An End-to-End Neighborhood-based Interaction Model for Knowledge-enhanced Recommendation

Yanru Qu*[1], Ting Bai*[2], Weinan Zhang[1], Jianyun Nie[3], Jian Tang[4]
kevinqu16@gmail.com

[1]Shanghai Jiao Tong University, [2]Beijing University of Posts and Telecommunications, [3]Université de Montréal, [4]Mila-Quebec Institute for Learning Algorithms

上海交通大学 Shanghai Jiao Tong University
北京郵電大學 Beijing University of Posts and Telecommunications
Université de Montréal
Mila

Outline

- Background & Motivation
- Knowledge-enhanced Neighborhood Interaction
- Experiments
- Case Study
- Conclusion

Background & Motivation

- Recommender systems have become increasingly important in various online services.
- Challenges:
  - Data sparsity
  - Cold start
  - etc.
- Feature-based, meta path-based, and graph-based models are proposed to incorporate side-information and enhance user-item interactions.

Background & Motivation

- Graph-based models facilitate recommendation with rich structural information. When the relation are unknown, a system is expected to distinguish useful patterns from noise.

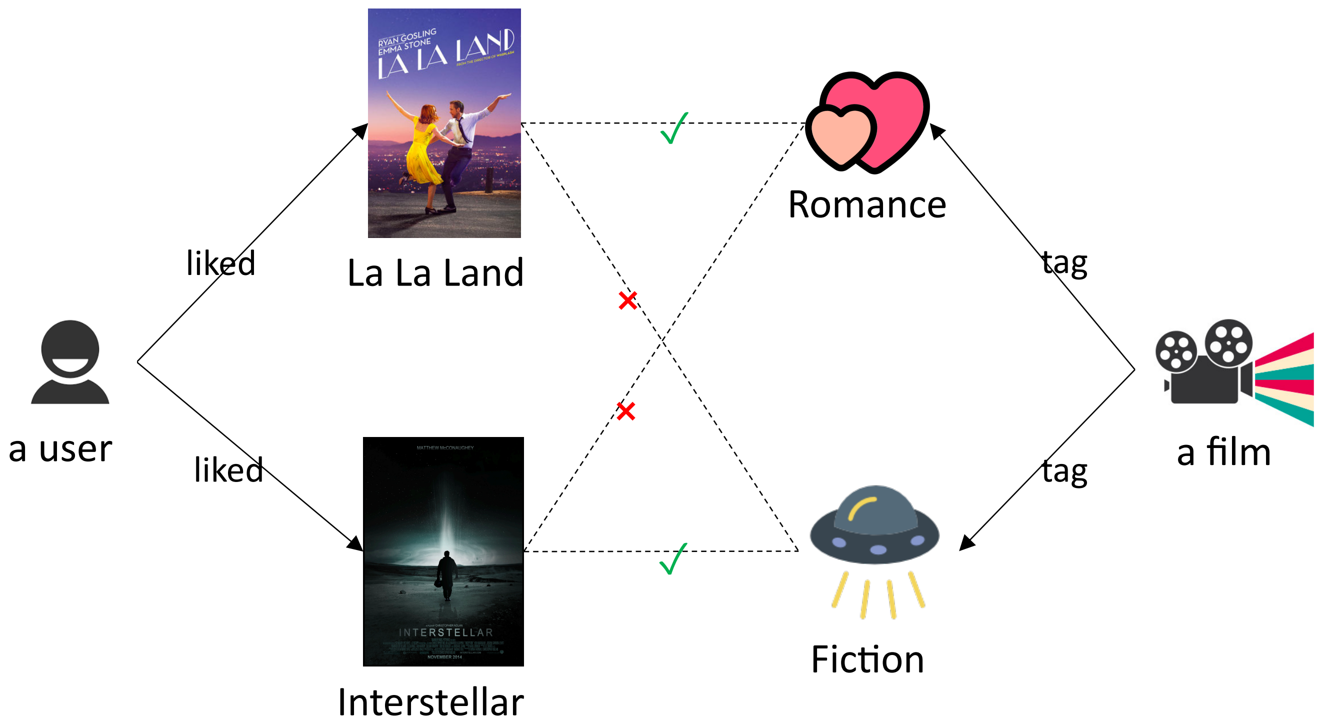

Fig.1 Recommending a film to a user

Background & Motivation

- Extending interactions between users and items to any two nodes, we observe an "*early summarization*" problem in graph-based methods:
  - User-neighbors are summarized into a single representation (similar for item) *before* final prediction.
  - The interactions among neighborhoods are mixed, no matter whether useful or noisy.
- We focus on the meticulous and valuable neighborhood structures.

Knowledge-enhanced Neighborhood Interaction

- Given a user u with neighbors N_u, an item v with neighbors N_v, most graph-based models share a general form ($\sigma()$: sigmoid, $\langle , \rangle$: inner product)

$$\hat{y}_{u,v} = \sigma(\langle agg(N_u), agg(N_v) \rangle)$$

- "agg" function aggregates neighbors into a single vector.
- It could be (x_i are embedding or feature vectors):
  - Average: $u = \frac{1}{|N_u|} \sum_{i \in N_u} x_i$
  - Attention: $\alpha_{u,i} = softmax_i(W[x_u, x_i] + b)$

    $u = \sum_{i \in N_u} \alpha_{u,i} x\_i$
  - etc.

Knowledge-enhanced Neighborhood Interaction

- Expanding average and attention "agg" functions, we find a shared form of graph-based models, i.e., weighted sum over inner products.

Average:

$$\hat{y}_{u,v} = \langle \frac{1}{|N_u|} \sum_{i \in N_u} x_i , \frac{1}{|N_v|} \sum_{j \in N_v} x_j \rangle$$

$$= \sum_{i \in N_u} \sum_{j \in N_v} \frac{1}{|N_u||N_v|} \langle x_i, x_j \rangle$$

Attention:

$$\hat{y}_{u,v} = \langle \sum_{i \in N_u} \alpha_{u,i} x_i , \sum_{j \in N_v} \alpha_{v,j} x_j \rangle$$

$$= \sum_{i \in N_u} \sum_{j \in N_v} \alpha_{u,i} \alpha_{v,j} \langle x_i, x_j \rangle$$

Shared Form:

$$\hat{y} = A \odot Z$$

$$\text{s.t. } \sum_{i,j} A_{i,j} = 1, Z_{i,j} = \langle x_i, x_j \rangle$$

- Z: modeling the interactions of each pair of nodes
- A: assigning proper weights for different terms

Knowledge-enhanced Neighborhood Interaction

- "average" and "attention" are special cases of the shared form, i.e.,
  - "average": A is a constant
  - "attention": A is the outer product of two vectors.
- We propose Neighborhood Interaction (NI) model to fully explore the shared form, which can learn different weights for different terms.

$$\alpha_{i,j} = softmax_{i,j}\, W\left[x_u, x_v, x_i, x_j\right] + b$$

$$\hat{y}_{u,v} = \sum_{i \in N_u} \sum_{j \in N_v} \alpha_{i,j} \langle x_i, x_j \rangle$$

Knowledge-enhanced Neighborhood Interaction

- To integrate external knowledge graphs, we construct Knowledge-enhanced Interaction Graph (KIG).
- To integrate high-order neighborhood information, we introduce graph convolution/attention networks to learn node embeddings (i.e., x_i) for NI model.
- Neighbor sampling technique is utilized to reduce complexity.

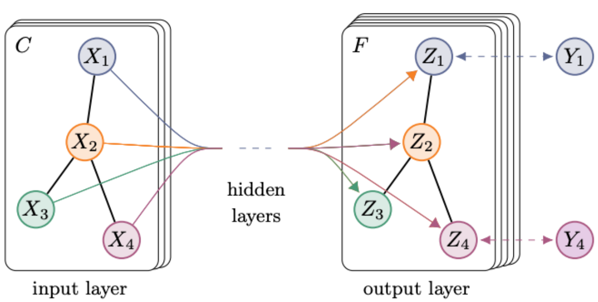

Fig.2 Graph Convolution Network

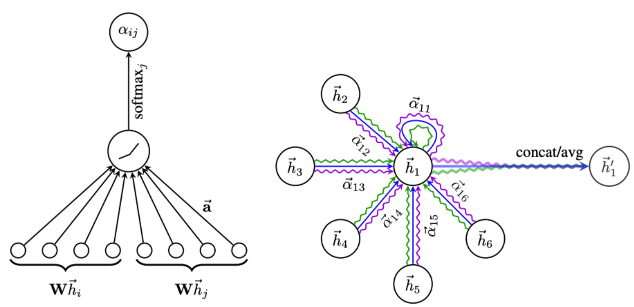

Fig.3 Graph Attention Network

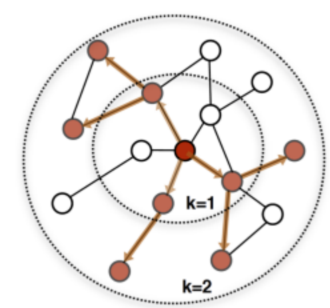

Fig.4 Neighbor Sampling

数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

Knowledge-enhanced Neighborhood Interaction

- Combining KIG, GNN, NI, we obtain Knowledge-enhanced Neighborhood Interaction (KNI) model
- Training objective

$$L(Y,\hat{Y}) = -\sum_{y_{u,v}=1} \log(\hat{y}_{u,v}) - \sum_{y_{u,v}=0} \log(1 - \hat{y}_{u,v}) + \lambda \|\theta\|_2^2$$

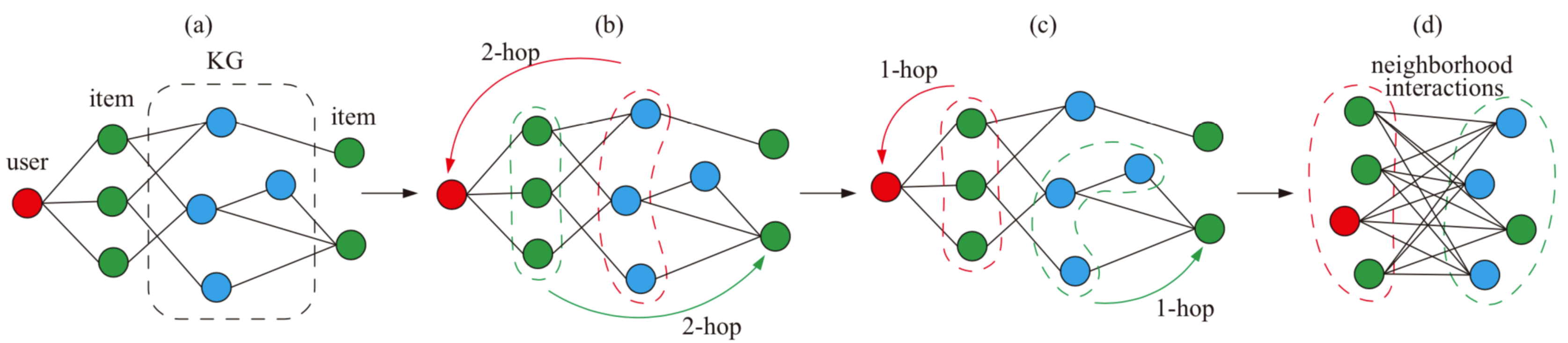

Fig.5 KNI overview

Experiments

- We combine 4 movie/book recommendation datasets with Microsoft Satori and Freebase.

Datasets	C-Book	Movie-1M	A-Book	Movie-20M
# users	17,860	6,036	78,809	59,296
# items	14,967	2,445	32,389	11,895
# interactions	139,746	753,772	1,181,684	9,104,038
# entities	77,881	182,011	265,478	64,067
# relations	10	12	22	38
# triples	71,628	923,718	1,551,554	1,195,391

Table.1 Dataset statistics

- We compare NI (w/o KG) and KNI models with 2 feature-based (libFM, Wide&Deep), 2 meta path-based (PER, MCRec), and 4 graph-based (CKE, DKN, PinSage, RippleNet) models. (red: SOTA)
- 2 tasks: CTR prediction, Top-N recommendation

Experiments

- CTR prediction

Model	C-Book		Movie-1M		A-Book		Movie-20M	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
libFM	0.6850	0.6390	0.8920	0.8120	0.8300	0.7597	0.9481	0.8805
Wide&Deep	0.7110	0.6230	0.9030	0.8220	0.8401	0.7684	0.9507	0.8831
PER	0.6230	0.5880	0.7120	0.6670	0.7392	0.6939	0.8161	0.7327
MCRec	0.7250	0.6707	0.9127	0.8331	0.8708	0.7930	0.9558	0.8872
CKE	0.6760	0.6422	0.8974	0.8171	0.8572	0.7839	0.9574	0.8940
DKN	0.6488	0.6333	0.8835	0.8070	0.8455	0.7679	0.9473	0.8787
PinSage	0.7102	0.6477	0.9213	0.8443	0.8634	0.7804	0.9597	0.8960
RippleNet	0.7290	0.6630	0.9210	0.8440	0.8736	0.7975	0.9579	0.8942
NI	0.7468	0.6796	0.9401	0.8679	0.9160	0.8362	0.9693	0.9110
KNI	**0.7723***	**0.7063***	**0.9449***	**0.8721***	**0.9238***	**0.8472***	**0.9704***	**0.9120***

Table.2 Main results. (Note: All models (except NI) utilize knowledge graphs. For feature-based models, we extract structural features with TransR as their input. Every experiment is repeated for 5 times.)

- NI (w/o KG) model outperforms SOTA methods (w/ KG), which means low-order neighborhood structures are more valuable.
- External KGs can further boost NI, and KNI outperforms others by 1.1%-8.4% AUC.

Experiments

- Top-N recommendation

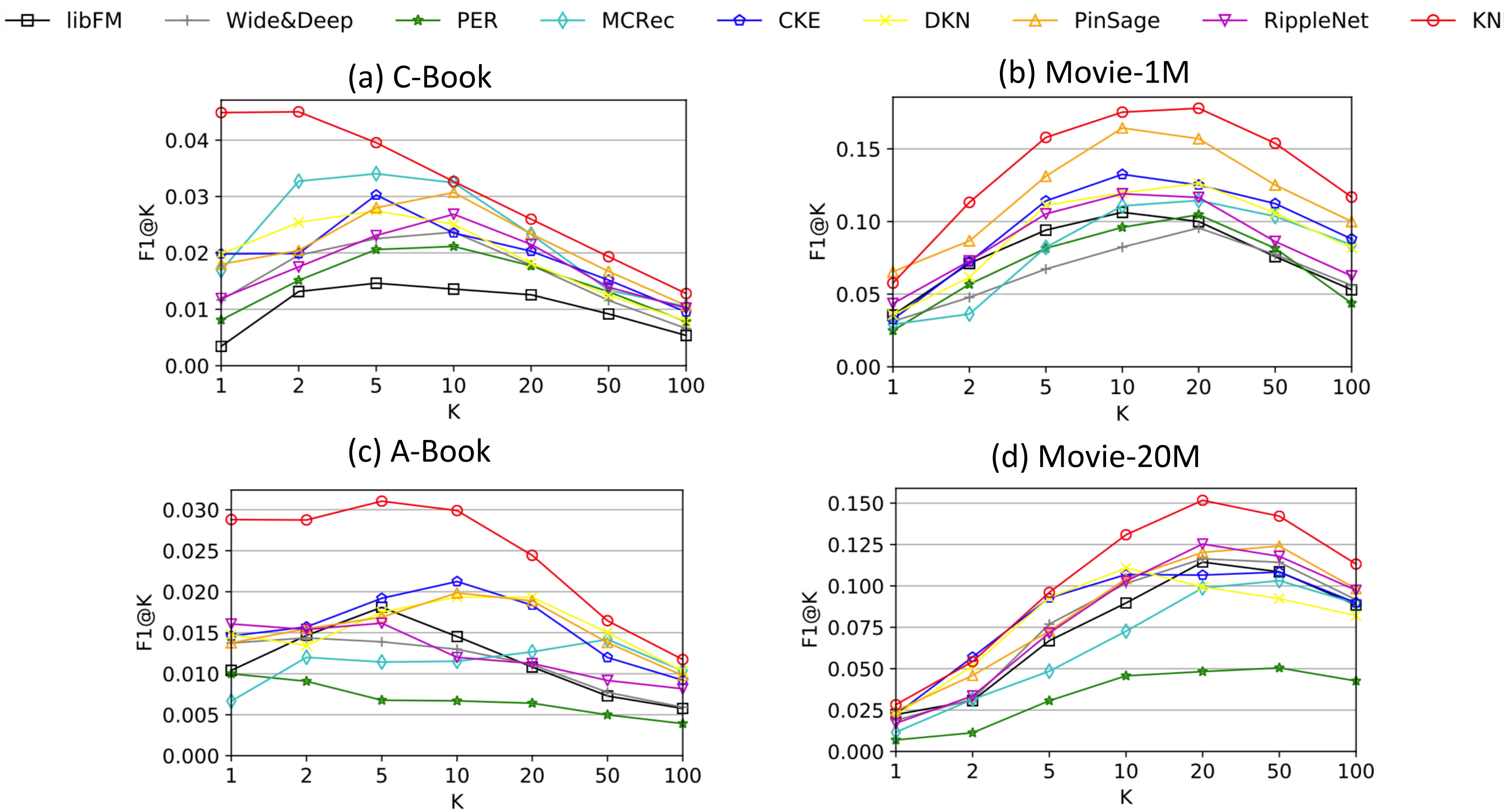

Fig.6 F1 scores of Top-N recommendation

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

Experiments

- KNI performs even better on sparser datasets

Datasets	1-hop	2-hop	3-hop	Sparsity	Improvement
C-Book	1	58	40	99.97%	4.33%
Movie-1M	14	42,227	35,534	97.45%	2.36%
A-Book	5	17,027	49,419	99.98%	5.02%
Movie-20M	17	40,547	14,966	99.35%	1.07%

Table.3 Absolute AUC improvements of KNI over best baselines

- KNI is efficient when using neighbor sampling technique on GPU

Models	C-Book	Movie-1M	A-Book	Movie-20M
RippleNet	17.75s	66.85s	120.38s	937.92s
KNI	2.05s	11.58s	21.52s	166.72s

Table.4 Training time (1 epoch) of RippleNet and KNI

Case Study

- Recall the general form of graph-based models.

$$\hat{y} = A \odot Z$$
$$\text{s.t. } \sum_{i,j} A_{i,j} = 1, Z_{i,j} = \langle x_i, x_j \rangle$$

- Regarding A as a distribution over all terms, the entropy of A reflects the information of the distribution.
- For simplicity, we remove KGs and compare NI with attention aggregation model (AAM). The average entropy is 3.18 (NI) > 2.12 (AAM), which confirms the "early summarization" problem, and NI can learn better interactions.

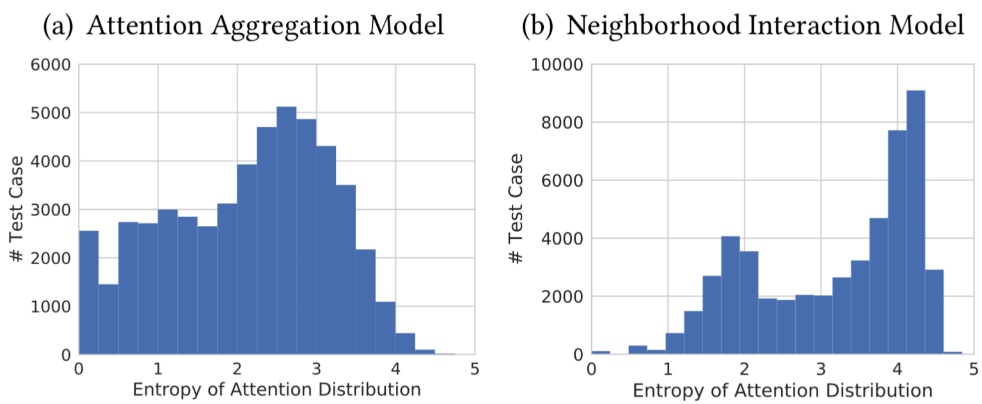

Fig.7 Histogram of entropy

Case Study

- We further visualize the attention matrices, which shows NI model captures more interactive patterns than traditional attention method.

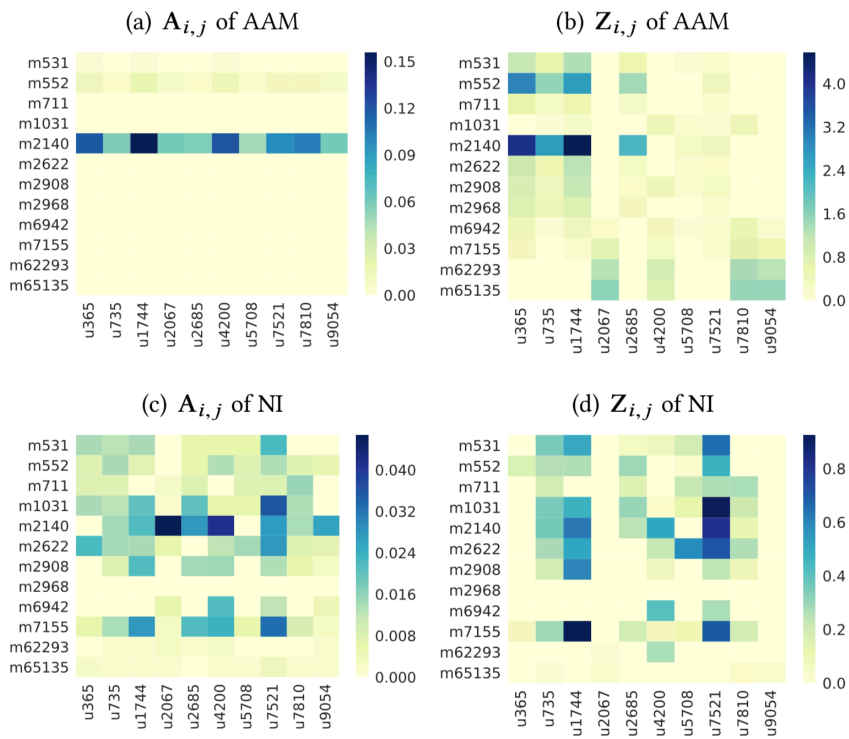

Fig.8 Test case (u46, m3993). (Note: y-axis: u46 neighbors, x-axis: m3993 neighbors)

Conclusion

- Traditional graph-based recommendation models usually encode user/item neighbors information into single embedding vectors before final prediction.
- The "early summarization" behavior mixes up the whole neighborhoods, making it harder to distinguish useful patterns from noise.
- Exploring the neighborhood structures is very promising to alleviate data sparsity and cold start problems.
- KNI is an end-to-end framework which can incorporate external knowledge graphs and fully utilize the valuable neighborhood information.

Q & A