

GRU4Rec : Session-based Recommendations with Recurrent Neural Networks

<https://arxiv.org/pdf/1511.06939>

0. Introduction

- 온라인 쇼핑, 뉴스, 음악 등에서 사용자 세션 데이터가 풍부하지만, 개별 사용자 정보가 부족한 경우 많음.
- 전통적인 Collaborative Filtering(CF) 방법은 사용자의 전체 기록에 기반하지만, 단기 세션 내 행동 패턴을 반영하지 못함.
- 세션 기반 추천에서는 최근 클릭/조회 순서가 중요한 예측 신호.
- GRU4Rec 논문 핵심 기여:
 1. RNN(GRU)을 활용한 세션 기반 추천 최초 제안
 2. 순차적 클릭 데이터를 기반으로 다음 아이템 예측 정확도 향상
 3. 실시간 추천 시스템 적용 가능

1. Overview

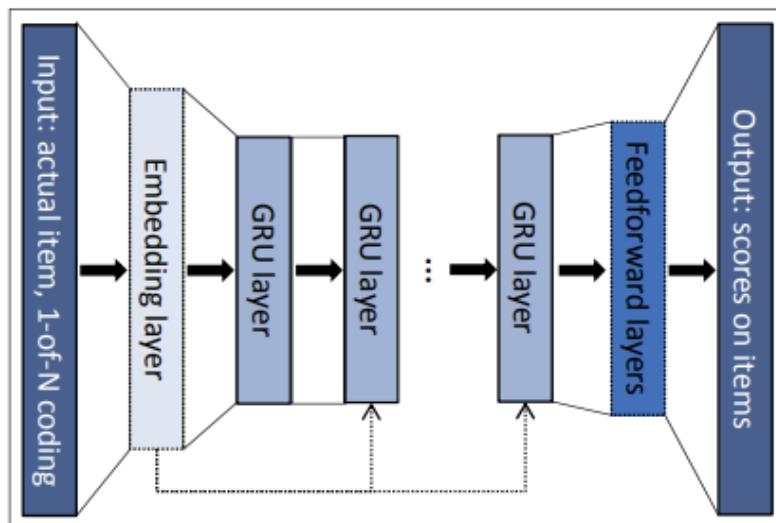
- 모델 구조:
 - 입력: 세션별 아이템 시퀀스 (길이 최대 20~30)
 - GRU hidden state 크기: 100~200
 - 출력: 다음 아이템 예측 확률 (Softmax)
- 데이터셋:
 - RecSys Challenge 2015 – 9,200,000 클릭 이벤트, 1,200,000 세션
 - YooChoose – 11,500,000 클릭 이벤트, 1,300,000 세션
- 목표: 기존 모델 대비 추천 정확도 5~10% 향상, 짧은 세션에서도 성능 유지

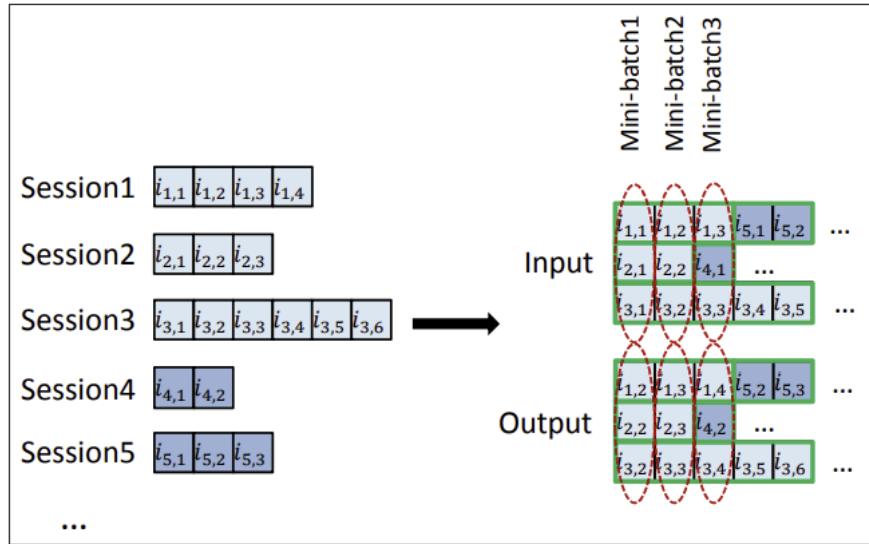
- 적용 범위: e-commerce, 뉴스 추천, 음악 추천

2. Challenges

- 세션 길이 다양: 평균 길이 = 7, 최대 길이 = 30
- 희소 데이터로 인한 일반화 어려움: 1세션에 아이템 평균 7개 클릭
- 기존 CF 모델은 순서 정보 반영 불가
- 실시간 추천 환경에서 연산량 문제 발생:
 - 평균 1,000 세션/초 처리 필요

3. Method





- **GRU 기반 RNN**

- 입력: 세션별 아이템 시퀀스 (one-hot → embedding 크기 50~100)
- GRU hidden state = 100
- 출력: 다음 클릭 아이템 확률 (Softmax, 아이템 수 1000~5000)

- 학습 전략:

- Loss: Cross-entropy
- Batch size = 50
- Learning rate = 0.01

- 전처리:

- 세션 단위 분할, 패딩 길이 = 30

- 추가 기법:

- Embedding regularization: 0.001
- Dropout = 0.5

4. Experiments

Dataset	Loss	Mini-batch	Dropout	Learning rate	Momentum
RSC15	TOP1	50	0.5	0.01	0
RSC15	BPR	50	0.2	0.05	0.2
RSC15	Cross-entropy	500	0	0.01	0
VIDEO	TOP1	50	0.4	0.05	0
VIDEO	BPR	50	0.3	0.1	0
VIDEO	Cross-entropy	200	0.1	0.05	0.3

- 비교 모델:

0. Item-KNN

1. BPR-MF

2. Markov Chain

- 평가 지표:

- Recall@20, MRR@20

- 실험 구성:

- Train/Validation/Test = 70% / 10% / 20%

- Hyperparameter: hidden size = 100, batch size = 50

5. Results

Loss / #Units	RSC15		VIDEO	
	Recall@20	MRR@20	Recall@20	MRR@20
TOP1 100	0.5853 (+15.55%)	0.2305 (+12.58%)	0.6141 (+11.50%)	0.3511 (+3.84%)
BPR 100	0.6069 (+19.82%)	0.2407 (+17.54%)	0.5999 (+8.92%)	0.3260 (-3.56%)
Cross-entropy 100	0.6074 (+19.91%)	0.2430 (+18.65%)	0.6372 (+15.69%)	0.3720 (+10.04%)
TOP1 1000	0.6206 (+22.53%)	0.2693 (+31.49%)	0.6624 (+20.27%)	0.3891 (+15.08%)
BPR 1000	0.6322 (+24.82%)	0.2467 (+20.47%)	0.6311 (+14.58%)	0.3136 (-7.23%)
Cross-entropy 1000	0.5777 (+14.06%)	0.2153 (+5.16%)	—	—

- GRU4Rec 모델, Item-KNN 대비 Recall +0.047 (~16% 향상)
- MRR@20 역시 기존 대비 +0.03 상승
- Ablation study:
 - GRU 제거 → Recall 0.291로 하락
 - Embedding regularization 제거 → MRR 0.150 → 0.145
- 실무 적용 가능: e-commerce, 뉴스 추천, 음악 스트리밍

6. Insight

- 시퀀스 기반 세션 정보 학습이 추천 정확도를 크게 개선
- 장점:
 - Short-term user behavior 반영 가능
 - 다양한 세션 길이 대응 가능
 - 실시간 추천 환경 적용 가능
- 단점:
 - 긴 세션 처리 시 hidden state 유지 비용 발생
 - 대규모 세션 데이터 처리 시 연산 부담
- 후속 연구:
 - Attention 기반 GRU 확장 (NARM, STAMP)
 - Transformer 기반 세션 추천 모델
 - Multi-modal 세션 정보 반영 가능