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

IIC 3633 Sistemas Recomendadores

Paula Navarrete Campos
Astrid San Martín Jiménez

Departamento Ciencias de la Computación
Facultad de Ingeniería
Pontificia Universidad Católica de Chile

20 noviembre 2018.

Outline

Introduction (2 min)

Related Work (3 min)

Preliminars (2 min)

Attentive Collaborative Filtering (4 min)

Experiments(3 min)

Conclusions (2 min)

Introduction
Context

Increasing need to **sift through massive multimedia contents** for users in a **highly dynamic environment** such as Web multimedia content.

Items are multimedia contents consumed by users (video, photo, song)

Absence of negative feedback (implicit feedback)

Two levels of implicit feedback (IF) are proposed: item and component level.

Introduction
Implicit Feedback Levels

Item-Level IF:

preference information on each item is not provided.

A positive set of user feedback can be biased and thus not necessarily indicate real item preference (ex. social likes to friends and family).

! Neighborhood context obtained fails to model the item-level implicit feedback

Introduction
Implicit Feedback Levels

Component-Level IF:

When **feedback for each component is not available**.

Play feedback on a video **does not implie like on all components** of it

! *Model user preferences with lower-level content components* (image features in different locations and video features of various frames).

Introduction

Attentive Collaborative Filtering (ACF) CF Framework

Automatically assigns weights to the two levels of feed back in a distant supervised manner.

Draws on the latent factor model transforming both items and users to the same latent factor space to make them directly comparable.

Can be efficiently **trained using Stochastic Gradient Decent**(SGD) on large user-item interactions of images and videos.

Related Work
Implicit Feedback

Dubbed the one-class problem due to lack of negative feedback

The remaining data is a mixture of real negative feed back and missing values. Coping approaches:

- **Sample based learning:** samples negative feedback from the missing data *!* More effective.
- **Whole-data based learning:** treats all the missing data as negative. *!* higher coverage.

Related Work
Implicit Feedback

Recent efforts focus on the **weighting scheme**, considering the confidence whether the unobserved samples are indeed negative ones.

Non-uniform weighting schemes are defined based on authors' assumptions *!* **may be biased**

Attention mechanism weights positive implicit signal automatically based on the user item interaction matrix and the content of the item.

Item-level attention and *component-level* attention can be seen as the **weighting strategy on positive samples**.

Related Work
Multimedia Recommendation

Classical CF is good for popular and frequently watched contents but **less applicable to fresh or tail contents** (due to the data sparsity).

Handling the coldstart scenario:

1. use different **context information** (multi-modal relevance, cross-domain knowledge and latent attributes feature)
2. **hybrid approaches** combine video content (topics mined from video metadata, related queries, etc.) with the co-view information.
3. Use a **latent factor model** for recommendation, and further predicting the latent factors from multimedia contents

Do not pay attention to the two levels of implicitness in the multimedia recommendation.

Related Work
Attention Mechanism

Effective in various machine learning tasks such as image/video captioning and machine translation.

Soft attention **learns to assign attentive weights** for a set of features ! higher (lower) weights indicate features are more informative (less informative) for the end task.

Reasonably assumes that human **recognition does not tend to process a whole signal at once**; instead focuses on selective parts when and where as needed.

Component-level attention: *soft spatial attention* model for images and *soft temporal attention* model for videos.

Preliminars

$\mathbf{R}\ 2\ \mathrm{R}^{MxN}$! user-item interaction matrix.

M, N ! users and items.

R_{ij} ! implicit feedback: 1 if interacted, 0 otherwise.

$R = f(i,j)jR_{ij} = 1g$! set of user-item pairs with implicit interactions.

goal ! exploit the entire $\mathbf{R}$ to estimate $\hat{R}_{ij}$ for the unobserved interactions.

Preliminars
Latent Factor Models

mapping of users and items to a joint low dimensional latent space where the user-item preference score is estimated by vector inner product.

$\mathbf{U} = [u_1, \ldots, u_M]$ 2 R^{DxM} ! user latent vectors

$\mathbf{V} = [v_1, \ldots, v_N]$ 2 R^{DxN} ! item latent vectors

D $min(M, N)$ latent feature dimension

$\hat{R_{ij}} = hu_i, v_j i = u_i^t v_j$! preference score

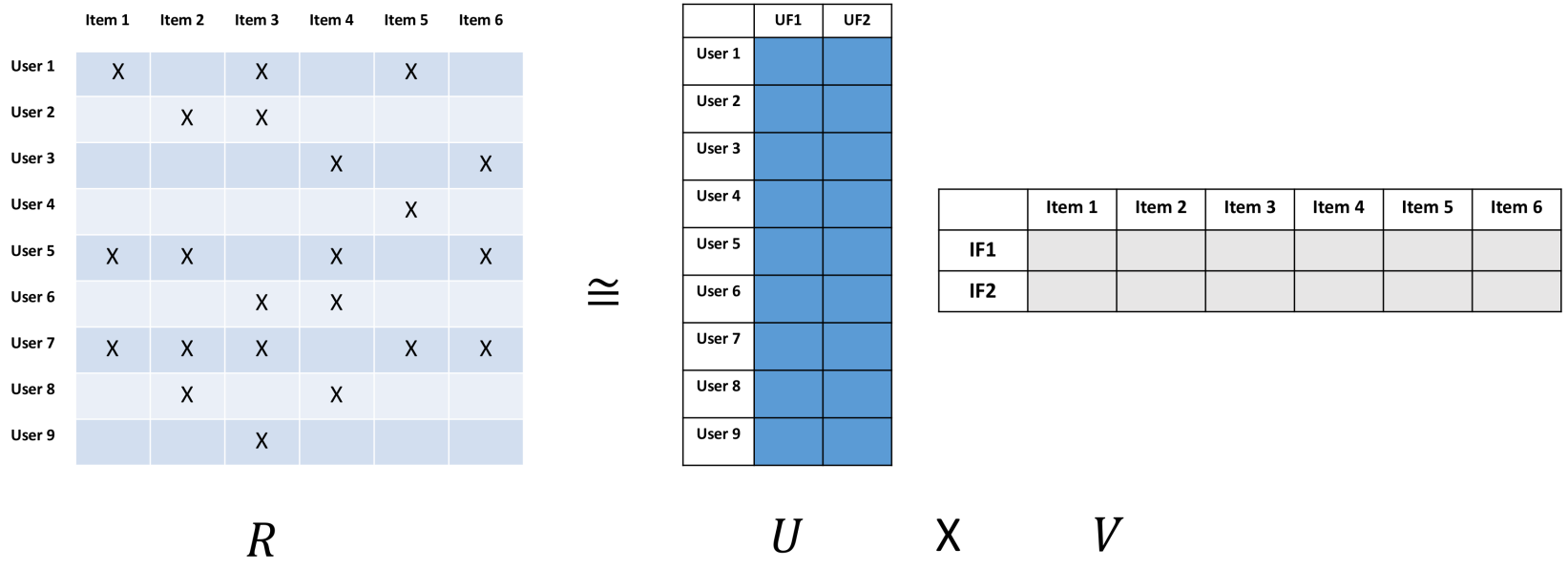
$$argmin_{U;V} \sum_{(i;j)2R} (R_{ij} \quad \hat{R_{ij}})^2 + \lambda(jjUjj^2 + jjVjj^2) \qquad (1)$$

λ controls the strength of regularization (usually an L2 norm)

Preliminars
Latent Factor Models

Recommendation is **reduced to a ranking problem** according to the estimated scores $\hat{R}_{ij}$.

Difficulties arise when **carelessly treating the unobserved entries** ! negative samples, it may introduce false negative samples in the training data.



Preliminars

Bayesian Personalized Ranking(BPR)

Models a triplet (i, j, k) of one user and two items - one of the items is observed and the other one is not.

When item j has been viewed by user i, assumes that i **prefers** j **over all the other unobserved items**.

I ! set of all items in the dataset

$R(i)$! set of items that are interacted by the i-th user.

$$argmin_{U;V} \sum_{(i;j;k)2R_B} ln\sigma(\hat{R_{ij}} \quad \hat{R_{ik}})^2 + \lambda(jjUjj^2 + jjVjj^2) \qquad (2)$$

$R_B = f(i,j,k)jj\ 2\ R(i)\ \hat{}\ k\ 2\ I\ n\ R(i)g$

$(i,j,k)\ 2\ R_B$! user i prefers item j over k.

effective in exploiting the unobserved user-item feedback.

Attentive Collaborative Filtering
Framework

Neural network to model user's preference score with respect to the item in item-level and content in component-level.

$\alpha(i, l)$! user i's preference degree in item l.

$\beta(i, l, m)$! user i's preference degree in the m-th component of item l.

Two attention sub-networks to learn these two preference scores jointly.

1. Component-level module generates content representations for each item.
2. item-level module obtains user representations.

Attentive Collaborative Filtering
Objective Function

$$\arg\min_{\mathbf{U},\mathbf{V},\mathbf{P},\Theta} \sum_{(i,j,k)\in\mathcal{R}_B} -\ln\sigma\left\{\left(\mathbf{u}_i + \sum_{l\in\mathcal{R}(i)} \alpha(i,l)\mathbf{p}_l\right)^T \mathbf{v}_j - \left(\mathbf{u}_i + \sum_{l\in\mathcal{R}(i)} \alpha(i,l)\mathbf{p}_l\right)^T \mathbf{v}_k\right\} + \lambda(||\mathbf{U}||^2 + ||\mathbf{V}||^2 + ||\mathbf{P}||^2)$$

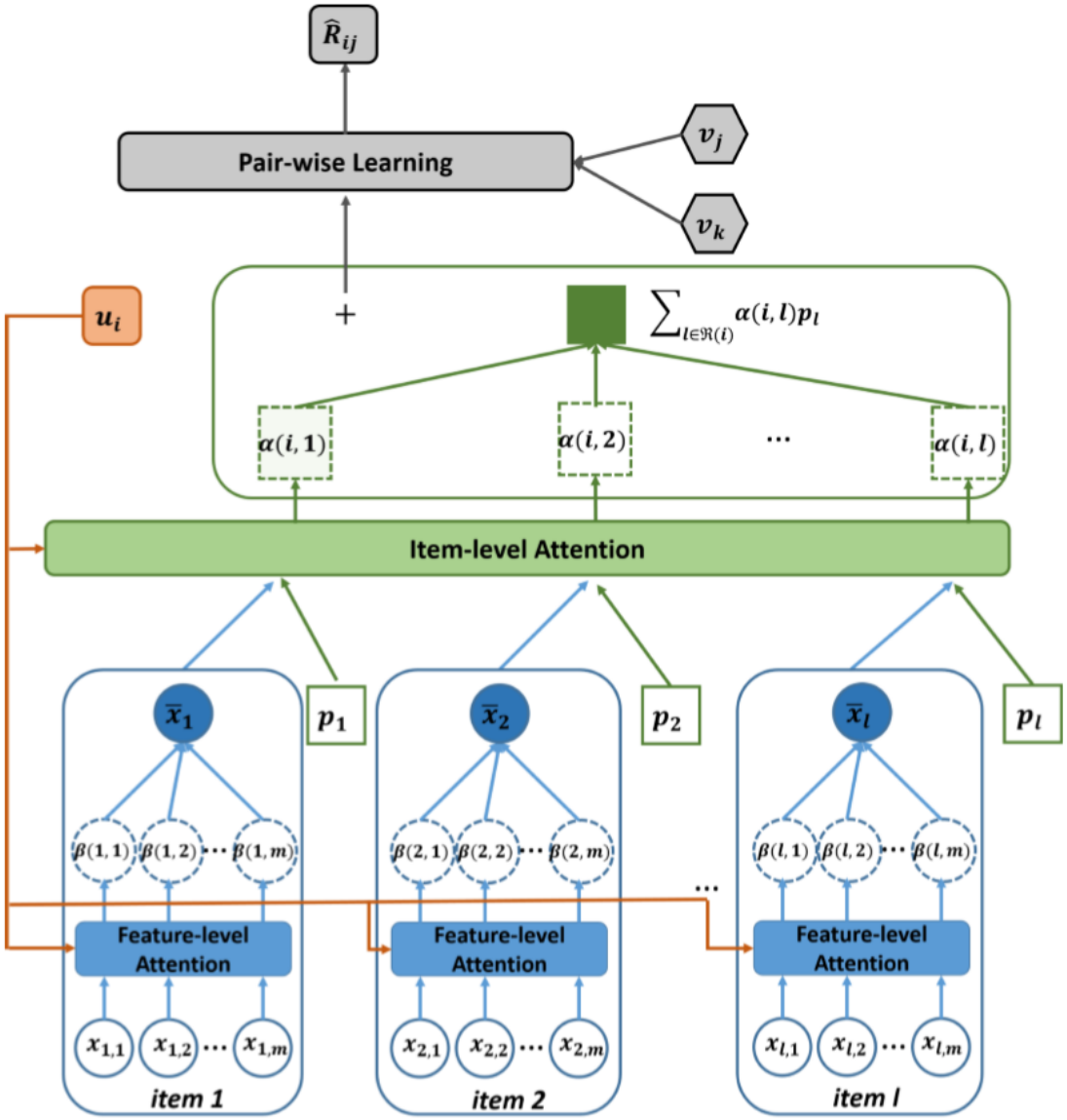
item I is associated with v_I (vector in latent factor model) and p_I that characterizes users based on the set of items they interacted with.

$u_i + \sum_{I2R(i)} \alpha(i,I)p_I$! user representation

Ranking on estimated score $\hat{R}_{ij} = \left(u_i + \sum_{I2R(i)} \alpha(i,I)p_I\right)^T v_j$

Attentive Collaborative Filtering

Architecture



Attentive Collaborative Filtering

Item-Level Attention

Select items that are representative to users' preferences and aggregate the representation to characterize users.

two-layer network to compute the attention score as:

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

matricesW_1 and bias b_1 are the first layer parameters,

vector w_1 and bias c_1 are second the layer parameters

$\phi(x) = max(0,x)$! is the ReLU function.

item-level weights are obtained by normalizing the attentive scores using Softmax.

$$\alpha(i,l) = \frac{exp(a(i,j))}{\sum_{n2R(i)} exp(a(i,n))}$$

Attentive Collaborative Filtering
Component-Level Attention

Assign components attentive weights that are consistent with user preference.
Weighted sum to construct the content representation.
Item l is coded into a variable-sized set of component features x_l .
two-layer network to compute the component score as:

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

Analogous to item-level.
Then the content representation of item l with the encoded preference of user i :

$$\bar{x}_l = \sum_{m=1}^{j f x_l \ \ g j} \beta(i,l,m) \ \ x_{lm}$$

Experiments
Two Research Questions

RQ1

Does ACF outperform state-of-art recommendation methods?

RQ2

How do the proposed item-level and component-level attentions perform?

Experiments
Datasets

Dataset 1

Pinterest

Dataset 2

Vine

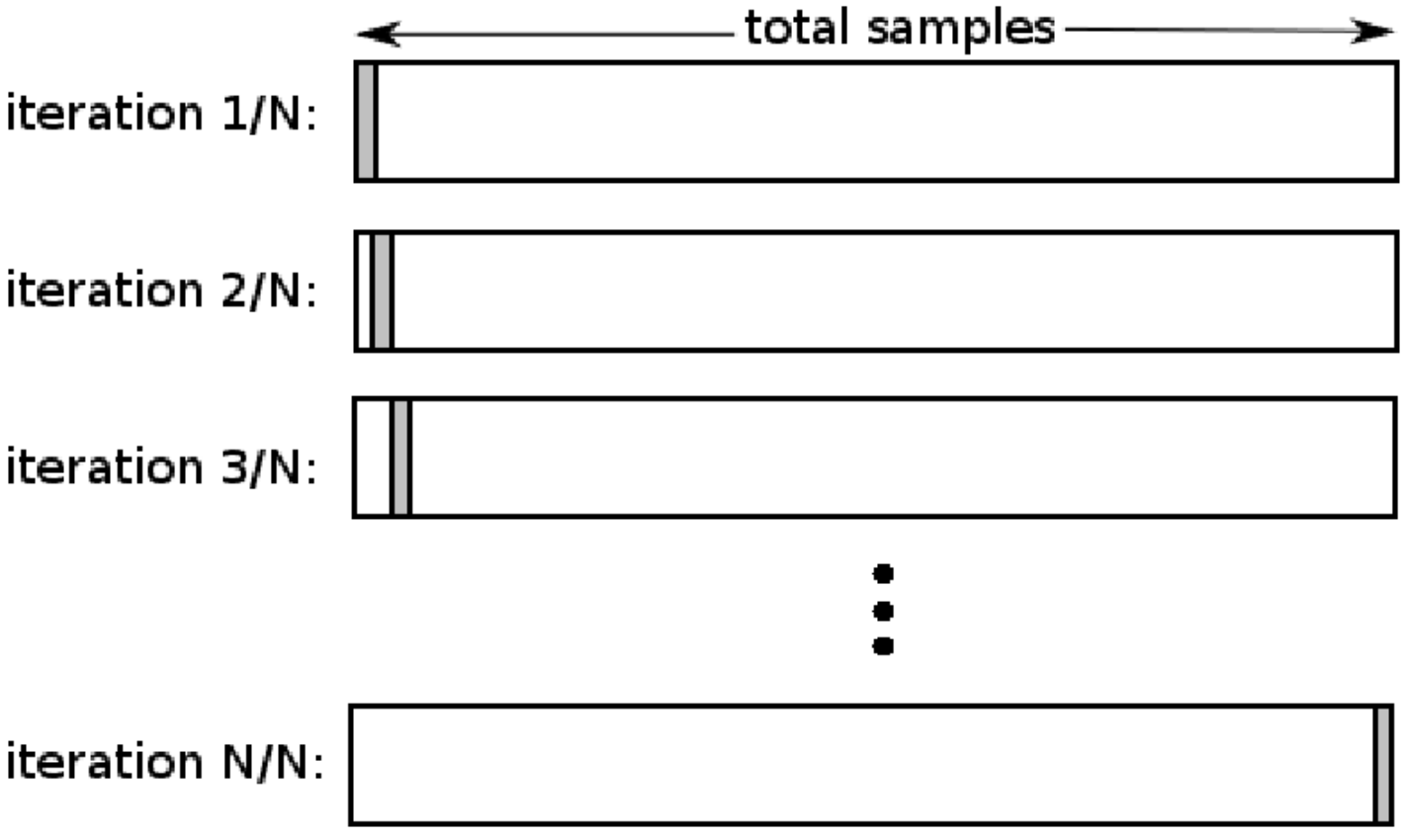
Experiments
Datasets

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

Table: Statistics datasets

Experiments
Evaluation

Leave-one-out for item recommendation



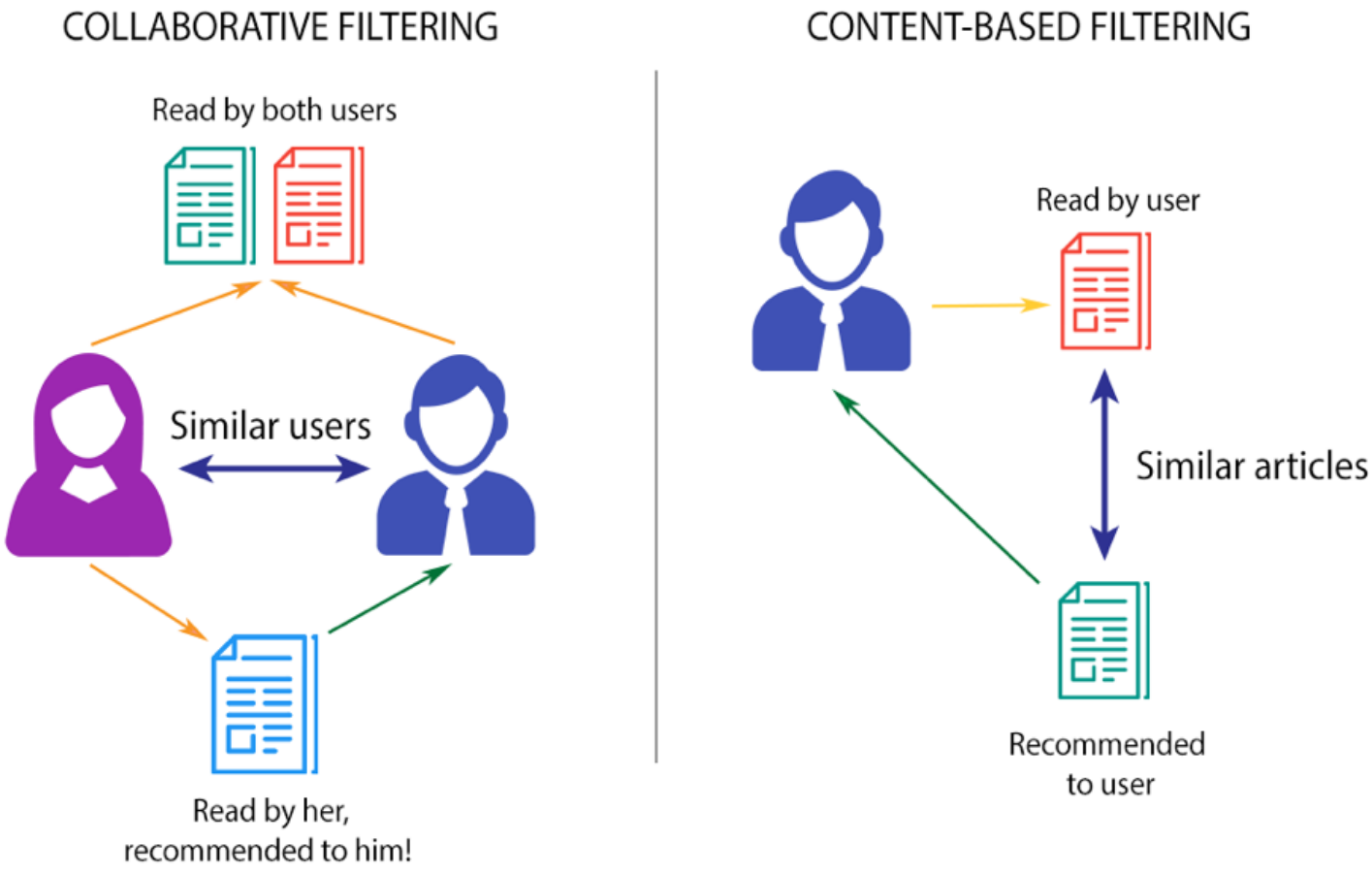
Experiments
Evaluation

HR Hit Ratio: measures whether the ground truth item is present on the ranked list.

NDCG Normalized Discounted Cumulative Gain: accounts for the position of hit.

Experiments
Baselines

Methods



Experiments
Baselines

Methods

CF-based

- **UCF**
- **ItemKNN**
- **BPR**
- **SVD++**

Content-based

- **CBF**

Hybrid

- **SVDFeature**
- **Deep Hybrid**

Experiments
Feature Extraction

ResNet-152

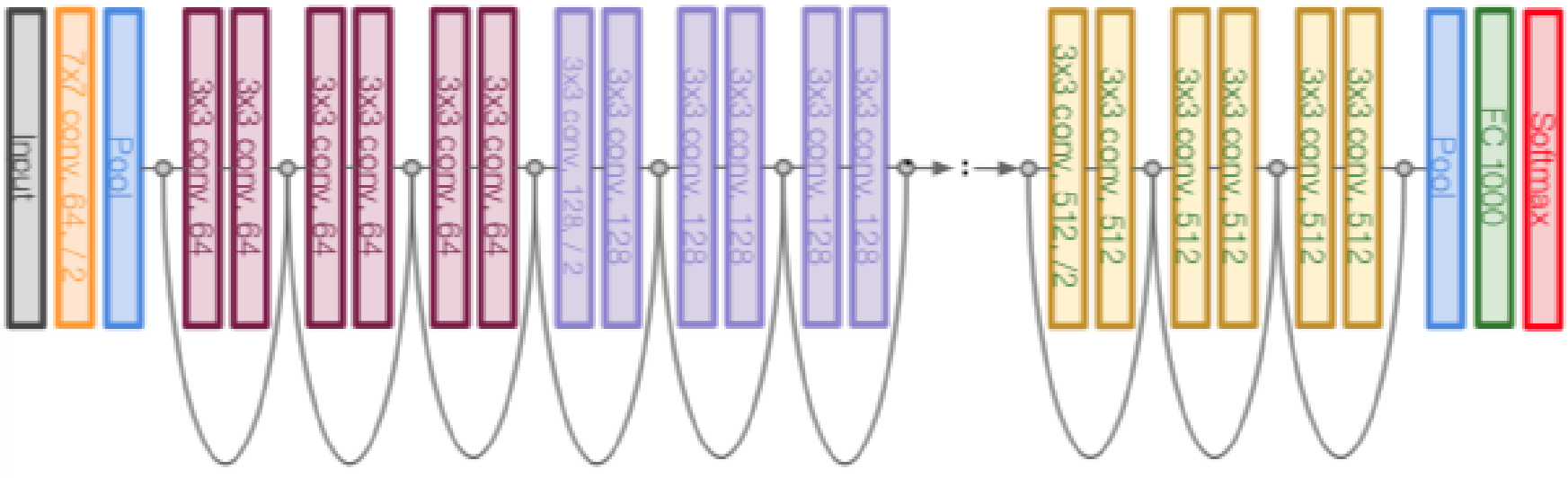


Figure: Sun et al.

Experiments
Parameters Settings

Initialization	Optimizer	Batch Size	Latent feat.	Learning rate	Regularizer
Gaussian dist.	SGD	256	32	0.001	0.00001
		512	64	0.005	0.0001
			128	0.01	0.001
				0.05	0.01
				0.1	0.1
				0.1	0

Table: Settings

Experiments
Model Answer RQ1

RQ1

Does ACF outperform state-of-art recommendation methods?

Experiments
Model Answer RQ1

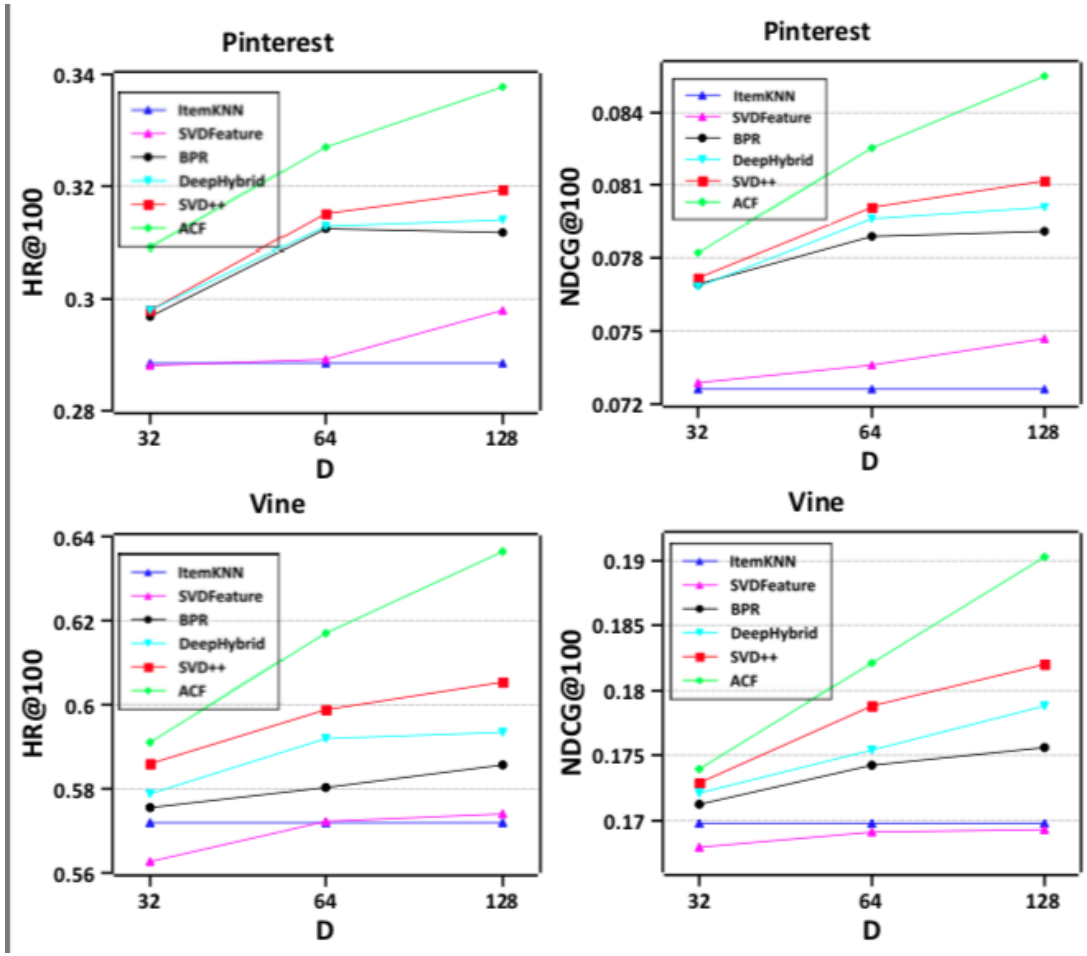


Figure: The performance of HR@100 and NDCG@100 with respect to the number of latent factors.

Experiments
Model Answer RQ1

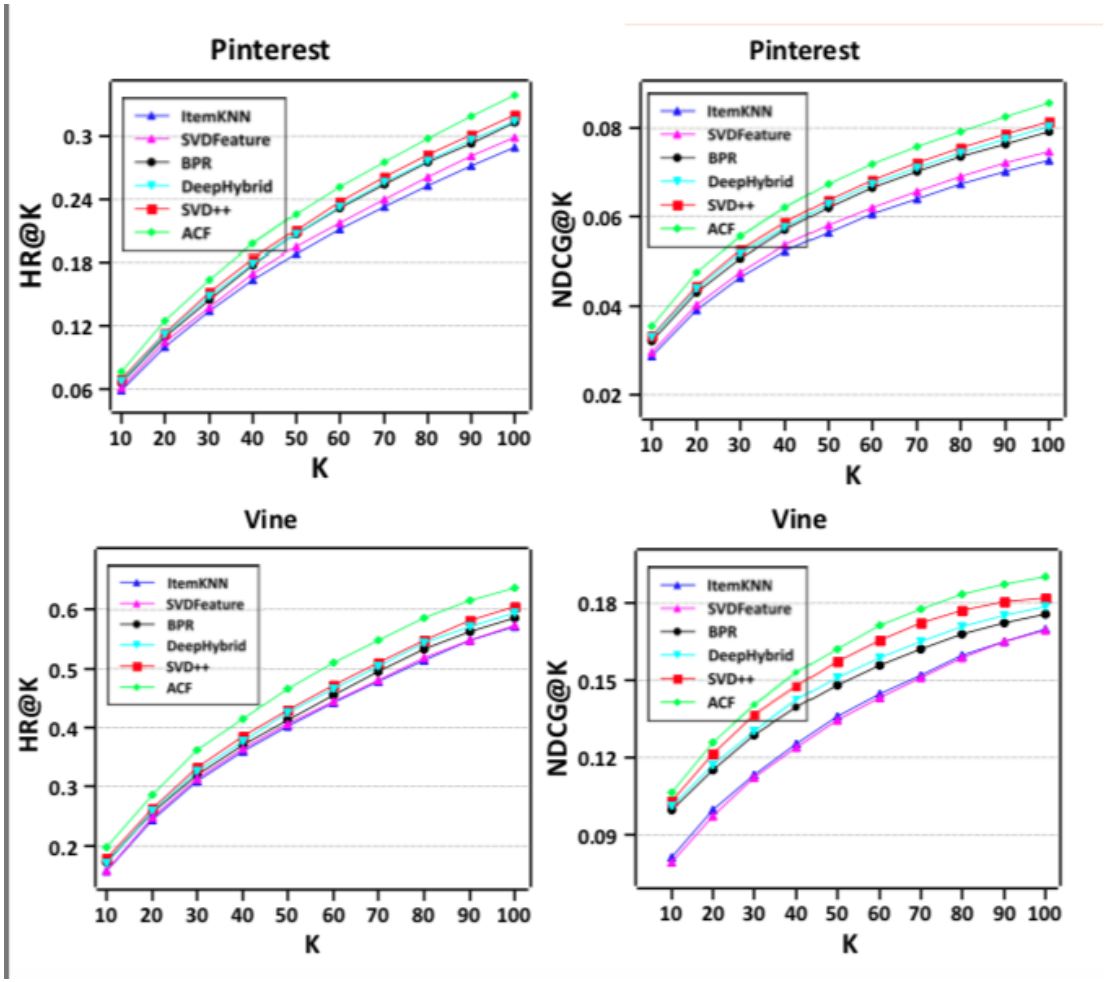


Figure: The performance of Top-K recommended lists where the ranking position K ranges from 10 to 100.

Experiments
Model Answer RQ1

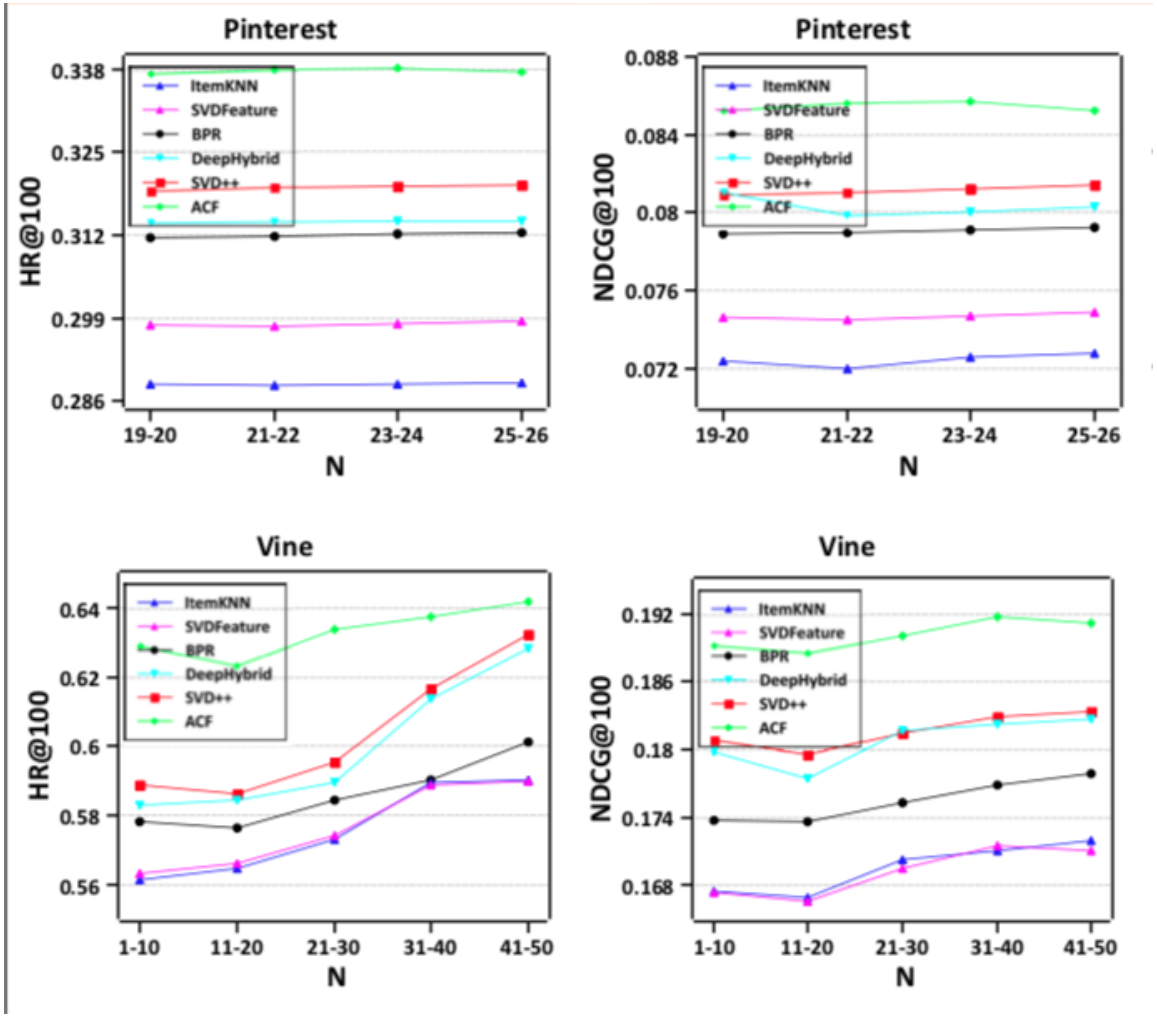


Figure: The performance with respect to the number of items a user has.

Experiments
Model Answer RQ1

ACF with attention mechanism outperforms others baseline methods.

ACF performs much better when the number of items per user is relatively small. Attention mechanism could improve recommendation quality with insufficient training data for each user.

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.

Experiments
Model Answer RQ2

RQ2

How do the proposed item-level and component-level attentions perform?

Experiments
Model Answer RQ2

Effect of Attention Mechanisms in Item- and Comp-Level

Level	Pinterest	Vine
Item — Comp	HR — NDCG	HR — NDCG
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%

Table: Model ACF

Experiments
Model Answer RQ2

Effect of User, Item and Content Information

Attention Type	Pinterest	Vine
Item — Comp	HR — NDCG	HR — NDCG
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%

Table: Model ACF

Experiments
Model Answer RQ2

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.

The information of both user and item contributes to our models compared to a constant weight model.

The information of users is more effective than the items to enhance recommendation.

Conclusions

In this paper is introduced a component- and item-level attention model to adress implicit feedback in multimedia recommendation.

In this paper is performed experiments on two real-world multimedia social networks: Pinterest and Vine, in order to demonstrate the effectiveness of ACF.

ACF is a generic attention-based CF fraework, so they plan to extend ACF to other CF models such Factorization Machine, Neural CF and Discrete CF.

References

Bahdanau, D., Cho, K., Bengio, Y. (2014).

Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

Chen, J., Song, X., Nie, L., Wang, X., Zhang, H., Chua, T. S. (2016)

Micro tells macro: predicting the popularity of micro-videos via a transductive model. In Proceedings of the 2016 ACM on Multimedia Conference (pp. 898-907). ACM.

Chen, L., Zhang, H., Xiao, J., Nie, L., Shao, J., Liu, W., Chua, T. S. (2017, July).

Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 6298-6306). IEEE.

Chen, X., Zhang, Y., Ai, Q., Xu, H., Yan, J., Qin, Z. (2017, August).

Personalized key frame recommendation. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 315-324). ACM.

References

Geng, X., Zhang, H., Bian, J., Chua, T. S. (2015).
Learning image and user features for recommendation in social networks. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4274-4282).

Chen, J., Zhang, H., He, X., Nie, L., Liu, W., Chua, T. S. (2017, August).
Attentive collaborative Itering: Multimedia recommendation with item-and component-level attention. In Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 335-344). ACM.

He, X., Zhang, H., Kan, M. Y., Chua, T. S. (2016, July).
Fast matrix factorization for online recommendation with implicit feedback. In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 549-558). ACM.

Hu, Y., Koren, Y., Volinsky, C. (2008, December).
Collaborative Itering for implicit feedback datasets. In Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on (pp. 263-272). Ieee.

References

Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L. (2009, June).

BPR: Bayesian personalized ranking from implicit feedback. In Proceedings of the twenty- fth conference on uncertainty in arti cial intelligence (pp. 452-461). AUAI Press.

You, Q., Jin, H., Wang, Z., Fang, C., Luo, J. (2016).

Image captioning with semantic attention. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4651-4659).

Zan r, M., Marinoiu, E., Sminchisescu, C. (2016).

Spatio-Temporal Attention Models for Grounded Video Captioning ACCV.

References

Chen, J., Zhang, H., He, X., Nie, L., Liu, W., Chua, T. (2017).

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

SIGIR.

Sun, L.

ResNet on Tiny ImageNet.

He, K., Zhang, X., Ren S., Sun, J. (2015).

Deep Residual Learning for Image Recognition.

arXiv:1512.03385v1

¡Gracias!