

# Priors for Diversity and Novelty on Neural Recommender Systems <sup>†</sup>

Alfonso Landin <sup>\*</sup>, Daniel Valcarce , Javier Parapar  and Álvaro Barreiro 

Information Retrieval Lab, Centro de Investigación en Tecnoloxías da Información e as Comunicaciós (CITIC), Universidade da Coruña, Elviña, 15071 A Coruña, Spain

\* Correspondence: alfonso.landin@udc.es; Tel.: +34-881-01-1276

<sup>†</sup> Presented at the 2nd XoveTIC Conference, A Coruña, Spain, 5–6 September 2019.

Published: 31 July 2019



**Abstract:** PRIN is a neural based recommendation method that allows the incorporation of item prior information into the recommendation process. In this work we study how the system behaves in terms of novelty and diversity under different configurations of item prior probability estimations. Our results show the versatility of the framework and how its behavior can be adapted to the desired properties, whether accuracy is preferred or diversity and novelty are the desired properties, or how a balance can be achieved with the proper selection of prior estimations.

**Keywords:** recommender systems; neural models; item priors; diversity; novelty

## 1. Introduction

In recent years, there has been a shift in the way users interact with the information services, from a proactive approach, where users looked actively for content, to one where users take a more passive role and content is suggested to them. Recommender Systems have played a pivotal role in this transformation. The advances in this field have driven increases in user engagement and revenue.

Among the properties of a recommender, accuracy is usually the more desirable one. However, there are other properties, such as novelty and diversity [1], that are also important and that, depending on the task, can be of higher priority. Diversity is the ability of the system to produce recommendations that include items from the whole catalog. This property is usually desirable for the vendors [2,3]. Novelty is the ability of the system to produce unexpected recommendations and is associated with serendipity, a property associated with higher user engagement and satisfaction [4]. These properties are related to accuracy insofar that increasing the later lowers the best achievable numbers in those metrics [5].

Recently a neural based recommendation method was presented, PRIN [6], that allows the incorporation of prior knowledge about the items into the recommendation process. This previous work was focused on the accuracy of the system. In this work we focus on studying the behavior of the system in terms of diversity and novelty and how the choice of prior information affects the performance of the system in these metrics.

## 2. Materials and Methods

We conducted a series of experiments in order to analyze the trade-off between accuracy, diversity and novelty when choosing different configurations of item prior probabilities with the PRIN recommender framework.

### 2.1. PRIN and Prior Probabilities of Items

PRIN [6] is a recommendation framework composed of two components: a neural model trained to predict the probability of a user given an item,  $p(u|i)$ , and a graph model of the data where centrality measures are applied to calculate the prior probability distribution of the items,  $p(i)$ . The objective of the system is to produce a ranking of items for a user. We cannot use  $p(u|i)$  to produce such ranking as  $p(u|i)$  and  $p(u|j)$  for two items  $i$  and  $j$ ,  $i \neq j$ , are not comparable as they are in different event spaces. Bayes' rule can be used to obtain  $p(i|u)$ ,  $\forall i \in \mathcal{I}$ , which is used to produce a ranking.

Previously reported results focused on the accuracy of the system [6]. In this work we analyze the behavior of the system with respect to diversity and novelty. We take the neural model that gets the best results in accuracy and explore how the system responds with different prior probability estimations. This showcases how the same model can be reused depending on the particular needs without the need to train a new model. We report the results obtained when using indegree, PageRank [7], Katz's Index [8], eigenvector, HITS [9] and closeness [10] centrality measures. We also report the result obtained when using an uniform prior and the complement of indegree. Lastly, we report the results of PRIN's dual model, PRN [6], that is trained to estimate  $p(i|u)$ , which is used directly to produce a ranking.

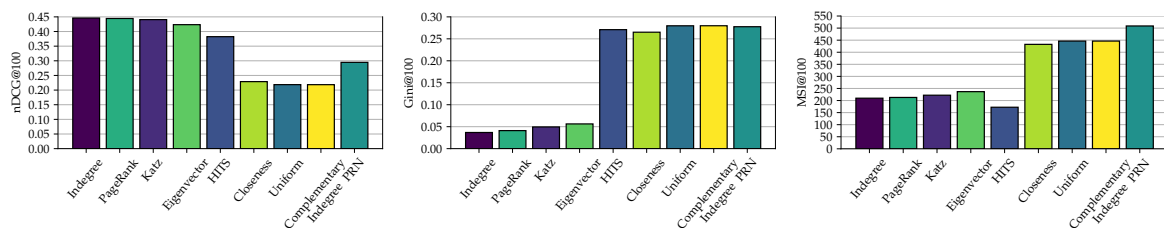
### 2.2. Evaluation Protocol

We report our results on the MovieLens 20M dataset, a popular public dataset. It contains 20,000,263 ratings that 138,493 users gave to 26,744 items. We split the data into training and test sets, with 80% of the ratings of each user in the train set and the remaining 20% in the test set.

To evaluate the accuracy of the recommendations we used the Normalized Discounted Cumulative Gain (nDCG), using the standard formulation as described in [11], with ratings as graded relevance judgments. We only considered items with a rating of 4.0 or more as relevant during evaluation. We used the inverse of the Gini index [2] to assess diversity and mean self-information (MSI) [12] to evaluate novelty. All metrics are evaluated at a cut-off of 100 because of its better robustness to sparsity and popularity biases and greater discriminative power with respect to shallower cut-offs [13].

## 3. Results

Results for all the centrality measures and other prior estimations, including also the results for PRN, can be seen in Figure 1. The systems are ordered by their accuracy result, except for PRN, that is the last bar in the graphs.



**Figure 1.** Values of nDCG@100, Gini@100 and MSI@100 for the MovieLens 20M dataset with different item prior configurations and for PRN.

We can see that as accuracy decreases we get better results in novelty and diversity. An exception to this in the HITS centrality measure, that has the worst results in novelty. Uniform prior and complementary indegree obtain similar results in all the metrics. This can be explained by the long tail distribution of the items that leads to similar estimations for most of the items.

The results of PRIN dual model, PRN, are also noteworthy. It obtains competitive results in diversity and has the best result in novelty while having better accuracy than the systems with similar behavior in those metrics.

#### 4. Discussion

We have shown how the properties of the recommendations produced by PRIN can be adjusted by the selection of the estimation of the prior probabilities of the items. This is done without the need to train the neural model with a different configuration of hyper-parameters. It should be possible to fine tune even more the trade-off between accuracy and novelty/diversity by tuning the hyper-parameters of the neural model independently, but this process is computationally intensive.

The behavior of the system when using HITS shows that while usually increases in diversity and novelty are paired, it is not always necessarily true. In this particular case we can see that the results in diversity are competitive with the best systems in this metric, while having greater accuracy. On the other hand, the results in terms of novelty are the worst of all the prior configurations.

**Funding:** This work was supported by project RTI2018-093336-B-C22 (MCIU & ERDF), project GPC ED431B 2019/03 (Xunta de Galicia & ERDF) and accreditation ED431G/01 (Xunta de Galicia & ERDF). The first and second authors also acknowledge the support of grants FPU17/03210 and FPU014/01724 (MCIU), respectively.

#### References

1. McNee, S.M.; Riedl, J.; Konstan, J.A. Being Accurate is Not Enough: How Accuracy Metrics have hurt Recommender Systems. In Proceedings of the CHI EA '06, CHI '06 Extended Abstracts on Human Factors in Computing Systems, Montréal, QC, Canada, 22–27 April 2006; ACM Press: New York, New York, USA, 2006; pp. 1097–1101, doi:10.1145/1125451.1125659.
2. Fleder, D.; Hosanagar, K. Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. *Manag. Sci.* **2009**, *55*, 697–712, doi:10.1287/mnsc.1080.0974.
3. Valcarce, D.; Parapar, J.; Álvaro Barreiro. Item-based relevance modelling of recommendations for getting rid of long tail products. *Knowl.-Based Syst.* **2016**, *103*, 41–51, doi:10.1016/j.knosys.2016.03.021.
4. Ge, M.; Delgado-Battenfeld, C.; Jannach, D. Beyond Accuracy: Evaluating Recommender Systems by Coverage and Serendipity. In Proceedings of the RecSys '10, Fourth ACM Conference on Recommender Systems, Barcelona, Spain, 26–30 September 2010; pp. 257–260, doi:10.1145/1864708.1864761.
5. Landin, A.; Suárez-García, E.; Valcarce, D. When Diversity Met Accuracy: A Story of Recommender Systems. *Proceedings* **2018**, *2*, 1178, doi:10.3390/proceedings2181178.
6. Landin, A.; Valcarce, D.; Parapar, J.; Álvaro Barreiro. PRIN: A Probabilistic Recommender with Item Priors and Neural Models. In Proceedings of the 41st European Conference on IR Research, ECIR 2019, Cologne, Germany, 14–18 April 2019; Part I, pp. 133–147, doi:10.1007/978-3-030-15712-8\_9.
7. Page, L.; Brin, S.; Motwani, R.; Winograd, T. *The PageRank Citation Ranking: Bringing Order to the Web*; Technical Report; Stanford InfoLab: Stanford, CA, USA, 1999.
8. Katz, L. A new status index derived from sociometric analysis. *Psychometrika* **1953**, *18*, 39–43.
9. Kleinberg, J.M. Authoritative Sources in a Hyperlinked Environment. *J. ACM* **1999**, *46*, 604–632.
10. Bavelas, A.; Barrett, D. *An Experimental Approach to Organizational Communication*; American Management Association: New York, NY, USA, 1951.
11. Wang, Y.; Wang, L.; Li, Y.; He, D.; Chen, W.; Liu, T.Y. A Theoretical Analysis of NDCG Ranking Measures. In Proceedings of the 26th Annual Conference on Learning Theory (COLT '13), Princeton, NJ, USA, 12–14 June 2013; pp. 1–30.
12. Zhou, T.; Kuscsik, Z.; Liu, J.G.; Medo, M.; Wakeling, J.R.; Zhang, Y.C. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4511–4515, doi:10.1073/pnas.1000488107.
13. Valcarce, D.; Bellogín, A.; Parapar, J.; Castells, P. On the robustness and discriminative power of information retrieval metrics for top-N recommendation. In Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18); Vancouver, BC, Canada, 2–7 October 2018; ACM: New York, NY, USA, 2018; pp. 260–268, doi:10.1145/3240323.3240347.

