Multi-order Attentive Ranking Model for Sequential Recommendation

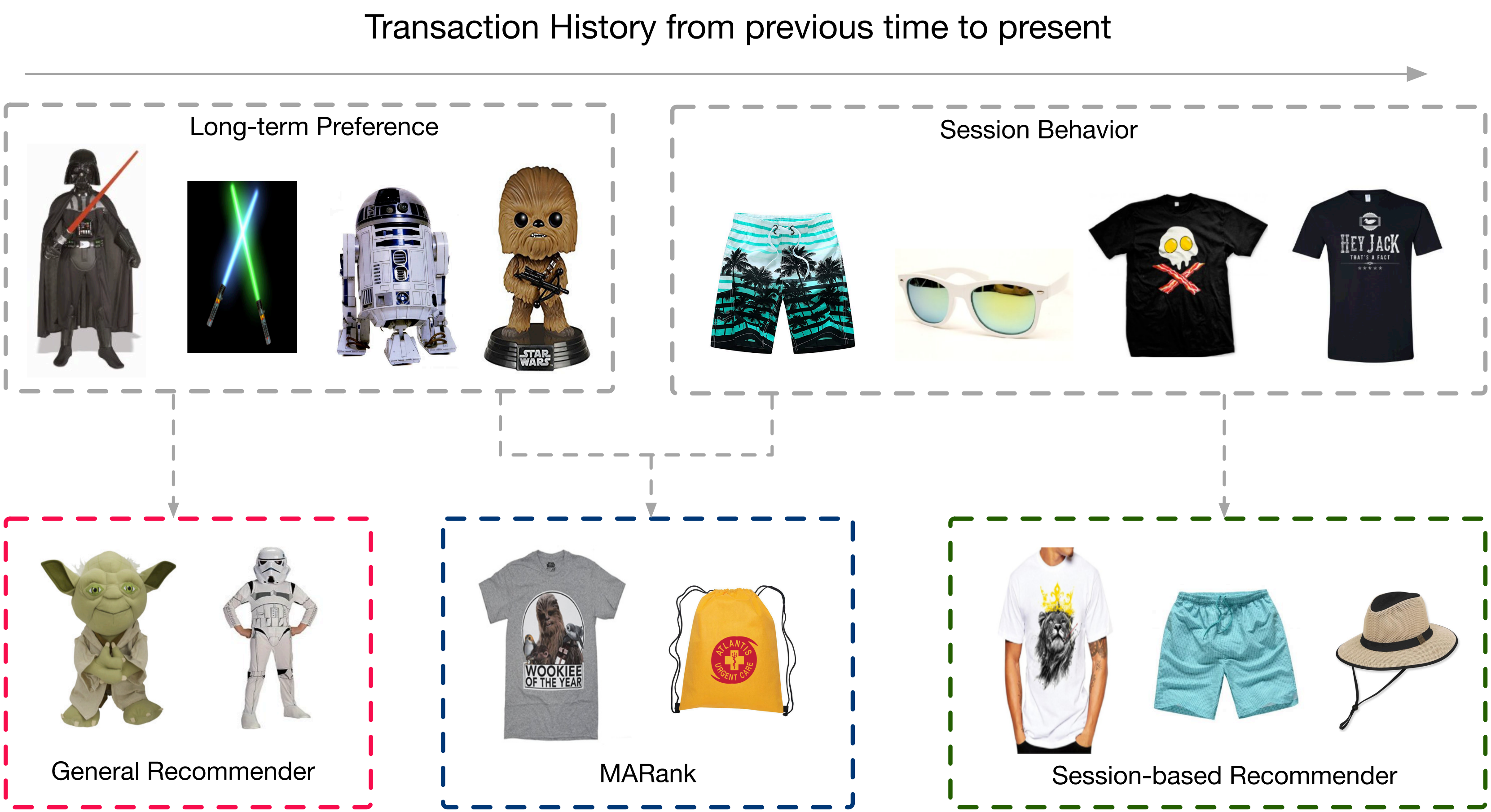
Lu Yu[1], Chuxu Zhang[2], Shangsong Liang[1,3], Xiangliang Zhang[1]

1. King Abdullah University of Science and Technology
2. University of Notre Dame
3. Sun Yat-Sen University

The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19),
Jan. 27 – Feb. 1, Honolulu, Hawaii, USA

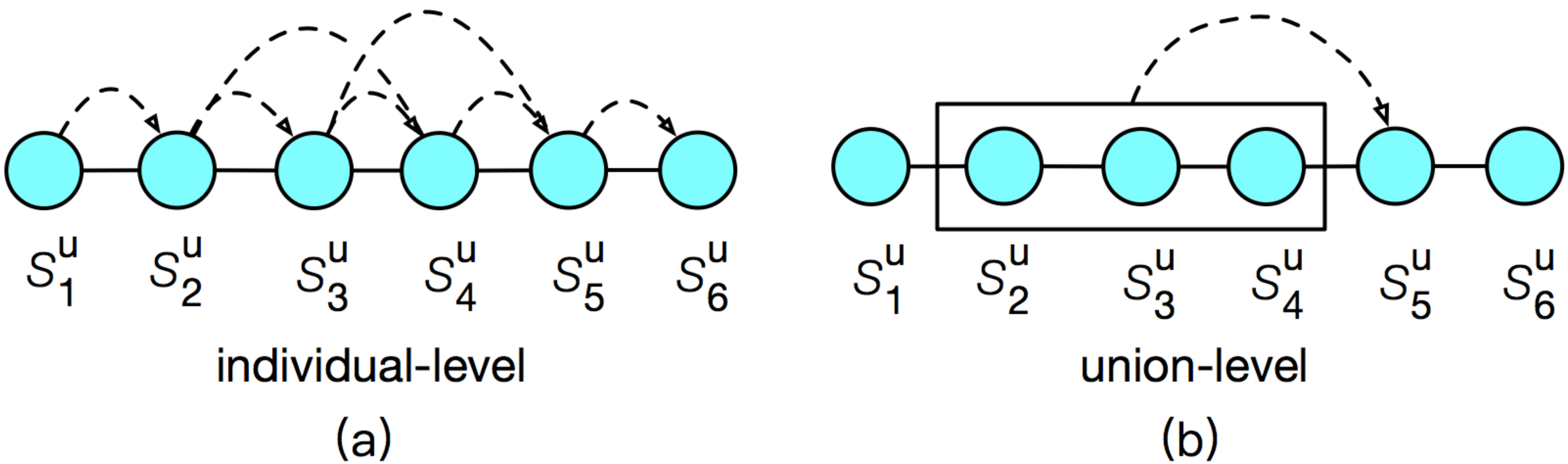
Goal: Understanding Dynamic User Behaviors for Next-step Prediction

E.g. Predicting what users are going to buy next.



Goal: Understanding Dynamic User Behaviors for Next-step Prediction

Augmenting User Preference with Item-item Transition.



Markov-chain Models VS Deep Network for Sequence Modeling

Modeling Markov-chain Transition Probability

$$\hat{y}_{u,j} = f(j, u, S^u_{t-1,l} | \mathbf{P}, \mathbf{M}, \mathbf{Q}, \Theta_f)$$

Previous Interacted Items

Model Parameters

Markov-chain Models VS Deep Network for Sequence Modeling

Factorizing Personalized Markov Chains (FPMC)

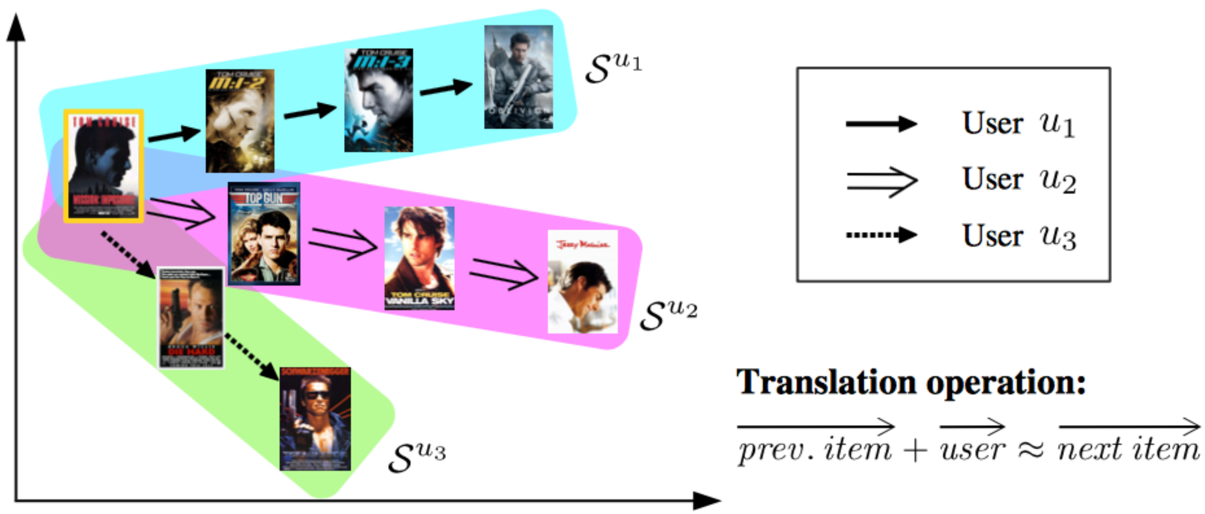
$$p(j|u, S^u_{t-1,l}) = \overbrace{<\mathbf{P}_u, \mathbf{Q}_j>}^{general\ preferences} + \overbrace{\frac{1}{|S^u_{t-1,l}|} \sum_{i \in S^u_{t-1,l}} <\mathbf{M}_i, \mathbf{N}_j>}^{sequential\ dynamics}$$

Only Modeling Individual-level Interaction

Steffen Rendle et al., Factorizing personalized Markov chains for next-basket recommendation. In WWW '10

Markov-chain Models VS Deep Network for Sequence Modeling

Translation-based Recommendation (TranRec)



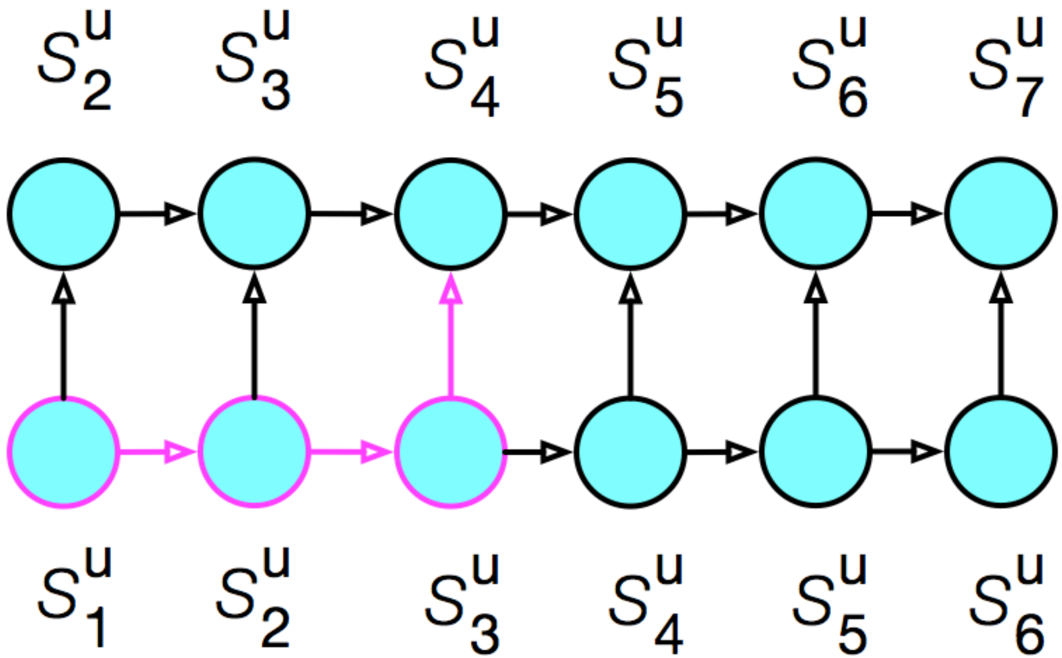
Only Modeling Individual-level Interaction

$$p(j|u,i) = \beta_j - \underbrace{d(\mathbf{M}_i + \gamma + \mathbf{P}_u, \mathbf{N}_j)}_{unified\ preference}$$

Ruining He et al., Translation-based Recommendation. In RecSys ‘17

Markov-chain Models VS Deep Network for Sequence Modeling

Gated Recurrent Neural Network for Recommendation



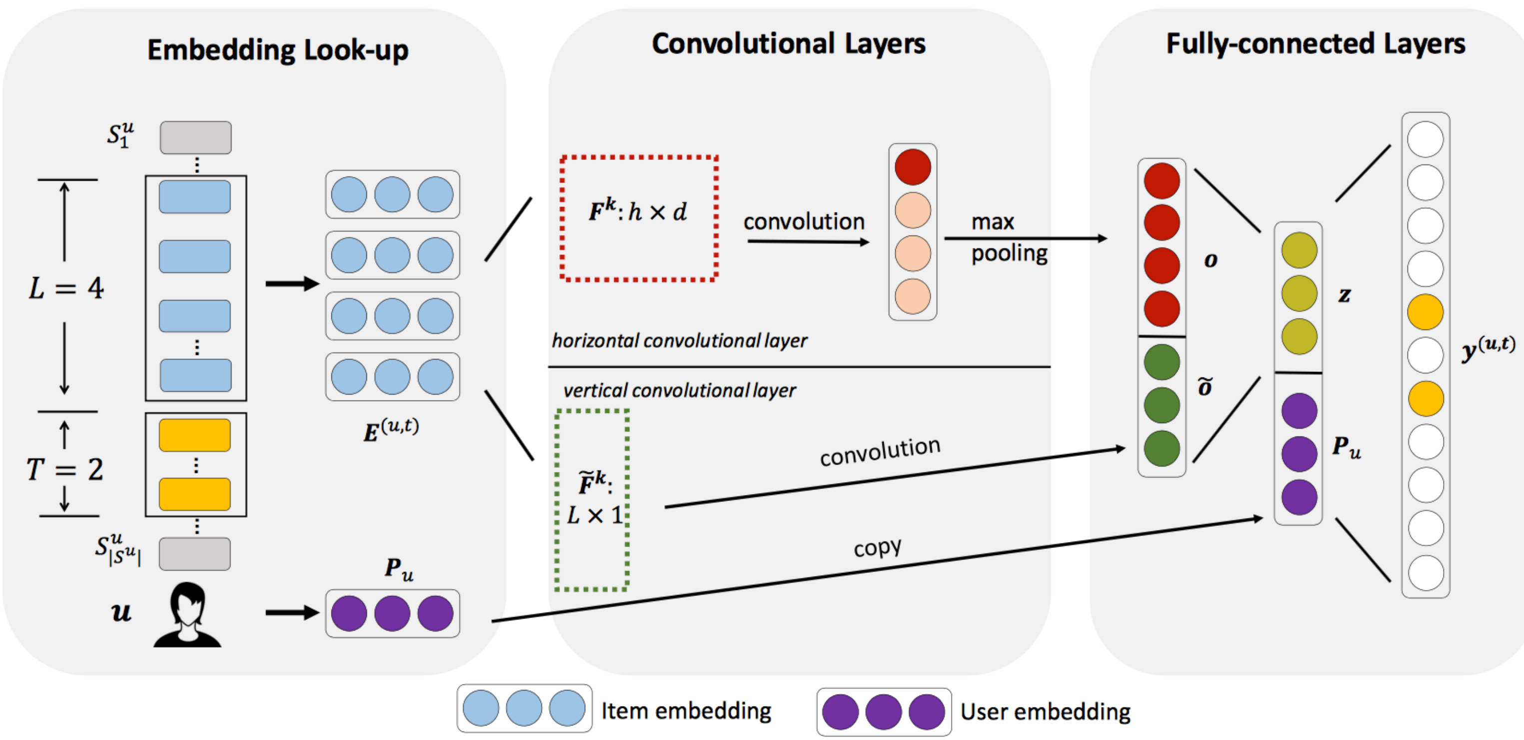
Union-level Interaction

$$\hat{y}_t = \mathrm{GRU4Rec}(s_{t-1}^u, h_{t-1})$$

Bala ́zs Hidasi et al., Session-based Recommendations with Recurrent Neural Networks. In ICLR ‘16

Markov-chain Models VS Deep Network for Sequence Modeling

Convolution Sequence Embedding

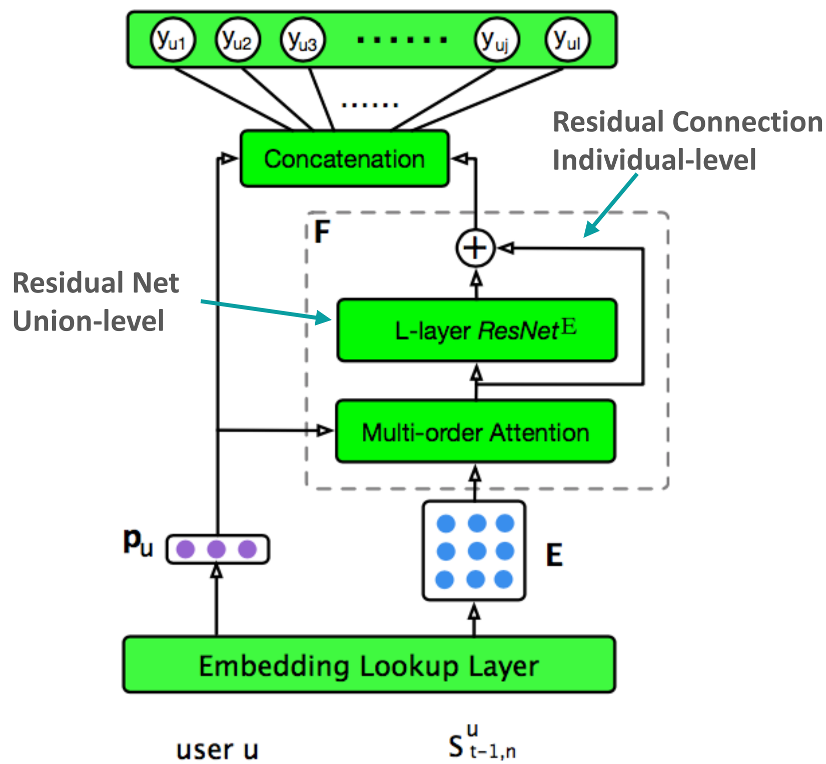


Both Individual- and Union-level Interaction

Temporal-Spatial Aggregation, Fails to Personalization

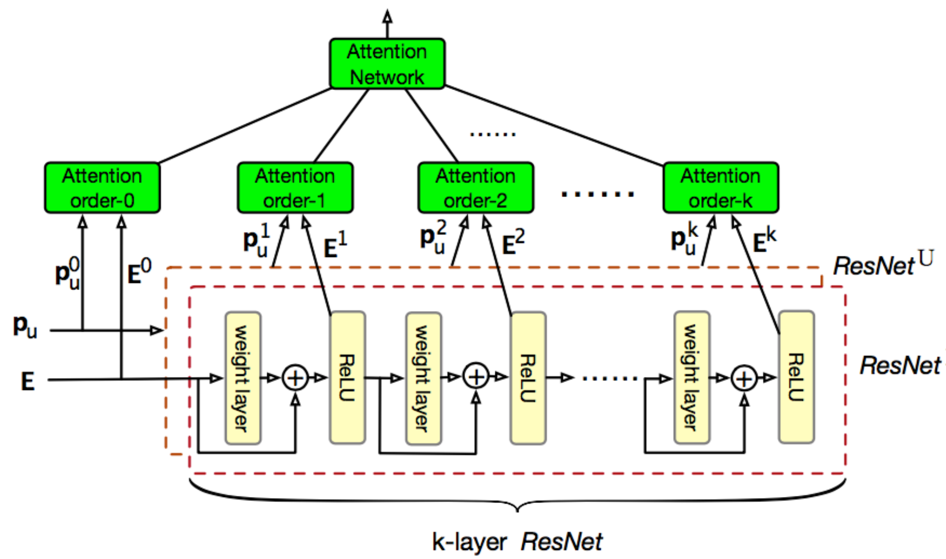
Jiaxi Tang et al., Personalized Top-N Sequential Recommendation via Convolutional Sequence Embedding. In WSDM '18

Our Solution: Multi-order Attention Model



Overall Architecture

$$\hat{y}_{u,j} = \underbrace{\mathbf{p}_u \mathbf{q}_j^T}_{long-term} + \underbrace{\overbrace{\mathbf{e}_c \mathbf{m}_j^T}^{Individual-level} + \overbrace{\mathbf{h}_L \mathbf{m}_j^T}^{Union-level}}_{short-term}$$



Multi-order Attention Network

Lu Yu et al., Multi-order Attentive Ranking Model for Sequential Recommendation. In AAAI '19

Experimental Analysis – Overall Comparisons

RQ1: Can our proposed method outperform the state-of-the-art baselines for sequential recommendation task?

RQ2: How does data sparsity influence MARank?

RQ3: How MARank is affected by each component?

http://jmcauley.ucsd.edu/data/amazon/

https://www.yelp.com/dataset/challenge

Table 1: Statistical information of datasets.

Data	# Users	# Items	# Observation	Sparsity
Yelp	25677	25815	731671	99.89%
Movies&TV	35168	51227	1070645	99.94%
CDs&Vinyl	26876	66820	770188	99.95%

Experimental Analysis – Overall Comparisons

Table 2: Ranking performance comparison (the best results of baseline are marked as * along with underline). The last row shows the improvement of MARank over the best baseline algorithm.

Methods	Yelp				Movies&Tv				CDs&Vinyl			
Measures@20	Rec	Pre	NDCG	MRR	Rec	Pre	NDCG	MRR	Rec	Pre	NDCG	MRR
BPR-MF	0.0618	0.0156	0.0382	0.0132	0.0404	0.0095	0.0248	0.0089	0.0512	0.0121	0.0323	0.0119
GRU4Rec	0.0586	0.0151	0.0373	0.0131	0.0372	0.0098	0.0235	0.0081	0.0358	0.0093	0.023	0.0082
NARM	0.0616	0.0159	0.0401	0.0145	0.0429	0.0111	0.0273	0.0096	0.0426	0.0108	0.028	0.0104
GRU4Rec+	0.0699	0.0179	0.0456	0.0166	0.0526	0.0138	0.0344	0.0124	0.0527	0.0138	0.0352	0.0129
NARM+	*0.0718	*0.0183	*0.0465	*0.0169	0.0587	*0.0145	0.0375	0.0136	0.0678	*0.0169	0.0449	0.0169
ACF	0.0661	0.0169	0.0425	0.0153	0.0458	0.0107	0.0286	0.0106	0.0554	0.0133	0.0357	0.0134
Caser	0.0653	0.0161	0.0415	0.0151	0.0494	0.0122	0.0318	0.0116	0.0495	0.0129	0.0346	0.0132
TranRec	0.0703	0.0176	0.0458	0.0169	0.0606	0.0135	0.0392	0.0155	*0.0704	0.0164	*0.0468	*0.0184
FPMC	0.0713	0.0178	0.0463	0.0169	*0.0608	0.0142	*0.0406	*0.0162	0.0646	0.0156	0.0450	0.0179
MARank	**0.0791**	**0.0199**	**0.0509**	**0.0183**	**0.0680**	**0.0162**	**0.0444**	**0.0170**	**0.0817**	**0.0201**	**0.0552**	**0.0212**
Improvement	10.1%	8.7%	9.4%	8.2%	11.8%	11.7%	9.3%	4.9%	16.0%	18.9%	17.8%	15.2%

Experimental Analysis – Overall Comparisons

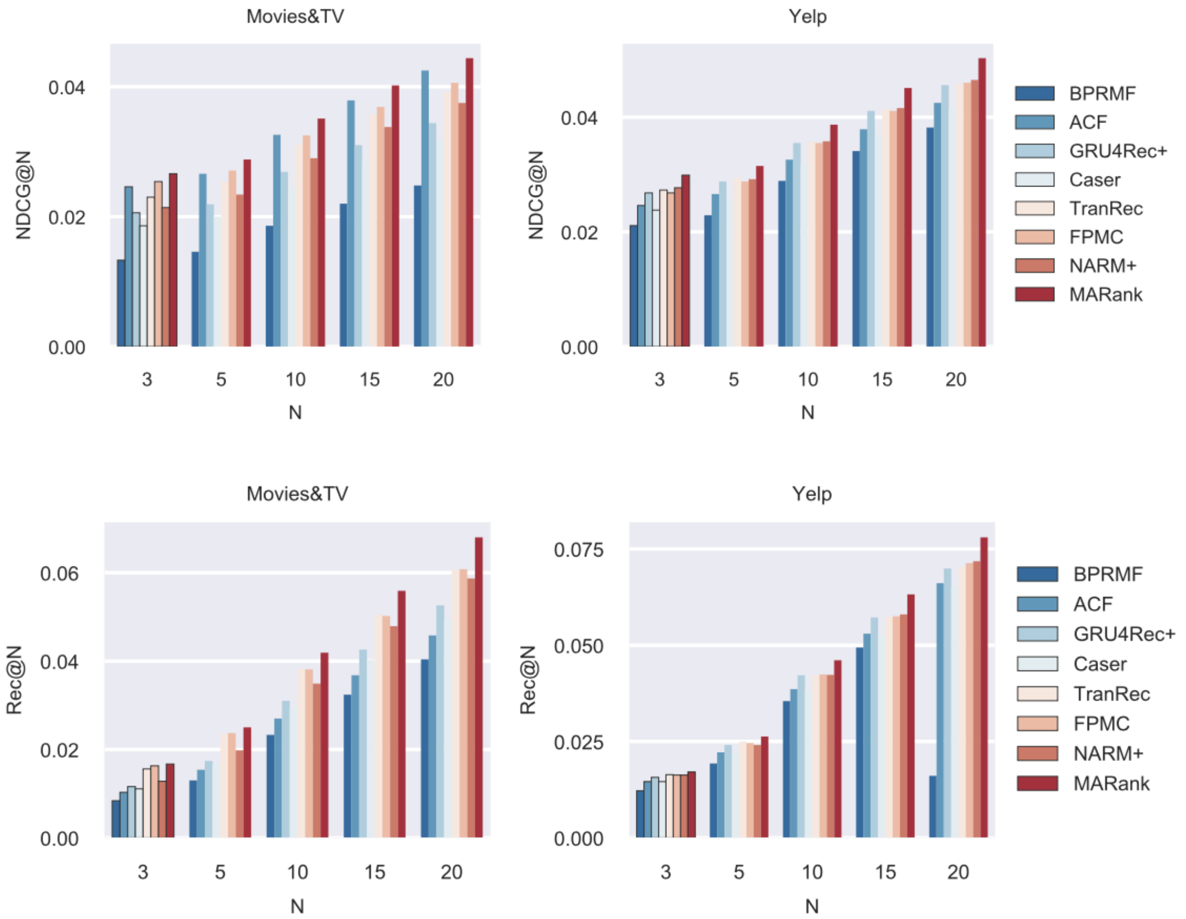


Figure 3: Top-N recommendation evaluation with different values of N on NDCG and Recall.

Experimental Analysis – Sparsity

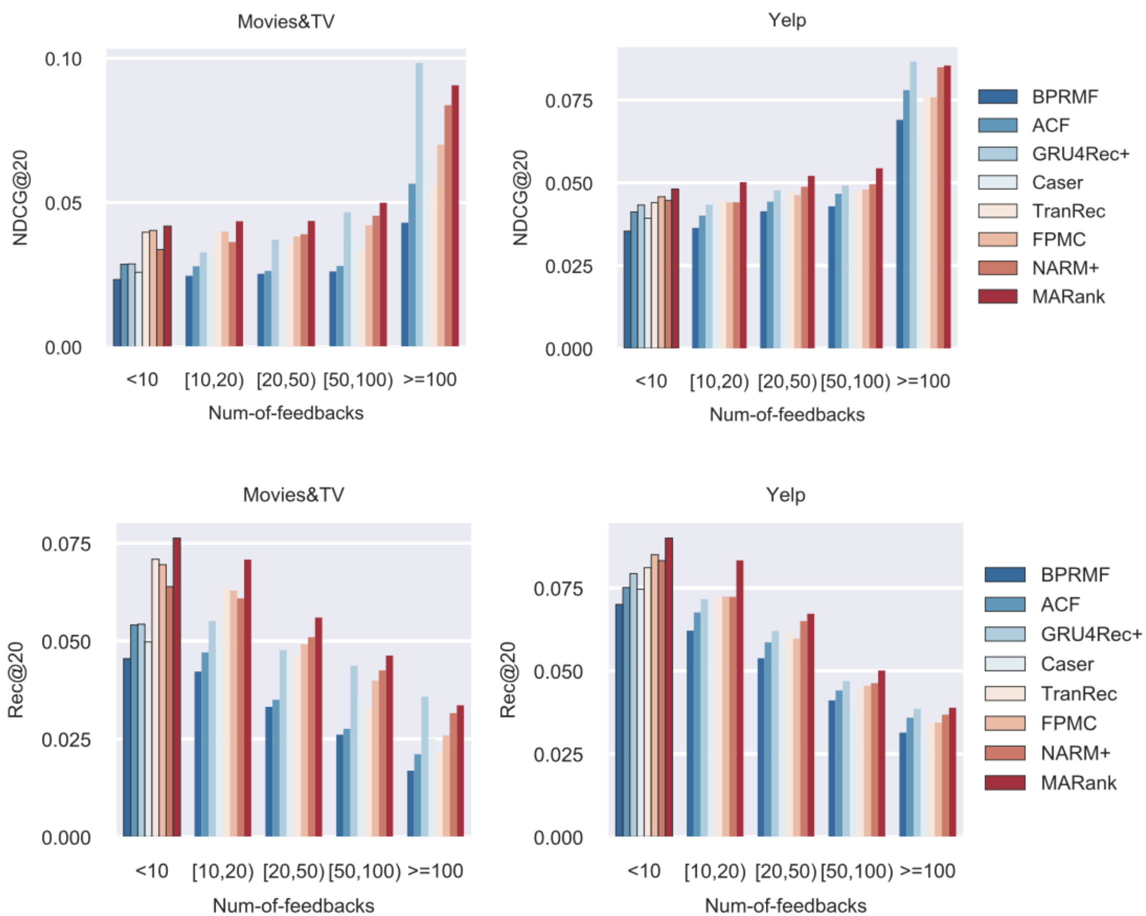


Figure 4: Recommendation evaluation on different user groups, which are separated according to the number of feedbacks.

Experimental Analysis – Sparsity

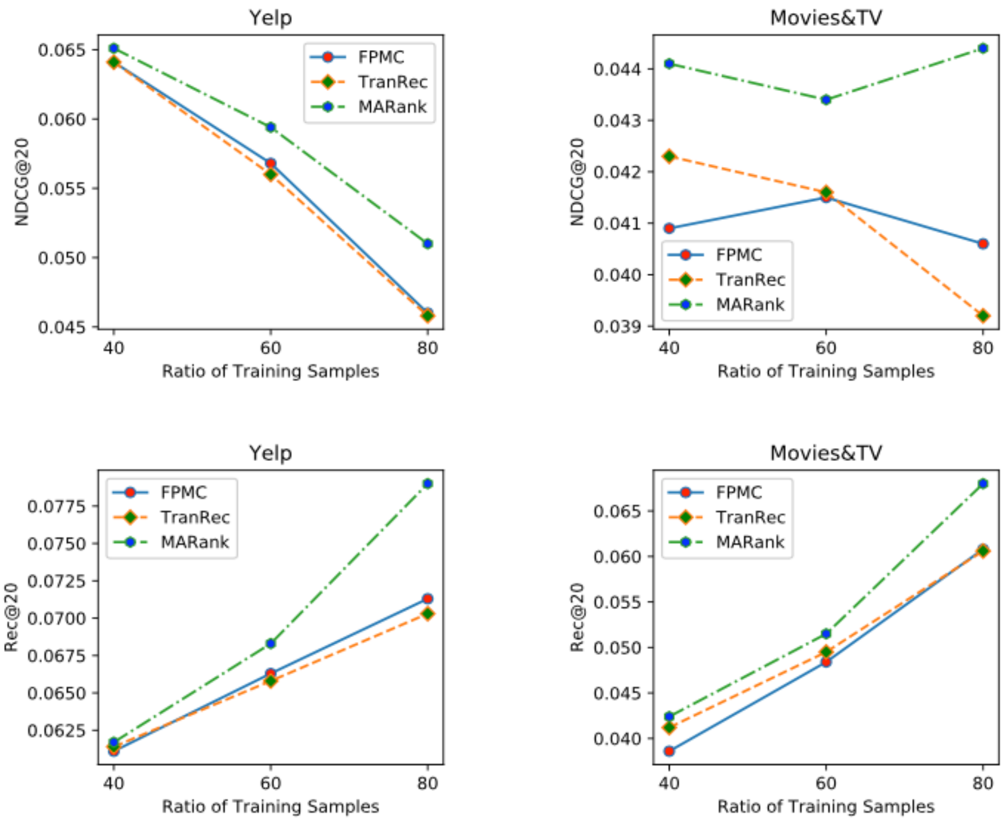


Figure 5: Performance on sparse training set. *x* axis stands for the proportion of training samples.

Experimental Analysis – Components

Table 3: Performance with different level of item-item interactions. MARank-I or MARank-U represents MARank with only individual- or union-level item-item dependency.

Methods	Yelp				Movies&TV			
Metrics@20	Rec	Pre	NDCG	MRR	Rec	Pre	NDCG	MRR
MARank-I	0.0755	0.0185	0.0475	0.0175	0.0635	0.0153	0.0405	0.0158
MARank-U	0.0731	0.0178	0.0456	0.0168	0.0616	0.0148	0.0374	0.0145
MARank	0.0785	0.0195	0.0509	0.0183	0.0680	0.0162	0.0441	0.0170

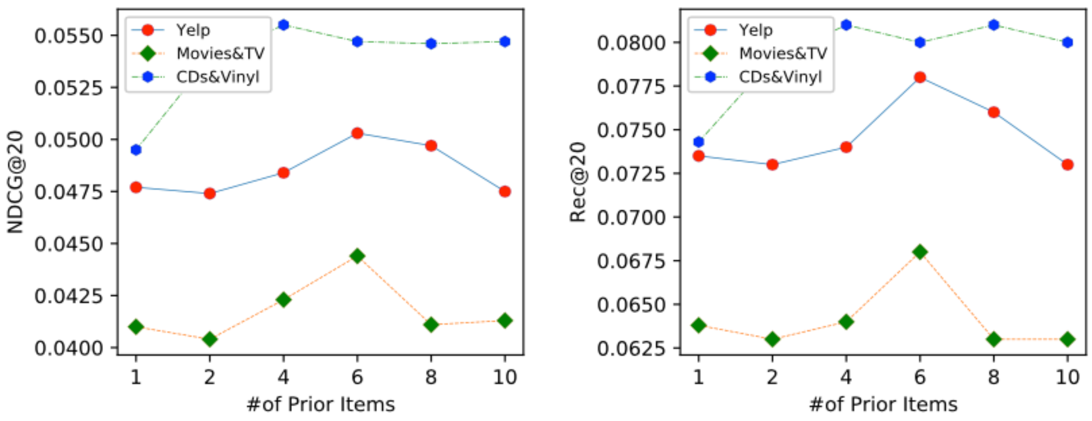


Figure 6: The impact of Markov chain length.

Question & Answer

Thank You

Contact Info: {lu.yu, xiangliang.zhang}@kaust.edu.sa
Lab Page: https://mine.kaust.edu.sa
Code Page: https://github.com/voladorlu/MARank