Session-Based Recommendations With Recurrent Neural Networks

ICLR 2016

Authors: Balazs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, Domonkos Tikk
Presenter: Hyunwoo Jung

SEOUL NATIONAL UNIVERSITY

Contents

- Motivation
- Prior Approaches & Background
- Technique Details
- Experiments & Results
- Summary & Critique

Motivation

- How to build a short session-based recommendation system?
    - Example of a session: A user access to an e-commerce without login, clicks multiple product pages, and leave the service.

- Constraints of session-based recommendation
    - Missing user profile
    - Small number of items in each session

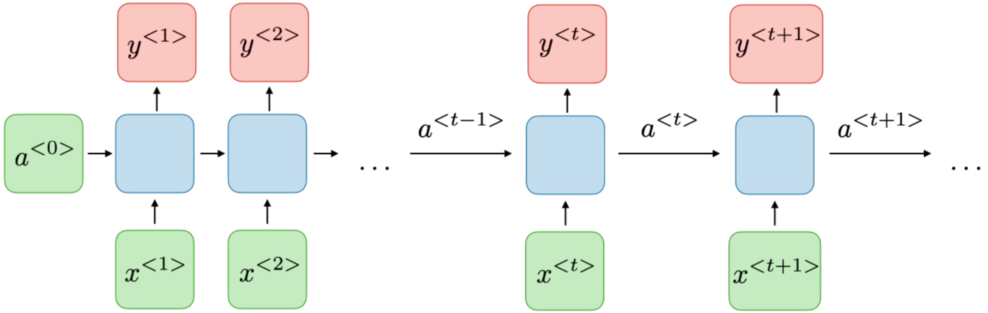
Prior Works

Approach	Limitation
Matrix Factorization	Infeasible because of the absence of a user profile.
Item-to-item Recommendation	Only takes into account the last click of the user.
Markov Decision Processes (MDP)	Unmanageable state spaces when includes all possible sequences of user selections.
General Factorization Framework (GFF)	Doesn't consider the ordering within the session.

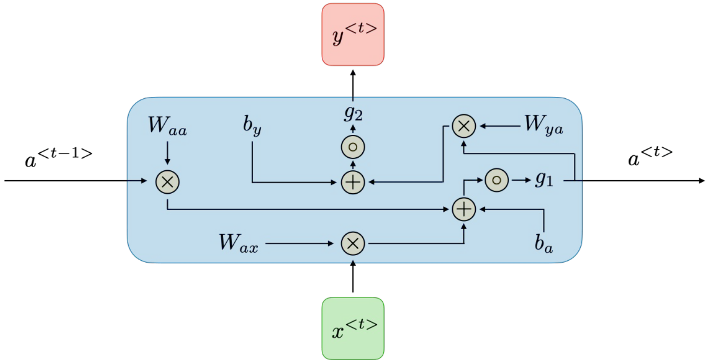
RNN: A New Trend of Time-series Model

- Recurrent Neural Network (RNN): Neural network based sequential model

RNN Architecture



RNN Cell Architecture

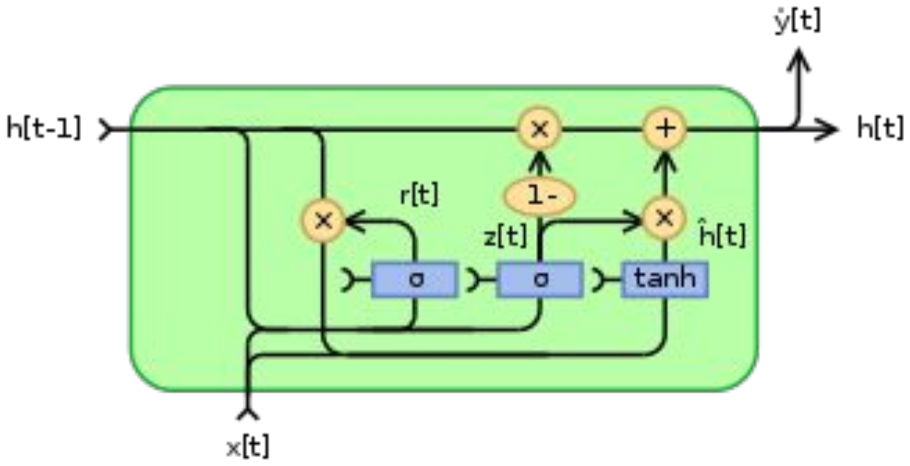


- Little attention has been paid to the area of recommender system.

Gated Recurrent Unit (GRU)

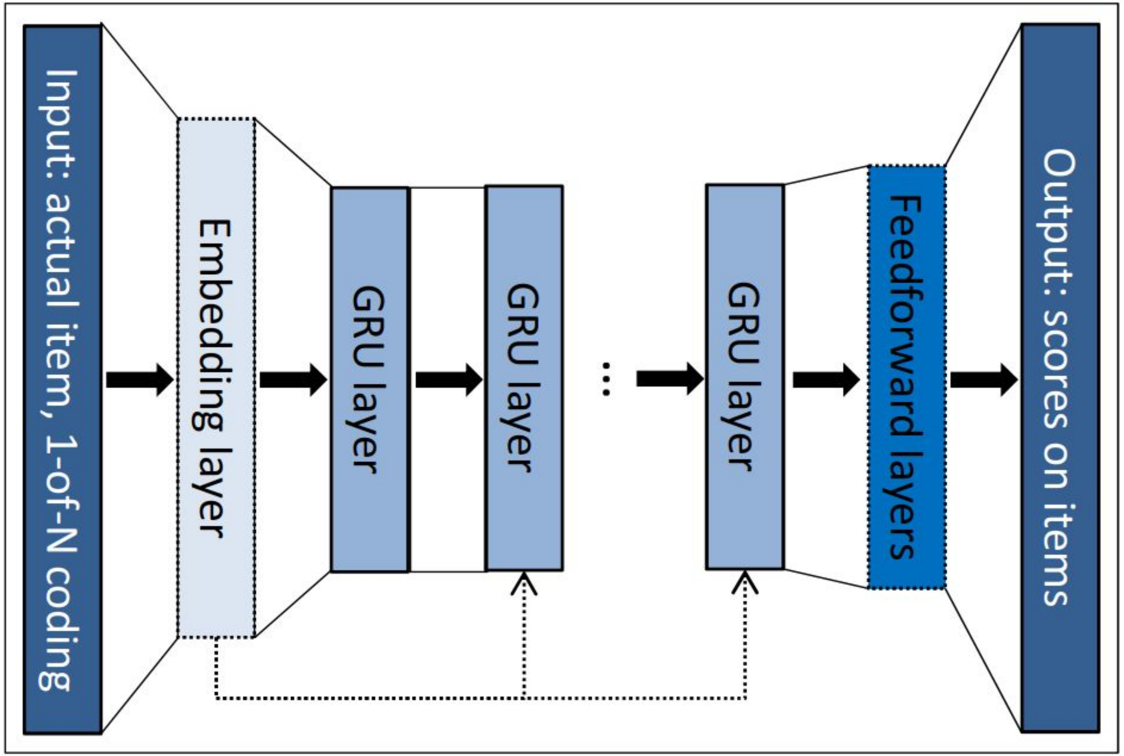
- Deal with vanishing gradient problem of RNN.
- Learns how much to update the hidden state of the unit.

GRU Cell Architecture



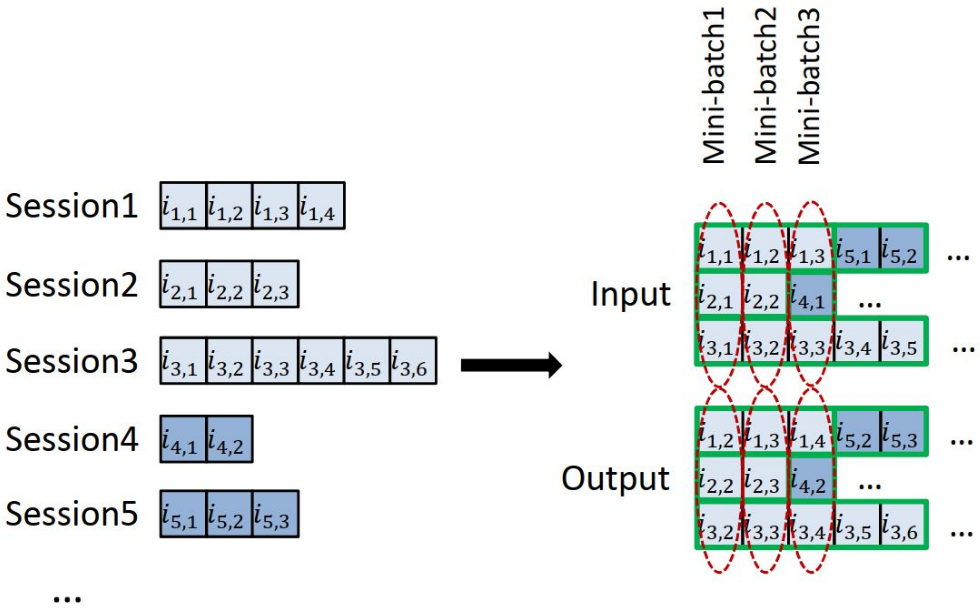
$$z_t = \sigma_g(W_z x_t + U_z h_{t-1} + b_z)$$
$$r_t = \sigma_g(W_r x_t + U_r h_{t-1} + b_r)$$
$$\hat{h}_t = \phi_h(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h)$$
$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t$$

GRU-based Model Architecture



Session-Parallel Mini Batches

- Efficient batching policy for variable-length sequence data
- A session should not be broken down into fragments.



Ranking Loss

- Pairwise ranking loss for stable training

Bayesian Personalized Ranking (BPR)

$$L_s = \frac{1}{N_s}\sum_{j=1}^{N_s}\sigma(\hat{r}_{s,i} - \hat{r}_{s.j}) + \sigma(\hat{r}_{s,j}^2)$$

TOP1 (Proposed)

$$L_s = -\frac{1}{N_s}\sum_{j=1}^{N_s} log(\sigma(\hat{r}_{s,i} - \hat{r}_{s,j}))$$

N_s : sample size

$\hat{r}_{s,k}$: score on item k at the given point of the session

i : desired item (next item in the session)

j : negative samples

Sampling on the Output

- Sampling items on the output in proportional of their popularity

	Presence of Interaction	No Interaction
Popular Item	Preferable	High probability of dislike
Unpopular Item	Preferable	**Hard to guess...**

Baselines

Model	Approach
POP	Recommend the most popular items in the training set
S-POP	Recommend the most popular items in the session
Item-KNN	Item-to-item recommendation
BPR-MF	A variation of Matrix Factorization

Metrics

Recall@k

The fraction of relevant instance from the top-k results

MRR@k (Mean Reciprocal Rank)

Reciprocal rank: The inverse of the rank of the correct answer from the top-k results.

Query	Proposed Results	Correct response	Rank	Reciprocal rank
cat	catten, cati, **cats**	cats	3	1/3
tori	torii, **tori**, toruses	tori	2	1/2
virus	**viruses**, virii, viri	viruses	1	1

MRR = mean(⅓, ½, 1) = 0.61

Results

- GRU-based approach surpasses the baselines

Baseline

Baseline	RSC15 Recall@20	RSC15 MRR@20	VIDEO Recall@20	VIDEO MRR@20
POP	0.0050	0.0012	0.0499	0.0117
S-POP	0.2672	0.1775	0.1301	0.0863
Item-KNN	0.5065	0.2048	0.5508	0.3381
BPR-MF	0.2574	0.0618	0.0692	0.0374

Proposed Model

Loss / #Units	RSC15 Recall@20	RSC15 MRR@20	VIDEO Recall@20	VIDEO 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%)	–	–

Summary

- Proposes GRU-based network for short-session based recommender system.

- Introduces methods to realize the system such as ***session-parallel mini batch*** , ***popularity-proportional sampling***, and ***TOP1 pairwise loss***.

Critique

- Strength
  - Studying session-based recommendation system is important, and not explored much before this paper is published.

- Weakness
  - CNN or Transformers based approaches show better performances for short sequences.
  - Insufficient ablation studies such as the effect of popularity-proportional sampling method.

Thank You
Q & A