



Multicriteria Evaluation for Top-k and Sequence-based Recommender Systems

PhD Candidate:

Diego Michele MONTI

1. Context

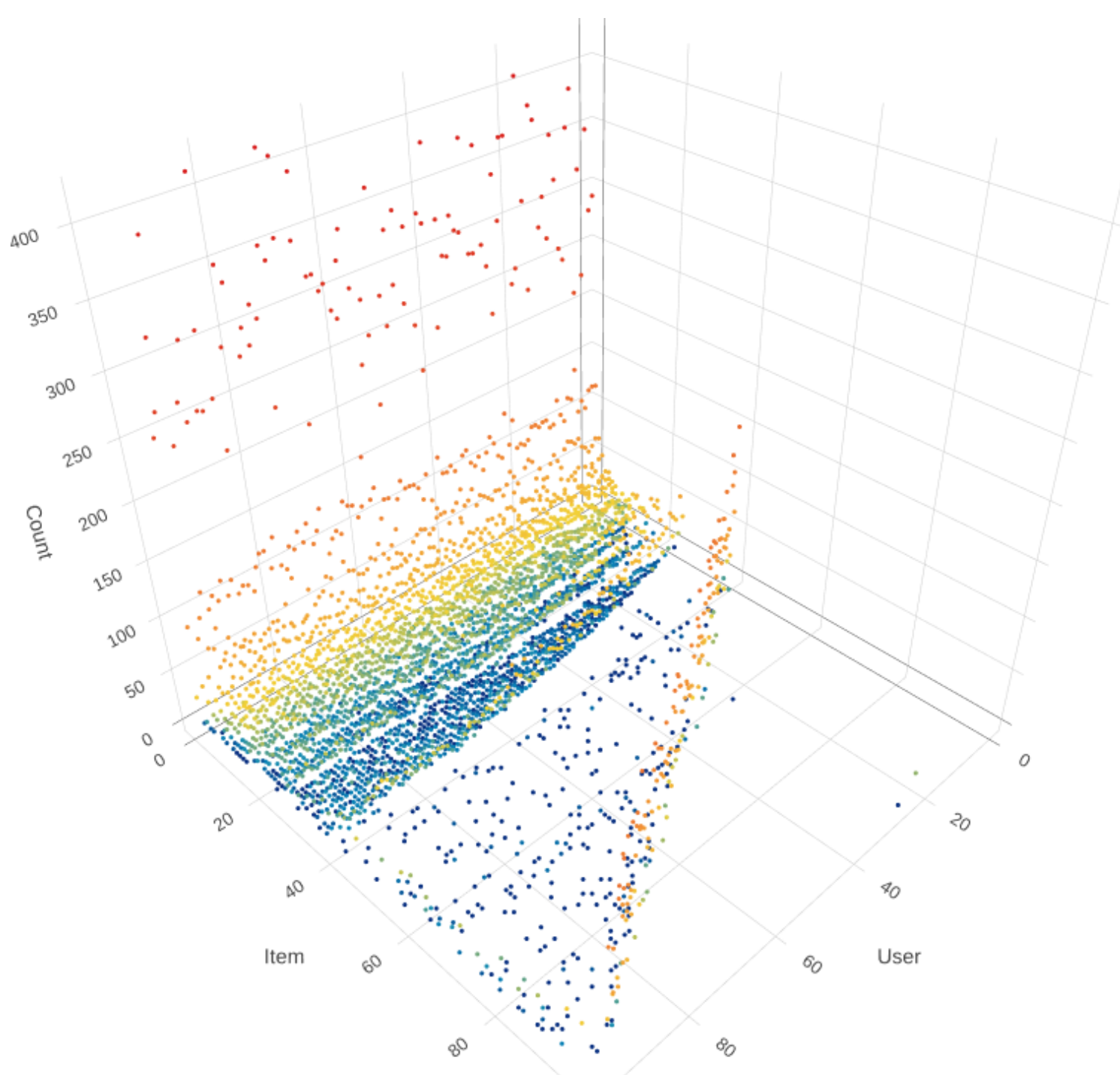
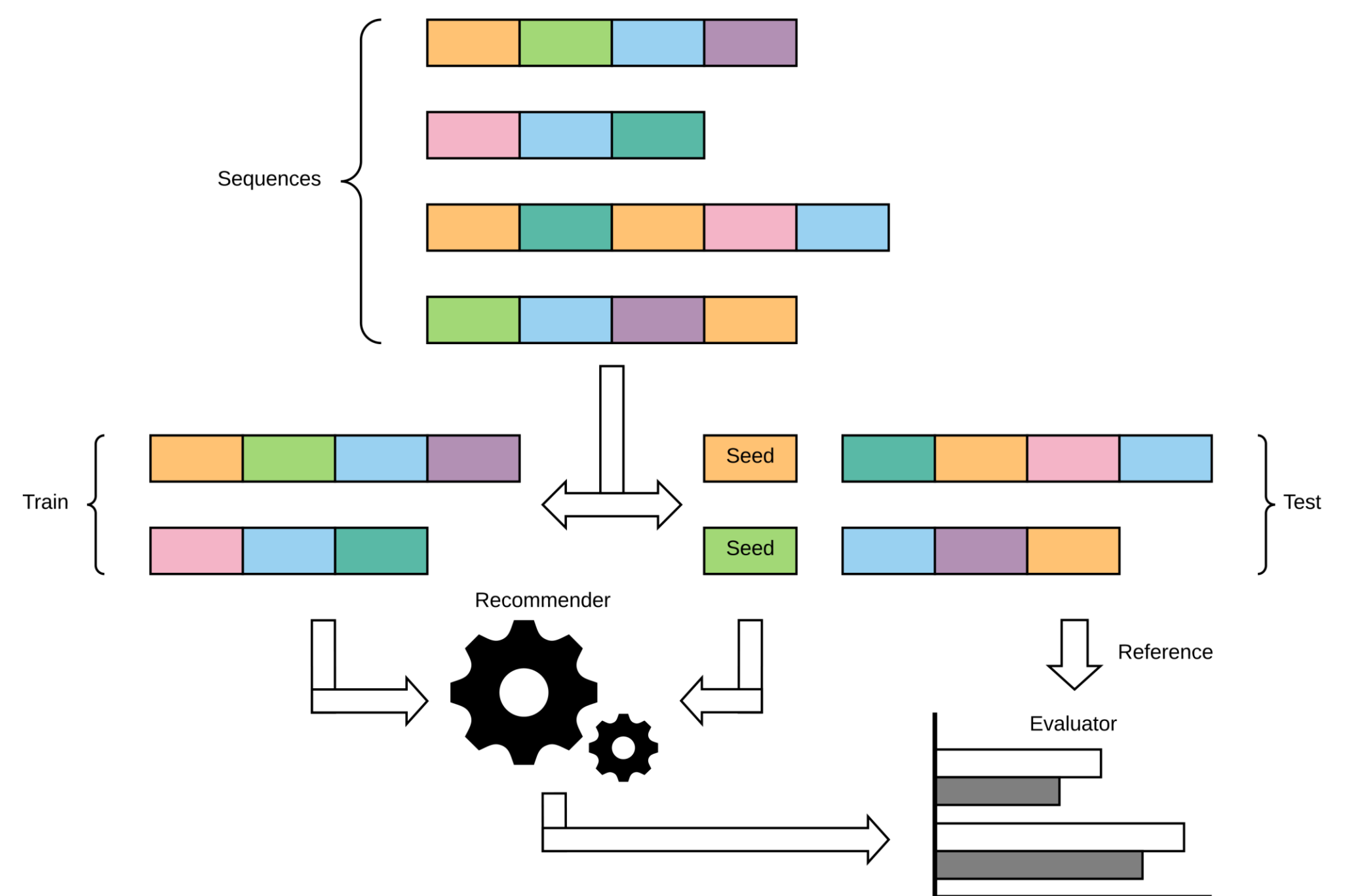
Due to the large variety of products and digital content available on the Web, an increasing number of people are interested in obtaining **personalized suggestions**. An automated tool capable of recommending items to users is defined as a **recommender system**. Nowadays, many recommender systems rely on **binary or implicit signals**. Furthermore, recommenders capable of suggesting temporally ordered sequences of items can be defined as **sequence-based**. In general, their **offline evaluation** is a challenging task, because the ground truth is always uncertain, as it is based on the subjective preferences of the users collected before the introduction of the system under evaluation.

2. Goals

1. What is the most suitable protocol for performing an offline evaluation of a top-k recommender system?
2. How already established metrics can be extended and adapted for evaluating a sequence-based recommender system?
3. To what extent the structure of a rating dataset can influence the performance of different recommenders?
4. Can the shortage of publicly available rating datasets be mitigated by a generative approach?

3. Approach

1. **RecLab**: a distributed evaluation framework for top-k recommender systems. The results of all experiments are permanently stored and publicly available.
2. **Sequeval**: an offline evaluation framework designed to compare any sequence-based recommender system using a multicriteria approach with eight metrics.
3. **RS-viz**: a data visualization technique for qualitatively comparing in an intuitive way different rating datasets.
4. **Synthetic datasets**: a clustering based method for creating new datasets starting from an existing one in a privacy aware way.

**Figure 1.** The LastFM dataset obtained with RS-viz.**Figure 2.** The Sequeval evaluation protocol.

4. Results

Metric	MP	Random	Unigram	Bigram	CRF	RNN
Coverage	0.0077	1.0000	0.9616	1.0000	0.9677	0.5069
Precision	0.2259	0.0080	0.0774	0.0607	0.0754	0.0962
nDPM	0.4998	0.5000	0.4994	0.4998	0.4993	0.4991
Diversity	0.9194	0.9971	0.9616	0.9777	0.9621	0.9469
Novelty	4.6056	12.300	7.1421	9.0216	7.3710	6.8374
Serendipity	0.0000	0.0060	0.0256	0.0230	0.0252	0.0365
Confidence	1.0000	0.0015	0.0171	0.0140	0.0179	0.0264
Perplexity	$+\infty$	651.00	141.41	122.99	147.49	140.39

Table 1. The results with Sequeval using the Foursquare dataset.

5. Conclusions

There is the need of adopting a multicriteria approach for performing the offline evaluation of a recommender system. In fact, different metrics can provide a more unbiased view of the results. Stronger emphasis should be put on the evaluation protocol, as it can have a dramatic impact on the results. Finally, it is necessary to know the main characteristics of the rating dataset exploited during an experiment in order to correctly interpret its outcome.

6. References

1. Monti, Diego; Rizzo, Giuseppe; Morisio, Maurizio. **A distributed and accountable approach to offline recommender systems evaluation**. In: 12th ACM Conference on Recommender Systems, Vancouver (CA), 2018.
2. Monti, Diego; Palumbo, Enrico; Rizzo, Giuseppe; Morisio, Maurizio. **Sequeval: An offline evaluation framework for sequence-based recommender systems**. In: Information, 2019, 10(5), 174.
3. Monti, Diego; Rizzo, Giuseppe; Morisio, Maurizio. **Visualizing ratings in recommender system datasets**. In: 13th ACM Conference on Recommender Systems, Copenhagen (DK), 2019.
4. Monti, Diego; Rizzo, Giuseppe; Morisio, Maurizio. **All you need is ratings: A clustering approach to synthetic rating datasets generation**. In: 13th ACM Conference on Recommender Systems, Copenhagen (DK), 2019.