Joint Relational Dependency Learning for Sequential Recommendation

Xiangmeng Wang, Qian Li, Wu Zhang∗, Guandong Xu∗ .etc

CONTENTS

PART ONE

Background

WHAT IS RECOMMENDATION TASK?

Input: A user + her profile; An item + its Profile

Output: How likely she will adopt the target item? -> Relevance Score

Traditional Recommendation System:

?

User Profile:
- User ID
- Rating history
- Age, Gender
- Clicks
- Income level

....

Item Profile:
- Item ID
- Description
- Image
- Category
- Price

....

ignore to consider the time information of each interaction

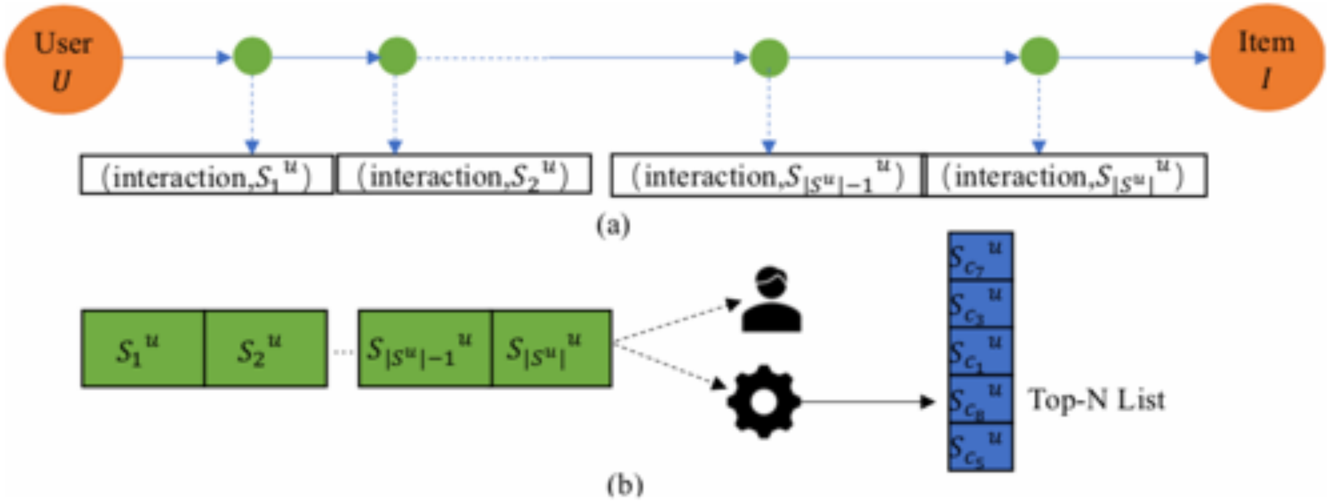
Long-term dependency
Short-term dependency

Sequential recommendation

WHY SEQUENTIAL RECOMMENDATION?

Input: A user + her profile; An item + its Profile + Time information -> behavior sequence

Sequential Recommendation System:



transaction-based behavior sequence

RELATED WORK

- BPR-MF[1]: long-term dependencies using matrix factorization method.
- GRU4Rec[2]: long-term dependencies using Gated Recurrent Units
- MARank[3]: models two types of short-term interactions with factorization model

References:
[1] Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.: BPR: Bayesian Personalized Ranking from Implicit Feedback (may 2012), http://arxiv.org/abs/1205.2618
[2]Hsu, K., Chou, S., Yang, Y., Chi, T.: Neural network based next-song recommen- dation. arXiv: Information Retrieval (2016)
[3]Yu, L., Zhang, C., Liang, S., Zhang, X.: Multi-Order Attentive Ranking Model for Sequential Recommendation. Proceedings of the AAAI Conference on Artificial In- telligence 33, 5709–5716 (2019). https://doi.org/10.1609/aaai.v33i01.33015709

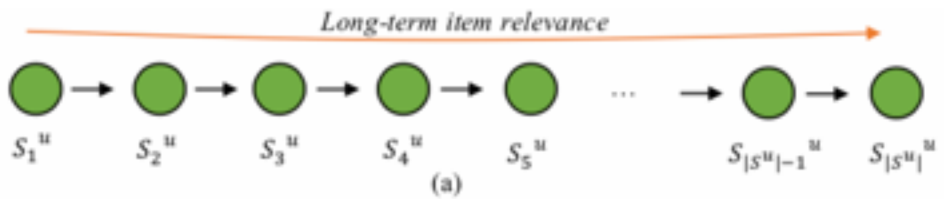
PART

Driven idea

TWO

LEVEL OF DEPENDENCIES

long-term dependency:



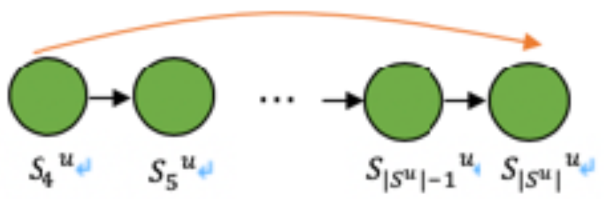
capturing holistic dependencies of user-item sequence

Collaborative Filtering (CF) is the most well-known technique for recommendation.

- Homophily assumption: a user preference can be predicted from his/her similar users.
- Pros: well-suited to train models that capture longer-term user preference profiles
- Cons:
  1) fail to exploit the rich information of transition dependencies of multiple items.
  2) enormous computing cost

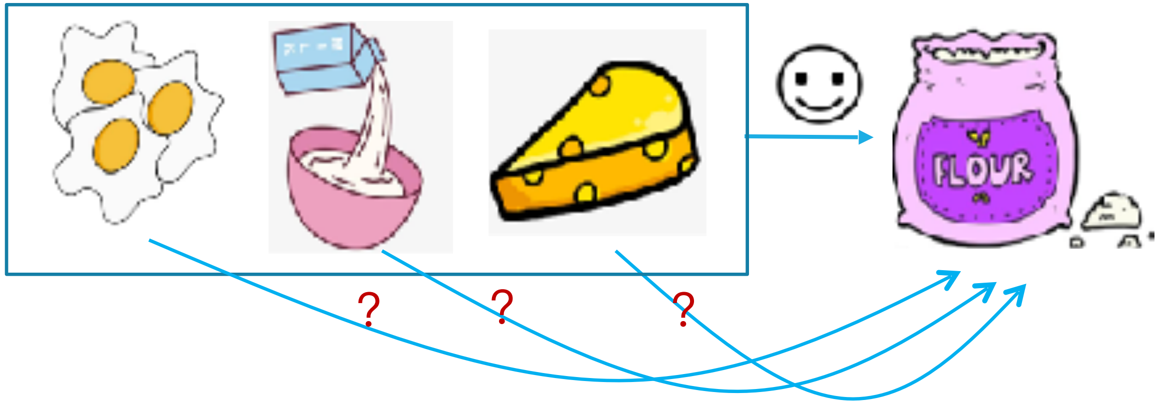
LEVEL OF DEPENDENCIES

Short-term dependency:

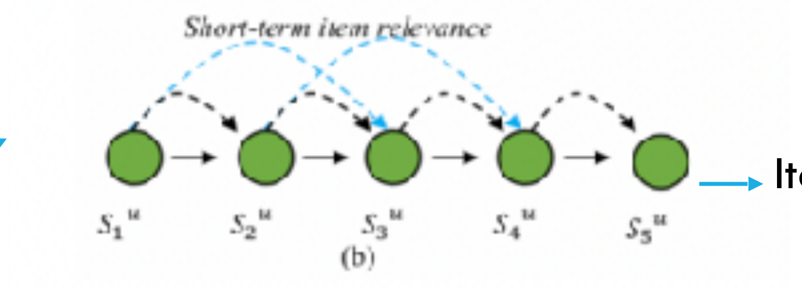


capturing partial dependencies within

A short time window.

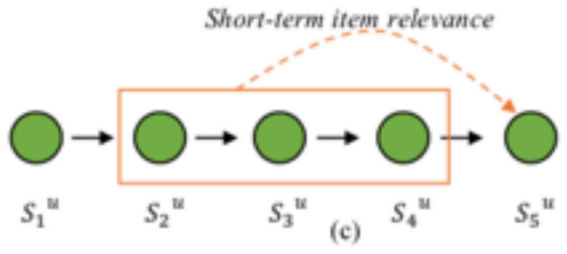


Individual-level dependencies:



Item-item pairs

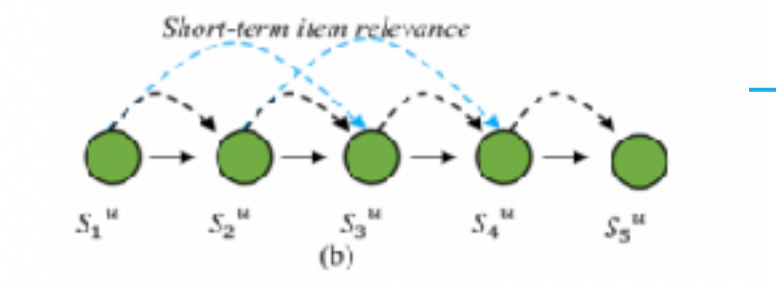
union-level dependencies



a group of items on a single item.

LEVEL OF DEPENDENCIES

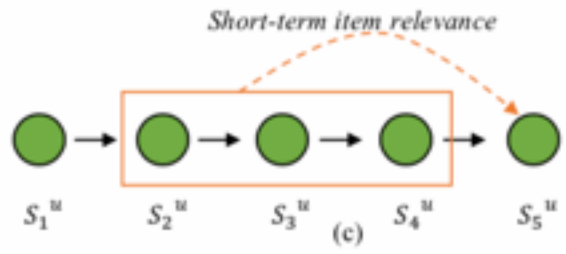
Individual-level dependencies:



Pros:
- capture individual influence between a pair of single item
- rich in quantity

Cons:
- neglect the collective influence among three or more items

union-level dependencies



Pros:
- capture collective influence

Cons:
- sparsity problem

WHAT WE DO?

Problems:

- Modeling Long-term dependency only fails to exploit the information of transition dependencies of multiple items
- enormous computing cost
- Individual-level dependency neglect the collective influence
- sparsity problem of union-level dependency

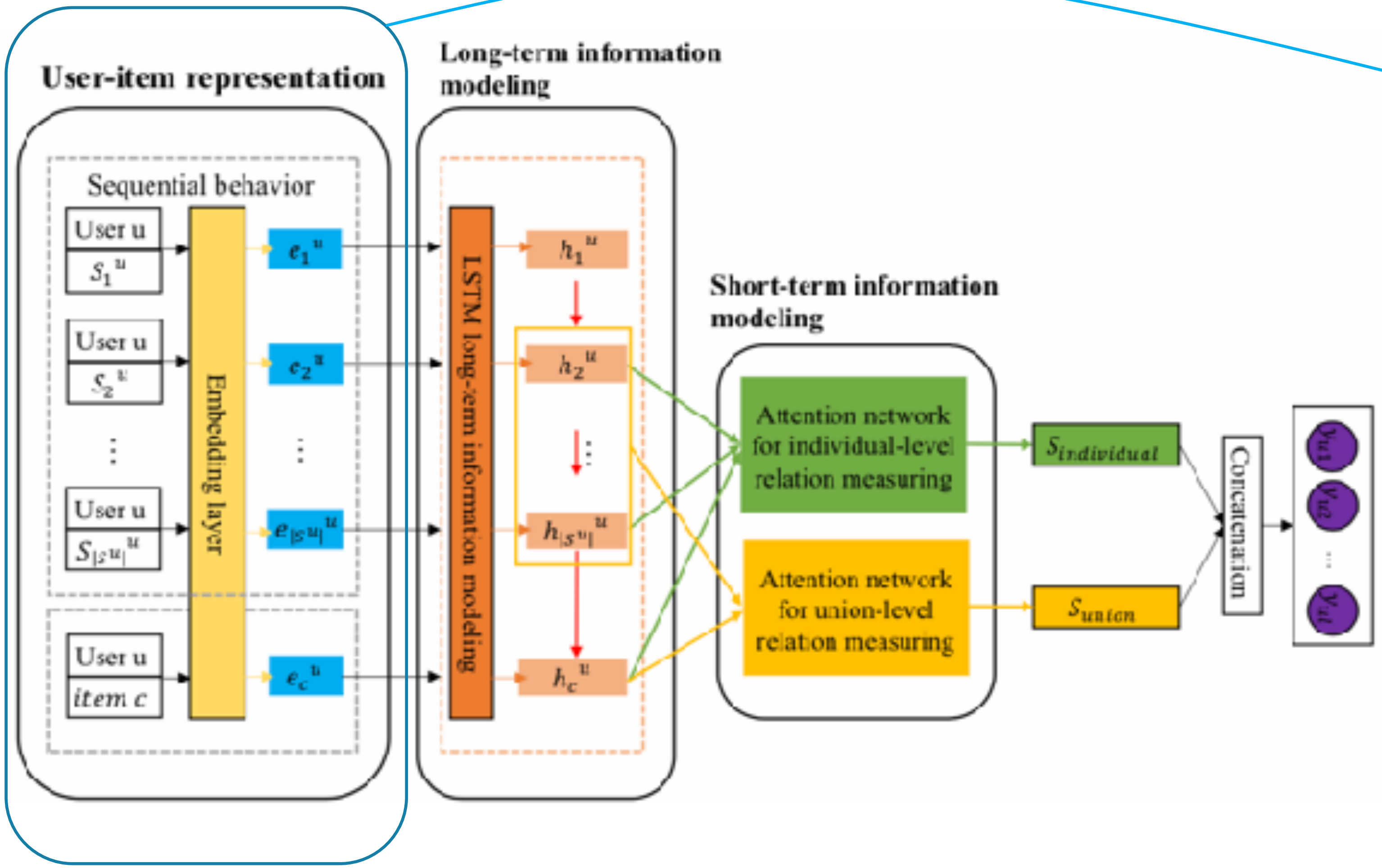
Solutions:

- Adding short-term dependencies
- Deep learning-based method
- Adding union-level dependencies
- Model individual-level relation and union-level relation together to address the sparsity

PART THREE

Methodology

UNIFIED FRAMEWORK:



1. **User-item embedding**

(Skip-gram Based Item Representation)

- apply skip-gram with negative sampling (SGNS) to generate representation for each item in user-item interaction sequence

Input:

A sequence of interactions between U and I

$$S = \{S_j^{u_i} : u_i \in U\}, S_j^{u_i} = (S_1^{u_i}, S_2^{u_i}, ..., S_{|S_j^{u_i}|}^{u_i})$$

Output:

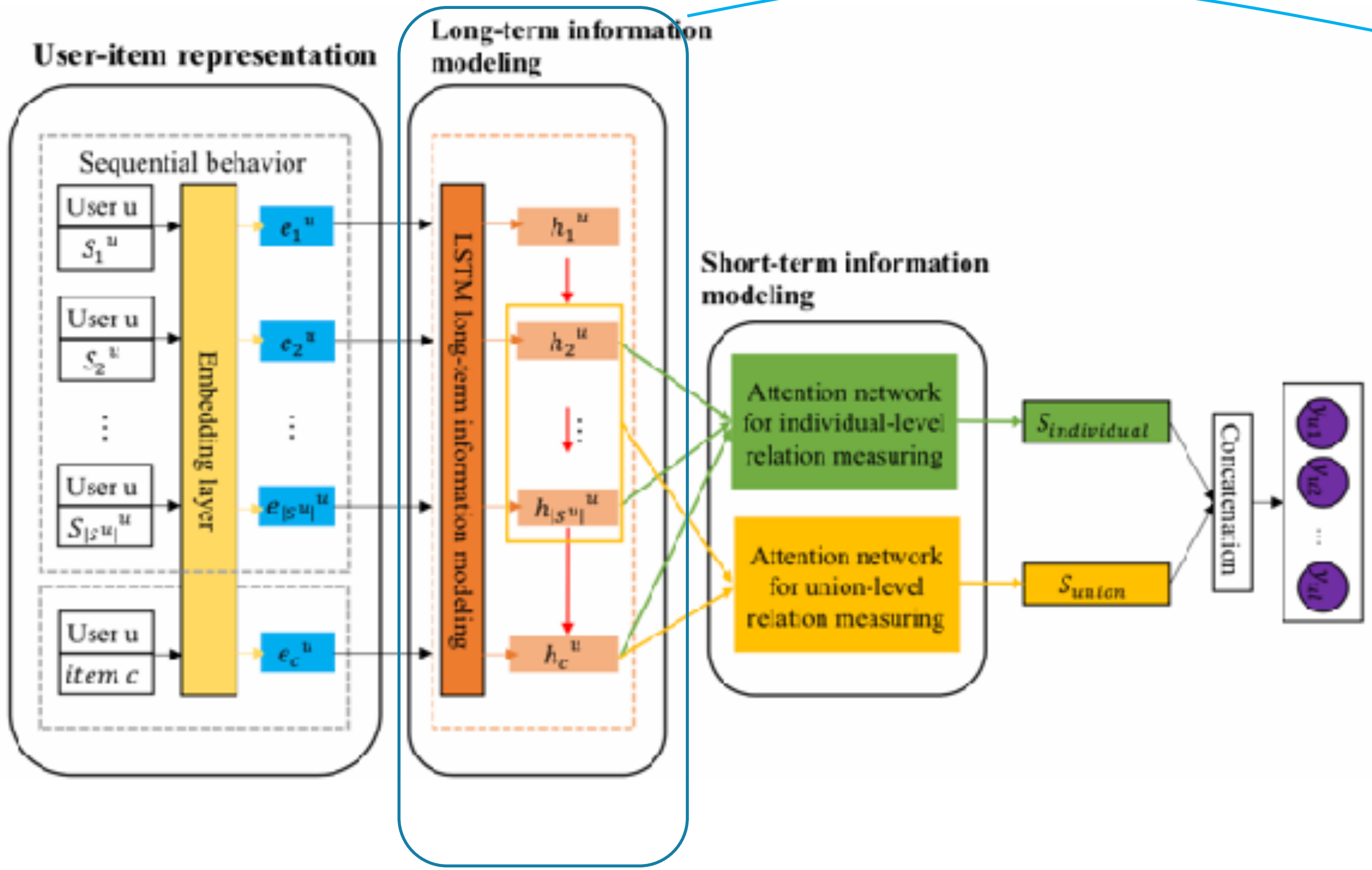
$$L_c^{u_i}$$

,

$$P_u = e_1^{u_i}, e_2^{u_i}, ..., e_{|S_j^{u_i}|}^{u_i}$$

$e_c^{u_i}$ $e_{c_i}^{u_i} \in e_c^{u_i}$

UNIFIED FRAMEWORK:



2. Long-term Pattern modeling

- apply standard LSTM to model the dependency over the whole user-item interaction sequence
- Main equation:

$e_1^{u_i}, e_2^{u_i}, \ldots, e_{|S_j^{u_i}|}$

$$h_i^u = g(e_i^u, h_{i-1}^u, W_{LSTM})$$

$h_1^{u_i}, h_2^{u_i}, \ldots, h_{|S_j^{u_i}|}$

Each $e_{c_i}^{u_i}$

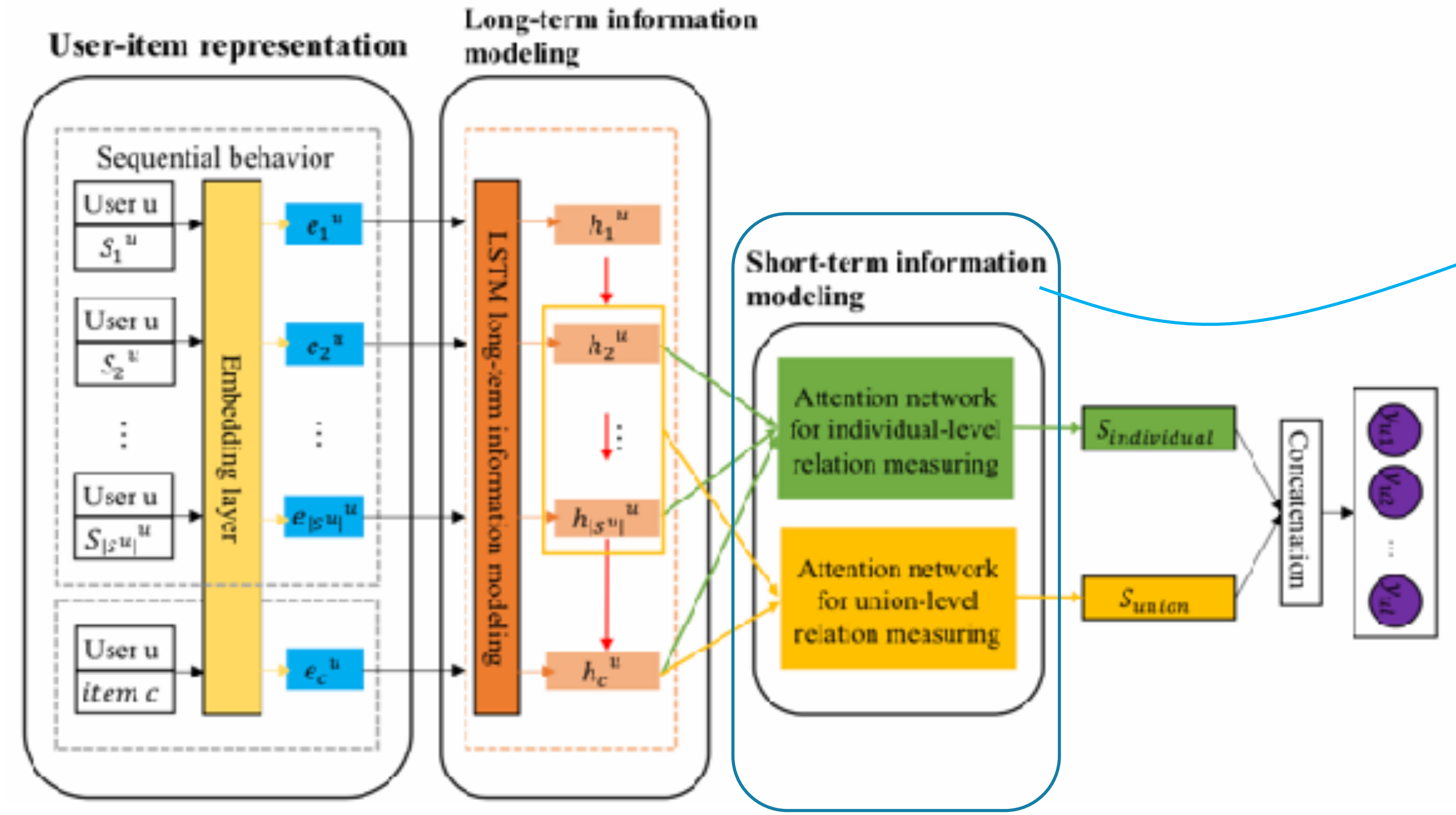
$$h_{c_i}^u = g(e_{c_i}^u, h_{|S^u|}^u, W_{LSTM})$$

$h_{c_i}^{u_i}$

- Output: long-term dependency - sensitiv

$h_1^{u_i}, h_2^{u_i}, \ldots, h_{|S_j^{u_i}|}$ $h_{c_1}^{u_i}, h_{c_2}^{u_i}, \ldots, h_{|c_i|}^{u_i}$

UNIFIED FRAMEWORK:



3. Multi-relational Dependency Modeling for Short-term Pattern

- Attention layer for individual-level relation measuring
- Attention layer for union-level relation measuring

$$h_1^{u_i}, h_2^{u_i}, \ldots, h_{|S_j^{u_i}|}^{u_i}$$

$$h^{u_i}{}_{t-1}, h^{u_i}{}_{t-2}, \ldots, h^{u_i}{}_{t-n} \; (t-n < |S_j^{u_i}|)$$

Individual-level relation measuring:

- Input:

$$h^{(1)} = h^{u_i}{}_{t-1}, h^{u_i}{}_{t-2}, \ldots, h^{u_i}{}_{t-n} \; (t-n < |S_j^{u_i}|)$$

$$h^{(2)} = (h_{c_1}^{u_i}, h_{c_2}^{u_i}, \ldots, h_{|c_l|}^{u_i})$$

eights of the items in interaction sequence;

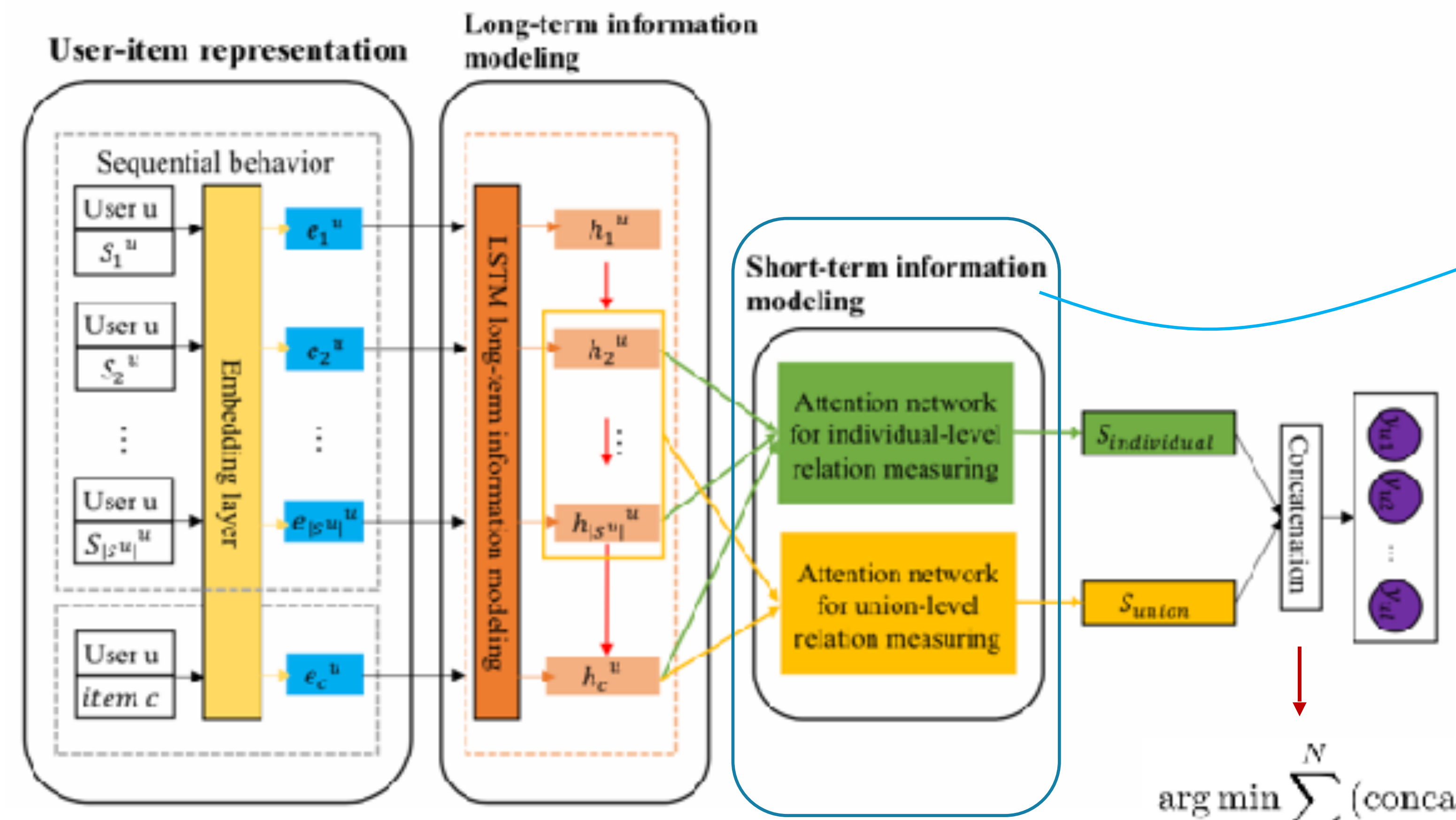
2. Calculate the likelihood of next items

$\alpha_i \in [0, 1]$ is the weight of $h_{t-j}^{u_i}$ and $h_{t-j}^{u_i} \in h^{(1)}$

$$s_k = \text{softmax}(\beta_1 h_{t-j}^{u} + \beta_2 h_{c_i}^{u} + b)$$

$$S_{individual} = \sum_{i=1}^{n-1} \alpha_i \cdot s_k$$

UNIFIED FRAMEWORK:



Union-level relation measuring:
1. Learn the attention weights of the items in interaction sequence;
2. Calculate the likelihood of next items

$$s_m = \mathrm{softmax}(\beta_3 W_i + \beta_4 h^u_{c_i} + b)$$

$$S_{union} = \sum_{i=1}^{n-1} \alpha_i \cdot s_m$$

$$\arg\min_{\Theta} \sum_{i=1}^{N} (\mathrm{concatenate}(S^{(i)}_{individual}, S^{(i)}_{union}) - y_i)^2 + \frac{\lambda}{2}\|\Theta\|^2$$

PART FOUR

Evaluation

EXPERIMENT

Movie&TV dataset :

Table 1. Statistical information of dataset.

Movies&TV	Users	Items	Interactions
before-processing	40929	51510	1163413
after-processing	35168	51227	1070645

Results:

Methods	Movie&TV	
Measures@20	MRR	NDCG
BPR-MF	0.0089	0.0248
TranRec	0.0155	0.0392
GRU4Rec	0.0124	0.0344
FPMC	0.0162	0.0406
MARank	0.0170	0.0444
JDR-L	0.0179	0.0518
Improvement	5.2%	16.7%

Baselines:

- BPR-MF: matrix factorization method for Sequential RS
- TranRec: typical representation-based Sequential RS
- GRU4Rec: Gated Recurrent Unit for sequential RS
- FPMC: Markov chain method modeling individual item interaction
- MARank: models both individual-level and union-level interactions with factorization model.

Methods	Movie&TV	
Measures@20	MRR	NDCG
LSTM-only	0.0154	0.0447
LSTM+ individual-level item interaction	0.0147	0.0442
LSTM+ union-level item interaction	0.0142	0.0423
JDR-L	0.0178	0.0518

PART FIVE

Conclusion

CONCLUSION

- We design a Joint Relational Dependency learning (JRD-L) for sequential recommendation.
- JDR-L model can unify both long-term dependencies and short-term dependencies from individual-level and union-level.
- Moreover, JDR-L can handle the sparsity problem when exploiting the individual-level relation information from the sequential behaviors.
- Extensive experiments on the benchmark dataset demonstrate the effectiveness of JRD-L.

THANKS