Leveraging Uniformity of Normalized Embeddings for Sequential Recommendation

Hyunsoo Chung and Jungtaek Kim

OMNIOUS.AI University of Pittsburgh

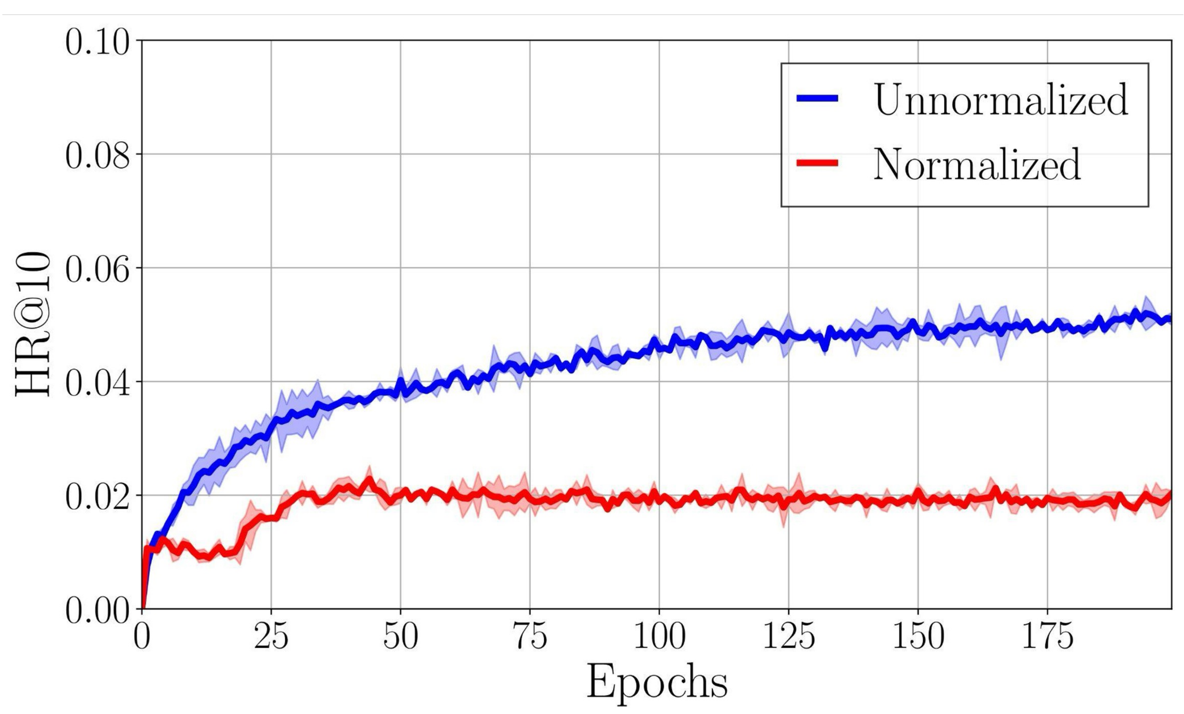
NEURAL INFORMATION PROCESSING SYSTEMS

Motivation

- Given a previous interaction **history** of a user, a sequential recommendation task aims to learn a strategy that recommends the most relevant **next item (highest score)** to the user.
- Despite recent successes by advanced architectures such as Transformers and GNNs, most studies focus on learning a policy with **unnormalized embeddings**.
- Inspired by the **uniformity** metric, we propose a new training objective that enhances recommendation performance with each history and item embeddings normalized.

Performance Drop with Normalization

Comparison between normalized & unnormalized embeddings



Architecture and training loss are fixed between two approaches.

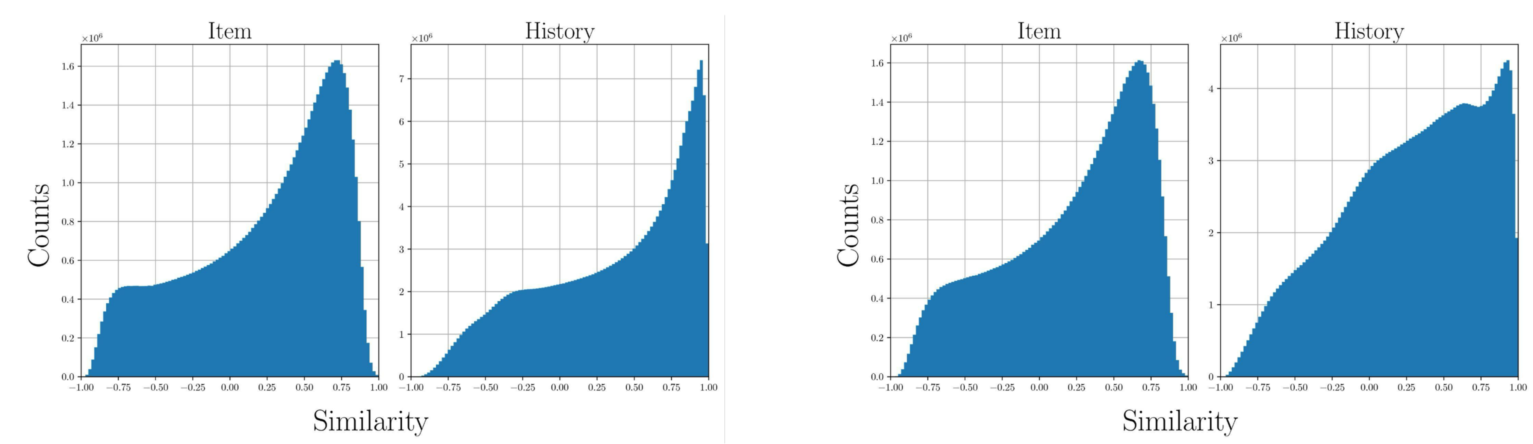
- Given two latent representations of a history and an item, we normalize each vectors such that they reside on the unit hypersphere.

$$\bar{h}'_u = \frac{h'_u}{\|h'_u\|_2} \quad \text{and} \quad \bar{i}'_k = \frac{i'_k}{\|i'_k\|_2}$$

- We then compare with unnormalized embeddings with other components such as the model pipeline, training objective, and hyperparameters fixed.
- Consequently, we see severe performance drop of normalized embeddings compared to the original approach.

Embedding Collapse

(Left) early stage of training, (Right) late stage of training



- We calculate pairwise cosine similarities of normalized history and item embeddings throughout the training.
- We observe that both history and item embeddings are extremely **clusterized** while training..

Uniformity-Inspired Objective

To alleviate the clusterization of embeddings, we introduce a new training objective that evenly distributes embeddings on the unit hypersphere while learning effective recommendation policy.

- Similar to the uniformity metric, we define a homogeneous term as the sum of pairwise Gaussian potentials for each history and item embedding sets.

$$\mathcal{L}_{\text{hom}} = \sum_{x,y \in \mathcal{D}_H} e^{-t\|\bar{h}'_x - \bar{h}'_y\|_2^2} + \sum_{x,y \in \mathcal{D}_I} e^{-t\|\bar{i}'_x - \bar{i}'_y\|_2^2}$$

- We then minimize a heterogeneous term which is a sum of Gaussian potential between each history and item embeddings.

$$\mathcal{L}_{\text{het}} = \sum_{x \in \mathcal{D}_H, y \in \mathcal{D}_J} e^{-t\|\bar{h}'_x - \bar{i}'_y\|_2^2}$$

- For the recommendation loss, we adopt the original pointwise (BCE) term.

$$\mathcal{L}_{\text{rec}}(\mathcal{D};\boldsymbol{\theta}) = -\sum_{(u,i,j)\in\mathcal{D}} \log \sigma(\hat{s}_{ui};\boldsymbol{\theta}) + \log(1-\sigma(\hat{s}_{uj};\boldsymbol{\theta}))$$

- Resulting loss are weighted sum between the three aforementioned terms with hyperparameters for balancing purposes.

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \beta_1 \mathcal{L}_{\text{hom}} + \beta_2 \mathcal{L}_{\text{het}}$$

- For convenience, we fix two hyperparameters to the same value.

Experimental Results

Recommendation performance comparison

Model	Method	Beauty HR	Beauty NDCG	Toys HR	Toys NDCG	Sports HR	Sports NDCG	Yelp HR	Yelp NDCG
GRU4Rec	BCE	0.0369	0.0185	0.0298	0.0164	0.0148	0.0076	0.0273	0.0136
	BPR	0.0472	0.0248	0.0492	0.0276	0.0283	0.0153	0.0368	0.0190
	InfoNCE	0.0365	0.0169	0.0282	0.0141	0.0314	0.0153	0.0482	0.0241
	Our Loss	**0.0702**	**0.0373**	**0.0705**	**0.0385**	**0.0365**	**0.0193**	**0.0665**	**0.0378**
Caser	BCE	0.0348	0.0172	0.0250	0.0125	0.0218	0.0110	0.0251	0.0121
	BPR	0.0332	0.0159	0.0298	0.0140	0.0165	0.0083	0.0644	0.0342
	InfoNCE	0.0421	0.0172	0.0371	0.0149	0.0245	0.0107	0.0643	0.0333
	Our Loss	**0.0544**	**0.0258**	**0.0416**	**0.0174**	**0.0261**	**0.0116**	**0.0658**	**0.0357**
SASRec	BCE	0.0522	0.0278	0.0604	0.0295	0.0301	0.0145	0.0507	0.0278
	BPR	0.0594	0.0261	0.0662	0.0309	0.0337	0.0150	0.0552	0.0326
	InfoNCE	0.0588	0.0261	0.0677	0.0305	0.0367	0.0165	0.0593	0.0334
	Our Loss	**0.0821**	**0.0371**	**0.0896**	**0.0411**	**0.0471**	**0.0214**	**0.0668**	**0.0405**

Bold metrics are best performing methods in each configuration.

Loss component analysis

Reference	Normalization	$\mathcal{L}_{\text{rec}}$	$\mathcal{L}_{\text{hom}}$	$\mathcal{L}_{\text{het}}$	Beauty HR	Beauty NDCG	Toys HR	Toys NDCG
(a)		✓			0.0522	0.0278	0.0604	0.0295
(b)	✓	✓			0.0233	0.0113	0.0181	0.0093
(c)	✓	✓	✓		0.0786	0.0362	0.0865	0.0394
(d)	✓	✓		✓	0.0416	0.0205	0.0399	0.0204
(e)	✓	✓	✓	✓	0.0821	0.0371	0.0896	0.0411
(f)	✓		✓	✓	0.0028	0.0015	0.0023	0.0011
(g)		✓	✓	✓	0.0534	0.0257	0.0614	0.0301

Effects of the regularization hyperparameter

β	Beauty HR	Beauty NDCG	Toys HR	Toys NDCG
0.1	0.0785	0.0358	0.0854	0.0389
0.05	0.0821	0.0371	0.0896	0.0411
0.01	0.0658	0.0314	0.0708	0.0331
0.005	0.0557	0.0268	0.0582	0.0280
0.001	0.0353	0.0168	0.0320	0.0158

Acknowledgement

HC acknowledges and thanks the OMNIOUS.AI team for their support and helpful discussions. JK acknowledges support from the MDS-Rely Center, which is funded by the National Science Foundation's Industry-University Cooperative Research Center (IUCRC) program through awards EEC-2052662 and EEC-2052776.