Package Recommendation with Intra- and Inter-Package Attention Networks

Chen Li
Beijing University of Posts and Telecommunications
WeChat Search Application Department, Tencent
Beijing, China
chenlipuleya@163.com

Yuanfu Lu
WeChat Search Application Department, Tencent
Beijing, China
luyfroot@gmail.com

Wei Wang
WeChat Search Application Department, Tencent
Beijing, China
unoywang@tencent.com

Chuan Shi*
Beijing University of Posts and Telecommunications
Beijing, China
shichuan@bupt.edu.cn

Ruobing Xie
WeChat Search Application Department, Tencent
Beijing, China
xrbsnowing@163.com

Haili Yang
WeChat Search Application Department, Tencent
Beijing, China
hailiyang@tencent.com

Cheng Yang
Beijing University of Posts and Telecommunications
Beijing, China
yangcheng@bupt.edu.cn

Xu Zhang
WeChat Search Application Department, Tencent
Beijing, China
xuonezhang@tencent.com

Leyu Lin
WeChat Search Application Department, Tencent
Beijing, China
goshawklin@tencent.com

Motivation

- Traditional recommendation scenario that only displays the candidate items to the user, in package recommendation, objects in the package (e.g., the publisher of the news and the friends who click the news) are explicitly shown to the user, which will highlight the influence of various objects on the user's behaviors.

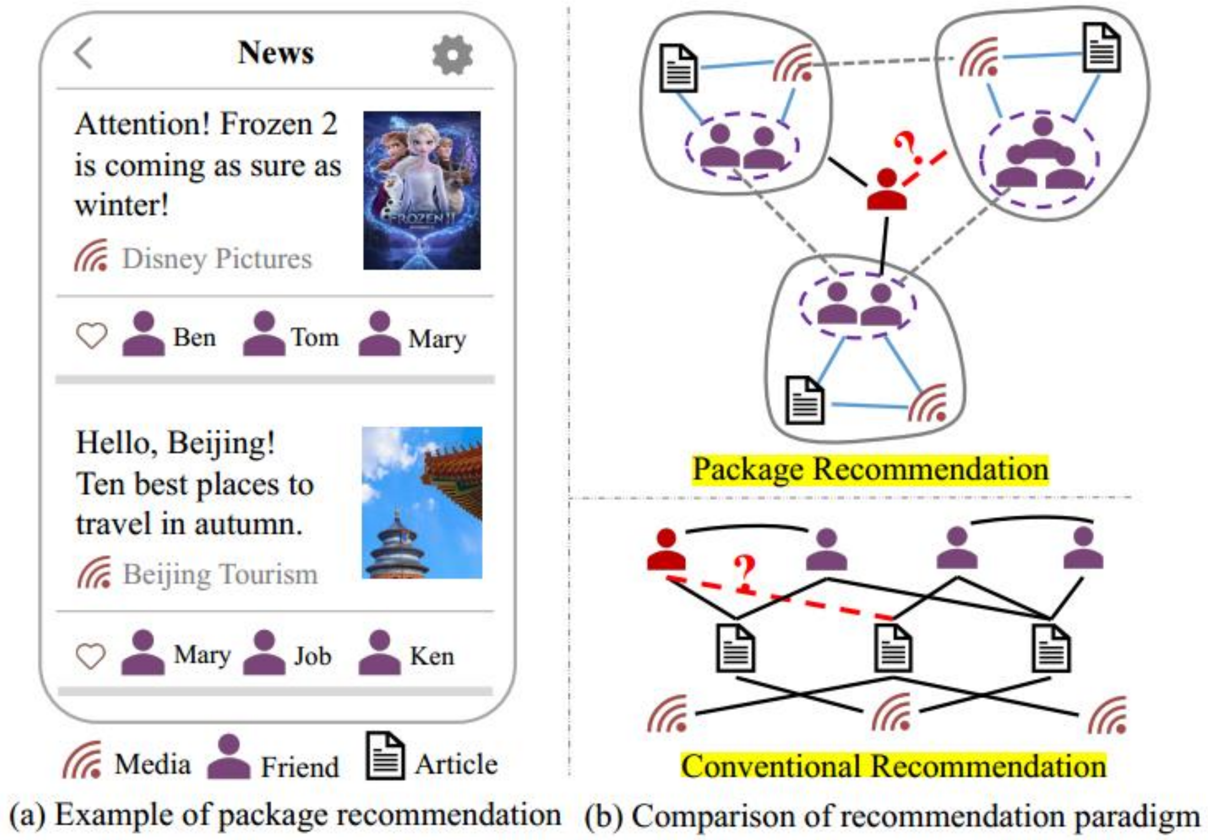


Figure 1: (a) A typical example of package recommendation in a real-world social platform. The publisher of the news and friends clicked the news are also explicitly displayed, which will have a certain influence to user. (b) Comparison between package recommendation and conventional recommendation.

Problem Formulation

- **Package Recommendation:** Denote a package as $\mathcal{P} = \mathcal{O}^1 \cup \mathcal{O}^2 \cup \cdots \cup \mathcal{O}^{\mathcal{T}}$, where $\mathcal{O}^t = \{o_1^t, o_2^t, \ldots, o_{|\mathcal{O}^t|}^t\}$ is the object set of $t \in \mathcal{T}$ types, and $\mathcal{T}$ is the set of objects types, i.e., $\mathcal{T} = \{Article, Media, Friend\}$, represents the object types set. $\{Frozen2-related\ article\} \cup \{DisenyPrictures\ media\} \cup \{Ben, Tom, Mary\ friends\}$ formulate a package $\mathcal{P}_1$. There is a user set $U = \{u_1, u_2, \ldots, u_m\}$, a package set $P = \{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n\}$ and their interaction matrix $Y \in \mathbb{R}^{m \times n}$, where $y_{u,\mathcal{P}} = 1$ indicate the user has interacted with the package $\mathcal{P}$, otherwise $y_{u,\mathcal{P}} = 0$. Given a user u w.r.t. a non-interaction package $\mathcal{P}$, the package recommendation aims to predict whether user u has a potential preference to the package $\mathcal{P}$.

Package Recommendation (IPRec)

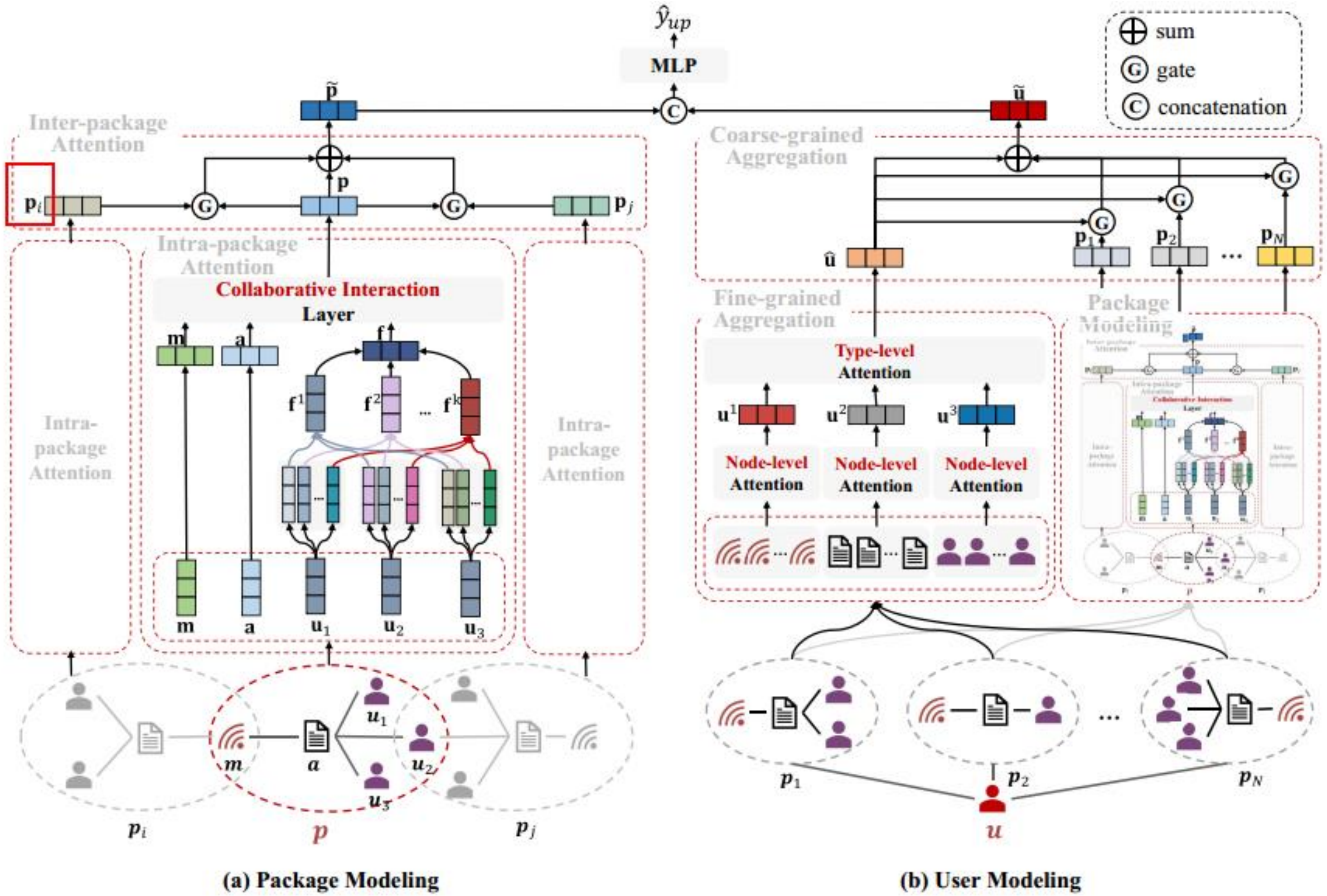


Figure 2: The framework of IPRec. It contains three components: (a) package modeling module, which learns the package embedding with intra- and inter-package attention networks to capture multifaceted influence and collaborative information; (b) user modeling module, which models the user preference at different granularities with fine-grained and coarse-grained aggregation networks; and rating prediction module, which predicts the rating with learned user and package embeddings.

Package Modeling

- **Intra-Package Attention Network**

Denote the representation of article a, media m and friends $\{u_1, u_2, \dots, u_m\}$ as $\mathbf{a}, \mathbf{m}$ and $\mathbf{f}$. Fusing the heterogeneous and diverse information as:

$$C = \{\mathbf{a}, \mathbf{m}, \mathbf{f}, \Gamma(\mathbf{a}, \mathbf{m}), \Gamma(\mathbf{a}, \mathbf{f}), \Gamma(\mathbf{m}, \mathbf{f}), \Gamma(\mathbf{a}, \mathbf{m}, \mathbf{f})\}$$

Where $\Gamma(\cdot)$ serves as a fusion function. Element-wise product is adopted in this work.

An attention mechanism to distill different importance of multifaceted information for the current user u and fuse them as,

$$\mathbf{p} = \sum\nolimits_{\mathbf{c} \in C} \gamma^c \mathbf{c}, \qquad \gamma^c = \frac{\exp(\mathbf{z}^T \cdot \tanh(\mathbf{w}[\mathbf{u}||\mathbf{c}]))}{\sum_{\mathbf{c}' \in C} \exp(\mathbf{z}^T \cdot \tanh(\mathbf{w}[\mathbf{u}||\mathbf{c}']))}$$

where $\mathbf{p}$ is the representation for the package $\mathcal{P}$ and $\mathbf{c} \in C$ is one combination of different objects.

Package Modeling

- **Intra-Package Attention Network**

Disentangling the friend influence into different areas and then make all objects interact with each other in a package to model the complex intention of user interacting with the package, first. Then, combining influence fromdisentangled spaces with an attention mechanism.

$$\mathbf{f}^k = \sum_{u_i \in \mathcal{O}^{Friend}} \alpha_i^k \mathbf{u}_i^k$$

$$\alpha_i^k = \frac{\exp(\mathbf{z}^T \cdot \tanh(\mathbf{w}[\mathbf{a}^k || \mathbf{u}_i^k]))}{\sum_{u_j \in \mathcal{O}^{Friend}} \exp(\mathbf{z}^T \cdot \tanh(\mathbf{w}[\mathbf{a}^k || \mathbf{u}_j^k]))}, \qquad \mathbf{u}_i^k = \mathbf{W}_f^k \mathbf{u}^i, \qquad \mathbf{a}^k = \mathbf{W}_a^k \mathbf{a}$$

$$\mathbf{f} = \sum_{k=1}^{K} \beta^k \mathbf{f}^k, \qquad \beta^k = \frac{\exp(\mathbf{z}^T \cdot \tanh(\mathbf{w}[\mathbf{u} || \mathbf{f}^k]))}{\sum_{k=1}^{K} \exp(\mathbf{z}^T \cdot \tanh(\mathbf{w}[\mathbf{u} || \mathbf{f}^k]))}$$

where $\mathbf{u}_i^k$ and $\mathbf{a}^k$ respectively represent the k-th disentangled embeddings for friend u_i and article a in the package, and$k \in [1, K]$. $\mathbf{f}^k$ represents the k-th disentangled embedding for social influence of friends. $\mathbf{f}$ is the final friends embedding encoding the complex social influence on the current user u.

Package Modeling

- **Inter-Package Attention Network**

Formally, given a package set $P = \{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_{|P|}\}$, where each package $\mathcal{P}_i \in P$ has an inter-connection with the current package $\mathcal{P}$, thus aggregating the package set with a gate filter to fuse the collaborative information as follows:

$$\widetilde{\mathbf{p}} = \mathbf{p} + \sum\nolimits_{\mathcal{P}_i \in P} \mathbf{g_i} \odot \mathbf{p_i}$$

$$\mathbf{g_i} = \sigma(\mathbf{W}_1 \mathbf{p} + \mathbf{W}_2 \mathbf{p_i} + \mathbf{b}_{g_1})$$

where $\mathbf{p_i}$ and $\mathbf{p}$ are the representations of package $\mathcal{P}_i$and $\mathcal{P}$ learned with intra-package attention network. $\widetilde{\mathbf{p}}$ is the final package representation.

User Modeling

- **Fine-Grained Aggregation Network**

Differentiate the contribution of multiple objectives with same type to the user preference aggregation, first. Then, aggregating multiple embeddings for the user u in various type spaces $\{\mathbf{u}^1, \ldots, \mathbf{u}^{|\mathcal{T}|}\}$ with a type-level attention.

$$\mathbf{u}^t = \sigma\left(\sum_{v \in \mathcal{A}_u^t} \eta_{uv}\mathbf{v}\right), \qquad \eta_{uv} = \frac{\exp(\sigma(\mathbf{w}_t^{\mathrm{T}}[\mathbf{u}||\mathbf{v}]))}{\sum_{v' \in \mathcal{A}_u^t} \exp(\sigma(\mathbf{w}_t^{\mathrm{T}}[\mathbf{u}||\mathbf{v}']))}$$

$$\hat{\mathbf{u}} = \sigma\left(\sum_{t \in \mathcal{T}} \epsilon_u^t \mathbf{u}^t\right), \qquad \epsilon_u^t = \frac{\exp(\mathbf{z}^T \cdot \tanh(\mathbf{w}[\mathbf{u}||\mathbf{u}^t]))}{\sum_{t' \in \mathcal{T}} \exp(\mathbf{z}^T \cdot \tanh(\mathbf{w}[\mathbf{u}||\mathbf{u}^{t'}]))}$$

where $\mathbf{u}^t$ is the t-induced representation of the user u, and $\mathbf{v}$ is the initial embeddings of object $v \in \mathcal{A}_u^t$. $\mathcal{A}_u$ is the different type objects (article, media) associated with u. $\hat{\mathbf{u}}$ is the fine-grained user representation.

User Modeling

- **Coarse-Grained Aggregation Network**

Aggregating historical interacted package of the user u with a gate attention mechanism as follows,

$$\tilde{\mathbf{u}} = \hat{\mathbf{u}} + \sum_{\mathcal{P}_i \in H} \mathbf{g_i} \odot \mathbf{p_i}$$

$$\mathbf{g_i} = \sigma(\mathbf{W}_3 \hat{\mathbf{u}} + \mathbf{W}_4 \mathbf{p_i} + \mathbf{b}_{g_2})$$

where H is set of packages that the user u interacted with, $\tilde{\mathbf{u}}$ is the final user representation.

Prediction and Optimization

- **Prediction**

Concatenating the representation of user u and package $\mathcal{P}$ and predict the probability score $\hat{y}_{up}$ of the interaction between$u$ and $\mathcal{P}$ with a two-layer MLP:

$$\hat{y}_{u\mathcal{P}} = \sigma(\mathrm{MLP}[\tilde{\mathbf{u}}||\tilde{\mathbf{p}}])$$

- **Objection function**

Optimizing the model with cross-entropy loss,

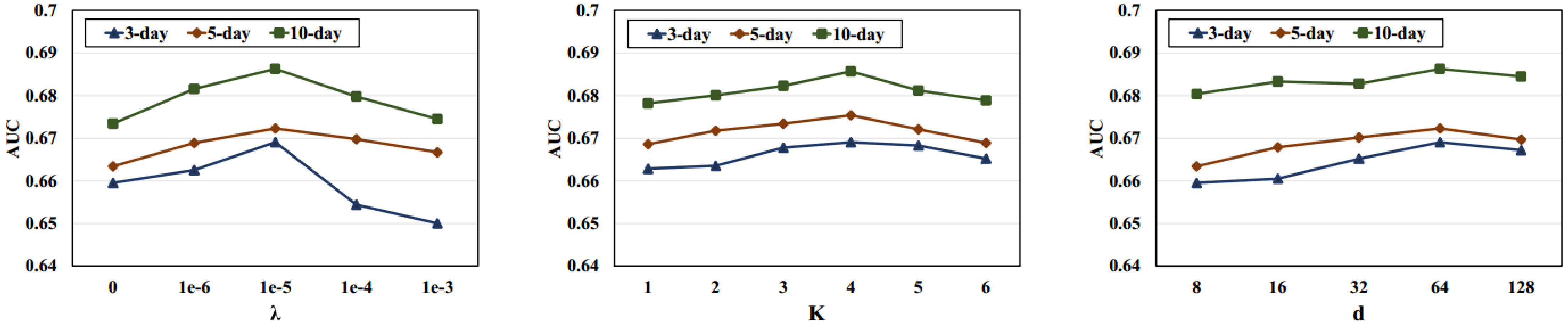
$$\mathcal{L} = -\sum_{<u,\mathcal{P}>\in Y} (y_{u\mathcal{P}} \cdot \log(\hat{y}_{u\mathcal{P}}) + (1 - y_{u\mathcal{P}}) \cdot \log(1 - \hat{y}_{u\mathcal{P}})) + \lambda||\Theta||$$

Experiments

Table 2: Experimental results on three datasets. The best method is bolded, and second best is underlined. The improvements of IPRec over the second best models are shown in the last column. '-' means that methods cannot obtain results due to out of memory.

Methods	Ratio	MF	DeepMF	NeuMF	TrustMF	DiffNet	GC-MC	NGCF	triple2vec	DAM	IPRec	Improve.
3-day	40%	0.5655	0.6292	<u>0.6363</u>	0.6026	0.6148	0.5987	0.6090	0.5971	0.6078	**0.6538**	2.75%
	60%	0.5683	0.6310	<u>0.6438</u>	0.6078	0.6319	0.6021	0.6123	0.5976	0.6089	**0.6572**	2.08%
	80%	0.5710	0.6360	<u>0.6486</u>	0.6122	0.6432	0.6054	0.6158	0.5982	0.6103	**0.6620**	2.07%
	100%	0.5728	0.6393	0.6497	0.6171	<u>0.6520</u>	0.6092	0.6220	0.5988	0.6129	**0.6691**	2.62%
5-day	40%	0.5624	0.6189	<u>0.6350</u>	0.5987	0.6211	0.6020	0.6144	0.5644	0.5930	**0.6558**	3.28%
	60%	0.5673	0.6228	<u>0.6415</u>	0.6012	0.6355	0.6065	0.6189	0.5637	0.5967	**0.6634**	3.41%
	80%	0.5685	0.6269	<u>0.6474</u>	0.6078	0.6465	0.6112	0.6250	0.5642	0.6012	**0.6701**	3.51%
	100%	0.5712	0.6313	0.6498	0.6095	<u>0.6580</u>	0.6152	0.6310	0.5648	0.6089	**0.6754**	2.64%
10-day	40%	0.5572	0.6101	<u>0.6370</u>	-	0.6335	-	-	0.5681	0.5834	**0.6667**	4.66%
	60%	0.5623	0.6168	0.6432	-	<u>0.6522</u>	-	-	0.5683	0.5892	**0.6732**	3.22%
	80%	0.5642	0.6180	0.6478	-	<u>0.6645</u>	-	-	0.5684	0.5922	**0.6781**	2.05%
	100%	0.5658	0.6203	0.6502	-	<u>0.6700</u>	-	-	0.5687	0.5960	**0.6853**	2.29%

Experiments



(a) Effect of the regularization weight λ.

(b) Effect of the number of disentangle spaces K.

(c) Effect of the embedding dimension d.

Table 3: Ablated models analysis.

Model	3-day	5-day	10-day
IPRec$_{w/o\ att}$	0.6598	0.6638	0.6745
IPRec$_{w/o\ dual}$	0.6638	0.6710	0.6800
IPRec$_{w/o\ inter}$	0.6645	0.6690	0.6814
IPRec	**0.6691**	**0.6754**	**0.6853**