protocol and the rating dataset exploited for conducting the experiments [2].

Different metrics have been proposed in literature to summarize the main characteristic of a rating dataset, i.e. sparsity or entropy. However, we argue that such metrics are not sufficient for comparing datasets in a reliable way, as many other facets should be taken into account. A possible solution to this problem could be represented by data visualization techniques [3].

We propose a novel qualitative approach based on data visualization for creating a graphical summary of any collection of user preferences. This method is useful for visually identifying similarities and differences among the available datasets. Furthermore, we present a Web-based tool, named RS-viz, for easily constructing the proposed visualization and comparing rating datasets in an intuitive way.

The code is freely available in a GitHub repository at `https://github.com/D2KLab/rs-viz`.

Visualization Approach

In order to visually represent the rating matrix associated with a generic dataset we opted for a 3D scatter plot. We decided to create a more compact representation of the rating matrix before constructing the visualization. As an example of the proposed method, we report in Figure 1 the scatter plot obtained from the MovieLens 100K dataset.

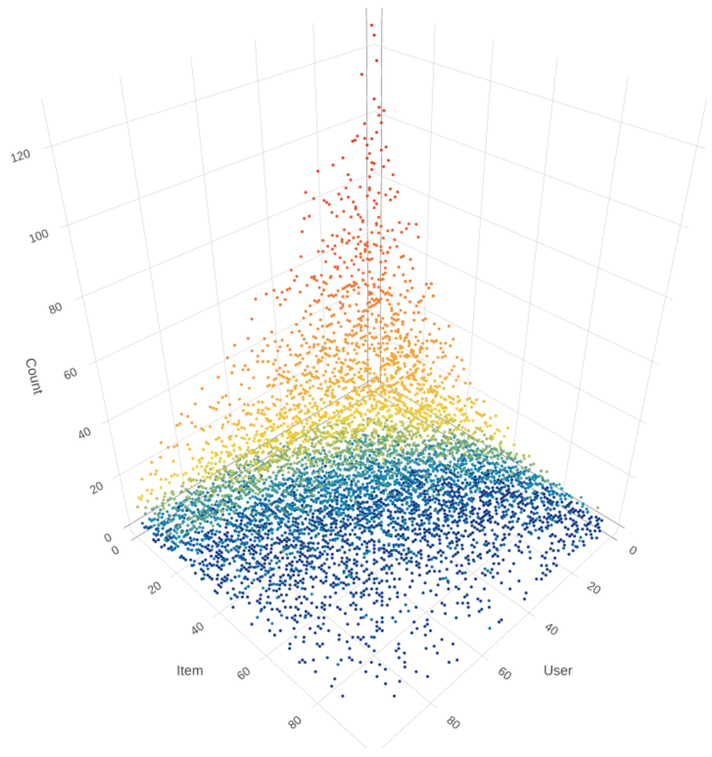


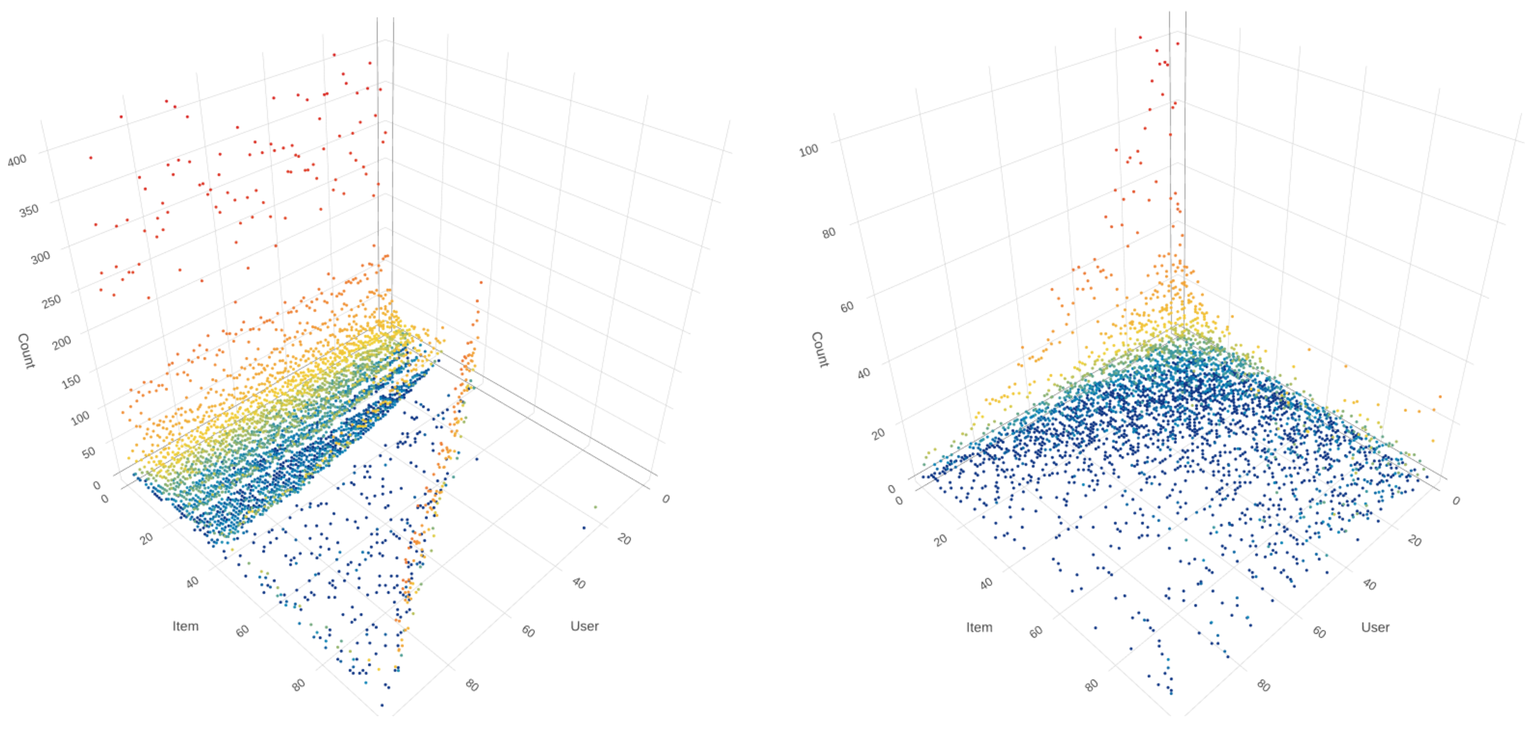
Figure 1: The MovieLens 100K dataset.

Evaluation Campaign

We report the numerical outcomes of an evaluation campaign conducted on the HetRec LastFM dataset using various recommendation approaches with the purpose of understanding if our visualization technique is capable of capturing the different characteristics of a rating dataset. The results of these experiments are reported in Table 1. The same datasets obtained from HetRec LastFM by modifying the rating threshold were exploited for creating two scatter plots using RS-viz, as displayed in Figure 2.

Table 1: The numerical results of the experimental comparison using the HetRec LastFM dataset.

	Rating threshold = 0				Rating threshold = 1,000			
Algorithm	Coverage	Precision	Recall	NDCG	Coverage	Precision	Recall	NDCG
Random	0.706679	0.000798	0.000745	0.000858	0.705562	0.000107	0.000622	0.000133
Most Popular	0.001692	0.071170	0.071480	0.079673	0.001684	0.022122	0.090233	0.027437
Item KNN	0.235321	0.129362	0.131967	0.145258	0.107233	0.002878	0.013012	0.002686
User KNN	0.030074	0.157234	0.160353	0.193121	0.049343	0.040672	0.160767	0.055013
BPRMF	0.022979	0.081277	0.082248	0.094737	0.003756	0.021695	0.088211	0.024366
WRMF	0.015558	0.159947	0.162332	0.195107	0.012886	0.039606	0.157484	0.053148



(a) Rating threshold = 0 (b) Rating threshold = 1,000

Figure 2: The 3D scatter plots obtained using the HetRec LastFM dataset with different rating thresholds.

Discussion

From the visualization provided in Figure 2(a), we can observe that the HetRec LastFM dataset has a very different structure from the one of the MovieLens datasets. In fact, a limited number of items are associated with the preferences of almost all users.

Also the strange area in the plot with almost no preferences is a direct result of the collection protocol, which relied on the LastFM website to obtain the top artists for a set of users. In fact, the list of artists available in the dataset is limited to 50 items per user.

If we increase the value of the rating threshold, we can observe that the resulting scatter plot represented in Figure 2(b) is more similar to the ones of the MovieLens datasets, resulting in a very typical long tail distribution with respect to both the items and the users. This outcome is due to the fact that we removed ratings produced by more casual listeners.

Software Demonstration

It is possible to experiment with the RS-viz tool at `http://datascience.ismb.it/rs-viz/`.

Conclusion and Future Work

We proposed a method for creating graphical summaries of any rating dataset for the purpose of enabling researchers and practitioners to better interpret the results of an offline evaluation campaign.

As future work, we would like to quantitatively characterize rating datasets according to different dimensions and place them in various categories, for example by analyzing the diversity of user preferences or the tendency to rate popular items only.

References

[1] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach.
Beyond accuracy: Evaluating recommender systems by coverage and serendipity.
In *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys '10, pages 257–260, New York, NY, USA, 2010. ACM.

[2] Alan Said and Alejandro Bellogín.
Comparative recommender system evaluation: Benchmarking recommendation frameworks.
In *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, pages 129–136, New York, NY, USA, 2014. ACM.

[3] Johannes Kunkel, Benedikt Loepp, and Jürgen Ziegler.
A 3d item space visualization for presenting and manipulating user preferences in collaborative filtering.
In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, IUI '17, pages 3–15, New York, NY, USA, 2017. ACM.