

# Towards Representation Alignment and Uniformity in Collaborative Filtering

title: Metadata

- \*\*CiteKey\*\*: DirectAU\_KDD2022
- \*\*Type\*\*: preprint
- \*\*Author\*\*: Wang, Chenyang; Yu, Yuanqing; Ma, Weizhi; Zhang, Min; Chen, Chong; Liu, Yiqun; Ma, Shaoping
- \*\*Editor\*\*: {{editor}}
- \*\*Translator\*\*: {{translator}}
- \*\*Publisher\*\*: {{publisher}}
- \*\*Location\*\*: {{place}}
- \*\*Series\*\*: {{series}}
- \*\*Series Number\*\*: {{seriesNumber}}
- \*\*Journal\*\*: {{publicationTitle}}
- \*\*Volume\*\*: {{volume}}
- \*\*Issue\*\*: {{issue}}
- \*\*Pages\*\*: {{pages}}
- \*\*Year\*\*: 2022
- \*\*DOI\*\*: 10.1145/3534678.3539253
- \*\*ISSN\*\*: {{ISSN}}
- \*\*ISBN\*\*: {{ISBN}}

[language-ad-info](#)

title: Abstract

Collaborative filtering (CF) plays a critical role in the development of recommender systems. Most CF methods utilize an encoder to embed users and items into the same representation space, and the Bayesian personalized ranking (BPR) loss is usually adopted as the objective function to learn informative encoders. Existing studies mainly focus on designing more powerful encoders (e.g., graph neural network) to learn better representations. However, few efforts have been devoted to investigating the desired properties of representations in CF, which is important to understand the rationale of existing CF methods and design new learning objectives. In this paper, we measure the representation quality in CF from the perspective of alignment and uniformity on the hypersphere. We first theoretically reveal the connection between the BPR loss and these two properties. Then, we empirically analyze the learning dynamics of typical CF methods in terms of quantified alignment and uniformity, which shows that better alignment or uniformity both contribute to higher recommendation performance. Based on the analyses results, a learning objective that directly optimizes these two properties is proposed, named DirectAU. We conduct extensive experiments on three public datasets, and the proposed learning framework with a simple matrix factorization model leads to significant performance improvements compared to state-of-the-art

CF methods. Our implementations are publicly available at <https://github.com/THUwangcy/DirectAU>.

language-ad-quote

```
title: Files and Links
- **Url**: http://arxiv.org/abs/2206.12811
- **Uri**: http://zotero.org/users/6550764/items/LQ8NPAG6
- **Eprint**: {{eprint}}
- **File**: [arXiv.org Snapshot]
(file://C:\Users\qyzl7\Zotero\storage\T8E2JQJZ\2206.html); [Wang et
al_2022_Towards Representation Alignment and Uniformity in Collaborative
Filtering.pdf]
(file://C:\Users\qyzl7\Zotero\storage\K3Y6SNWD\Wang%20et%20al_2022_Towards%20
- **Local Library**: [Zotero](zotero://select/library/items/LQ8NPAG6)
```

language-ad-abstract

```
title: Tags and Collections
- **Keywords**: Computer Science - Information Retrieval
- **Collections**: 21.图神经网络; 211.图神经网络理论
```

language-ad-note

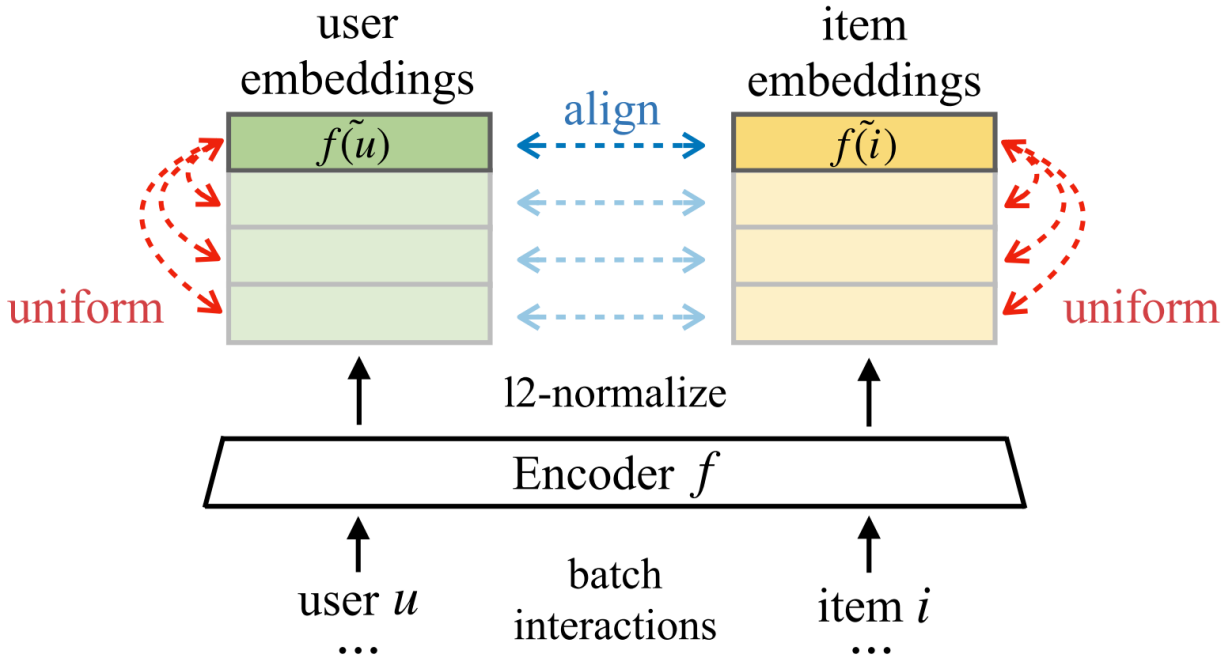
## Comments

本文主要针对协同过滤方法中的数据表征展开研究，现有的研究主要集中在设计更强大的编码器（例如，图神经网络）以学习更好的表征。而缺乏对 CF 中表征的期望属性进行研究，这对于理解现有 CF 方法的基本原理和设计新的学习目标很重要。本文从一致性和超球面上的均匀性的角度来衡量 CF 中的表征质量。

- 从理论上揭示了 BPR 损失与一致性和均匀性之间的联系。
- 从量化一致性和均匀性的角度对典型 CF 方法的学习动态进行了分析，表明更好的一致性|性或均匀性都有助于提高推荐性能。
- 根据分析结果，提出了一个直接优化这两个属性的学习目标，称为 DirectAU。

## Extracted Annotations

### 一致性与均匀性



给定数据分布  $p_{data}(\cdot)$  和正样本对  $p_{pos}(\cdot)$  的分布，一致性损失定义为用户  $u$  与商品  $i$  正样本对的标准化 embedding 之间的距离的期望，公式如下，其中  $\tilde{f}(\cdot)$  为 L2 标准化表征

$$l_{align} \triangleq \mathbb{E}_{(u,i) \sim p_{pos}} \|\tilde{f}(u) - \tilde{f}(i)\|^2$$

均匀性损失为用户-用户间，商品-商品间成对高斯函数的均值的对数，公式如下，

$$l_{uniform} = \log \mathbb{E}_{u, u' \sim p_{user}} e^{-2\|\tilde{f}(u) - \tilde{f}(u')\|^2} + \log \mathbb{E}_{i, i' \sim p_{item}} e^{-2\|\tilde{f}(i) - \tilde{f}(i')\|^2}$$

一致性损失提高了正相关用户-项目对的 embedding 之间的相似性，而均匀性损失衡量了 embedding 在超球面上的分散程度。这两个指标与表征学习的目标非常一致：**正实例应该彼此靠近，而随机实例应该分散在超球面上。**

## 当且仅当满足一致性与均匀性时 BPR 损失最小

$$\begin{aligned} \mathcal{L}_{BPR} &= \mathbb{E}_{(u,i) \sim p_{pos}} -\log \text{sigmoid}(s(u, i) - s(u, i^-)) \\ &= \mathbb{E}_{(u,i) \sim p_{pos}} -\log \left( \frac{e^{f(u)^T \tilde{f}(i)}}{e^{\tilde{f}(u)^T \tilde{f}(i)} + e^{\tilde{f}(u)^T \tilde{f}(i^-)}} \right) \\ &= \mathbb{E}_{(u,i) \sim p_{pos}} -\tilde{f}(u)^T \tilde{f}(i) + \log \left( e^{\tilde{f}(u)^T \tilde{f}(i)} + e^{\tilde{f}(u)^T \tilde{f}(i^-)} \right) \end{aligned}$$

当嵌入满足完美一致性时，即  $\tilde{f}(u) = \tilde{f}(i)$ ，可以有

$$\mathcal{L}_{BPR} \geq \mathbb{E}_{(u,i) \sim p_{pos}} \left[ -1 + \log \left( e^1 + e^{\tilde{f}(u)^T \tilde{f}(i^-)} \right) \right]$$

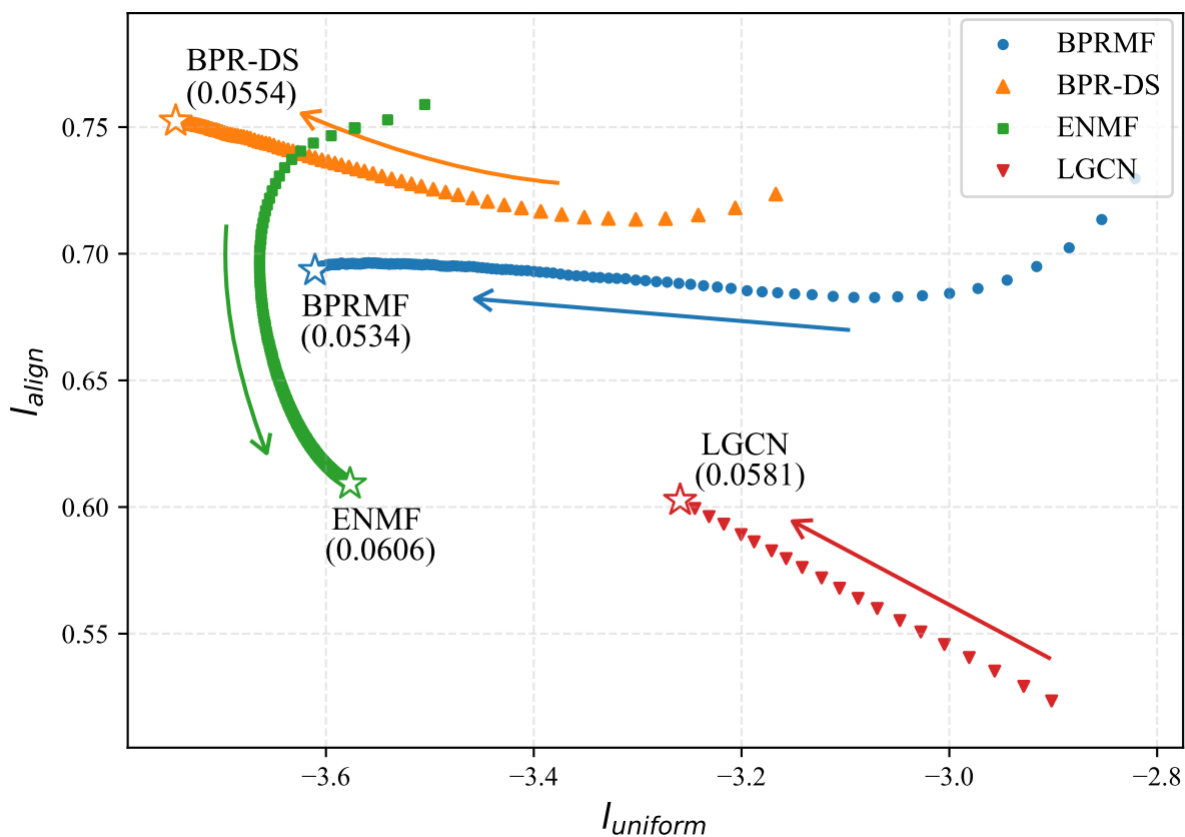
进一步地，当嵌入满足完美均匀性时，即  $f$  诱导出的特征分布服从  $\sigma_{d-1}$  时，有

$$\mathcal{L}_{BPR} \geq -1 + \int_{\mathcal{S}^{d-1}} \int_{\mathcal{S}^{d-1}} \log(e + e^{x^T y}) d\sigma_{d-1}(x) d\sigma_{d-1}(y)$$

可知， $\mathcal{L}_{BPR}$  大于等于一个独立于  $f$  的常数，当且仅当  $f$  完全满足一致性且均匀时，可以相等。

## 各模型在一致性与均匀性上的表现

为了验证BPR损失以及相关方法在优化的过程中会优化一致性和均匀性，作者在不同的方法上进行了实验，随着优化的不断进行，一致性和均匀性也在发生着变化。从图中可以发现，随着优化的进行，一致性和均匀性会得到相应的优化和改善。这也说明 CF 中的用户和商品表征确实偏爱这两个属性。实现更好的一致性或均匀性都有助于提高推荐性能，同时优化它们可能是有益的。



- BPR 的优化更侧重于一致性（区分正负交互），但不会不断地将正用户-项目对推得更近。
- BPR-DS 对更难的负样本进行采样，从而导致更低的均匀性损失和更好的性能。但是难负样本也使得对齐正用户项对变得困难（更高的一致性损失）。
- LGCN 聚合邻域信息，因此即使在开始时也能实现出色的一致性。但可见训练过程伤害了一致性，导致在一致性上相交 BPRMF 略差。但最终的一致性仍然显著，这相应地导致了更好的性能
- ENMF 直接优化 MSE 将正用户-项目对的分数推至 1，而不是使用 BPR 时的仅仅大于负对。这种基于整体数据的训练在很大程度上有利于一致性的优化，同时保持有希望的均匀性，从而产生卓越的性能。但是可见后期训练，由于逐点优化也损害了均匀性。

## 实验

Method	Beauty		Gowalla	
	Recall@20	NDCG@20	Recall@20	NDCG@20
BPRMF	0.1153	0.0534	0.1263	0.0736
+DirectAU	<b>0.1400**</b>	<b>0.0686**</b>	<b>0.2014**</b>	<b>0.1170**</b>
LGCN-1	0.1211	0.0560	0.1769	0.1033
+DirectAU	<b>0.1444**</b>	<b>0.0700**</b>	<b>0.2036**</b>	<b>0.1184**</b>
LGCN-2	0.1201	0.0581	0.1871	0.1097
+DirectAU	<b>0.1455**</b>	<b>0.0707**</b>	<b>0.2043**</b>	<b>0.1191**</b>

注意到 BPRMF 在使用了本文提到的损失之后取得了与 LightGCN 相似的性能。

Setting		Baseline Methods								Ours	
Dataset	Metric	BPRMF	BPR-DS	ENMF	RecVAE	LGCN	DGCF	BUIR	CLRec	DirectAU	Improv.
Beauty	Recall@10	0.0806	0.0816	0.0915	0.0824	0.0863	0.0897	0.0816	<u>0.0937</u>	<b>0.1002**</b>	6.94%
	Recall@20	0.1153	0.1181	0.1282	0.1145	0.1201	0.1283	0.1204	<u>0.1337</u>	<b>0.1400**</b>	4.74%
	Recall@50	0.1763	0.1745	0.1914	0.1712	0.1819	0.1958	0.1866	<u>0.1996</u>	<b>0.2062**</b>	3.33%
	NDCG@10	0.0444	0.0459	0.0511	0.0486	0.0484	0.0501	0.0457	<u>0.0547</u>	<b>0.0582**</b>	6.44%
	NDCG@20	0.0534	0.0554	0.0606	0.0570	0.0581	0.0600	0.0556	<u>0.0651</u>	<b>0.0686**</b>	5.38%
	NDCG@50	0.0658	0.0670	0.0736	0.0686	0.0699	0.0738	0.0692	<u>0.0786</u>	<b>0.0820**</b>	4.33%
Gowalla	Recall@10	0.0866	0.1132	0.1149	0.1211	0.1289	<u>0.1301</u>	0.0798	0.1215	<b>0.1394**</b>	7.15%
	Recall@20	0.1263	0.1637	0.1671	0.1771	0.1871	<u>0.1889</u>	0.1164	0.1755	<b>0.2014**</b>	6.63%
	Recall@50	0.2040	0.2593	0.2675	0.2768	<u>0.2934</u>	0.2919	0.1917	0.2813	<b>0.3127**</b>	6.56%
	NDCG@10	0.0622	0.0814	0.0797	0.0845	0.0930	<u>0.0939</u>	0.0570	0.0868	<b>0.0991**</b>	5.56%
	NDCG@20	0.0736	0.0961	0.0953	0.1007	0.1097	<u>0.1099</u>	0.0676	0.1022	<b>0.1170**</b>	6.44%
	NDCG@50	0.0926	0.1196	0.1200	0.1251	0.1356	<u>0.1358</u>	0.0858	0.1281	<b>0.1442**</b>	6.20%
Yelp2018	Recall@10	0.0416	0.0533	<u>0.0596</u>	0.0495	0.0508	0.0519	0.0444	0.0547	<b>0.0684**</b>	14.83%
	Recall@20	0.0693	0.0864	<u>0.0957</u>	0.0820	0.0833	0.0849	0.0737	0.0890	<b>0.1096**</b>	14.55%
	Recall@50	0.1293	0.1572	<u>0.1710</u>	0.1494	0.1534	0.1575	0.1386	0.1606	<b>0.1935**</b>	13.16%
	NDCG@10	0.0335	0.0423	<u>0.0482</u>	0.0395	0.0406	0.0409	0.0349	0.0436	<b>0.0553**</b>	14.77%
	NDCG@20	0.0428	0.0534	<u>0.0603</u>	0.0504	0.0514	0.0521	0.0448	0.0551	<b>0.0691**</b>	14.53%
	NDCG@50	0.0602	0.0740	<u>0.0821</u>	0.0698	0.0717	0.0732	0.0636	0.0758	<b>0.0933**</b>	13.67%

此处的 DirectAU 仅使用了 MF 作为特征提取器。