Attentive Collaborative Filtering: Multimedia Recommendation with Item- and Component-Level Attention

9 Aug, 2017

Jingyuan Chen, Hanwang Zhang, Xiangnan He, Liqiang Nie, Wei Liu, Tat-Seng Chua
National University of Singapore / Columbia University / Shandong University/ Tencent

OUTLINE

- Introduction
- Motivation
- Attentive Collaborative Filtering
- Experimental Result
- Conclusion

Value of recommendation

Value for the customer

- Find things that are interesting
- Narrow down the set of choices
- Help to explore the space of options
- Discover new things

Value for the provider

- Provide personalized service for the customer
- Increase sales
- Obtain more knowledge about customers

Amazon: 35% of sales result from recommendations

Netflix: 75% of views result from recommendations

Google News: 38% of clickthrough results from recommendations

Why multimedia recommendation

- Information overload

Amount of videos uploaded: 300 hours/min

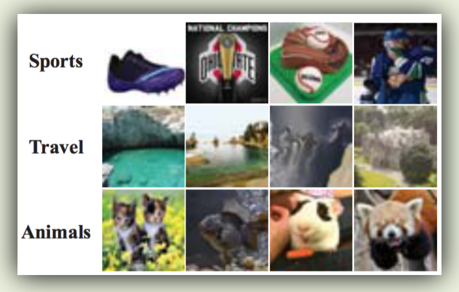
Average number of monthly searches: 2 billion

Average number of photos and videos shared daily: 95 million

- Exploratory user behavior

Multimedia information seeking is often for entertainment. Users explore the multimedia space with no clear end goal.

- Diversity

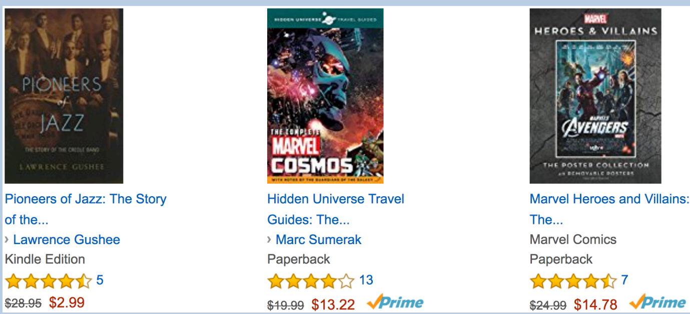


Different from products, multimedia contents are hard to be categorized or described.

The intention is there, but cannot be explicitly expressed.

Implicit Feedback in MM

- Explicit Feedback



As implicit feedback lacks evidence on how users like and dislike items, it is a major challenge that MM recommender systems should tackle.

User preference is known

- Implicit Feedback

Pin on Pinterest

Revine on Vine

Watch history on Youtube

No preference information

OUTLINE

- Introduction
- Motivation
- Attentive Collaborative Filtering
- Experimental Result
- Conclusion

Two levels of implicitness – Item-level

NO RATINGS

- ✓ A positive set of user feedback does not necessarily indicate equal item preferences.
- ✓ ***Item-level implicitness***: user's preference on each item is unknown.

Two levels of implicitness – Component-level

Two levels of implicitness – Component-level

- ✓ Positive feedback on multimedia content is merely in the whole content level. However, multimedia content usually contains diverse semantics.
- ✓ ***Component-level implicitness***: user’s preference on different components of the item is unknown.

Limitations of previous work

Item-level implicitness

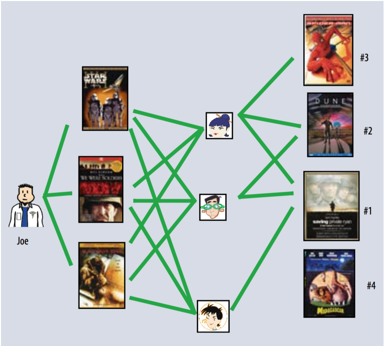
- ✓ Most efforts are focused on how to select the negative items (popularity-based [He et al.])
- ✓ As for positive item, only constant weight for each item is considered [Koren et al.]

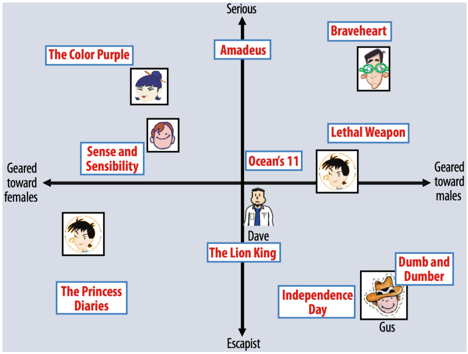
Component-level implicitness

OUTLINE

- Introduction
- Motivation
- Attentive Collaborative Filtering
  - Preliminaries
  - Model
- Experimental Result
- Conclusion

Latent Factor Models [Koren et al. 2009]





$$
\begin{array}{c|cccc|}
 & i_1 & i_2 & i_3 & i_4 \\
\hline
u_1 & ? & 1 & 2 & ? \\
u_2 & 2 & ? & ? & 4 \\
u_3 & 2 & 1 & ? & ? \\
u_4 & ? & 2 & 2 & 1 \\
\hline
\end{array}
$$

user

item

$$
= \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_n \end{bmatrix} \approx \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix} \times \begin{bmatrix} v_1 & v_2 & v_3 & \cdots & v_m \end{bmatrix}
$$

$$\hat{R}_{ij} =< u_i, v_j >$$

User-Item Matrix **Latent Space**

Bayesian Personalized Ranking (BPR) -- Implicit Feedback

user × item:

	i_1	i_2	i_3	i_4
u_1	?	+	+	?
u_2	+	?	?	+
u_3	+	+	?	?
u_4	?	+	+	+

u_1: $i >_{u1} j$

item × item:

	i_1	i_2	i_3	i_4
j_1		+	+	?
j_2	-		?	-
j_3	-	?		-
j_4	?	+	+	

…

u_4: $i >_{u4} j$

item × item:

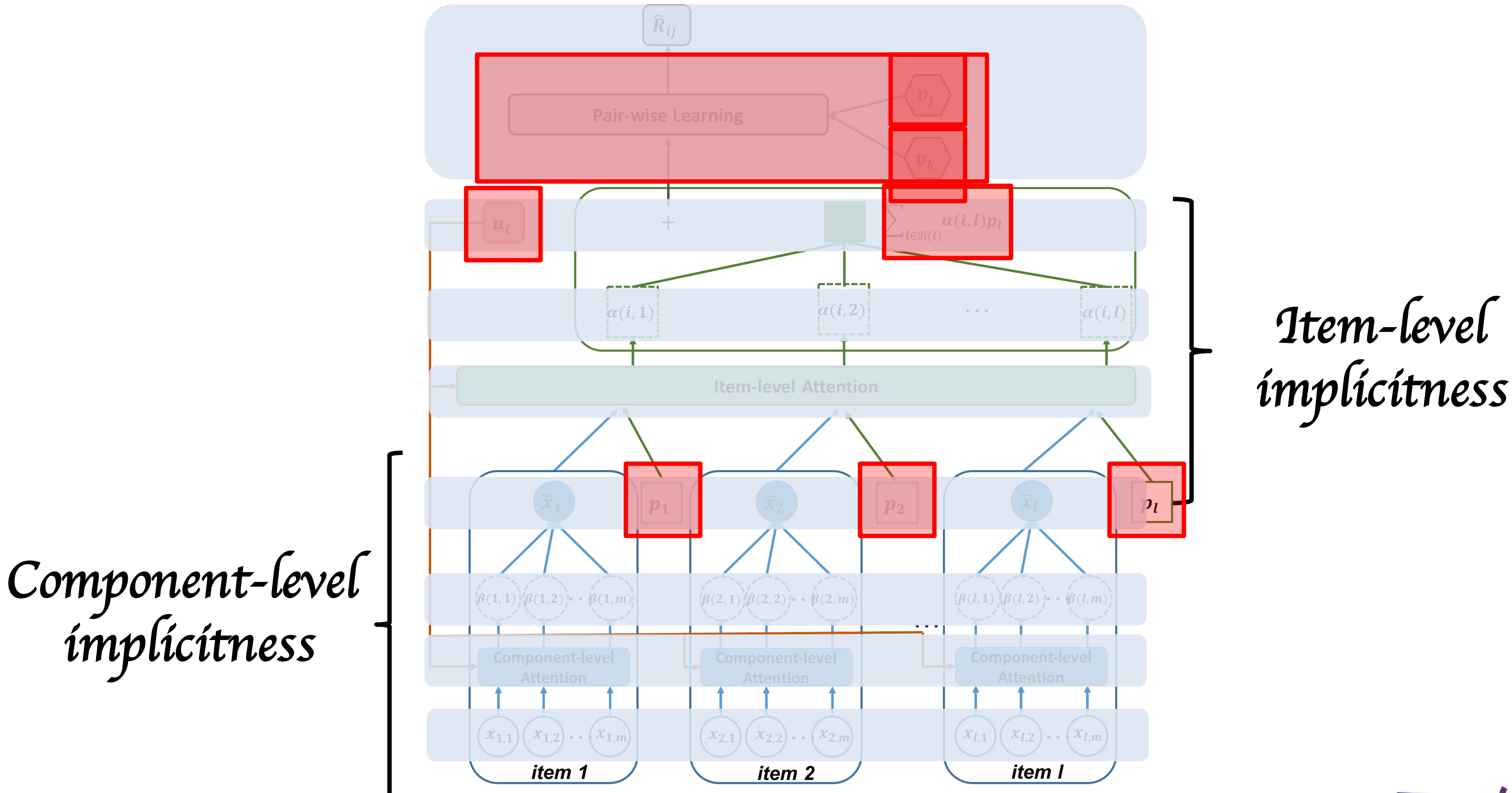
	i_1	i_2	i_3	i_4
j_1		+	+	+
j_2	-		?	?
j_3	-	?		?
j_4	-	?	?	

$$\arg\min_{\mathbf{U},\mathbf{V}} \sum_{(i,j,k)\in\mathcal{R}_B} -\ln\sigma(\hat{R}_{ij}-\hat{R}_{ik}) + \lambda(||\mathbf{U}||^2 + ||\mathbf{V}||^2)$$

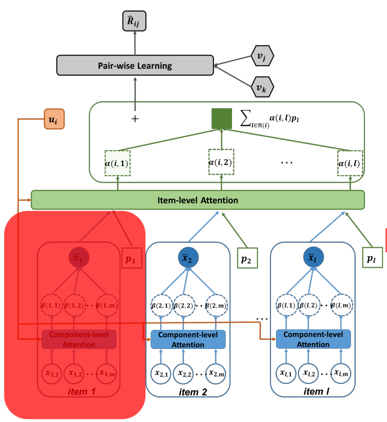
OUTLINE

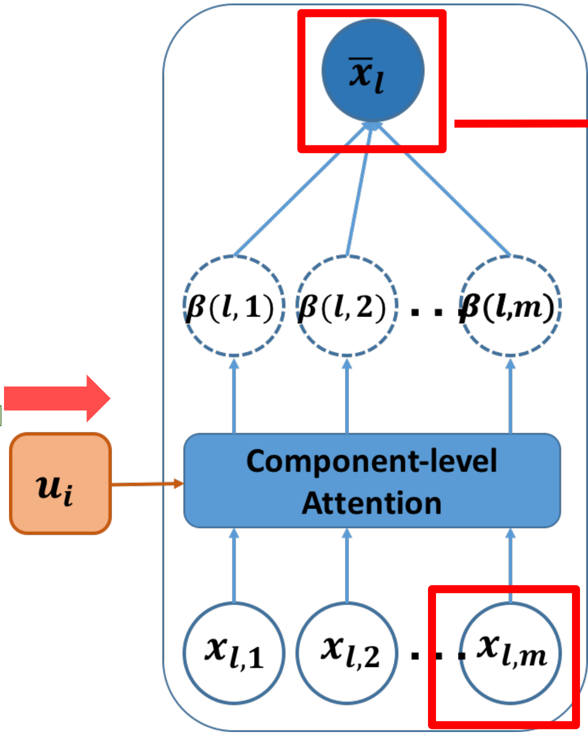
- Introduction
- Motivation
- Attentive Collaborative Filtering
  - Preliminaries
  - Model
- Experimental Result
- Conclusion

Attentive Collaborative Filtering
-- General Framework



Component-Level Attention





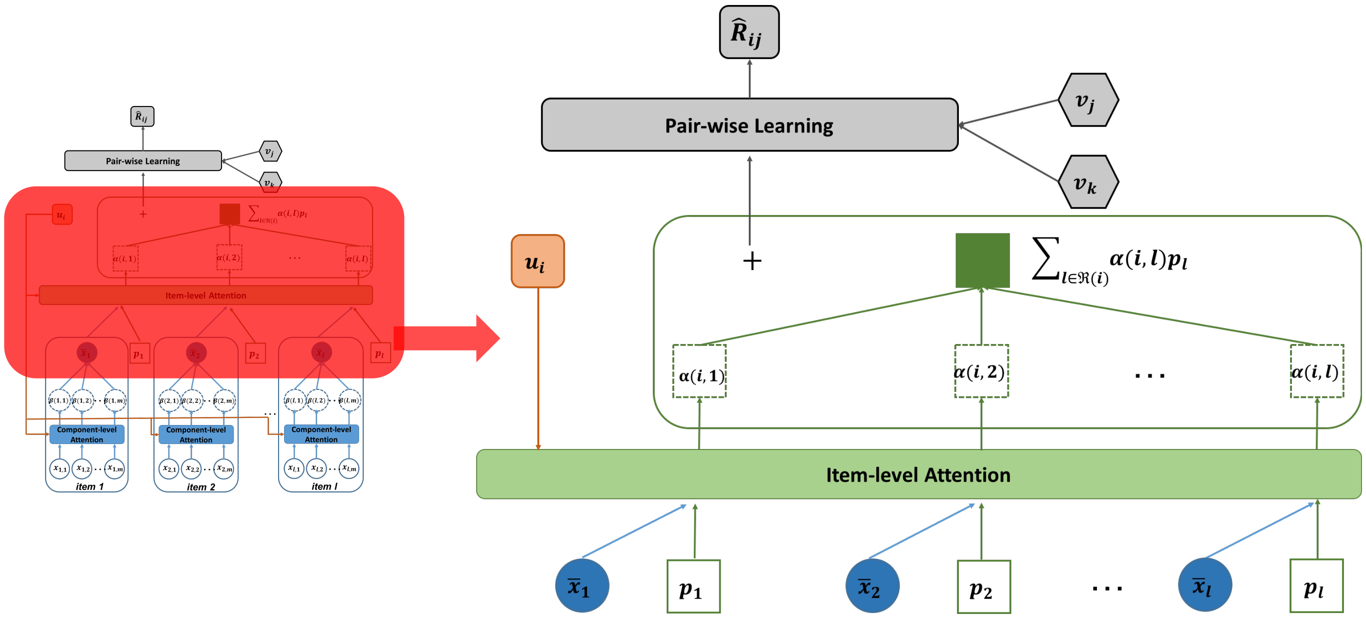
$$\overline{\mathbf{x}}_l = \sum_{m=1}^{|\{\mathbf{x}_{l*}\}|} \beta(i,l,m)\cdot \boxed{\mathbf{x}_{lm}}$$

$$b(i,l,m) = \mathbf{w}_2^T \phi(\mathbf{W}_{2u}\mathbf{u}_i + \mathbf{W}_{2x}\mathbf{x}_{lm} + \mathbf{b}_2) + \mathbf{c}_2$$

$$\phi(x) = max(0,x)$$

$$\beta(i,l,m) = \frac{exp\,(b(i,l,m))}{\sum_{n=1}^{|\{\mathbf{x}_{l*}\}|} exp\,(b(i,l,n))}$$

Item-Level Attention



$$\hat{R}_{ij} = \left(\mathbf{u}_i + \sum_{l\in\mathcal{R}(i)} \alpha(i,l)\mathbf{p}_l\right)^T \mathbf{v}_j$$

Attention score: $$a(i,l) = \mathbf{w}_1^T \phi(\mathbf{W}_{1u}\mathbf{u}_i + \mathbf{W}_{1v}\mathbf{v}_l + \mathbf{W}_{1p}\mathbf{p}_l + \mathbf{W}_{1x}\overline{\mathbf{x}}_l + \mathbf{b}_1) + \mathbf{c}_1$$

$$\phi(x) = max(0, x)$$

Normalization: $$\alpha(i,l) = \frac{exp(a(i,l))}{\sum_{n=1}^{|\mathcal{R}(i)|} exp(a(i,n))}$$

OUTLINE

- Introduction
- Motivation
- Attentive Collaborative Filtering
  - Preliminaries
  - Model
- Experimental Results
- Conclusion

Experimental Settings

- Datasets

Dataset	Interaction#	Item#	User#	Sparsity
Pinterest	1,091,733	14,965	50,000	99.85%
Vine	125,089	16,243	18,017	99.96%

- Evaluation Protocols
  - **Hit Ratio (HR):** measures whether the ground truth item is present on the ranked list
  - **NDCG:** accounts for the position of hit.

- Component-level Feature Extraction
  - **Image**: *res5c* layer feature map in *ResNet* $(7\times7\times2048)$
  - **Video**: *pool5* layer in *ResNet* (2048)

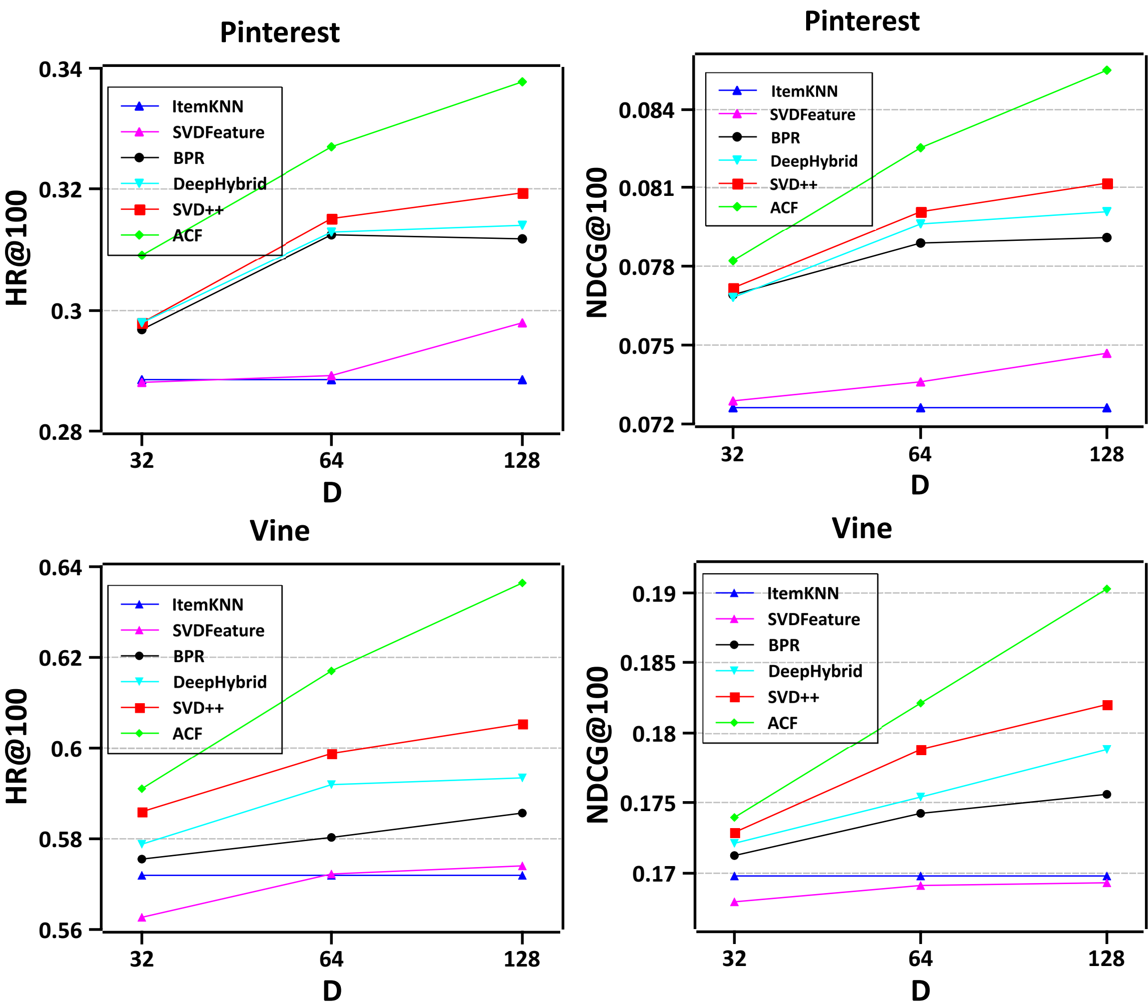
Baselines

- **CF-based Methods**
  - UCF: user-based collaborative filtering [Zhao et al.]
  - ItemKNN: item-based collaborative filtering [Hu et al.]
  - BPR: [Rendle et al.]
  - SVD++: a merged model of latent factor and neighborhood models [Koren et al.]

- **Content-based Methods**
  - CBF: content-based filtering [Pazzani et al.]

- **Hybrid Methods**
  - SVDFeature: is a generic model for feature-based collaborative filtering [Chen et al.]
  - Deep Hybrid: uses convolution neural network to regress multimedia content to the item latent vectors [Oord et al.]

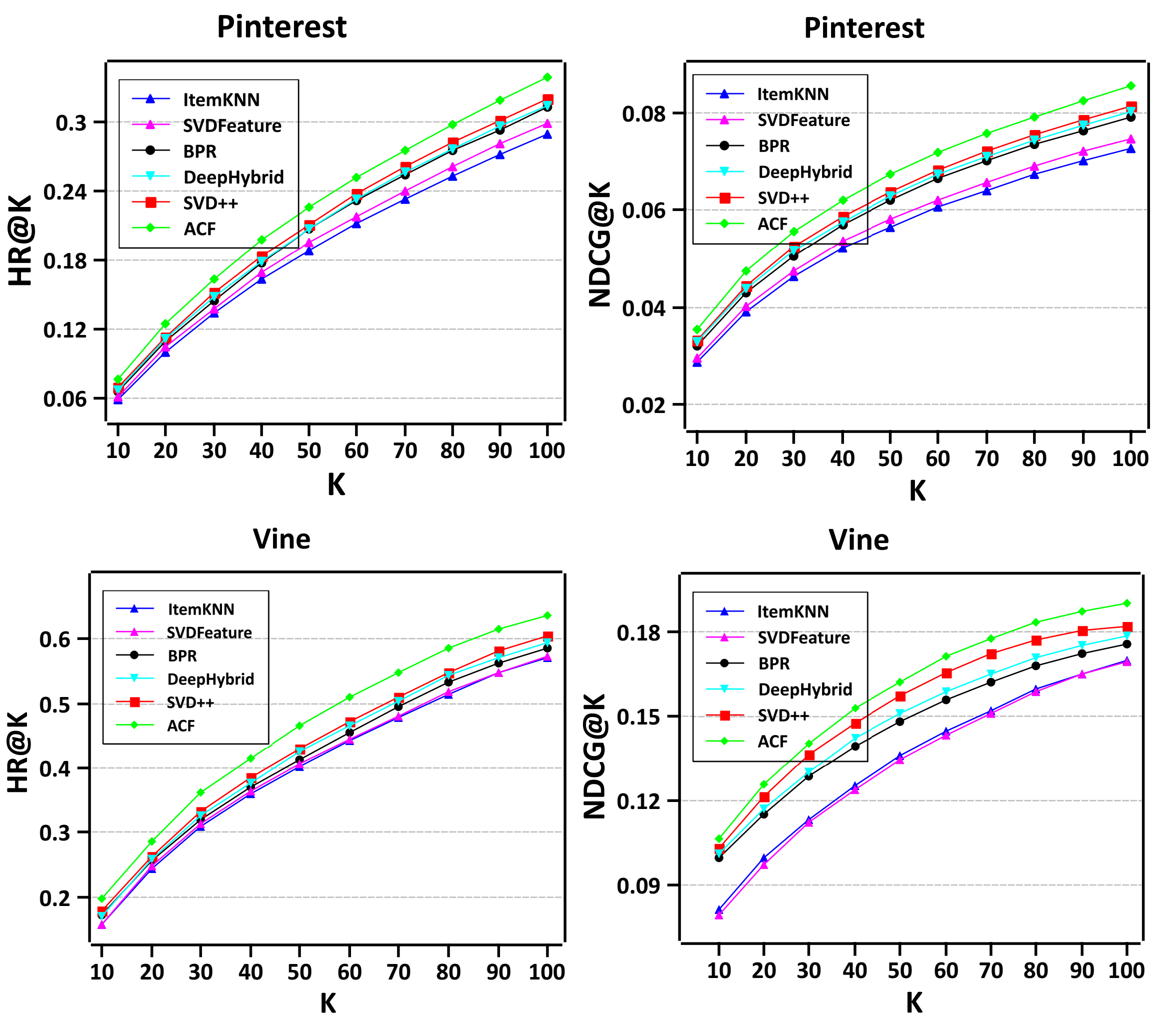
Model Comparison



The performance of HR@100 and NDCG@100 with respect to the number of latent factors.

- ACF achieves the best performance.
- Although the Vine dataset is more sparse than Pinterest, the performance is much better.
- With the increase of the number of latent factors, the performance improvement of ACF compared with other baseline methods also increases.

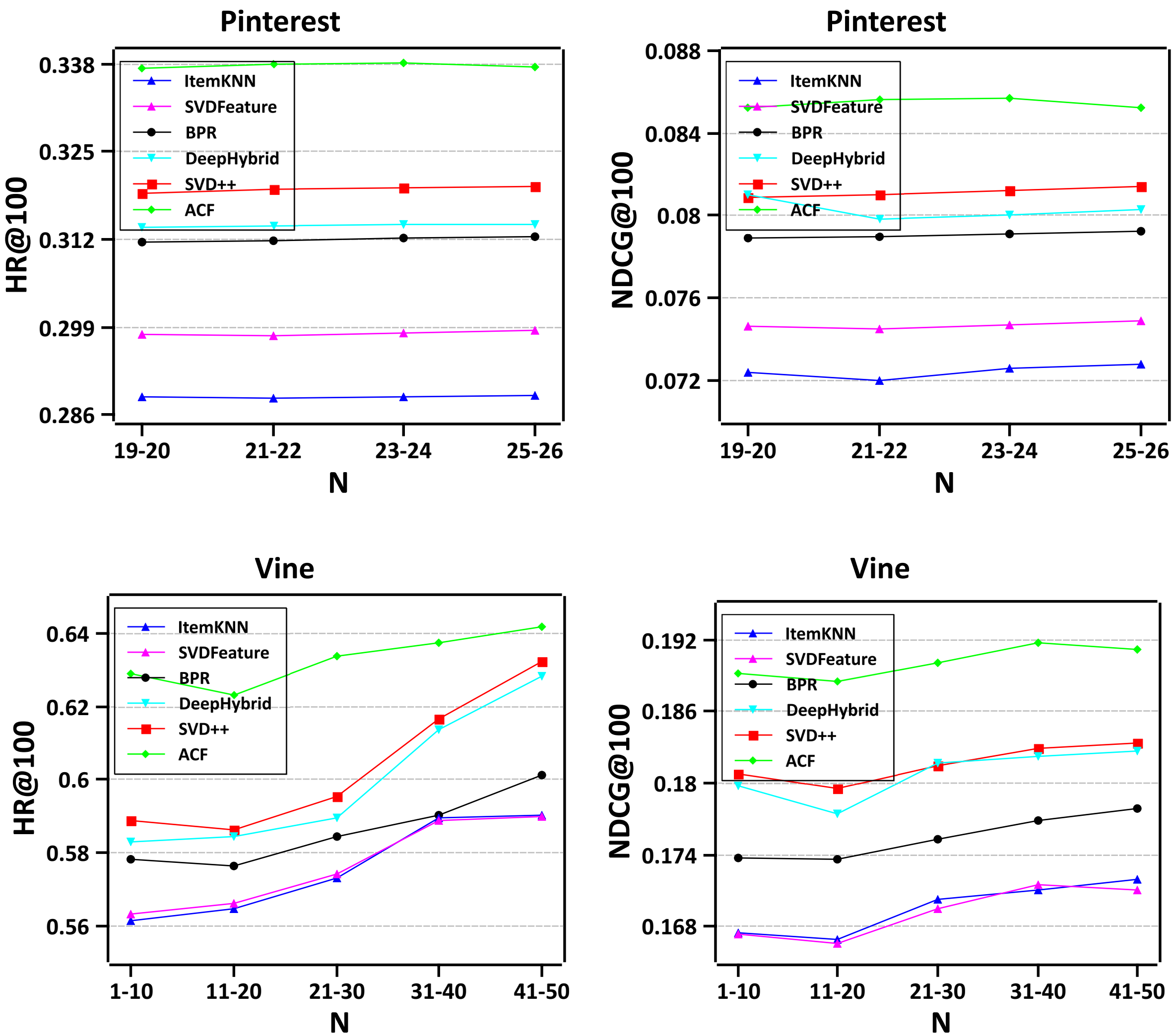
Model Comparison



The performance of Top-K recommended lists where the ranking position K ranges from 10 to 100.

- ACF demonstrates consistent improvements over other methods across positions.

Model Analysis: Performance over Users of Different Sparsity Levels



The performance with respect to the number of items a user has.

- ACF consistently outperforms other baseline methods for all the number of item settings.
- When the number of items per user is relatively small, ACF performs much better than the other methods.

Model Ablation

- Effect of Attention Mechanisms in Item- and Comp-Level

Model	Level		Pinterest		Vine	
	Item	Feature	HR	NDCG	HR	NDCG
ACF	AVG	–	31.95%	8.12%	60.54%	18.20%
	ATT	AVG	33.21%	8.42%	62.81%	18.75%
	ATT	ATT	**33.78%***	**8.55%***	**63.65%***	**19.03%***

  – Both attention mechanisms applied in item- and component- level improve the performance for multimedia recommendation compared with utilizing average pooling in each level.
  – The attention mechanism in item-level contributes more for our model as compared to that in component-level.

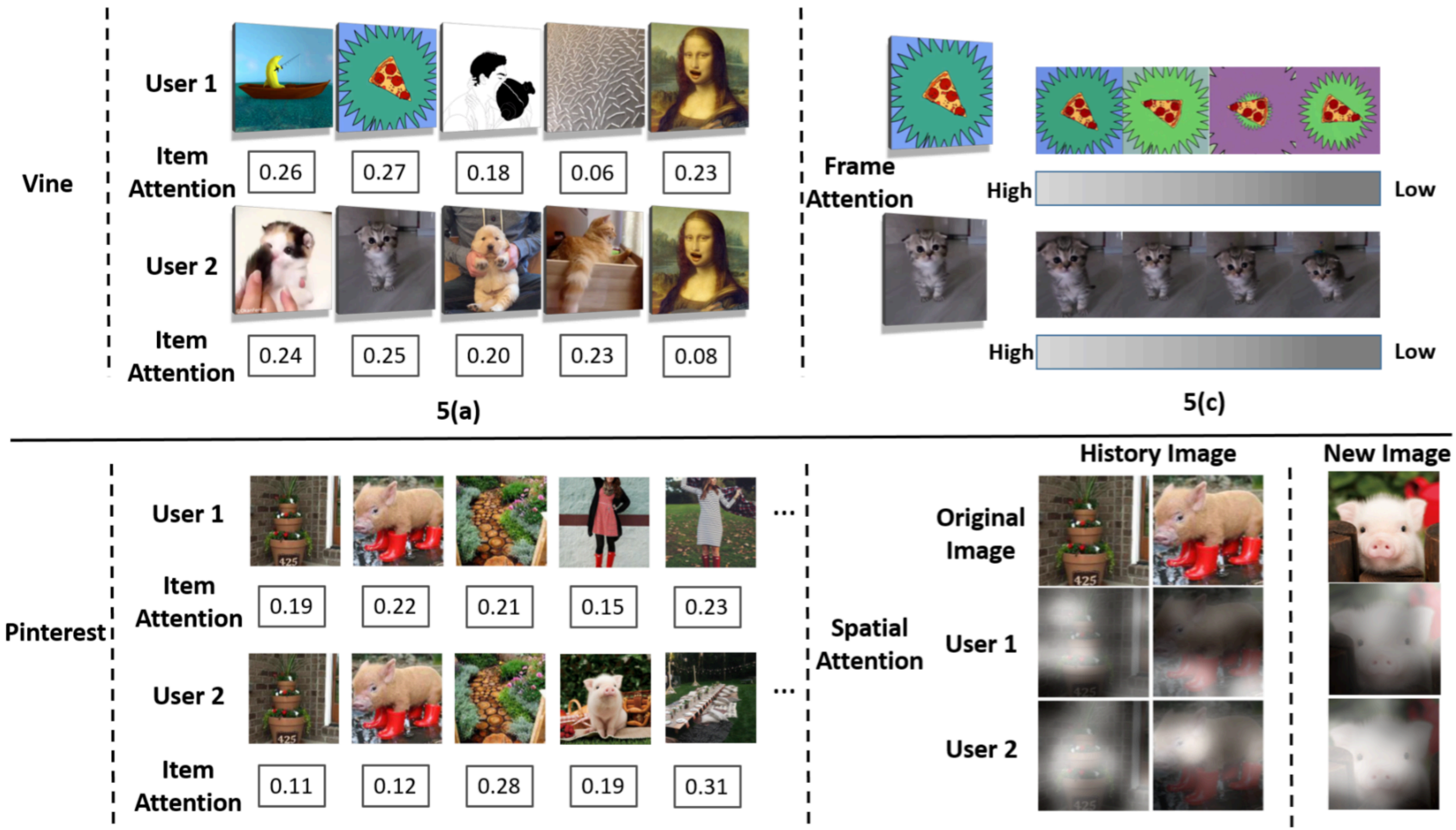
Model Ablation

- Effect of User, Item and Content Information

Model	Attention Type	Pinterest		Vine	
		HR	NDCG	HR	NDCG
ACF	None	31.95%	8.12%	60.54%	18.20%
	U+V	32.17%	8.31%	61.68%	18.36%
	U+P	32.69%	8.34%	62.37%	18.65%
	U+V+P	32.96%	8.32%	62.60%	18.71%
	U+V+P+X	**33.78%***	**8.55%***	**63.65%***	**19.03%***

  - The information of both user and item contributes to our model as compared to a constant weight model.
  - The information of users is more effective than the items to enhance recommendation.

Attention Visualization



OUTLINE

- Introduction
- Motivation
- Attentive Collaborative Filtering
  - Preliminaries
  - Model
- Experimental Results
- Conclusion

Conclusion

- We have introduced the component- and item-level attention model to assign attentive weights to infer the underlying user preference encoded in the implicit user feedback.
- We have conducted extensive experiments on two real-world multimedia social networks: Vine and Pinterest, to demonstrate the effectiveness of ACF.

- ***Key take-way insight***: inferring the underlying user preference encoded in the implicit feedback in a distant supervised manner should be explored towards *Explainable Recommendation*.

NUS-Tsinghua Centre for Extreme Search
A Joint Research Collaboration Between NUS & Tsinghua University

THANK YOU

arigatou gozaimasu

ありがとうございます

谢　谢　！

NExT research is supported by the National Research Foundation,
Prime Minister's Office, Singapore under its IRC@SG Funding Initiative.