

NISER: Normalized Item and Session Representation to Handle Popularity Bias_

논문 링크 : <https://arxiv.org/pdf/1909.04276.pdf>

ABSTRACT

- session-based recommendation (SR) 모델
 - 유저가 다음에 클릭할 아이템을 추천하기 위해 과거 행동(아이템/상품 클릭)으로부터의 정보를 활용하는 것이 목표
- GNN
 - 복잡한 아이템 변화를 설명하기 위해 세션에서의 아이템 상호작용의 시퀀스는 그래프화된 데이터로 모델링 될 수 있음.
 - 세션 그래프를 효과적으로 학습할 수 있고, RNN과 같은 시퀀셜 모델로 개선을 보임.
 - popularity bias (모델이 인기있는 아이템을 추천하는 쪽으로 편향되는 것) → 롱테일 아이템(인기가 덜하거나 덜 빈도를 갖는 아이템)에 대해서 추천하는 것에 실패
 - 온라인 세팅(매일 새로운 아이템이 도착하는)에서는 이런 모델들이 좋지 않은 성능

위 문제가 학습된 아이템과 세션 그래프의 표현(임베딩 벡터)의 크기(norm)에 관련 있음을 설명

표현을 정규화함으로써 이러한 문제를 완화시키는 학습 절차를 제안

정규화된 아이템/세션 그래프 표현을 사용하는 것이 모델이 더 좋은 성능을 보인다.

1. 오프라인 세팅의 덜 인기있는 롱테일 아이템
2. 온라인 세팅의 덜 인기있는 새로 도입된 아이템

INTRODUCTION

SR 모델의 목표

- 전에 클릭했던 아이템의 시퀀스를 기반으로 top-K 아이템을 유저에게 추천하는 것

딥러닝 모델에서는 아이템이나 세션에 대한 표현을 학습하는 능력이 중요

- RNN(시퀀셜 모델)
 - 입력 : 과거 세션에서 클릭했던 아이템
 - 타겟 : 아이템 목록안에서 사용가능한 아이템
- STAMP
 - 세션을 한 아이템 세트로 고려
 - 어텐션 모델 사용
- NARM, CSRM 접근 법은 시퀀셜과 어텐션 모델의 조합을 사용
- SR-GNN
 - 세션을 그래프 구조로 제안
 - 아이템 간의 복잡한 to-and-from을 설명
 - 아이템 - 노드, 순서 - 엣지로 설정
 - 각 노드는 이웃의 표현을 사용하여 업데이트(gnn 아이디어, 문맥을 인지하는 표현을 얻을 수 있음)
 - 이웃의 정보를 캡처하는 방법은 SKNN, STAN에서도 의미있다고 확인

(온라인 플랫폼)

인기있는 아이템 → 더 잦은 노출 → 더 많은 상호작용 → 상호작용 쏠림 분포 → popularity bias

위와 같은 문제를 완화하기 위해 한 세션 그래프의 표현이 다음 클릭할 아이템의 표현과 비슷해지는 것을 목표로 학습 메커니즘 구성

아이템과 세션 그래프 표현을 **단위 추구(?)**위에 놓도록 제한하는 이점을 취한다.

NISER: Normalized Item and Session Representations model for SR

- popularity bias를 다루는데 강화된 능력을 보임
- 시퀀셜 특성을 포함하기위해 포지션 임베딩 적용
- 시퀀스(RNN)와 그래프(GNN)를 모두 인지하는 이점을 가진다.

PROBLEM DEFINITION

세션이 주어졌을 때, 그 다음 아이템을 구하는 것.

가장 높은 스코어를 갖는 K개 아이템으로 top-K 추천 리스트를 구성

LEARNING ITEM AND SESSION REPRESENTATIONS

세션 내 각 아이템 임베딩 → (타겟 아이템의 임베딩과 가까운) 세션 임베딩 학습 → 전체 아이템 중 가장 값이 큰 인덱스를 선택

DNN 기반의 모델 f에서 다음 아이템이 될 i_k 의 확률은 미분가능한 소프트맥스 함수로 근사된다.

$$p_k(s) = \hat{y}_k = \frac{\exp(\mathbf{i}_k^T \mathbf{s})}{\sum_{j=1}^m \exp(\mathbf{i}_j^T \mathbf{s})}. \quad (1)$$

Radial Property of Softmax Loss

소프트맥스 로스를 최적화 하는 것은 피쳐의 방사형 분포를 타겟 클래스로 이끈다고 알려져 있다.

소프트 맥스 로스는 쉽게 구분할 수 있도록 놈을 크게하는 것을 선호한다

큰 값의 \hat{y}_k 는 벡터 \mathbf{i}_k 에 스칼라 값 시그마를 곱하거나 아이템 임베딩 벡터의 놈을 크게 함으로써 얻을 수 있다.

Normalizing the Representation

쉽게 예측되는 아이템은 높은 l2놈을 가지고 있을 수 있다.

인기있는 아이템이 더 클릭 될 가능성이 있고, 따라서 파라미터는 더 자주 추천하는 것에 확신을 갖는 값으로 학습된다.

인기도가 높은 아이템이 더 큰 놈을 갖고, 인기도가 하락하는 만큼 줄어들었다.

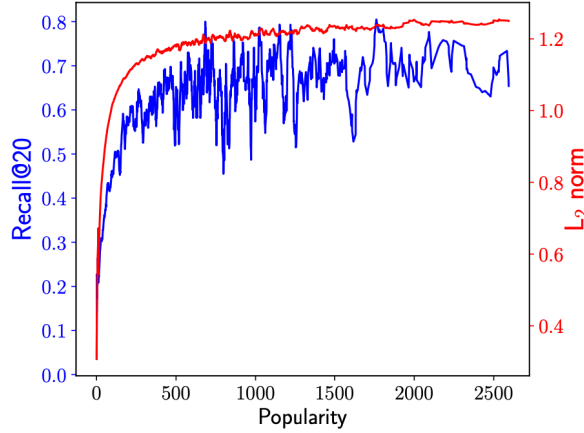


Figure 3: Recall@20 and L_2 norm of learned item embeddings decreases with decreasing popularity in GNN [14].

NISER

마지막 분류와 추천 의사결정 에서 임베딩 놈의 영향을 최소화하는 것을 고려

아이템과 세션 임베딩 사이의 유사도를 측정하는 것 코사인 유사도 최적화 하는 것을 제안

아이템 임베딩이 $i_k / \|i_k\|_2$ 가 되고, 이것은 세션 임베딩 또한 정규화 된 값으로 얻게 된다.

$$\hat{y}_k = \frac{\exp(\sigma \tilde{i}_k^T \tilde{s})}{\sum_{j=1}^m \exp(\sigma \tilde{i}_j^T \tilde{s})}. \quad (3)$$

코사인 유사도는 -1과 1 사이 값으로 강제된다.

더 나은 로스에 수렴하기 위해 스케일링 팩터 시그마를 사용

LEVERAGEING NISER WITH GNN

아이템과 세션의 임베딩이 학습에 GNN을 고려할 수 있다.

메세지 전파 절차는 다음과 같다.

$$\mathbf{a}_{s,j}^t = [\mathbf{A}_{s,j}^{in} \tilde{\mathbf{i}}_s^{t-1} \mathbf{H}_1, \mathbf{A}_{s,j}^{out} \tilde{\mathbf{i}}_s^{t-1} \mathbf{H}_2]^T + \mathbf{b}, \quad (4)$$

$$\mathbf{z}_{s,j}^t = \sigma(\mathbf{W}_z \mathbf{a}_{s,j}^t + \mathbf{U}_z \tilde{\mathbf{i}}_{s,j}^{t-1}), \quad (5)$$

$$\mathbf{r}_{s,j}^t = \sigma(\mathbf{W}_r \mathbf{a}_{s,j}^t + \mathbf{U}_r \tilde{\mathbf{i}}_{s,j}^{t-1}), \quad (6)$$

$$\hat{\mathbf{i}}_{s,j}^t = \tanh(\mathbf{W}_o \mathbf{a}_{s,j}^t + \mathbf{U}_o (\mathbf{r}_{s,j}^t \odot \tilde{\mathbf{i}}_{s,j}^{t-1})), \quad (7)$$

$$\tilde{\mathbf{i}}_{s,j}^t = (1 - \mathbf{z}_{s,j}^t) \odot \tilde{\mathbf{i}}_{s,j}^{t-1} + \mathbf{z}_{s,j}^t \odot \hat{\mathbf{i}}_{s,j}^t, \quad (8)$$

(그래프 연산 생략)

아이템 상호작용의 시퀀셜 정보를 만들기 위해, 선택적으로 포지션 임베딩을 학습하고 그것을 아이템 임베딩에 더함으로써 포지션을 고려한 아이템 임베딩(결국 세션 임베딩)을 얻을 수 있다.

소프트 어텐션 웨이트를 사용해 각 아이템에 대한 임베딩을 구하고, 그것을 통해 세션 임베딩을 구한다.

그리고 가장 최근 아이템과 결합하는 형식으로 최종 임베딩을 구한다.

최종 추천 스코어는 식 3과 같이 계산된다. 세션 그래프 임베딩은 세션그래프, 시퀀스를 인지한 아이템 임베딩을 통해 얻지만, 추천 점수를 계산할 때는 특정 세션과는 독립적인 정규화된 아이템 임베딩을 사용한다.

EXPERIMENTAL EVALUATION

Dataset Details

- Yoochoose(YC)
- Diginetica(DN)
- RetailRocket(RR)

Offline and Online setting

- Offline setting
 - 학습/ 추론 데이터 고정
- Online setting
 - 매 2주 마다 재학습

NISER and its variants

- Normalized Item Representation (NIR)
 - only item embeddings are normalized and scale factor sigma is not used
- Normalized Item and Session Representation (NISER)
 - both item and session embeddings are normalized
- NISER +
 - NISER with position embeddings and dropout applied to input item embeddings

Hyperparameter Setup

Evaluation Metrics

- Recall@K
- MRR@K
- Average Recommendation Popularity(ARP)

Result and Observations

(1) NISER+ reduces popularity bias in GNN+ in

Table 2: Offline setting evaluation: NISER+ versus GNN+ in terms of Average Recommendation Popularity (ARP). Lower values of ARP indicate lower popularity bias.

Method	DN	RR	YC-1/64	YC-1/4
GNN+	495.25±2.52	453.39±8.97	4128.54±27.80	17898.10±126.93
NISER+	487.31±0.30	398.53±3.09	3972.40±41.04	16683.52±120.74

ARP : NISER+ < GNN+

popularity bias를 줄여 덜 인기있는 아이템도 더 자주 추천할 수 있다.

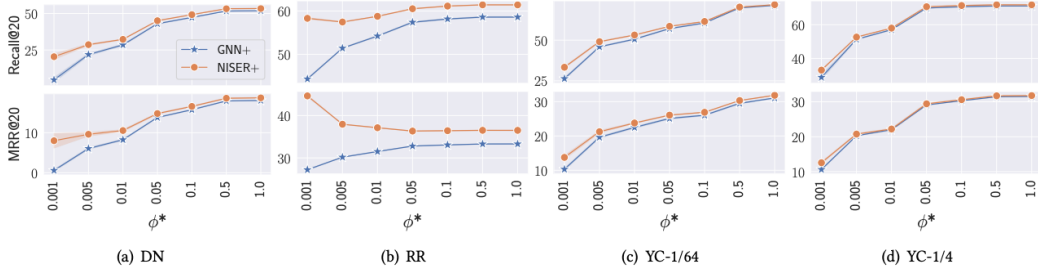


Figure 4: Offline setting evaluation: Recall@20 and MRR@20 with varying ϕ^* indicating larger gains by using NISER+ over GNN+ for less popular items.

파이가 작은 덜 인기있는 아이টে 대해서 더 좋은 성능을 보였다.

파이가 큰 값에서도 적어도 GNN+보다 좋은 성능을 보였다.

(2) NISER+ improves upon GNN+ in online setting

(새로 도입된 아이টে 대해서)

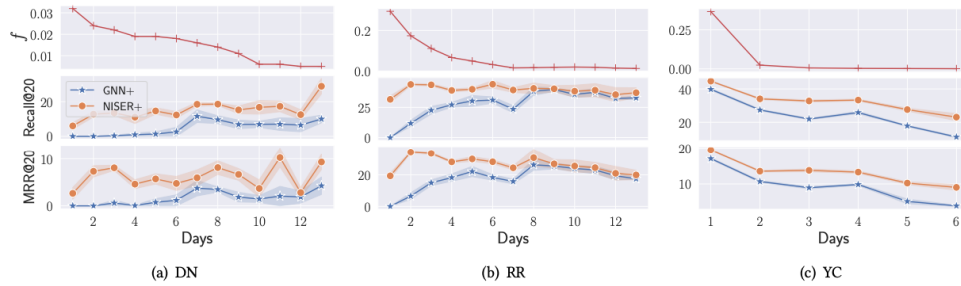


Figure 5: Online setting evaluation: Recall@20 and MRR@20 for sessions where target item is one of the less popular newly introduced items from the previous day. f denotes the fraction of such sessions in the training set for $\phi^* = 0.01$. Standard deviations over five models are shown in lighter-shaded region around the solid lines.

새로 도입된 덜 인기있는 아이টে 바로 다음날 타겟 아이টে 되는 세션에 대해서 NISER+가 더 좋은 성능을 보인다.

NISER+의 새로운 아이টে 대해서 바로 다음날 추천할 수 있는 능력을 보여주고, popularity bias를 줄였기 때문

학습데이터가 적은 상황(초기 일자들)에 GNN+는 성능이 좋지 않았고, NISER+는 비교적 일관된 성능을 보였다. 이는 NISER+의 잠재적 정규화 효과를 나타냄

날이 지나면서 학습 데이터가 많아지면서 GNN+의 성능도 올라가지만 NISER+보다 낮았다.

(3) NISER and NISER+ outperform GNN and GNN+ in offline setting

Table 3: NISER+ versus other benchmark methods in offline setting. Numbers after \pm are standard deviation values over five models.

Method	DN	RR	YC-1/64	YC-1/4
Recall@20				
SKNN [5]	48.06	56.42	63.77	62.13
STAN [3]	50.97	59.80	69.45	70.07
GRU4REC [4]	29.45	-	60.64	59.53
NARM [6]	49.70	-	68.32	69.73
STAMP [8]	45.64	53.94	68.74	70.44
GNN [14]	51.39 \pm 0.38	57.63 \pm 0.15	70.54 \pm 0.14	70.95 \pm 0.04
GNN+	51.81 \pm 0.11	58.59 \pm 0.10	70.85 \pm 0.08	71.10 \pm 0.07
NIR	52.40 \pm 0.06	60.67 \pm 0.08	71.12 \pm 0.05	71.32 \pm 0.11
NISER	52.63 \pm 0.09	60.85 \pm 0.06	70.86 \pm 0.15	71.69 \pm 0.03
NISER+	53.39\pm0.06	61.41\pm0.09	71.27\pm0.05	71.80\pm0.09
MRR@20				
SKNN [5]	16.95	33.16	25.22	24.82
STAN [3]	18.48	35.32	28.74	28.89
GRU4REC [4]	8.33	-	22.89	22.60
NARM [6]	16.17	-	28.63	29.23
STAMP [8]	14.32	28.49	29.67	30.00
GNN [14]	17.79 \pm 0.16	32.74 \pm 0.09	30.80 \pm 0.09	31.37 \pm 0.13
GNN+	18.03 \pm 0.05	33.29 \pm 0.03	30.84 \pm 0.10	31.51 \pm 0.05
NIR	18.52 \pm 0.06	35.57 \pm 0.05	30.99 \pm 0.10	31.73 \pm 0.11
NISER	18.27 \pm 0.10	36.09 \pm 0.03	31.50 \pm 0.11	31.80\pm0.12
NISER+	18.72\pm0.06	36.50\pm0.05	31.61\pm0.02	31.77 \pm 0.10

Recall, MRR 관점에서 NISER+가 가장 좋은 성능을 보인다.

ablation study

Table 4: Ablation results for NISER+ indicating that normalization of embeddings (L_2 norm) contributes the most to performance improvement. Here PE: Position Embeddings.

Method	DN	RR	YC-1/64	YC-1/4
Recall@20				
NISER+	53.39\pm0.06	61.41\pm0.09	71.27\pm0.05	71.80 \pm 0.09
- L_2 norm	52.23 \pm 0.10	59.16 \pm 0.10	71.10 \pm 0.09	71.46 \pm 0.19
-Dropout	52.81 \pm 0.12	60.99 \pm 0.09	71.07 \pm 0.13	71.90\pm0.03
-PE	53.11 \pm 0.12	61.22 \pm 0.03	71.13 \pm 0.04	71.70 \pm 0.11
MRR@20				
NISER+	18.72\pm0.06	36.50\pm0.05	31.61 \pm 0.02	31.77 \pm 0.10
- L_2 norm	18.11 \pm 0.05	33.78 \pm 0.04	30.90 \pm 0.07	31.49 \pm 0.07
-Dropout	18.43 \pm 0.11	35.99 \pm 0.02	31.56 \pm 0.06	31.93\pm0.17
-PE	18.60 \pm 0.09	36.32 \pm 0.03	31.68\pm0.05	31.71 \pm 0.06

L_2 norm이 가장 중요했다.

DISCUSSION

부분적으로 소프트맥스 로스의 방사형 특성에 관련된 것으로 인기도에 따라 높의 크기가 정해질 수 있는 것을 의미한다.

내적 대신 코사인 유사도를 사용해 아이템과 세션그래프를 최적화 함으로써 이러한 확장 이슈를 막는 표현을 학습할 수 있다.

이것이 popularity bias를 줄이고, 새로운 아이템이 도입되는 온라인 세팅에서 유용하다.

아이템과 세션그래프 표현에 대해서 정규화함으로써 성능 향상을 관측했다.

STAMP와 같은 알고리즘을 개선하기 위해 NISER를 탐색해보는 것도 의미 있을 것 같다.