

# 하이퍼파라미터 최적화를 통한 SASRec 추천 모델 성능 개선 연구

성다훈<sup>1</sup>, 임유진<sup>2</sup>

<sup>1</sup> 숙명여자대학교 IT 공학과 석사과정

<sup>2</sup> 숙명여자대학교 인공지능공학부 교수

ekgns324@sookmyung.ac.kr, yujin91@sookmyung.ac.kr

## A Study on the Performance Improvement of the SASRec Recommendation Model by Optimizing the Hyperparameters

Da-Hun Seong<sup>1</sup>, Yujin Lim<sup>2</sup>

<sup>1</sup>Dept. of Information Technology Engineering, Sookmyung Women's University

<sup>2</sup>Dev. of Artificial Intelligence Engineering, Sookmyung Women's University

### 요 약

최근 스마트폰과 같은 디지털 기기의 보급과 함께 개인화, 맞춤형 서비스의 수요가 늘어나면서 추천 서비스가 주목을 받고 있다. 세션 기반(Session based) 추천 시스템은 사용자의 아이템 선호에 따른 순서 정보를 고려한 학습 추천 모델로, 다양한 산업 분야에서 사용되고 있다. 세션 기반 추천 시스템 중 SASRec(Self-Attentive Sequential Recommendation) 모델은 MC/CNN/RNN 기반의 기존 여러 순차 모델들에 비하여 효율적인 성능을 보인다. 본 연구에서는 SASRec 모델의 하이퍼파라미터 중 배치 사이즈(Batch Size), 학습률 (Learning Rate), 히든 유닛(Hidden Unit)을 조정하여 실험함으로써 하이퍼파라미터에 의한 성능 변화를 분석하였다.

### 1. 서론

최근 스마트폰과 같은 디지털 기기의 보급과 함께 개인화, 맞춤형 서비스의 수요가 늘어나면서 추천 서비스가 주목을 받고 있다. 추천 시스템은 사용자에게 상품이나 콘텐츠 등을 추천하는 시스템으로, 온라인 쇼핑, OTT 서비스의 콘텐츠 추천 등 다양한 산업 분야에서 적용되고 있다. 이는 크게 콘텐츠 기반 필터링과 협업 필터링 방식으로 구분된다. 협업 필터링은 유사한 취향을 가진 다른 사용자가 좋아하는 아이템을 추천하는 방식으로, 최근에는 잠재 요인(Latent Factor) 협업 필터링 방식이 자주 사용된다. 이는 평가되지 않은 항목들의 평점까지 예측하여 추천하는 방식이다. 그러나 이 방식은 순서 정보를 고려하지 않아, 어떤 아이템 다음으로 선호되는 아이템의 순서 정보에 따른 추천이 어렵다는 문제가 있다. 이러한 문제를 해결하기 위해 순서 정보를 고려한 세션 기반 추천 시스템이 등장하였다.

세션 기반 추천 시스템은 사용자의 이전 상호작용 데이터에 기반해 다음 아이템을 추천하는 방식이다. SASRec(Self-Attentive Sequential Recommendation)[1]은 그중 하나로, 세션 기반 추천 기법들 중에서 높은 성능을 보여 많은 주목을 받고 있는 기법이다. 본 연구에서는 SASRec 모델의 하이퍼파라미터에 의한 성능 변화를 실험하고 이를 기반으로 하이퍼파라미터 최적화를 통한 추천 시스템의 성능을 개선하고자 한다.

### 2. 세션 기반 추천 시스템 모델

세션 기반 추천 시스템 모델의 종류로는 BPR(Bayesian Personalized Ranking)[2], FMC(Factorized Markov Chains)[3], FPMC(Factorizing Personalized Markov Chains)[3], TransRec[4], GRU4Rec(Gated Recurrent Unit for Recommendation)[5], Caser(Convolutional Sequence Embedding)[6], SASRec 등이 있다. BPR, FMC, FPMC, TransRec은 Markov Chains(MC) 기반 모델로, 이는 사용자의 사용 기록을 모두 반영하기에는 한계가 있다.

이후, 시퀀스 또는 시계열 데이터를 모델링하는 데 효과적인 신경망 클래스 RNN(Recurrent Neural Network) 기반 모델 GRU4Rec 이 등장하였다. 그러나 RNN 은 시퀀스의 각 요소를 처리할 때마다 이전 상태를 입력으로 받아들이는 순차적인 계산을 수행하기 때문에 연산이 느려진다. SASRec 은 Transformer 기반 모델로, Transformer 모델에서는 RNN 의 재귀적 구조를 없애고, self-attention 메커니즘을 사용해 입력 시퀀스 내의 모든 요소를 동시에 처리한다. 이러한 병렬적 처리로 인해 대규모 데이터를 다루는 딥러닝 모델에서 학습 속도와 성능이 향상되었으며, CNN 기반 Caser 모델처럼 N 개의 추천리스트를 제공한다. 지금까지 살펴본 세션 기반 추천 시스템의 발전 흐름에 따라, 최근 들어 SASRec 모델이 많은 주목을 받고있다. 이에 본 논문에서는 SASRec 모델에 집중하고자 한다.

### 3. SASRec 모델

RNN 방식은 모델 복잡성이 높은 고밀도 데이터셋에서 잘 수행되나 장기적인 사용자 행동 데이터가 사용되기 때문에, 많고 조밀한 데이터가 필요하다. MC 는 모델 간결성이 중요한 희박한 데이터셋에서 잘 수행되나, 고밀도 데이터셋에서의 성능은 저하된다. 두 모델의 장점을 결합하여 단점을 해소한 SASRec 은 RNN 과 같이 고밀도의 데이터셋에서는 장기적인 의존성을 고려할 수 있고, self-attention 매커니즘을 사용하여 MC 와 같이 희소 데이터셋에서는 최근 활동에 초점을 맞추므로써, 다양한 밀도의 데이터셋을 적응적으로 처리하여 예측을 수행하는 추천 모델이다.

Self-attention 매커니즘은 하나의 시퀀스 또는 벡터 데이터 내에서 다른 위치 간의 관계를 고려해 가중치를 계산하는 방법이다. 이를 통해 시퀀스나 벡터 데이터 내에서 중요한 패턴 등의 정보를 추출하거나, 다른 연산을 수행할 수 있다. 따라서 SASRec 는 시퀀스 내 각 원소들 간의 상호작용을 고려할 수 있고, 각 레이어 단계의 사용자 기록에서 어떤 항목이 관련성이 있는지 식별하여 다음 항목을 예측한다[7].

SASRec에서는 임베딩 계층, 여러 self-attention 블록, 예측 계층을 통한 세션 기반 추천 모델을 구축한다. self-attention 모델은 컨볼루션 또는 반복 모듈을 포함하지 않아 이전 항목의 위치를 알지 못한다. 따라서 임베딩 계층에서는 학습 가능한 위치를 입력 임베딩에 주입한다. 또한 self-attention 모델은 선형 모델이기에, 여러 개의 self-attention 블록에서는 입력값에 대한 비선형적인 변환을 수행함으로써 인공 신경망이 복잡한 함수를 학습할 수 있게 하여, 이전에 사용된 항목의 정보를 계층적이고 적응적으로 추출한다. 예측 계층에서는 Matrix Factorization 를 사용하여 사용자의 선

호도와 항목의 속성을 나타내고, 항목 간의 관련성을 구하며 점수를 순위로 매겨 추천 목록을 생성한다.

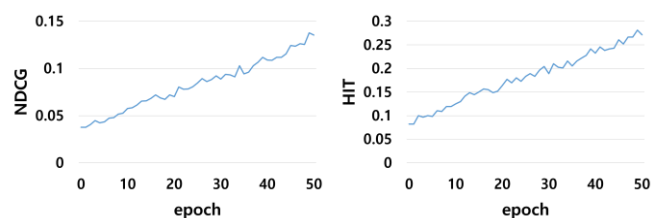
결과적으로 SASRec 의 성능은 MC/CNN/RNN 기반인 Caser, GRU4Rec, TransRec, FPMC, FMC, BPR 와 같은 기존의 다양한 순차 모델과 비교하여 Hit Rate@10 및 NDCG(Normalized Discounted Cumulative Gain)@10 성능지표로 평가했을 때 희소/고밀도 데이터셋 모두에서 가장 높은 성능을 보인다.

### 4. 하이퍼파라미터 튜닝

본 연구는 SASRec 모델의 성능에 하이퍼파라미터가 미치는 영향을 분석함으로써, 이를 기반으로 모델의 성능을 개선하고자 한다. 실험을 위하여 데이터셋으로는 MovieLens100K 를 사용하였다. 이는 영화 평점 데이터셋 중 하나로, 추천 시스템 모델을 학습하고 평가하는 데 사용된다. 하이퍼파라미터는 모델 학습 과정에 반영되는 값으로, 모델의 성능에 영향을 준다. 본 연구에서는 하이퍼파라미터로 딥러닝 기반 추천시스템의 성능에 영향을 끼치는 배치 사이즈, 학습률, 히든 유닛 수를 조정하였다. 그리고 이에 따른 성능을 NDCG 와 적중률(Hit Rate)를 통해 평가하였다. NDCG 는 추천 리스트의 상위 k 개 아이템 중 실제 선호도가 높은 아이템들이 얼마나 많이 포함되어 있는지를 나타낸다. 적중률은 전체 사용자 수 대비 적중한 사용자 수의 비율이다. NDCG 와 적중률은 0 에서 1 사이의 값을 가지고, 높을 수록 추천 성능도 높다. 에포크(epoch)는 50, 배치 사이즈는 128, 학습률은 0.001, 히든 유닛 수는 50 으로 설정했을 때 SASRec 의 초기 성능 평가 결과는 다음과 같다.

<표 1> SASRec 의 초기 성능 평가 결과값

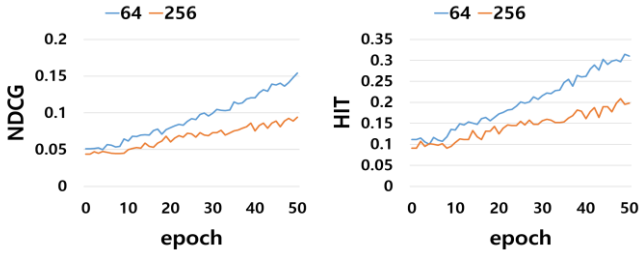
성능지표	결과값
NDCG@10	0.11212
HIT@10	0.23100



(그림 1) SASRec 의 초기 성능 평가 결과

<표 2> 배치 사이즈 64 와 256 의 성능 평가 결과값

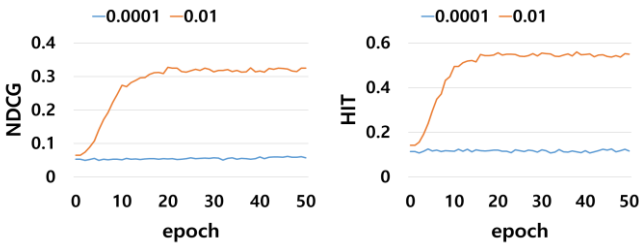
성능지표	배치 사이즈	결과값
NDCG@10	128 → 64	0.154
	128 → 256	0.09430
HIT@10	128 → 64	0.30999
	128 → 256	0.19821



(그림 2) 배치 사이즈 64 와 256 일 때의 성능 평가 결과

<표 3> 학습률 0.0001 과 0.01 의 성능 평가 결과값

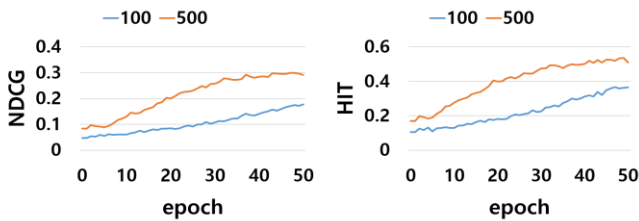
성능지표	학습률	결과값
NDCG@10	0.001 → 0.0001	0.05681
	<b>0.001 → 0.01</b>	<b>0.32522</b>
HIT@10	0.001 → 0.0001	0.11624
	<b>0.001 → 0.01</b>	<b>0.54993</b>



(그림 3) 학습률 0.0001 과 0.01 일 때의 성능 평가 결과

<표 4> 히든 유닛 100 과 500 의 성능 평가 결과값

성능지표	히든 유닛	결과값
NDCG@10	50 → 100	0.17672
	<b>50 → 500</b>	<b>0.36513</b>
HIT@10	50 → 100	0.29102
	<b>50 → 500</b>	<b>0.50969</b>



(그림 4) 히든 유닛 수 100 과 500 일 때의 성능 평가 결과

x 축은 학습 횟수인 에포크이고, 결과값이 진하게 표시된 부분은 하이퍼파라미터별로 각 성능지표에서 높은 값이며, 밑줄 표시는 세 개의 하이퍼파라미터에서 가장 높은 NDCG 와 HIT 값이다. 종합적인 성능을 고려하였을 때, 학습률을 0.01 로 늘려서 설정하였을 때와 히든 유닛을 500 으로 늘려서 설정했을 때 두 성능지표에서 높은 성능 향상을 보였다. 학습률과 히든 유닛의 수를 늘렸을 때 성능이 좋아지는 이유는,

학습률이 증가함에 따라 유의미한 가중치가 갱신되어 손실함수가 최소화되고, 히든 유닛이 증가함에 따라 학습의 모델이 더 복잡한 패턴을 학습하여 모델의 표현력이 높아지기 때문일 것이다.

## 5. 결론

본 연구에서는 SASRec 모델의 배치 사이즈, 학습률 및 히든 유닛 수를 조절함으로써 향상된 성능을 확인할 수 있었다. 위의 실험은 MovieLens100K 데이터셋을 활용하여 실험해서 얻은 결과이므로, 향후 다양한 데이터셋에서 좋은 성능을 얻을 수 있는 하이퍼파라미터 결정 기법에 대해 연구를 진행할 예정이다.

## 사사문구

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2021R1F1A1047113).

## 참고문헌

- [1] W.-C. Kang and J. McAuley, "Self-Attentive Sequential Recommendation," Proceedings of IEEE International Conference on Data Mining (ICDM), pp. 197-206, Nov. 17-20, Singapore, 2018.
- [2] S. Rendle, C. Freudenthaler, Z. Gantner and L. S-T "BPR: Bayesian Personalized Ranking from Implicit Feedback," Proceedings of ArXiv, Preprint ArXiv:1205.2618, 2012.
- [3] S. Rendle, C. Freudenthaler and L. S-T, "Factorizing Personalized Markov Chains for Next-basket Recommendation," Proceedings of the 19th International Conference on World Wide Web, pp. 811-820, Apr. 26 - 30, Raleigh, 2010.
- [4] H. Ruining, K. W-C and J. McAuley. "Translation-based Recommendation," Proceedings of the eleventh ACM Conference on Recommender Systems. pp.161-169, Aug. 27-31, Italy, 2017.
- [5] B. Hidasi, A. Karatzoglou, L. Baltrunas and D. Tikk "Session-based Recommendations with Recurrent Neural Networks," Proceedings of International Conference on Learning Representations (ICLR), Mar. 2-4, Puerto Rico, 2016.
- [6] J. Tand and K. Wang, "Personalized Top-N Sequential Recommendation Via Convolutional Sequence Embedding," Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM '18), pp. 565-573, Feb. 5-9, New York, 2018.
- [7] 안상준, 유원준, "딥 러닝을 이용한 자연어 처리 입문," WikiDocs, 2023.
- [8] 김성필, "딥러닝 첫걸음," 한빛미디어, 2016.