The 23rd AAAI Conference on Artificial Intelligence (AAAI-19)

Session-based Recommendation with Graph Neural Networks

Presented by **Yanqiao ZHU** sxkdz@tongji.edu.cn

Microsoft Research 微软亚洲研究院

Center for Research on Intelligent Perception and Computing
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences

Joint work with Shu WU, Yuyuan TANG, Liang WANG, Xing XIE, and Tieniu TAN

Outline

1. Preamble
2. The Proposed Method
3. Experiments and Analysis
4. Conclusions

1

Preamble

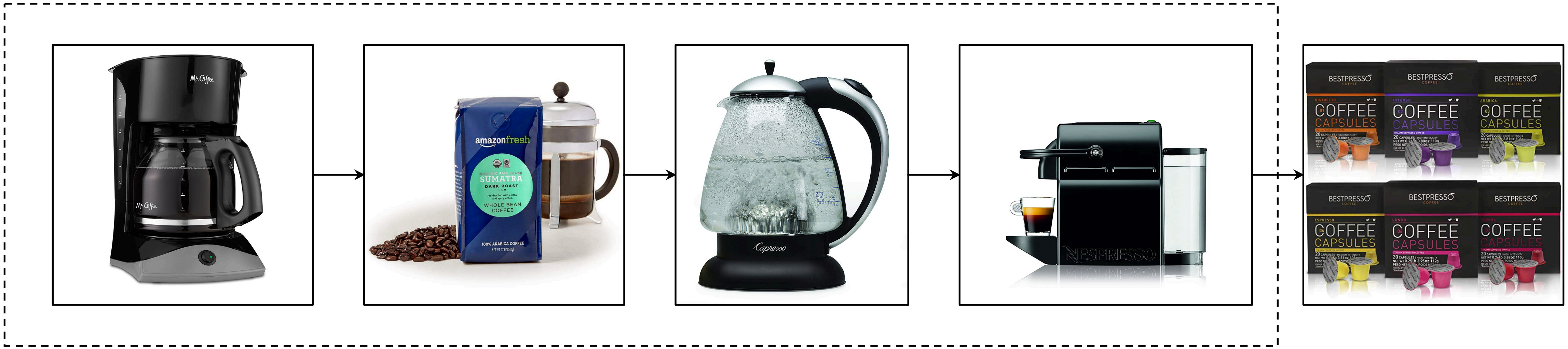
Session-based Recommendation with Graph Neural Networks

The 23rd AAAI Conference on Artificial Intelligence (AAAI-19)

Session-based Recommendation

- Recommendation systems help users find relevant items that meet their interests.
- Previous recommendation systems rely on long-term user profiles to make recommendations.
  - However, in real-world applications, long-term profiles may not exist.
  - Only user behavior during an ongoing session is available.

Session-based Recommendation (cont.)



Previously visited items

Item to be recommended

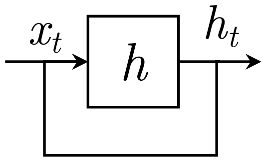
- No information about the actual user.
- Only **timestamp** and (possibly limited) clicked **items** available.

Recurrent Neural Networks (RNNs)

- Recently, many proposals based on RNNs have been developed for session-based recommendation.

- Hidden state
  - Next hidden state depends on the input and the current hidden state

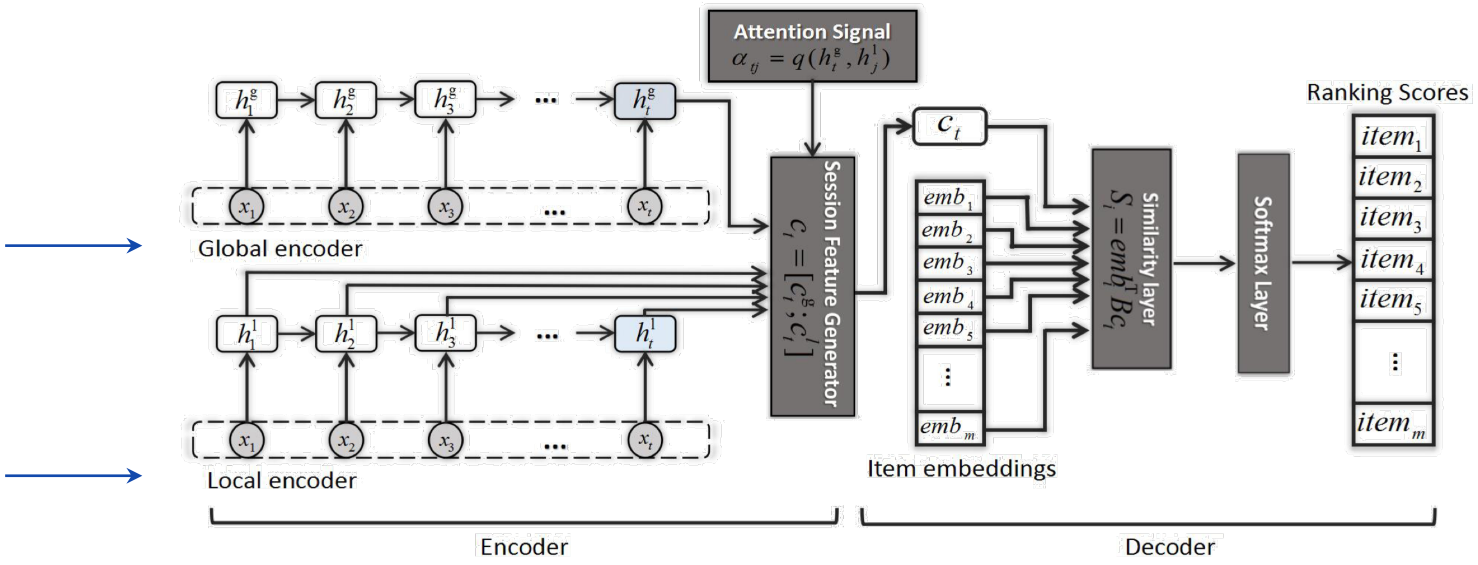
$$h_t = \tanh(Wx_t + Uh_{t-1})$$



- RNNs can be of arbitrary (infinite) depths
- Optimizing via back-propagation through time (BPTT)

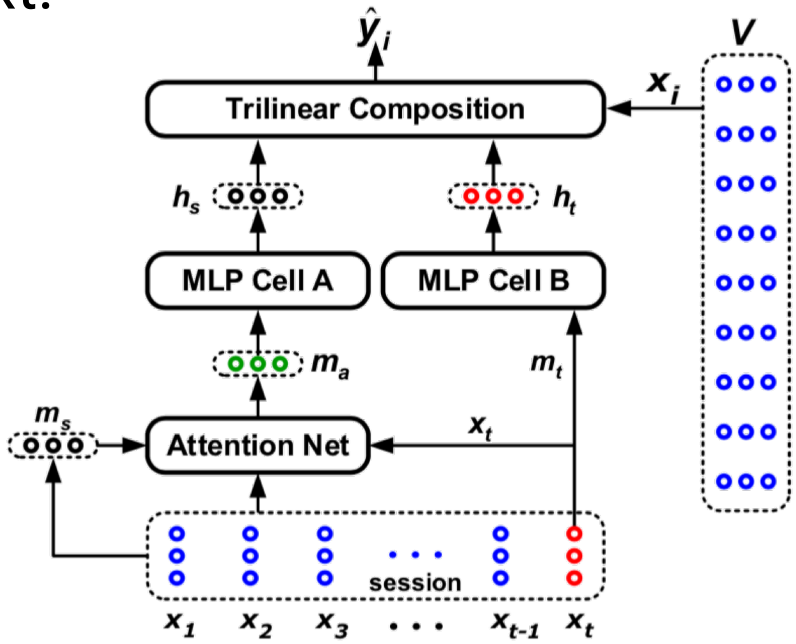
Recent Progress

- NARM: Neural Attentive Recommendation Machine [Li et al. 2017a]
  - For the global recommender, the user behavior in one session is inadequate and estimating user representations may not be sufficient.



Recent Progress (cont.)

- STAMP: Short-Term Attention/Memory Priority Model [Liu et al. 2018]
  - An attentive model for next-click prediction
  - Only models single-way transitions between consecutive items and neglects the context.

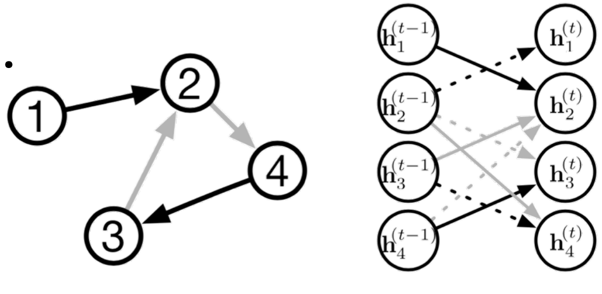


Motivations

- How to effectively capture the item transitions in session sequences?
- To facilitate recommendation, how to obtain accurate item embeddings and session embeddings?

Graph-based Neural Networks

- Graph Neural Networks (GNNs) [Scarselli et al. 2009]
    - Propagation: computes representation for each node.
    - Output mapping: maps from node representations and corresponding labels to an output.
    - Model training via Almeida-Pineda algorithm
- Gated Graph Neural Networks (GGNNs) [Li et al. 2016]
    - Uses gated recurrent units.
    - Unrolls the recurrence for **a fixed number of steps**.
    - Computes gradients through Backpropagation through time.



(a) (b)

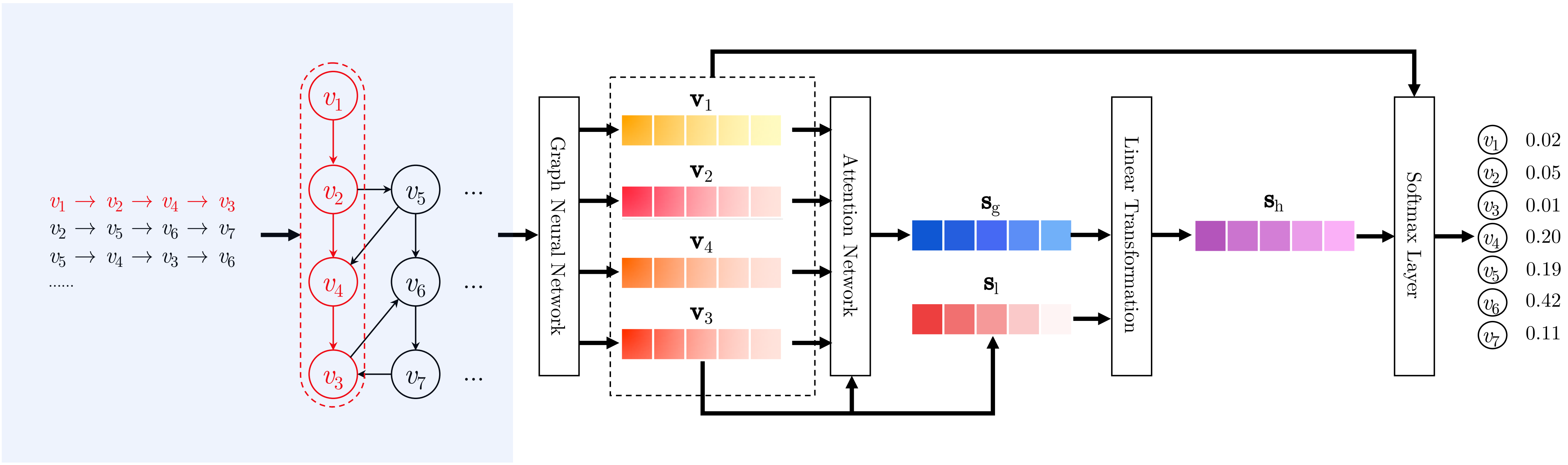
2

The Proposed Method

Session-based Recommendation with Graph Neural Networks

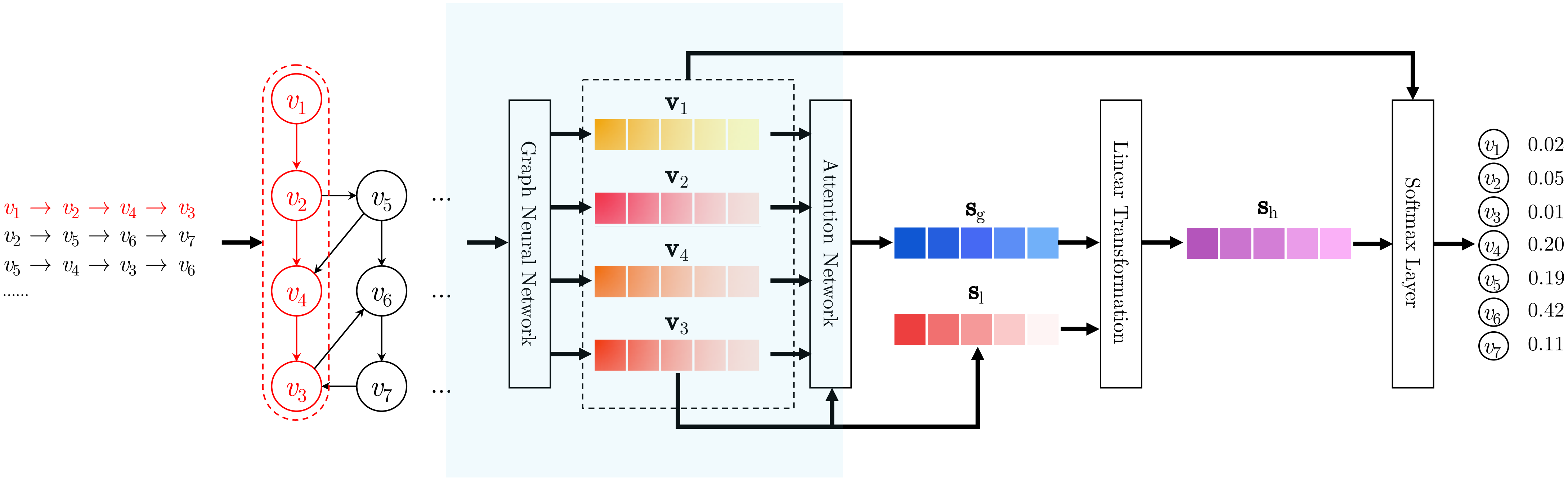
The 23rd AAAI Conference on Artificial Intelligence (AAAI-19)

An Overview of Our Approach



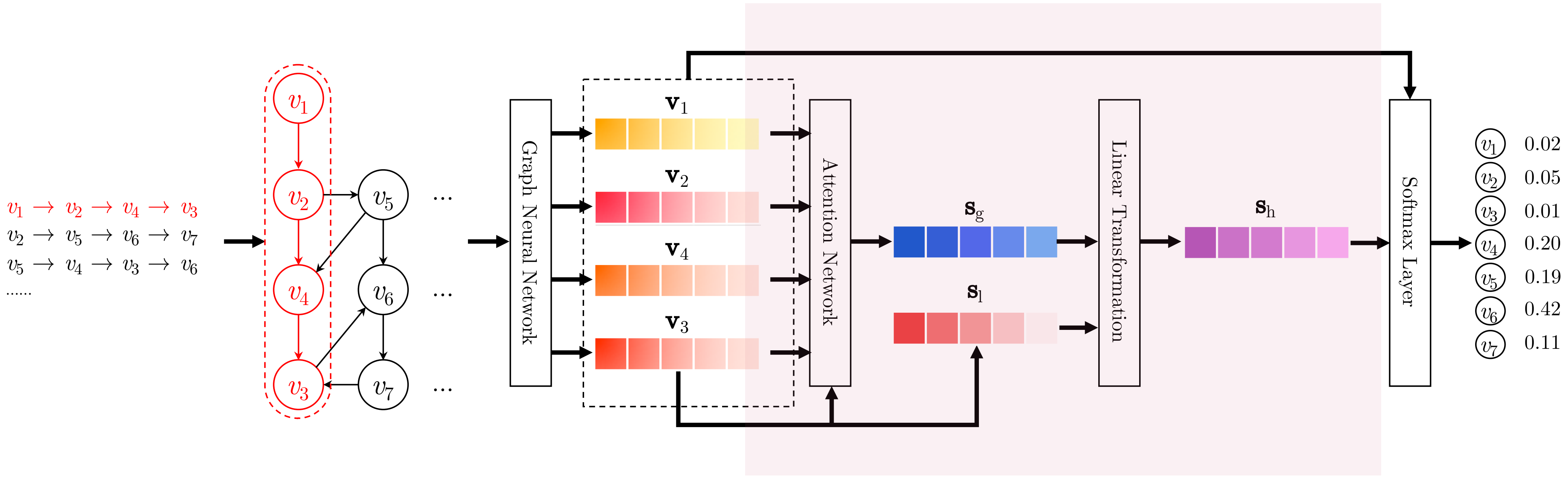
(a) **Session graph modeling**

An Overview of Our Approach



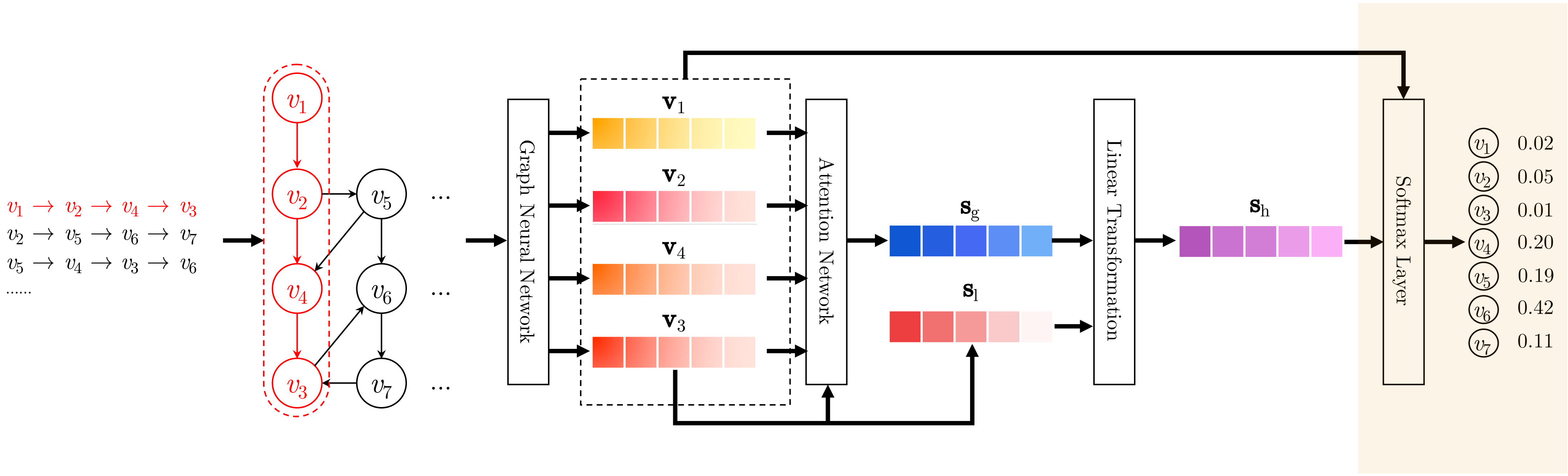
(b) **Node representation learning**

An Overview of Our Approach



(c) **Session representation generating**

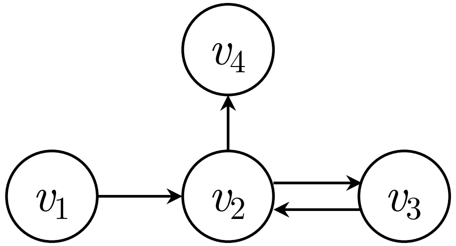
An Overview of Our Approach



(d) **Making recommendation**

Constructing Session Graphs

- Each session sequence s is modeled as a directed graph $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$.
- Edge weight normalization: the occurrence of the edge divided by the outdegree of that edge's start node



$$s = [v_1, v_2, v_3, v_2, v_4]$$

normalization $\longrightarrow$

	1	2	3	4
1	0	1	0	0
2	0	0	1/2	1/2
3	0	1	0	0
4	0	0	0	0

Learning Item Embeddings on Graphs

- We adopt GGNNs for learning unified representations for all nodes in session graphs.

- Propagation rules:

connection matrix

$$\mathbf{a}_{s,i}^{t} = \mathbf{A}_{s,i:} \left[\mathbf{v}_1^{t-1}, \ldots, \mathbf{v}_n^{t-1}\right]^{\top} \mathbf{H} + \mathbf{b},$$

$$\mathbf{z}_{s,i}^{t} = \sigma\left(\mathbf{W}_z \mathbf{a}_{s,i}^{t} + \mathbf{U}_z \mathbf{v}_i^{t-1}\right),$$ Reset gate

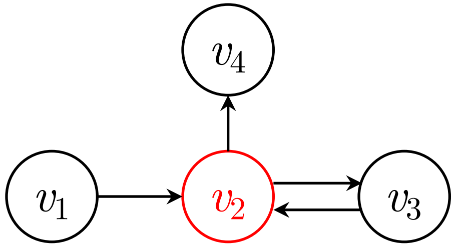
$$\mathbf{r}_{s,i}^{t} = \sigma\left(\mathbf{W}_r \mathbf{a}_{s,i}^{t} + \mathbf{U}_r \mathbf{v}_i^{t-1}\right),$$ Update gate

$$\widetilde{\mathbf{v}_i^{t}} = \tanh\left(\mathbf{W}_o \mathbf{a}_{s,i}^{t} + \mathbf{U}_o \left(\mathbf{r}_{s,i}^{t} \odot \mathbf{v}_i^{t-1}\right)\right),$$ Candidate

$$\mathbf{v}_i^{t} = \left(1 - \mathbf{z}_{s,i}^{t}\right) \odot \mathbf{v}_i^{t-1} + \mathbf{z}_{s,i}^{t} \odot \widetilde{\mathbf{v}_i^{t}}.$$ Final representation

The Connection Matrix

- The connection matrix $\mathbf{A}_s \in \mathbb{R}^{n \times 2n}$ determines how nodes within the graph communicate with each other, which is defined as a concatenation of two adjacency matrices $\mathbf{A}_s^{(\mathrm{out})}$ and $\mathbf{A}_s^{(\mathrm{in})}$.
- $\mathbf{A}_{s,i:} \in \mathbb{R}^{1 \times 2n}$ are the two columns of blocks in $\mathbf{A}_s$ corresponding to node $v_{s,i}$.



$\mathbf{A}_{s,i:} \in \mathbb{R}^{1 \times 2n}$

	Outgoing edges				Incoming edges			
	1	2	3	4	1	2	3	4
1	0	1	0	0	0	0	0	0
2	0	0	1/2	1/2	1/2	0	1/2	0
3	0	1	0	0	0	1	0	0
4	0	0	0	0	0	1	0	0

Generating Session Embeddings

- A session is represented directly by node embedding involved in that session.
- **Local embedding**

$$\mathbf{s}_{\mathrm{l}} = \mathbf{v}_n$$

- **Global embedding**

$$\alpha_i = \mathbf{q}^\top \sigma(\mathbf{W}_1 \mathbf{v}_n + \mathbf{W}_2 \mathbf{v}_i + \mathbf{c}),$$

$$\mathbf{s}_{\mathrm{g}} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$$

- **Hybrid embedding**

$$\mathbf{s}_{\mathrm{h}} = \mathbf{W}_3 \left[\mathbf{s}_l; \mathbf{s}_g\right]$$

Making Recommendation

- Compute the score for each candidate item by dot product session embeddings with item embeddings:

$$\hat{\mathbf{y}} = \mathrm{softmax}\left(\mathbf{s}_{\mathrm{h}}^{\top} \mathbf{v}_i\right)$$

- The cross-entropy loss function:

$$\mathcal{L}(\hat{\mathbf{y}}) = -\sum_{i=1}^{m} \mathbf{y}_i \log\left(\hat{\mathbf{y}}_i\right) + \left(1 - \mathbf{y}_i\right) \log\left(1 - \hat{\mathbf{y}}_i\right)$$

3

Experiments and Analysis

Session-based Recommendation with Graph Neural Networks

The 23rd AAAI Conference on Artificial Intelligence (AAAI-19)

Experiment Setup

- Datasets
  - **Yoochoose 1/64** and **Yoochoose 1/4** from RecSys Challenge 2014
  - **Diginetica** from CIKM Cup 2016
- Baselines
  - **POP** and **S-POP**
  - **Item-KNN** [Sarwar et al. 2001]
  - **BPR-MF** [Rendle et al. 2009]
  - **FPMC** [Rendle et al. 2010]
  - **GRU4REC** [Hidasi et al. 2016]
  - **NARM** [Li et al. 2017a]
  - **STAMP** [Liu et al. 2018]

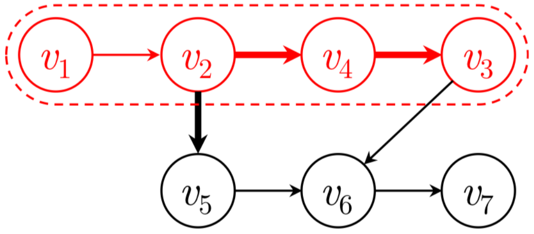
Comparison with Baselines

Method	Yoochoose 1/64		Yoochoose 1/4		Diginetica	
	P@20	MRR@20	P@20	MRR@20	P@20	MRR@20
POP	6.71	1.65	1.33	0.30	0.89	0.20
S-POP	30.44	18.35	27.08	17.75	21.06	13.68
Item-KNN	51.60	21.81	52.31	21.70	35.75	11.57
BPR-MF	31.31	12.08	3.40	1.57	5.24	1.98
FPMC	45.62	15.01	–	–	26.53	6.95
GRU4REC	60.64	22.89	59.53	22.60	29.45	8.33
NARM	68.32	28.63	69.73	29.23	49.70	16.17
STAMP	68.74	29.67	70.44	30.00	45.64	14.32
SR-GNN	**70.57**	**30.94**	**71.36**	**31.89**	**50.73**	**17.59**

Variants of Connection Schemes

- Since user behavior in sessions is limited, we propose two connection schemes to augment relationships between items in each session graph:
  - (a) **SR-GNN-NGC** aggregates all session sequences together and model them as a directed global item graph.

$s_1 = [v_1, v_2, v_4, v_3]$
$s_2 = [v_2, v_4, v_6, v_7]$
$s_3 = [v_2, v_5, v_3, v_6]$
……

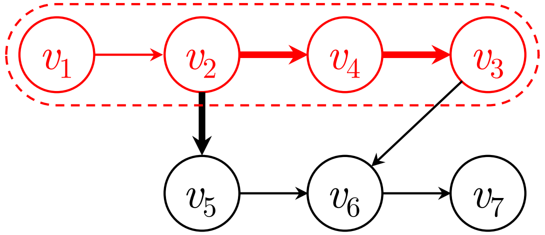


	$\mathbf{A}_g^{(out)}$							$\mathbf{A}_g^{(in)}$						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	1/2	1/2	0	0	1	0	0	0	0	0	0
3	0	0	0	0	0	1	0	0	0	0	1	0	0	0
4	0	0	1	0	0	0	0	0	1	0	0	0	0	0
5	0	0	0	0	0	1	0	0	1	0	0	0	0	0
6	0	0	0	0	0	0	1	0	0	1/2	0	1/2	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	1	0

Variants of Connection Schemes (cont.)

- Since user behavior in sessions is limited, we propose two connection schemes to augment relationships between items in each session graph:
  - (b) **SR-GNN-FC** models all high-order relationships between items within one session as direct connections explicitly.

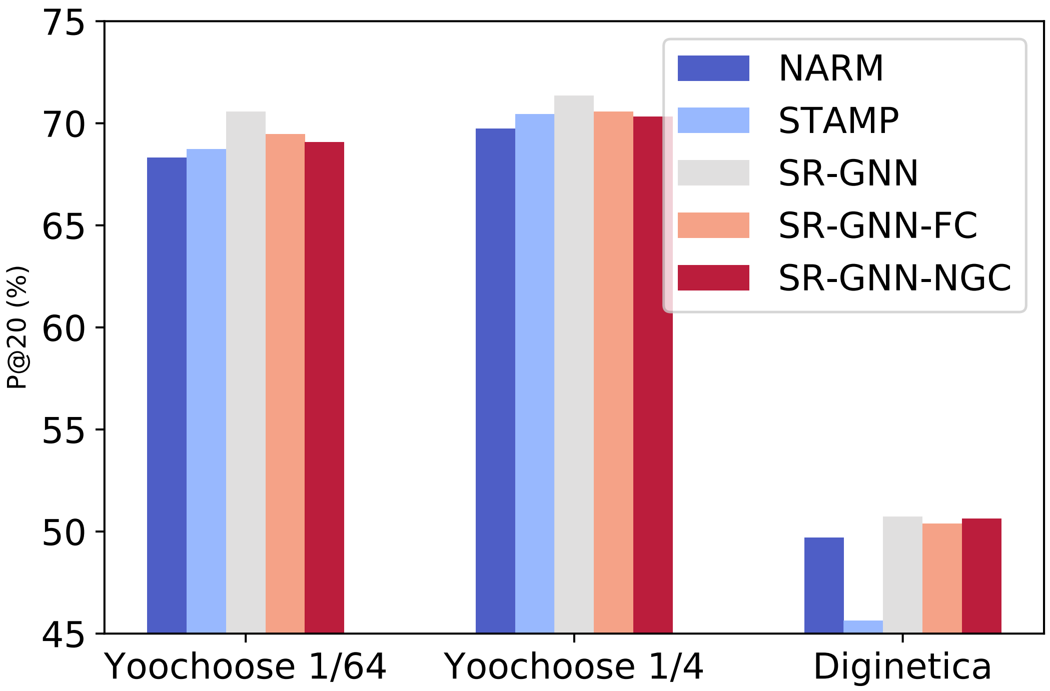
$$s_1 = [v_1,\ v_2,\ v_4,\ v_3]$$
$$s_2 = [v_2,\ v_4,\ v_6,\ v_7]$$
$$s_3 = [v_2,\ v_5,\ v_3,\ v_6]$$
$$\ldots\ldots$$



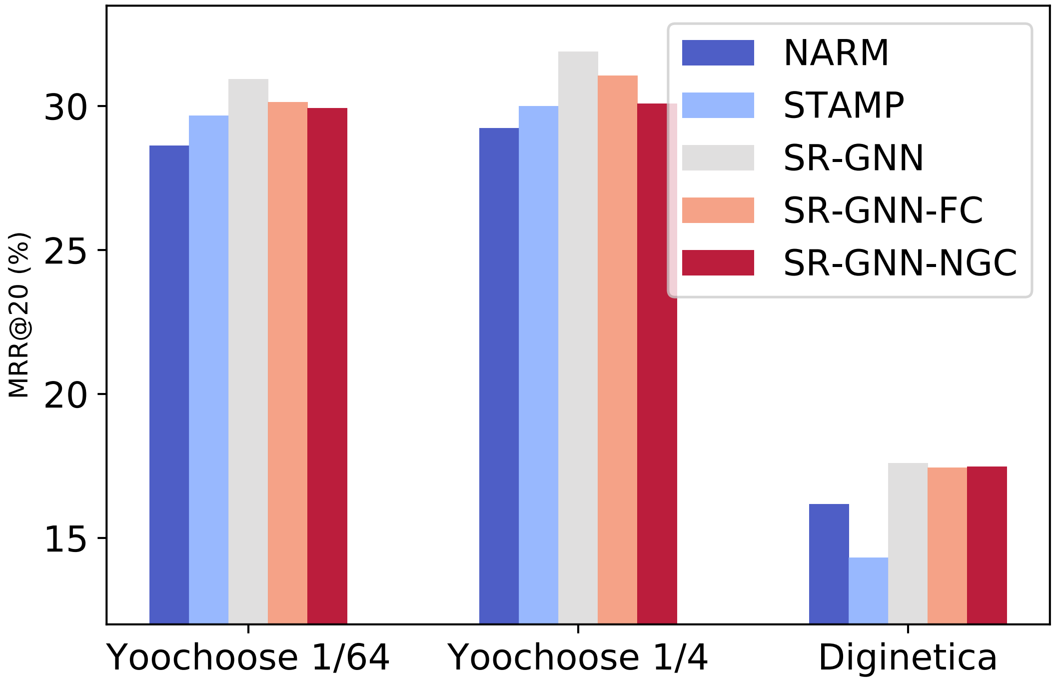
	$\mathbf{A}_g^{(out)}$				$\mathbf{A}_g^{(in)}$				$\mathbf{A}_g^{(FC)}$			
	1	2	3	4	1	2	3	4	1	2	3	4
1	0	1	0	0	0	0	0	0	1	0	1	1
2	0	0	0	1	1	0	0	0	1	1	1	0
3	0	0	0	0	0	0	0	1	1	1	1	1
4	0	0	1	0	0	1	0	0	1	1	0	1

Comparison with Connection Schemes

Precision@20



MRR@20

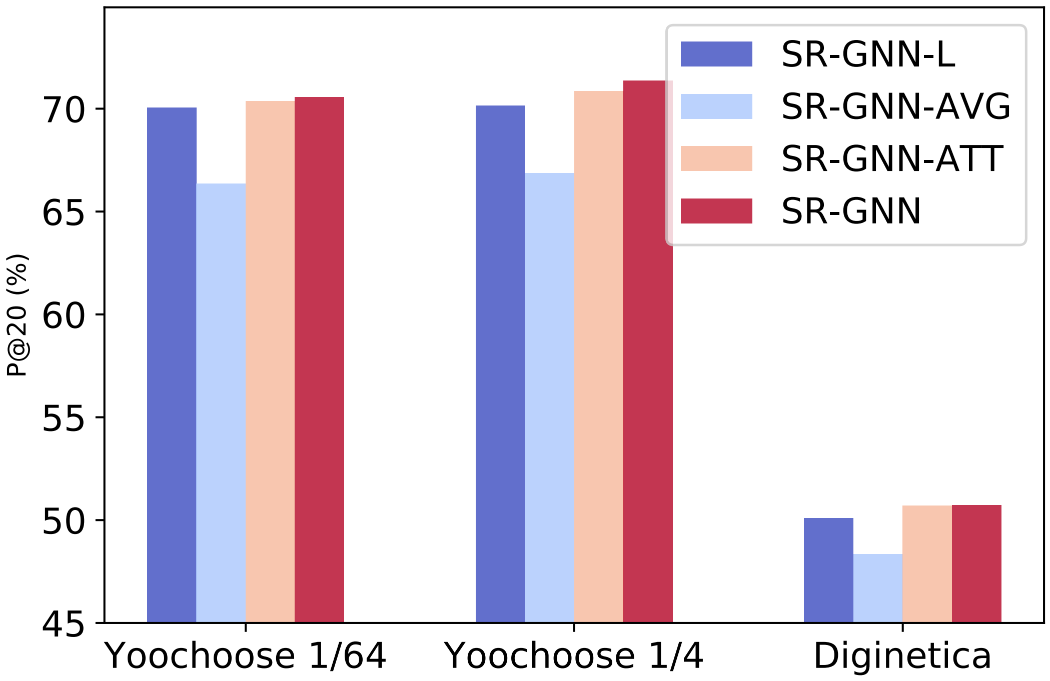


Variants of Session Representations

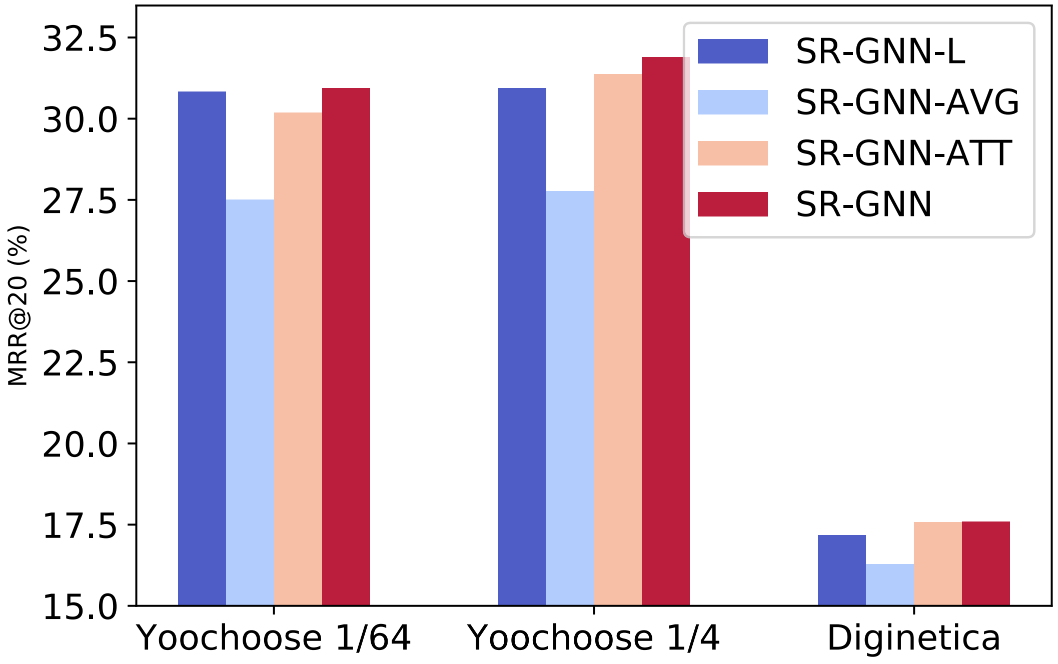
- Ablation study on session representations:
  - (a) **SR-GNN-L**: local embedding only
  - (b) **SR-GNN-AVG**: global embedding with average pooling
  - (c) **SR-GNN-ATT**: global embedding with attention networks

Comparison of Session Representations

Precision@20



MRR@20



Comparison of Sequence Lengths

- **Short** group: session lengths ≤ 5
- **Long** group: session lengths > 5

- Yoochoose 1/64
  - Short (70.1%)
  - Long (29.9%)
- Diginetica
  - Short (76.4%)
  - Long (23.6%)

Comparison of Sequence Lengths (cont.)

Method	Yoochoose 1/64		Diginetica	
	Short	Long	Short	Long
NARM	**71.44**	60.79	**51.22**	45.75
STAMP	70.69	64.73	47.26	40.39
SR-GNN-L	70.11	69.73	49.04	50.97
SR-GNN-ATT	70.31	70.64	50.35	51.05
SR-GNN	70.47	**70.70**	50.49	**51.27**

Precision@20

4

Concluding Remarks

Session-based Recommendation with Graph Neural Networks

The 23rd AAAI Conference on Artificial Intelligence (AAAI-19)

Wrapping Up

1. Session-based recommendation is indispensable where users' preference and historical records are hard to obtain.
2. We present a novel architecture for session-based recommendation that incorporates graph models into representing session sequences.
3. The proposed method not only considers the complex structure and transitions between items of session sequences, but also develops a strategy to combine long-term preferences and current interests of sessions to better predict users' next actions.
4. Comprehensive experiments confirm that the proposed algorithm can consistently outperform other state-of-art methods.

Bibliographies

- [Hidasi et al. 2016] B. Hidasi, A. Karatzoglou, L. Baltrunas, D. Tikk. Session-based Recommendations with Recurrent Neural Networks. In *ICLR*, 2016.
- [Li et al. 2016] Y. Li, D. Tarlow, M. Brockschmidt, R. Zemel. Gated Graph Sequence Neural Networks. In *ICLR*, 2016.
- [Li et al. 2017] J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian, and J. Ma, Neural Attentive Session-based Recommendation. In *CIKM*, 2017, 1419–1428.
- [Liu et al. 2018] Q. Liu, Y. Zeng, R. Mokhosi, and H. Zhang. STAMP: Short-Term Attention/Memory Priority Model for Session-based Recommendation. In *KDD*, 2018, 1831–1839.
- [Rendle et al. 2009] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *UAI*, 2009, 452–461.
- [Rendle et al. 2010] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme. Factorizing Personalized Markov Chains for Next-basket Recommendation. In *WWW*, 2010, 811–820.
- [Sarwar et al. 2001] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *WWW*, 2001, 285–295.
- [Scarselli et al. 2009] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini. The Graph Neural Network Model. In *T-NN*, 20 (1), 61–80, 2009.

Paper

Code

Thank You

Yanqiao ZHU sxkdz@tongji.edu.cn

Center for Research on Intelligent Perception and Computing
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences