Multiobjective Evaluation of Reinforcement Learning Based Recommender Systems

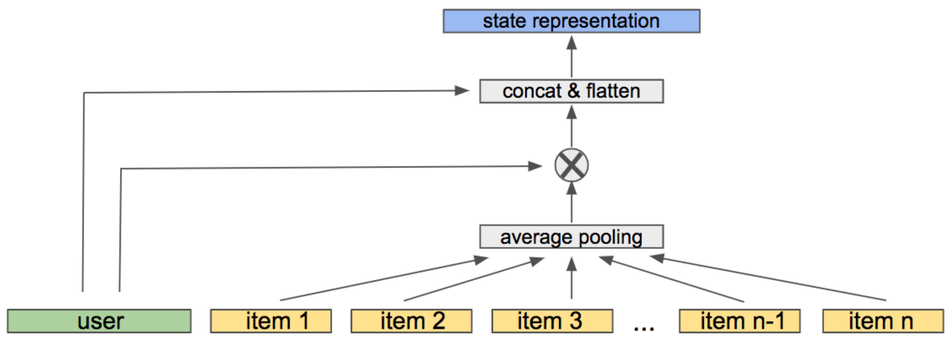
Alexey Grishanov[1] (grishanov@phystech.edu), **Anastasia Ianina**[2] (yanina@phystech.edu) and **Konstantin Vorontsov**[2] (vokov@forecsys.ru)
[1] AI Sber Lab, [2] Moscow Institute of Physics and Technology

Abstract

Movielens dataset has become a default choice for recommender systems evaluation. In this paper we analyze the best strategies of a Reinforcement Learning agent on Movielens (1M) dataset studying the balance between precision and diversity of recommendations. We found that trivial strategies are able to maximize ranking quality criteria, but useless for users of the recommendation system due to the lack of diversity in final predictions. Our proposed method stimulates the agent to explore the environment using the stochasticity of Ornstein-Uhlenbeck processes. Experiments show that optimization of the Ornstein-Uhlenbeck process drift coefficient improves the diversity of recommendations while maintaining high nDCG and HR criteria. To the best of our knowledge, the analysis of agent strategies in recommendation environments has not been studied excessively in previous works.

Recommender Environment

State We follow the state representation scheme from [3], where $s \in \mathcal{S}$ is a user's current preference which is expressed by concatenated products of user vector u with item vectors i of n previously viewed relevant items: $s = [u, u \otimes \{w_l i_l \mid l = 1, ..., n\}, \{w_l i_l \mid l = 1, \ldots, n\}] \in \mathbb{R}^{3k}$, where w_l represents importance of object i_l.
Actions. We specify the agent's actions using the $a \in \mathbb{R}^k$ vector, following [1].
Rewards. $r_t = \{1, \text{if } r_{ui} > 3; 0, \text{otherwise}\}$, where $r_t \in \mathcal{R}$, $r_{ui} \in R$ is rating that user u assigned to item i.



Ornstein-Uhlenbeck process

Ornstein-Uhlenbeck process O_t can be described with the stochastic differential equation:

$$dO_t = \theta(\mu - O_t)\, dt + \sigma dW_t,$$

where W_t is a Wiener process [4], $\theta > 0$, $\sigma > 0$ and $\mu \in \mathbb{R}$.

In this equation parameter μ describes the equilibrium state, σ is the variance caused by Brownian particles collisions, and θ controls "gravitation" to the initial state. Further in the experiment, we fix $\mu = 0$ and vary parameters θ and σ, maximizing ranking quality criteria and diversity of recommendations.

Main contributions

1. We leverage **DDPG with Ornstein-Uhlenbeck noise** for solving recommendation problem and analyze its strategies on **MovieLens-1M dataset**.
2. We indicate that Movielens dataset contains trivial strategies allowing to maximize relevance metrics with minimal diversity.
3. We propose **Diverse DDPG (D3PG) noise injection strategy** to find the right balance between exploration and exploitation.
4. We discuss different noise injection schemes and show that optimizing OU process drift coefficient allows an agent to discover both diverse and relevant strategies, significantly decreasing percentage of degraded strategies with high nDCG@10 and HR@10, yet useless for users (due to almost identical predictions).

Evaluation and Experiments

- **HitRate@k (HR@k).**
  $HR@k = \sum_{i=1}^{k} rel_i$, where $rel_i \in \{0, 1\}$, $rel_i = 1$ if $r_{ui} > 3, 0$ otherwise;
- **Normalized Discounted Cumulative Gain (nDCG@k).**
  $DCG@k = \sum_{i=1}^{k} \frac{2^{rel_i}-1}{\log_2 i+1}$.
  $nDCG@k = \frac{DCG@k}{IDCG@k}$, where IDCG is ideal discounted cumulative gain.
- **Coverage@p** $= \frac{\left|\bigcup_{u \in U} y_u\right|}{|I|}$ Neither HR nor nDCG metric directly evaluates variety of recommendations. For evaluating diversity numerically we used $Coverage@p$.

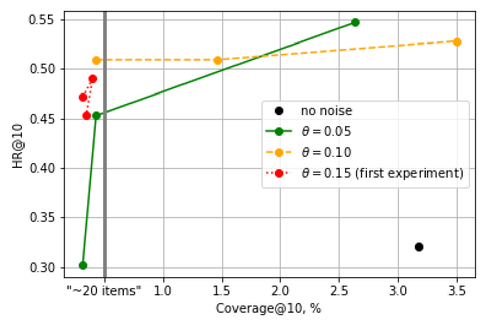
Model	nDCG@10	HR@10	Coverage@10, %
RandomRec	~ 0.05	~ 0.1	~ 100
PopRec	~ 0.2	~ 0.45	~ 0
NCF (non-RL) [2]	0.238	0.460	**6.1**
Our approach (trained with DDPG)			
no noise	0.181	0.320	3.1
noise in parameters	0.086	0.394	3.3
$\theta = 0.15,\ \sigma = 0.2$	0.251	0.472	0.3
$\theta = 0.15,\ \sigma = 0.4$	0.261	0.490	0.4
$\theta = 0.15,\ \sigma = 0.6$	0.261	0.453	0.3
$\boldsymbol{\theta = 0.10, \sigma = 0.2}$	**0.286**	0.528	**3.5**
$\theta = 0.10,\ \sigma = 0.4$	0.268	0.514	0.4
$\theta = 0.10,\ \sigma = 0.6$	0.280	0.509	1.4
$\boldsymbol{\theta = 0.05, \sigma = 0.2}$	0.281	**0.547**	**2.7**
$\theta = 0.05,\ \sigma = 0.4$	0.223	0.453	0.4
$\theta = 0.05,\ \sigma = 0.6$	0.262	0.302	0.4

Table 1: Comparison in terms of ranking quality (nDCG@10 and HR@10) and diversity (Coverage@10) between recommendations with no noise, different noise injection schemes (OU with fixed/varying θ and different σ), and non-RL baseline (NCF [2])

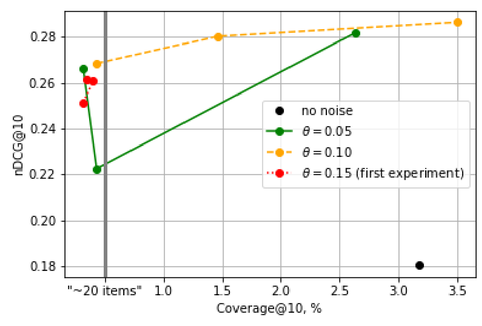
OU noise stimulates agent's explorative behaviour better resulting in substantial growth of HR@k and nDCG@k metrics. However, many strategies with OU noise degenerate to PopRec-like strategies (Coverage@10 < 0.5). The possible explanation is that the inertia of the OU process is too high for a recommendation problem.

Enforcing Diversity of Recommendations

To further enforce diversity of recommendations and decrease the inertia of the OU process, we propose to fine-tune θ coefficient which expresses the "gravitation" to the initial state. After it recommendations became much more diverse approaching high Coverage@10 values while retaining good ranking quality in terms of nDCG@10 and HR@10 (**yellow** and **green** lines in Pareto fronts).

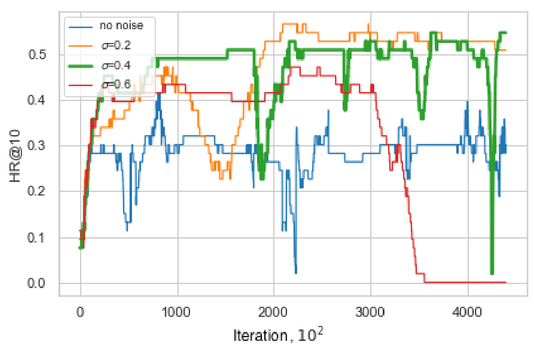


Pareto front Coverage@10 and nDCG@10

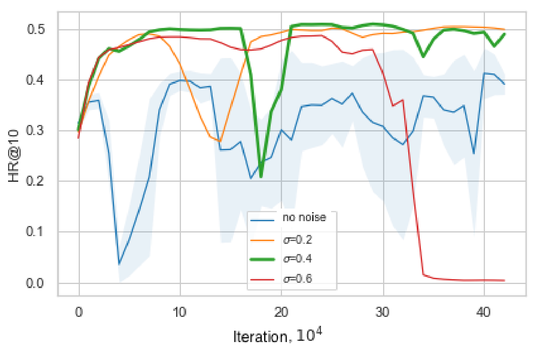


Pareto front Coverage@10 and HR@10

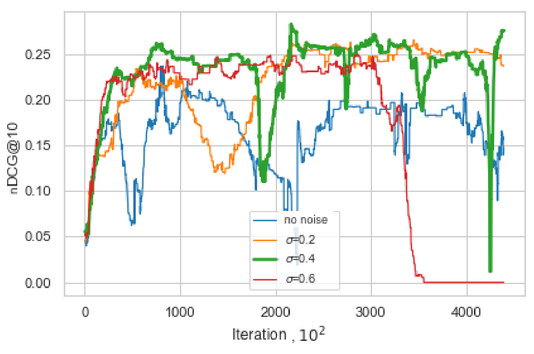
Learning Curves



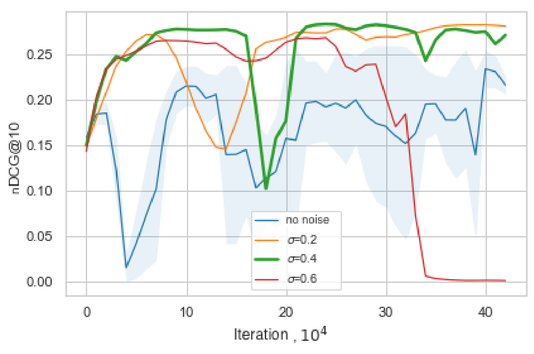
HR@10, randomly selected user



HR@10, averaged over all the users



nDCG@10, randomly selected user



nDCG@10, averaged over all the users

References

[1] Gabriel Dulac-Arnold, Richard Evans, Hado van Hasselt, Peter Sunehag, Timothy Lillicrap, Jonathan Hunt, Timothy Mann, Theophane Weber, Thomas Degris, and Ben Coppin. Deep reinforcement learning in large discrete action spaces, 2015.

[2] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In Rick Barrett, Rick Cummings, Eugene Agichtein, and Evgeniy Gabrilovich, editors, *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, pages 173–182. ACM, 2017.

[3] Feng Liu, Ruiming Tang, Xutao Li, Weinan Zhang, Yunming Ye, Haokun Chen, Huifeng Guo, and Yuzhou Zhang. Deep reinforcement learning based recommendation with explicit user-item interactions modeling. *arXiv preprint arXiv:1810.12027*, 2018.

[4] Norbert Wiener. *Collected works with commentaries*. Mit Press, 1976.