ACM SIGIR Conf. on Information Retrieval (SIGIR 2020)

Target Attentive Graph Neural Networks for Session-based Recommendation

Presented by **Yanqiao ZHU**

yanqiao.zhu@cripac.ia.ac.cn

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

Joint work with Feng YU, Qiang LIU, Shu WU, Liang WANG, and Tieniu TAN

Outline

1. Preamble
2. The Proposed Method
3. Experiments
4. Concluding Remarks

Outline

1. **Preamble**
2. The Proposed Method
3. Experiments
4. Concluding Remarks

Session-based Recommendation

Previously visited items

Item to be recommended

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

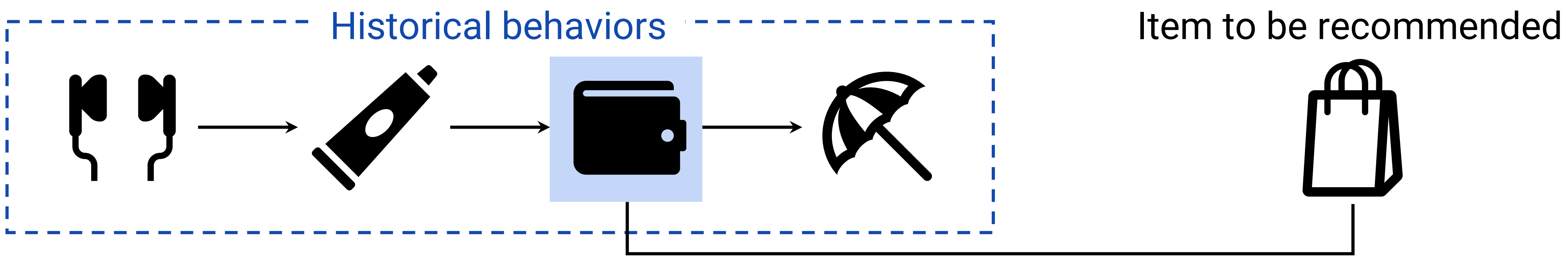
Graph-based Recommender Models

- Limitation of previous sequential methods:
  - Sequence-based methods only model sequential transitions between consecutive items, with complex transitions neglected.
  - Suppose a session is $s = v_1 \rightarrow v_2 \rightarrow v_1 \rightarrow v_3$. It is hard to capture such a to-and-fro relationship between item v_1 and items (v_2, v_3).
- SR-GNN [Wu et al. 2019] proposes to model sessions as **graphs**.
  - Discover the complex transitional patterns underneath sessions through session graphs.
  - This natural means of encoding rich temporal patterns within sessions produces more accurate representations for items.

[Wu et al. 2019] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, Session-based Recommendation with Graph Neural Networks, in *AAAI*, 2019, pp. 346–353.

Representing Diverse User Interests

- Representing one session using one fixed embedding vector cannot express users' **diverse** interests, considering abundant candidate items and user behaviors.
- The interests of a user with rich behaviors can be **specifically activated** given a target item [Zhou et al. 2018].

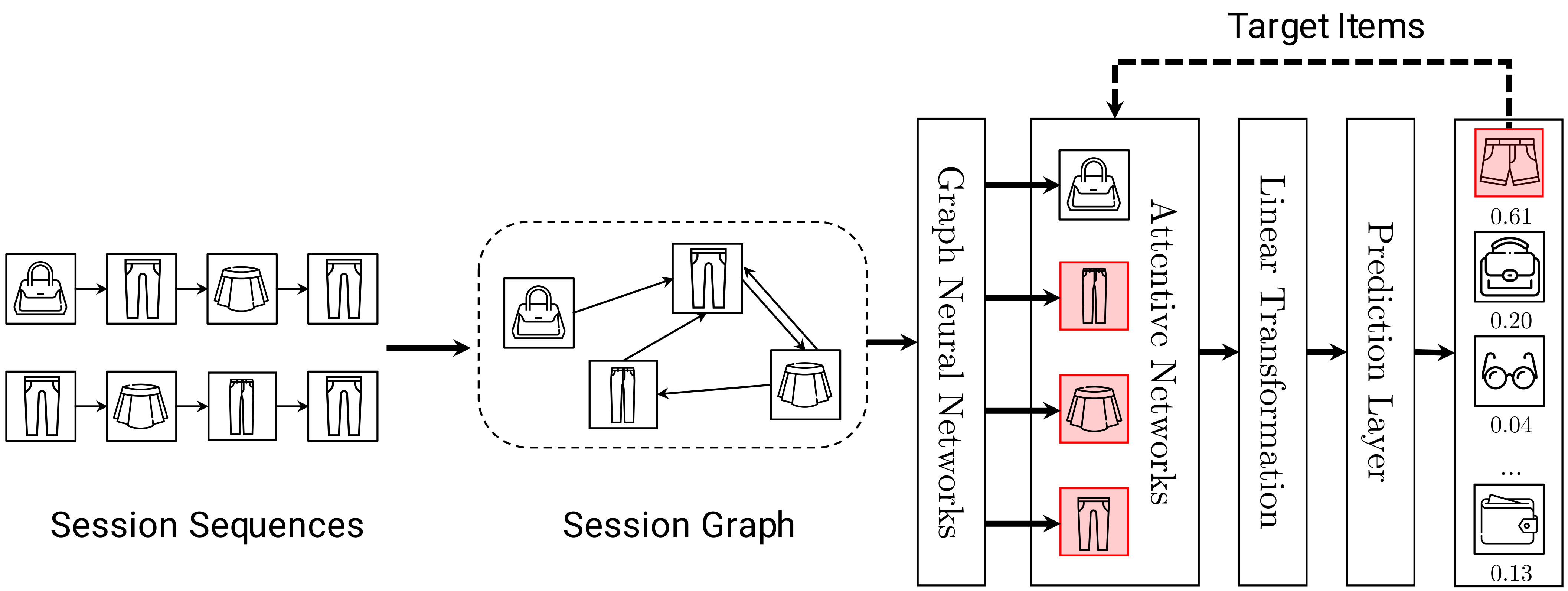


[Zhou et al. 2018] G. Zhou, X. Zhu, C. Song, Y. Fan, H. Zhu, X. Ma, Y. Yan, J. Jin, H. Li, and K. Gai, Deep Interest Network for Click-Through Rate Prediction, in *KDD*, 2018, pp. 1059—1068.

Outline

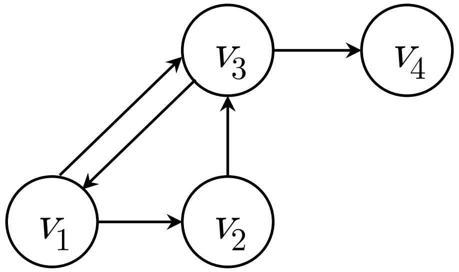
1. Preamble
2. **The Proposed Method**
3. Experiments
4. Concluding Remarks

Model: Target Attentive GNN



Graph Modeling

- Each session sequence is modeled as a directed graph.
  - Edge weight normalization: the occurrence of the edge divided by the outdegree of that edge's start node.
  - Treat incoming and outgoing edges separately to reflect bidirectional relationships.



Session graph $\mathcal{G}_s$

	Outgoing edges				Incoming edges			
	1	2	3	4	1	2	3	4
1	0	1/2	1/2	0	0	0	1	0
2	0	0	1	0	1	0	0	0
3	1/2	0	0	1/2	1/2	1/2	0	0
4	0	0	0	0	0	0	1	0
	$\boldsymbol{A}_s^{(\text{in})}$				$\boldsymbol{A}_s^{(\text{out})}$			

Adjacency matrix $\boldsymbol{A}_s$

Learning Item Embeddings

- We use Gated GNNs [Li et al. 2016] to learn item representations for all nodes in session graphs.
  - The GGNN model aggregates features from neighboring nodes and adopts memory units to better preserve useful information.

$$
\begin{aligned}
\boldsymbol{a}_{s,i}^{(t)} &= \boldsymbol{A}_{s,i:}\left[\boldsymbol{v}_1^{(t-1)},\ldots,\boldsymbol{v}_{s_n}^{(t-1)}\right]^\top \boldsymbol{H}+\boldsymbol{b},\\
\boldsymbol{z}_{s,i}^{(t)} &= \sigma\left(\boldsymbol{W}_z\boldsymbol{a}_{s,i}^{(t)}+\boldsymbol{U}_z\boldsymbol{v}_i^{(t-1)}\right),\\
\boldsymbol{r}_{s,i}^{(t)} &= \sigma\left(\boldsymbol{W}_r\boldsymbol{a}_{s,i}^{(t)}+\boldsymbol{U}_r\boldsymbol{v}_i^{(t-1)}\right),\\
\widetilde{\boldsymbol{v}_i^{(t)}} &= \tanh\left(\boldsymbol{W}_o\boldsymbol{a}_{s,i}^{(t)}+\boldsymbol{U}_o\left(\boldsymbol{r}_{s,i}^{(t)}\odot\boldsymbol{v}_i^{(t-1)}\right)\right),\\
\boldsymbol{v}_i^{(t)} &= \left(1-\boldsymbol{z}_{s,i}^{(t)}\right)\odot\boldsymbol{v}_i^{(t-1)}+\boldsymbol{z}_{s,i}^{(t)}\odot\widetilde{\boldsymbol{v}_i^{(t)}},
\end{aligned}
$$

Constructing Target-Aware Embeddings

- We introduce a novel target attention mechanism to calculate soft attention scores over all items with respect to each target item to **adaptively** capture relevant historical behaviors.
- The obtained target embedding for representing users' interests varies with different target items.

$$\beta_{i,t} = \operatorname{softmax}(e_{i,t}) = \frac{\exp\left(\boldsymbol{v}_t^\top \boldsymbol{W} \boldsymbol{v}_i\right)}{\sum_{j=1}^{m} \exp\left(\boldsymbol{v}_t^\top \boldsymbol{W} \boldsymbol{v}_j\right)}$$

$$\boldsymbol{s}_{\text{target}}^{t} = \sum_{i=1}^{s_n} \beta_{i,t} \boldsymbol{v}_i$$

Generating Session Embeddings

- We represent a session by node embeddings involved in that session, along with the user's target embedding.
  - **Local embedding** (short-term preference)

$$\boldsymbol{s}_{\text{local}} = \boldsymbol{v}_{s_n}$$

  - **Global embedding** (long-term preference)

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

$$\boldsymbol{s}_{\text{global}} = \sum_{i=1}^{s_n} \alpha_i \boldsymbol{v}_i.$$

  - **Final representation**

$$\boldsymbol{s}_t = \boldsymbol{W}_3 [\boldsymbol{s}^t_{\text{target}}; \boldsymbol{s}_{\text{local}}; \boldsymbol{s}_{\text{global}}]$$

Making Recommendation

- We compute the recommendation score for each item by taking inner-product of item embeddings and session embeddings.

$$\hat{z}_t = \boldsymbol{s}_t^\top \boldsymbol{v}_t,$$
$$\hat{\boldsymbol{y}} = \mathrm{softmax}\left(\hat{\boldsymbol{z}}\right).$$

- For training the model, we define the loss function as the cross-entropy of the prediction and the ground truth.

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

Outline

1. Preamble
2. The Proposed Method
3. **Experiments**
4. Concluding Remarks

Experiment Setup

- Dataset statistics

Statistics	Yoochoose 1/64	Diginetica
# Clicks	557,248	982,961
# Training sessions	369,859	719,470
# Test sessions	55,898	60,858
# Unique items	17,377	43,097
Average length	6.16	5.12

Experiment Setup (cont.)

- Baselines
  - Frequency-based POP and S-POP
  - Similarity-based Item-KNN [Sarwar et al. 2001]
  - Bayesian personalized ranking (BPR-MF) [Rendle et al. 2009]
  - Factorizing personalized Markov chain model (FPMC) [Rendle et al. 2010]
  - RNN-based GRU4REC [Hidasi et al. 2016]
  - Neural attentive recommender model (NARM) [Li et al. 2017]
  - Short-term attention/memory priority model (STAMP) [Liu et al. 2018]
  - Graph-based SR-GNN [Wu et al. 2019]

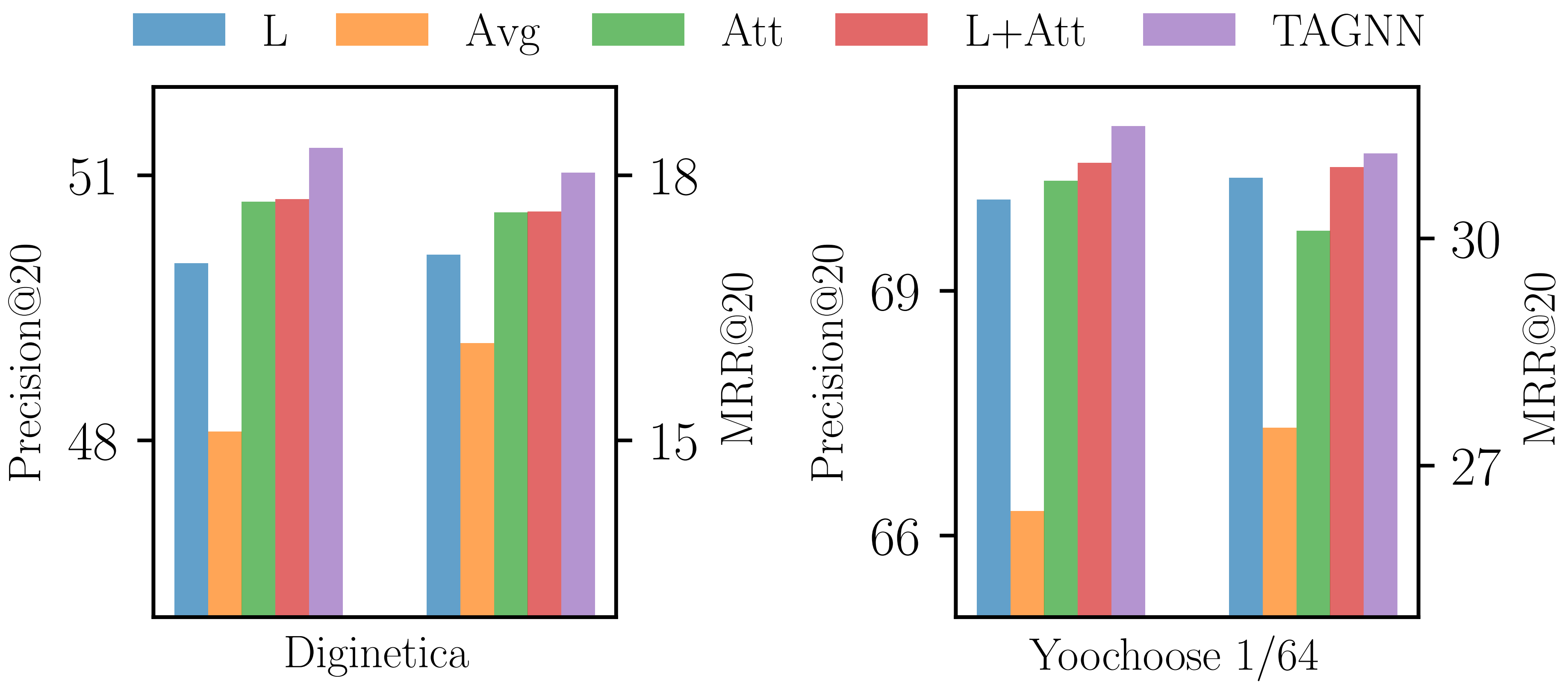
Comparison with Baselines

Method	Diginetica		Yoochoose 1/64	
	Recall@20	MRR@20	Recall@20	MRR@20
POP	0.89	0.20	6.71	1.65
S-POP	21.06	13.68	30.44	18.35
Item-KNN	35.75	11.57	51.60	21.81
BPR-MF	5.24	1.98	31.31	12.08
FPMC	26.53	6.95	45.62	15.01
GRU4REC	29.45	8.33	60.64	22.89
NARM	49.70	16.17	68.32	28.63
STAMP	45.64	14.32	68.74	29.67
SR-GNN	50.73	17.59	70.57	30.94
TAGNN	**51.31**	**18.03**	**71.02**	**31.12**

Ablation Studies

- Model variants to analyze how different representations for user preference affect performance:
  - (a) local embedding only (TAGNN-L)
  - (b) global embedding using average pooling only (TAGNN-Avg)
  - (c) attentive global embedding (TAGNN-Att)
  - (d) local embedding plus attentive global embedding (TAGNN-L+Att)

Ablation Studies (cont.)



Outline

1. Preamble

2. The Proposed Method

3. Experiments

4. **Concluding Remarks**

Wrapping Up

1. Session-based recommendation is indispensable where users' preference and historical records are hard to obtain.
2. We have developed a novel target attentive graph neural network model for session-based recommendation.
3. By incorporating graph modeling and a target-aware attention module, the proposed TAGNN jointly considers user interests given a certain target item as well as complex item transitions in sessions.
4. Extensive experiments on real-world benchmark datasets demonstrate the effectiveness of our model.

Acknowledgements

- This work is jointly supported by
  - National Key Research and Development Program under grants No. 2018YFB1402600 and No. 2016YFB1001000
  - National Natural Science Foundation of China under grants No. U19B2038 and No. 61772528

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, pp. 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, pp. 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, pp. 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, pp. 811–820.
- [Sarwar et al. 2001] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, Item-based Collaborative Filtering Recommendation Algorithms, in *WWW*, 2001, pp. 285–295.
- [Wu et al. 2019] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, Session-based Recommendation with Graph Neural Networks, in *AAAI*, 2019, pp. 346–353.
- [Zhou et al. 2018] G. Zhou, X. Zhu, C. Song, Y. Fan, H. Zhu, X. Ma, Y. Yan, J. Jin, H. Li, and K. Gai, Deep Interest Network for Click-Through Rate Prediction, in *KDD*, 2018, pp. 1059—1068.


THANKS
Code
Paper
Slides