Social Media Based Recommendation: Accuracy and Interpretability

Le Wu
Lab for Media Computing
Hefei University of Technology

Outline

- Research Background
- Influence Diffusion for Social Recommendation
- Explainable Multimedia based Recommendation
- A Unified Model for Social Multimedia Recommendation
- Conclusions and Future Work

Outline

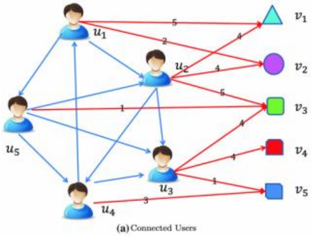
- Research Background
- Influence Diffusion for Social Recommendation
- Explainable Multimedia based Recommendation
- A Unified Model for Social Multimedia Recommendation
- Conclusions and Future Work

What is Recommender System

- **Recommender Systems (RS)**
  - A subclass of information filtering system that seeks to **predict the preference** that **a user** would give to **an item**.
  - Based on
    - Users' historical behavior

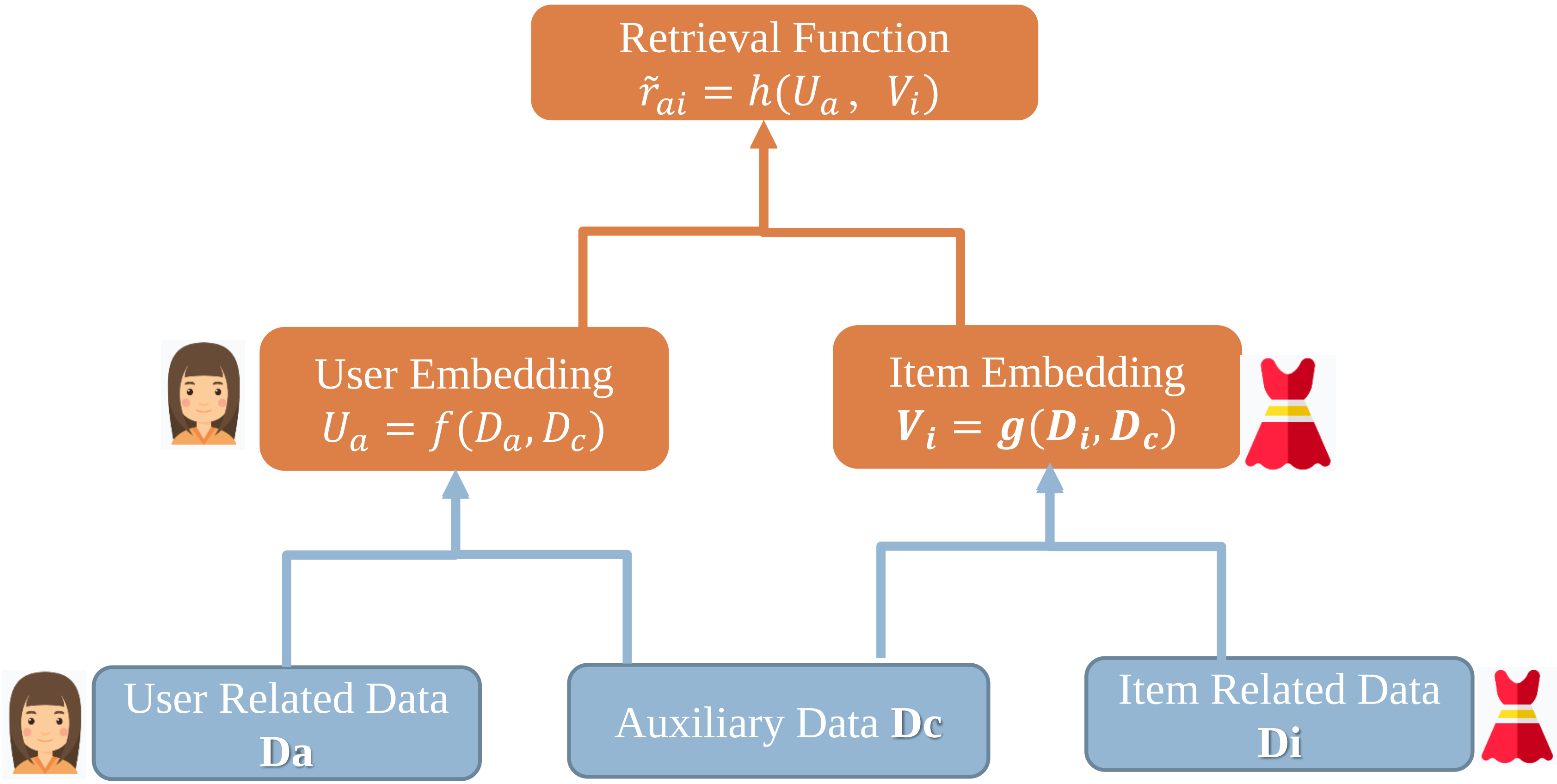
    - Item content, Item similarity

    - Relations to other users

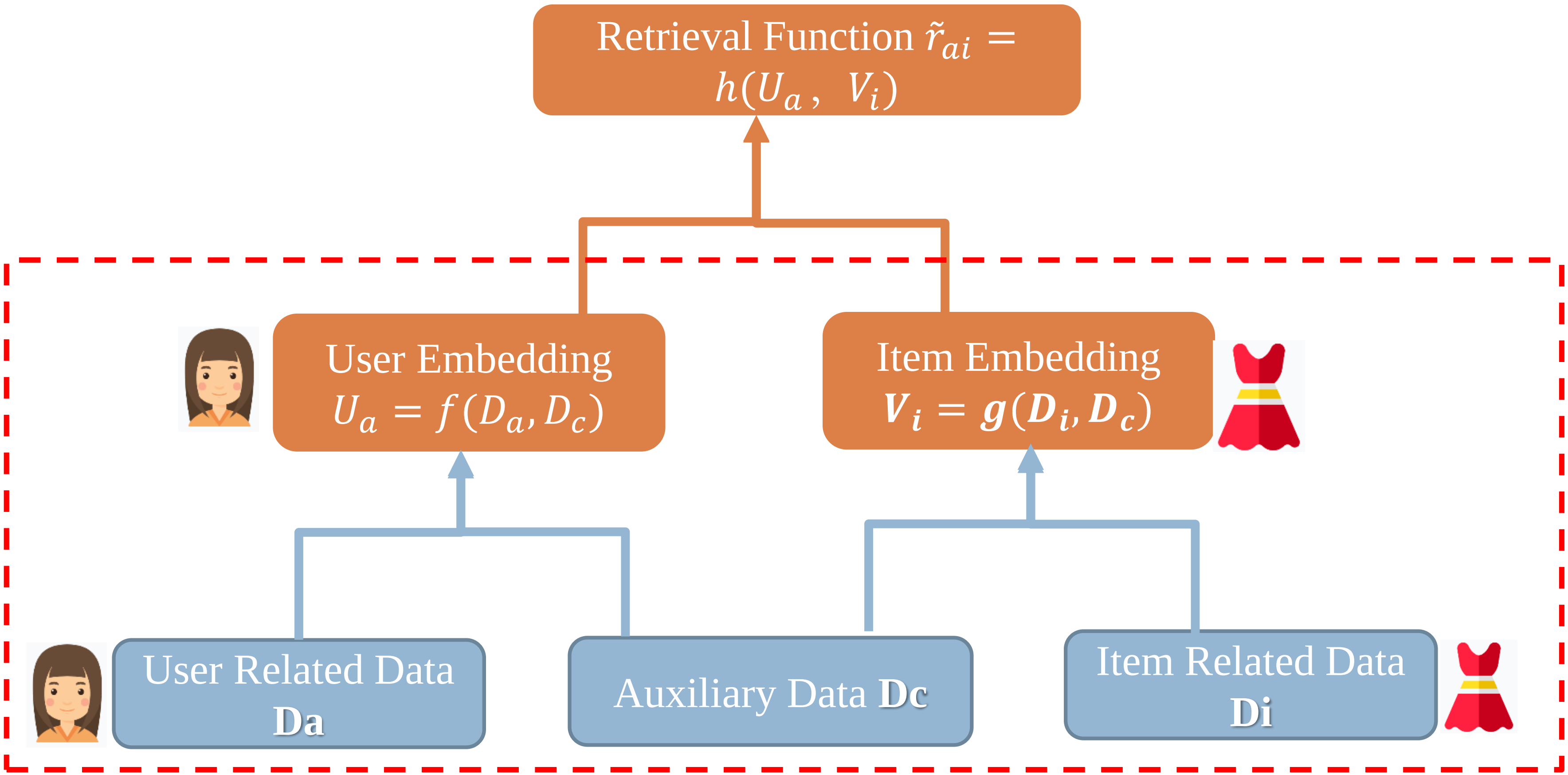


    - Context….

Modern Architecture of RS

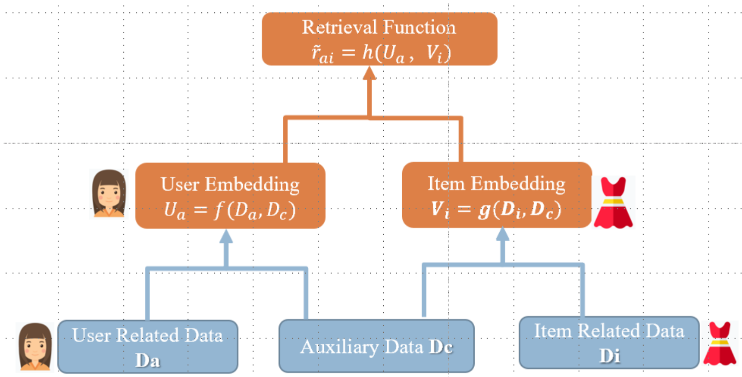


Modern Architecture of RS



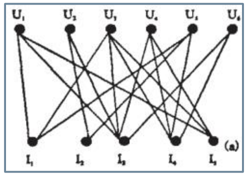
Categories of RS

- Content based RS



  - Recommendation based on content
  - Input: user (item) content Da (Di)
  - Model: Ua=f(Da, Dc), Vi=g(Di, Dc)

- Collaborative Filtering



  - Recommendation based on collaborative behaviors without content information
  - Input: user-item interaction behavior (a,i, r_{ai})
  - BPR: Ua = f(ID embedding of a), Vi=f(ID embedding of V)
  - SVD++: Ua = f(ID embedding of a, Rated items of a)

Desirable Properties of RS

- Data availability
  - User behavior data is available but suffers from cold-start
  - Content data: partially available
- Recommendation accuracy
- Recommendation interpretability
  - Explanations serve as a bridge between recommender systems and users
    - Increase trust, help users make better decisions, and persuade users to buy

Properties of Current RSs

	Data Availability	Cold Start	Accuracy	Interpretability
Content based Models	★☆☆	★★★	★☆☆	★★☆
Collaborative Filtering	★★★	★☆☆	★★☆	★☆☆

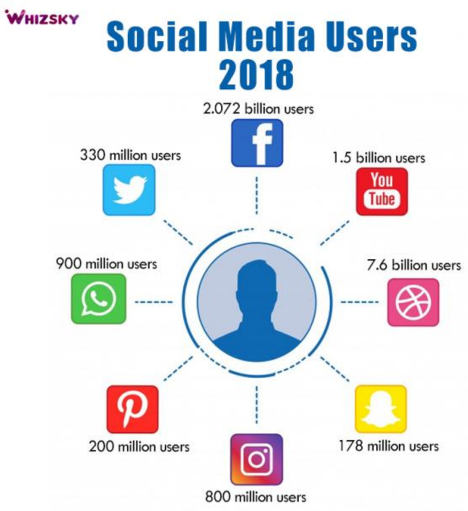
Desirable Properties of Current RSs

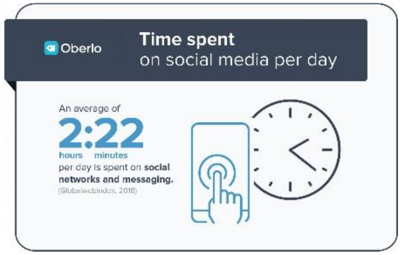
	Data Availability	Cold Start	Accuracy	Interpretability
Content based Models	★☆☆	★★★	★☆☆	★★☆
Collaborative Filtering	★★★	★☆☆	★★☆	★☆☆
Desirable Recommendation Models	★★★	★★★	★★★	★★★

New Opportunities for RSs

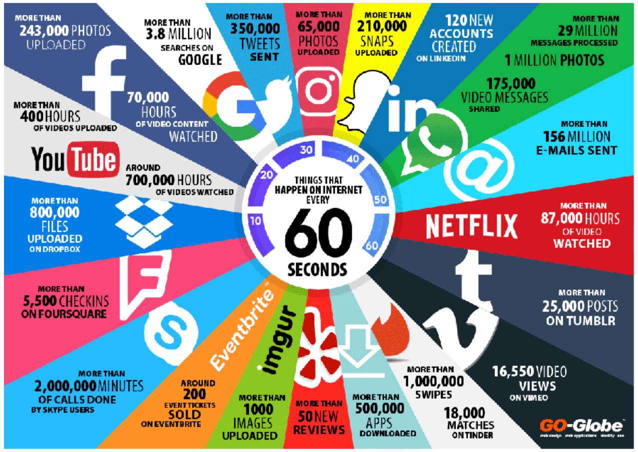
Social Media Era

Users in the Social Network

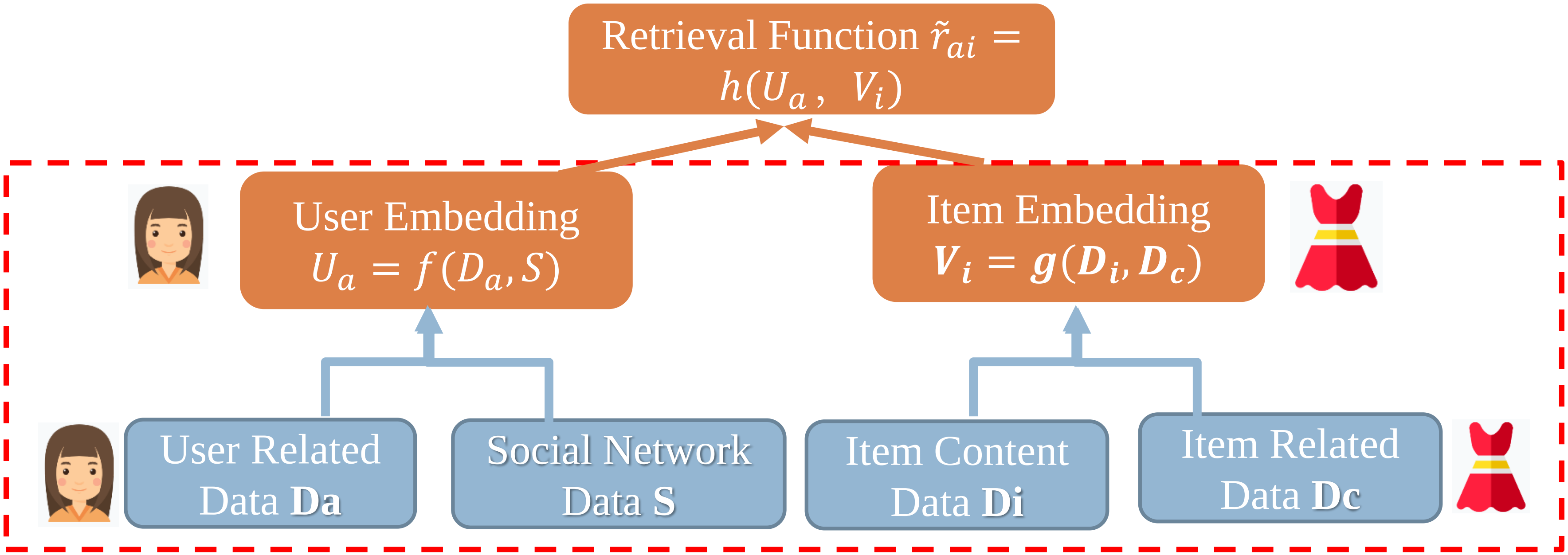




Multimedia Items



Soundness of Social Media based Recommendation



- User embedding learning $U_a = f(D_a, S)$
  - Social influence propagates in the social network, leading to similar social behaviors of connected users.
- Item embedding learning $V_i = g(D_i, D_c)$
  - Advances in deep learning models provide rich opportunities to learn item semantic representation.
  - Item semantics could better describe item content for recommendation.

New Opportunities for RSs

- Social Media based Recommendation

	Data Availability	Cold Start	Accuracy	Interpretability
Content based Models	★☆☆	★★★	★☆☆	★★☆
Collaborative Filtering	★★★	★☆☆	★★☆	★☆☆
Social Media based Recommendation	★★★	★★★	★★★	★★★

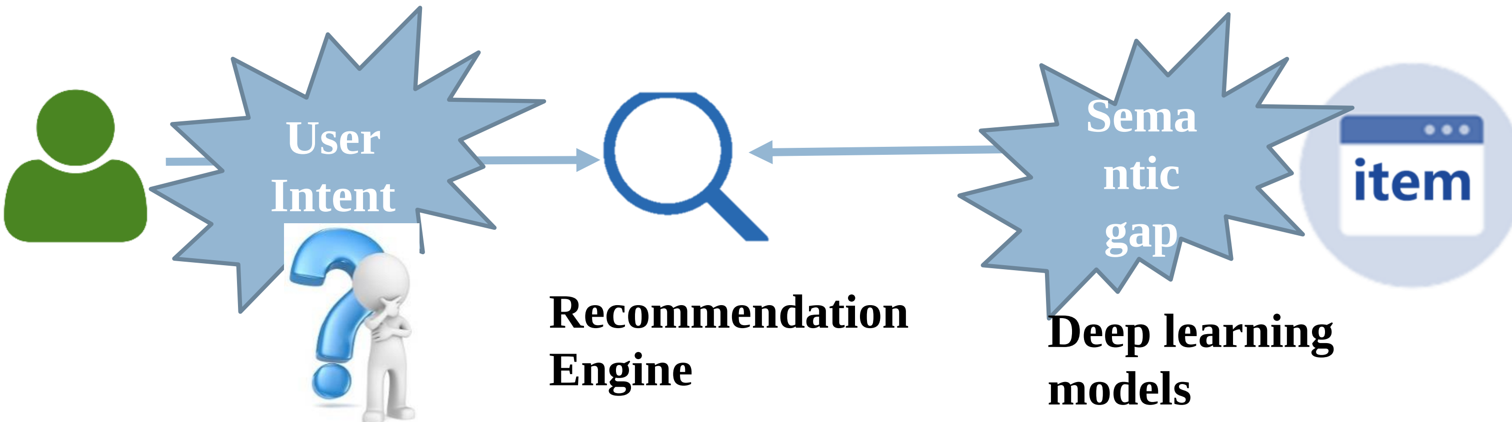
Social media data

Hybrid recommendation

Social network or multimedia as explainable components

Social Multimedia Recommendation: Challenges

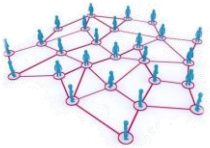
- Unobservable social diffusion effect for social recommendation.
  - Could not observe the influence in the social network
- From multimedia sematic gap to user intent gap



- Complex decision process with various decision process from complex heterogeneous data.

Research Roadmap

- Social diffusion recommendation model



  - Modeling the unobserved social diffusion process in social networks [SIGIR 2019]
  - Focus on improving model **accuracy**
- Explainable multimedia based recommendation

  - Semantic attribute guided fashion recommendation[IJCAI 2019]
  - Personalized multimedia item and key frame recommendation[IJCAI 2019]
  - Improved accuracy and multimedia **explanation**
- Social contextual recommendation
  - A unified recommendation model in social multimedia platforms[TKDE 2019]
  - Improved **accuracy** with **explanation** of each contextual factor

Outline

- Research Background
- Influence Diffusion for Social Recommendation
- Explainable Multimedia based Recommendation
- A Unified Model for Social Multimedia Recommendation
- Conclusions and Future Work

Le Wu(吴乐), Peijie Sun, Yanjie Fu, Richang Hong, Xiting Wang, Meng Wang . A Neural Influence Diffusion Model for Social Recommendation. ***SIGIR***, 2019.

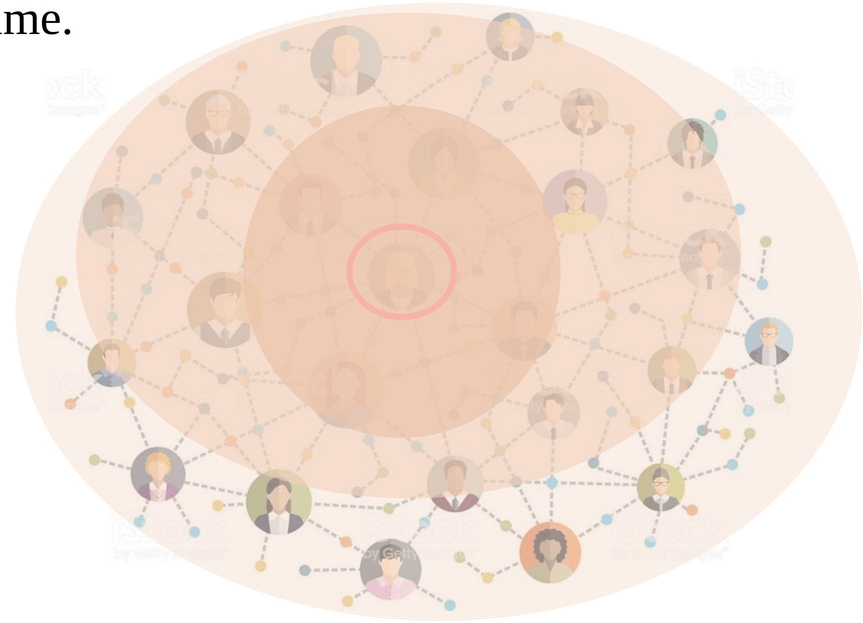
Social Recommender Systems

- Online social networking services make it possible to study social recommender system.
  - Increase the user participation with social connections
  - Alleviate the data sparsity issue in CF with social network

*Definition 2.1 (**SOCIAL RECOMMENDATION**).* Given a rating matrix $\mathbf{R}$, a social network $\mathbb{S}$, and associated real-valued feature matrix $\mathbf{X}$ and $\mathbf{Y}$ of users and items, our goal is to predict users' unknown preferences to items as: $\hat{R} = f(\mathbf{R}, \mathbb{S}, \mathbf{X}, \mathbf{Y})$, where $\hat{R} \in \mathbb{R}^{M \times N}$ denotes the predicted preferences of users to items.

Social Recommender Systems

- **<u>Social influence</u>** in social networks
  - [Wikipedia] Social influence occurs when one's emotions, opinions, or behaviors are affected by others.
  - In social recommender systems, the social influence exists and diffuses in the social networks
    - If user a follows user b, then *a*'s preference is influenced by b .
    - The social influences diffuses in the social network, thus iteratively influence users' preferences over time.



Existing Models for Social Recommendation

- Social regularization based models
  - Direct social influence of users' latent preferences [TKDE 2014， Recsys2010]

$$\hat{\mathbf{u}}_a = \sum_{b \in S_a} t(a,b)\mathbf{u}_b, \mathbf{U} \sim \mathcal{N}(\hat{\mathbf{U}}, \sigma^2)$$

Influences from direct social neighbors

  - Social influence would lead to correlated preferences among connected users [TSMC 2018, WSDM2011]

$$\sum_{a=1}^{M}\sum_{b=1}^{M} s_{ab}||\mathbf{u}_a - \mathbf{u}_b||_F^2 = \mathbf{U}(\mathbf{D}-\mathbf{S})\mathbf{U}^T$$

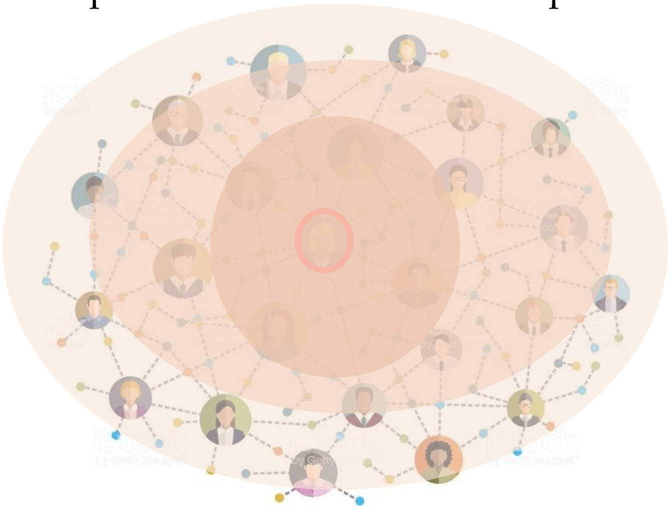
- Social behavior enhanced models[AAAI 2015]

$$\hat{r}_{ai} = \mathbf{v}_i^T(\mathbf{u}_a + \sum_{b \in S_a} \frac{\mathbf{u}_b}{|S_a|})$$

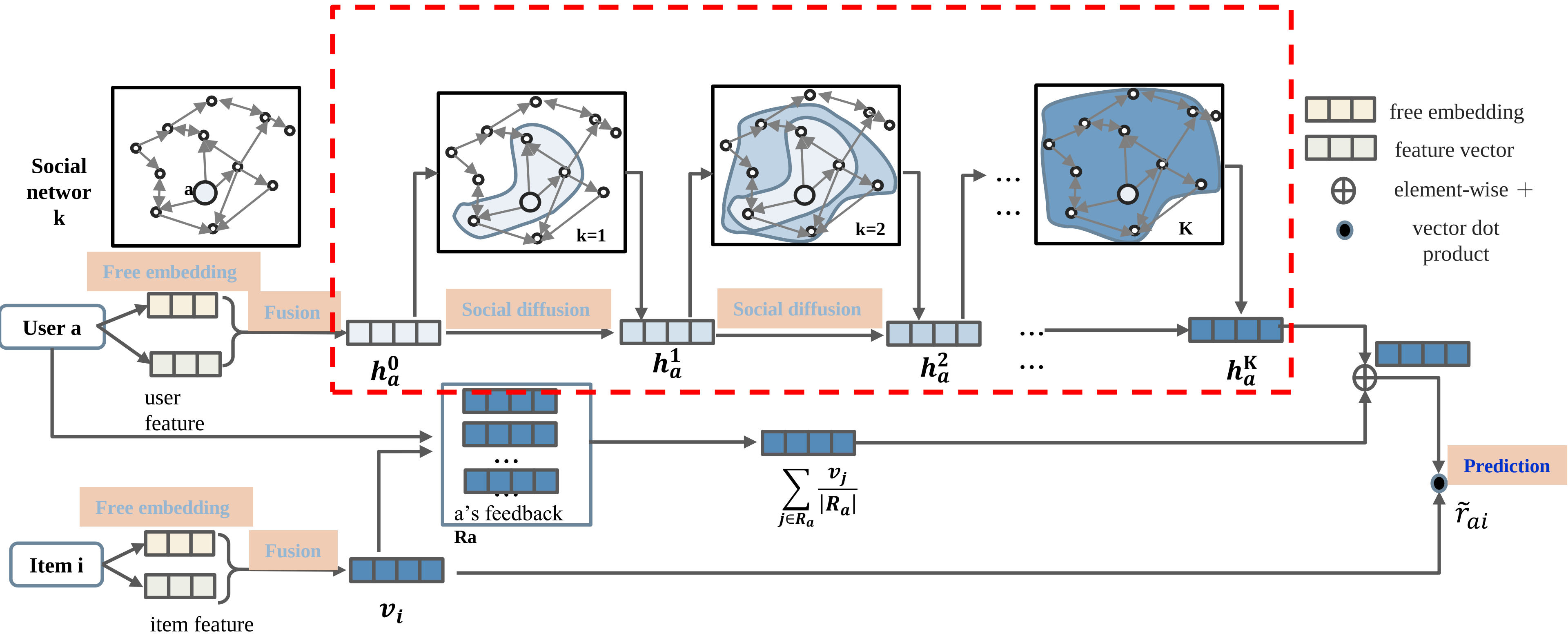
Challenges

- Nearly all previous models leverage the ***first-order direct social neighbors*** to alleviate the data sparsity and boost recommendation performance.

- In social networks, the ***recursive social influence diffuses in the global social network*** from time to time.
  - Each user's latent embedding changes over time due to the recursive social diffusion.
  - Precise stimulating the recursive diffusion process in the global social network would better model each user's embedding, thus improve the recommendation performance.

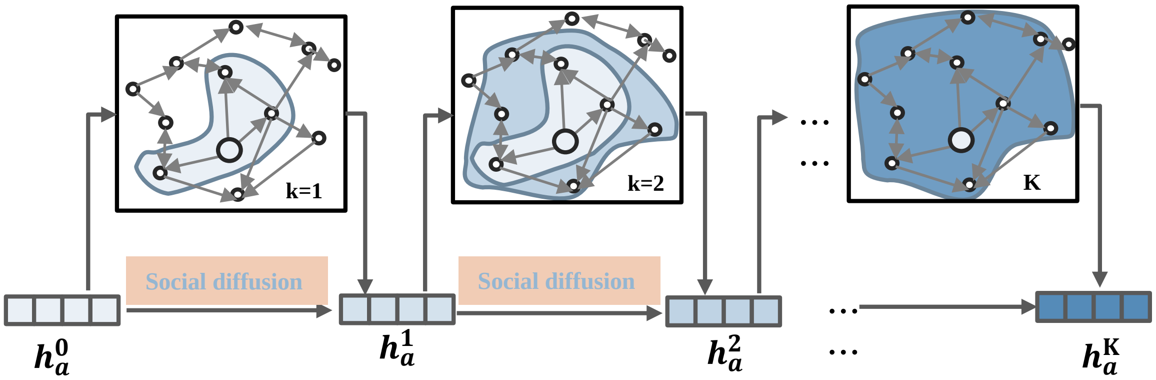


The Proposed DiffNet Architecture



- The fusion layer fuses the free latent embedding and the (item) user features
- The Influence diffusion layer models the dynamics of users latent preference diffusion in the social network

The Proposed DiffNet Architecture



- Influence diffusion layer with **recursive social diffusion** process.
  - For each user a, her fusion embedding h_a^0 is sent to the diffusion layer as k=0 .
  - **Recursively** update the diffusion in the social network from k to k+1 as:

Influences from social connections: $$\mathbf{h}_{Sa}^{k+1} = Pool(\mathbf{h}_b^k | b \in S_a)$$

The updated embedding: $$\mathbf{h}_a^{k+1} = s^{(k+1)}(\mathbf{W}^k \times [\mathbf{h}_{S_a}^{k+1}, \mathbf{h}_a^k])$$

- Prediction layer $$\mathbf{u}_a = \mathbf{h}_a^K + \sum_{i \in R_a} \frac{\mathbf{v}_i}{|R_a|}, \qquad \hat{r}_{ai} = \mathbf{v}_i^T \mathbf{u}_a$$

Model Complexity

- Space complexity: $\Theta = [\Theta_1, \Theta_2]$
  - $\Theta_1 = [\mathbf{P}, \mathbf{Q}]$: User and item latent free embeddings as most embedding models
  - $\Theta_2 = [\mathbf{F}, [\mathbf{W}^k]_{k=0}^{K-1}]$ transformation matrices shared among users(items)
  - Therefore, the space complexity is the same as classical embedding based models

- Time complexity
  - The additional time cost lies in **dynamic social diffusion: O(MKL)**
    - M: users, K: diffusion depth (small), L: average neighbors per user (L<< M)
    - The additional time cost is linear with the number of users

Model Generalization

- The recursive influence diffusion layer is inspired by Graph Convolutional Networks (GCN).
  - A concrete practice of how to apply GCNs to model social diffusion in the social network.
  - Previous works on GCNs for recommendation focus on modeling the structure of user-item or item-item correlation graph.

- DiffNet is generally applicable when the user and (or) item features are not available.
  - X=0: the user fusion layer disappears $\mathbf{h}_a^0 = g(\mathbf{W}^0 \times [\mathbf{x}_a, \mathbf{p}_a])$
  - Y=0: the item fusion layer disappears $\mathbf{v}_i = \sigma(\mathbf{F} \times [\mathbf{q}_i, \mathbf{y}_i])$

- When we omit the influence diffusion layers, DiffNet degenerates to SVD++.

$$\mathbf{u}_a = \mathbf{h}_a^K + \sum_{i \in R_a} \frac{\mathbf{v}_i}{|R_a|}, \qquad \mathbf{h}_a^K = \mathbf{h}_a^0$$

Experiments

- Datasets
  - Yelp: an online location based social network with item reviews
    - The item feature is learned by averaging the word2vec of each word in this item.
    - The user feature is averaged from her rated item features
  - Flickr: a directed online image social sharing platform
    - The item feature is represented as the last layer representation in VGG16.
    - The user feature is averaged from her rated images.
- Baselines
  - Classical CF models:

    BPR[UAI 2009], FM [TIST 2012]
  - Social recommendation models:

    TrustSVD[AAAI 2017], ContextMF[TKDE 2014]
  - GCN-based models

    GC-MC[KDD Workshop 2018], PinSage[KDD 2018]
- Evaluation metrics: HR, NDCG

Table 1: The statistics of the two datasets.

Dataset	*Yelp*	*Flickr*
Users	17237	8358
Items	38342	82120
Total Links	143765	187273
Ratings	204448	314809
Link Density	0.048%	0.268%
Rating Density	0.031%	0.046%

Overall Comparison

Table 2: HR@10 and NDCG@10 comparisons for different dimension size D.

Models	*Yelp*						*Flickr*					
	HR			NDCG			HR			NDCG		
	D=16	D=32	D=64	D=16	D=32	D=64	D=16	D=32	D=64	D=16	D=32	D=64
BPR	0.2443	0.2632	0.2617	0.1471	0.1575	0.155	0.0851	0.0832	0.0791	0.0679	0.0661	0.0625
SVD++	0.2581	0.2727	0.2831	0.1545	0.1632	0.1711	0.0821	0.0934	0.1054	0.0694	0.0722	0.0825
FM	0.2768	0.2835	0.2825	0.1698	0.1720	0.1717	0.1115	0.1212	0.1233	0.0872	0.0968	0.0954
TrustSVD	0.2853	0.2880	0.2915	0.1704	0.1723	0.1738	0.1372	0.1367	0.1427	0.1062	0.1047	0.1085
ContextMF	0.2985	0.3011	0.3043	0.1758	0.1808	0.1818	0.1405	0.1382	0.1433	0.1085	0.1079	0.1102
GC-MC	0.2876	0.2902	0.2937	0.1657	0.1686	0.174	0.1123	0.1155	0.1182	0.0883	0.0945	0.0956
PinSage	0.2952	0.2958	0.3065	0.1758	0.1779	0.1868	0.1209	0.1227	0.1242	0.0952	0.0978	0.0991
DiffNet	**0.3366**	**0.3437**	**0.3477**	**0.2052**	**0.2095**	**0.2121**	**0.1575**	**0.1621**	**0.1641**	**0.1210**	**0.1231**	**0.1273**

Table 3: HR@N and NDCG@N comparisons for different top-N values.

Models	*Yelp*						*Flickr*					
	HR			NDCG			HR			NDCG		
	N=5	N=10	N=15	N=5	N=10	N=15	N=5	N=10	N=15	N=5	N=10	N=15
BPR	0.1713	0.2632	0.3289	0.1243	0.1575	0.1773	0.0657	0.0851	0.1041	0.0607	0.0679	0.0737
SVD++	0.1868	0.2831	0.3492	0.1389	0.1711	0.1924	0.0827	0.1054	0.1257	0.0753	0.0825	0.0895
FM	0.1881	0.2835	0.3463	0.1359	0.1720	0.1895	0.0918	0.1233	0.1458	0.0845	0.0968	0.1046
TrustSVD	0.1906	0.2915	0.3693	0.1385	0.1738	0.1983	0.1072	0.1427	0.1741	0.0970	0.1085	0.1200
ContextMF	0.2045	0.3043	0.3832	0.1484	0.1818	0.2081	0.1095	0.1433	0.1768	0.0920	0.1102	0.1131
GC-MC	0.1932	0.2937	0.3652	0.1420	0.1740	0.1922	0.0897	0.1182	0.1392	0.0795	0.0956	0.1002
PinSage	0.2099	0.3065	0.3873	0.1536	0.1868	0.2130	0.0925	0.1242	0.1489	0.0842	0.0991	0.1036
DiffNet	**0.2276**	**0.3477**	**0.4232**	**0.1679**	**0.2121**	**0.2331**	**0.1210**	**0.1641**	**0.1952**	**0.1142**	**0.1273**	**0.1384**

Our proposed DiffNet shows the best performance under different ranking metrics

Outline

- Research Background
- Influence Diffusion for Social Recommendation
- Explainable Multimedia based Recommendation
  - A Semantic Attribute Region Guided Approach for Fashion Recommendation
  - Personalized Multimedia Item and Key Frame Recommendation
- A Unified Model for Social Multimedia Recommendation
- Conclusions and Future Work

Min Hou, **Le Wu(吴乐)**, Enhong Chen, Zhi Li, Vincent W. Zheng, Qi Liu. Explainable Fashion Recommendation: A Semantic Attribute Region Guided Approach . ***IJCAI***, 2019.

Background

- The ubiquity of online fashion shopping has led to information explosion in the fashion industry.
- When purchasing clothing products, it is intuitive that **we often have preferences for detailed semantic attribute** (such as neckline, heel height, skirt length) in addition to global impressions.

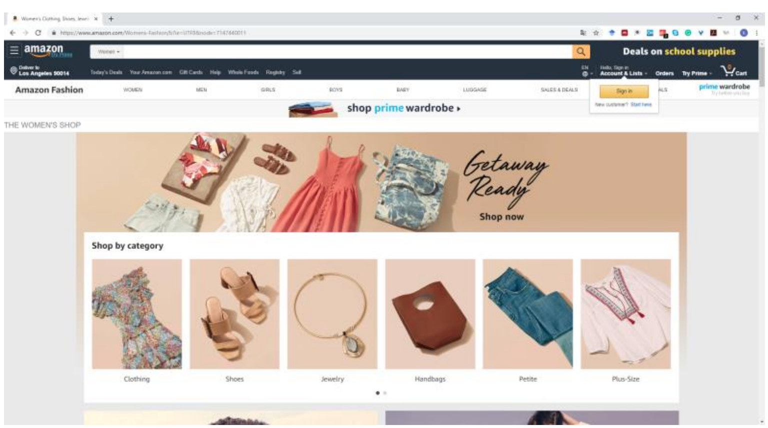


Figure 1: An example of user preferences for semantic attributes.

Motivation

- User decisions are **affected by the fine-grained semantic attributes**. However, traditional approaches hampered at understanding the fashion items from a holistic perspective.
  - Content category features from pretrained CNNs
  - Aesthetics features from an aesthetic network
  - Style features that are complementary to the content features

- Users prefer the visual explanations of the semantic attributes for recommendations, whereas most fashion recommendation models are latent black box models.

Challenges

- It is difficult to obtain clothing semantic attribute features without the manual attribute annotation.

- How to visualize the semantic attribute regions for explainable recommendation?

The Proposed SAERS Framework

- *S*emantic *A*ttribute *E*xplainable *R*ecommender System (*SAERS*)
  - Semantic attributed guided
  - Explainable

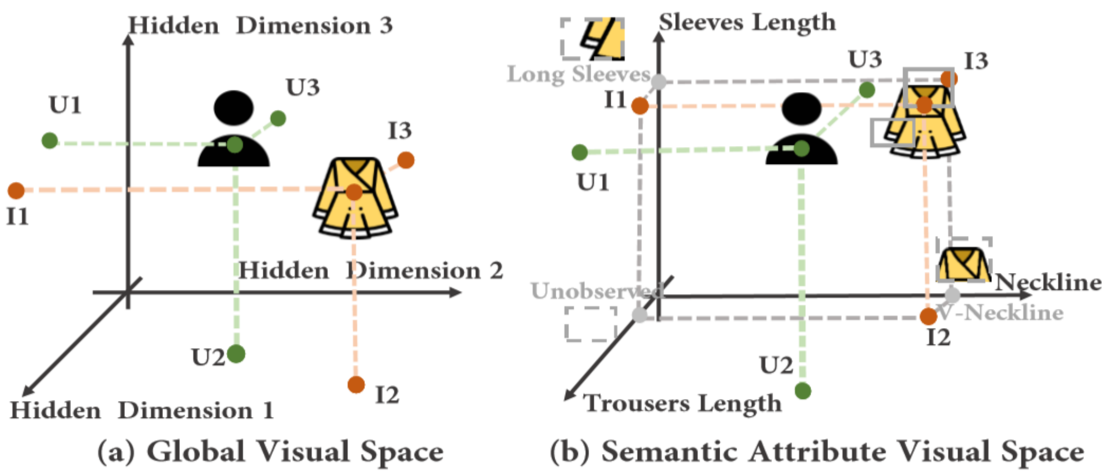
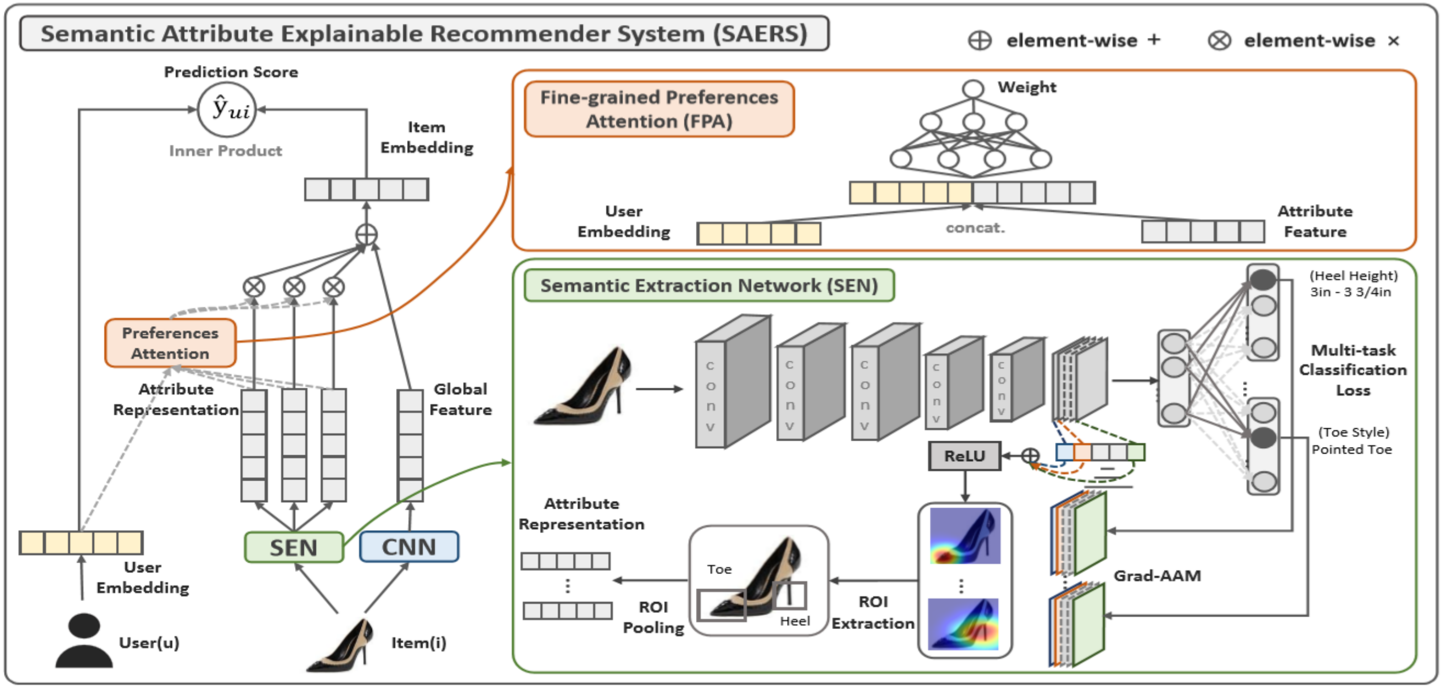


Figure 2: Difference between the conventional (a) Global Visual Space and our (b) Semantic Attribute Visual Space.

The Proposed SAERS Framework

Semantic Attribute Explainable Recommender System (SAERS)

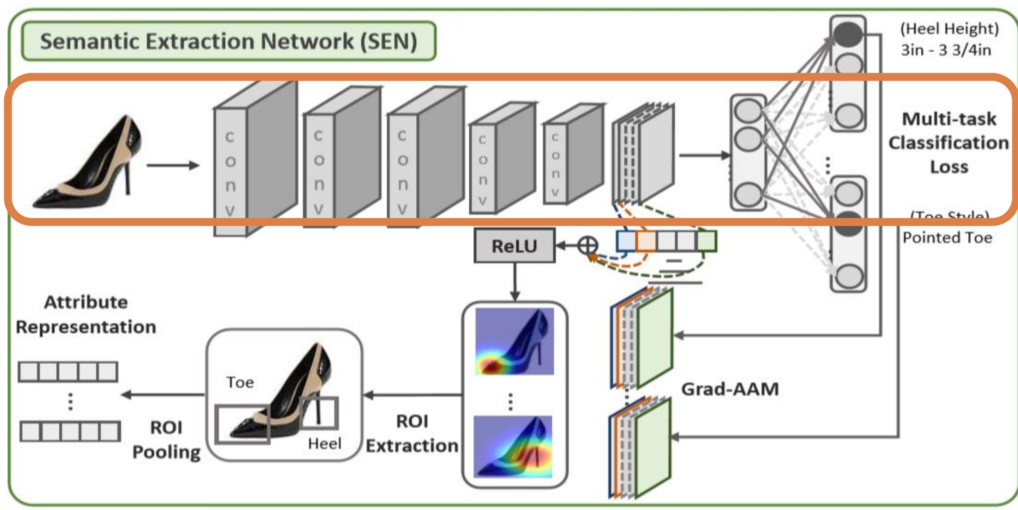
- 1.Projecting Item into Semantic Attribute Space
- 2. Projecting User into Semantic Attribute Space



The Proposed SAERS Framework

- Projecting Item into Semantic Attribute Space

We borrow an image-level fine-grained labeled data and pre-train a Semantic Extraction Network (SEN), which is used to extract the region-specific attribute representations and simultaneously locate and classify attributes.



Classify attributes

$$L_C = -\sum_{I=1}^{N}\sum_{a=1}^{A}\log\left(p\left(\hat{y}_{Ia}|y_{Ia}\right)\right)$$

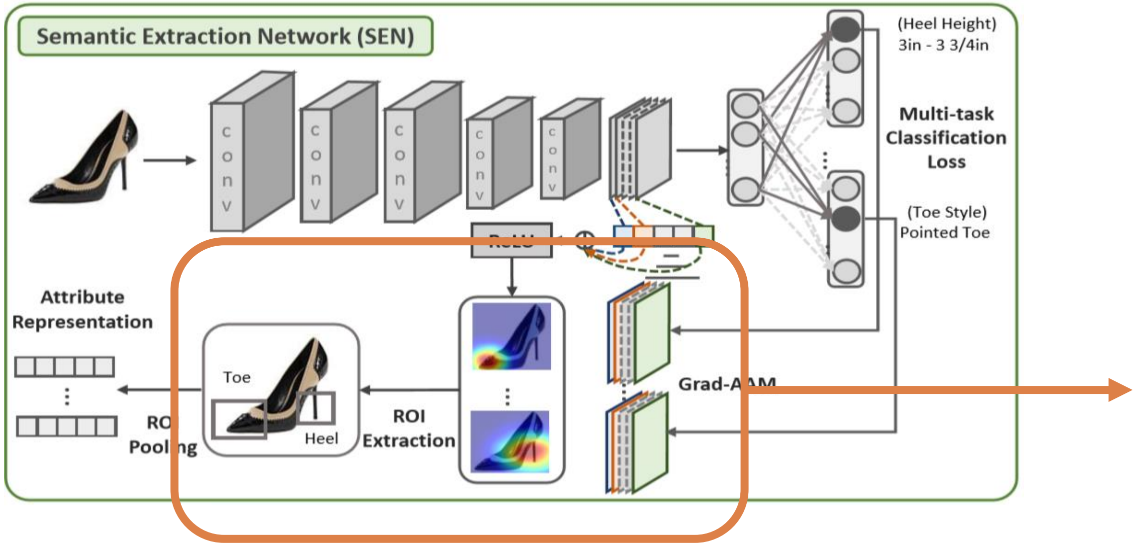
Category	Attribute: Class
Top	*high neck*: ruffle semi-high, turtle,... *collar*: rib collar, puritan collar,... *lapel*: notched, shawl, collarless,... *neckline*: V, square, round,... *sleeves length*: sleeveless, cap, short,... *body length*: high waist, long, regular,...
Bottom	*skirt length*: short, knee, midi, ankle,... *trousers length*: short, mid, 3/4, cropped,...
Shoes	*heel height*: flat, 1 in-7/4 in, under 1 in,... *boots height*: ankle, knee-high, mid-calf,... *closure*: lace-up, slip-on, zipper,... *toe style*: round, pointed, peep, open,...

Table 1: List of semantic attributes used in our method

The Proposed SAERS Framework

Projecting Item into Semantic Attribute Space

We borrow a image-level fine-grained labeled data and pre-train a Semantic Extraction Network (SEN), which is used to extract the region-specific attribute representations and simultaneously locate and classify attributes.



Locate attributes

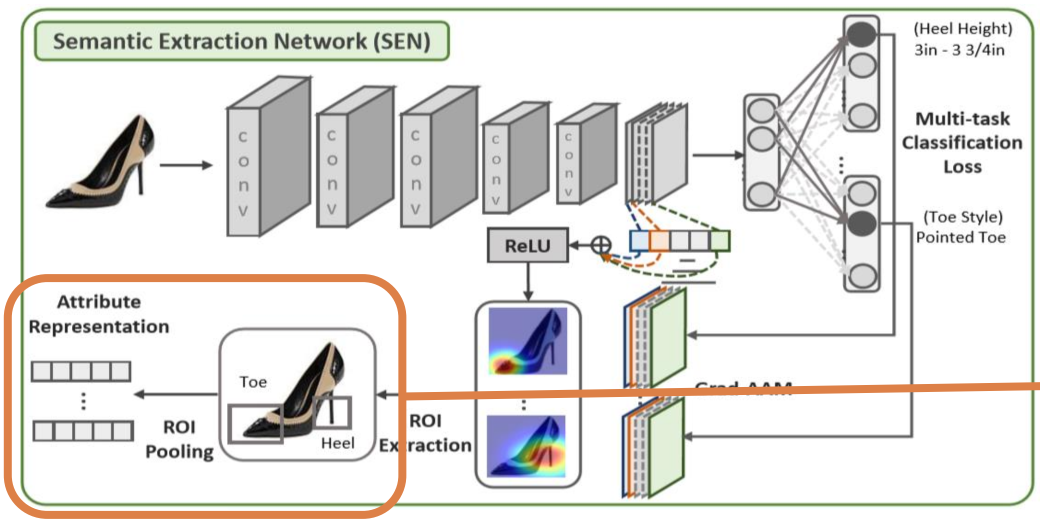
$$\alpha_t^{a_c} = \overbrace{\frac{1}{Z}\sum_m \sum_n}^{\text{global average pooling}} \underbrace{\frac{\partial y^{a_c}}{\partial F_{mn}^t}}_{\text{gradients via backpop}}.$$

$$M_{\text{Grad-AAM}}^{a_c} = ReLU \underbrace{\left(\sum_t \alpha_t^{a_c} F^t\right)}_{\text{linear combination}}.$$

The Proposed SAERS Framework

Projecting Item into Semantic Attribute Space

We borrow a image-level fine-grained labeled data and pre-train a Semantic Extraction Network (SEN), which is used to extract the region-specific attribute representations and simultaneously locate and classify attributes.



Extract attributes

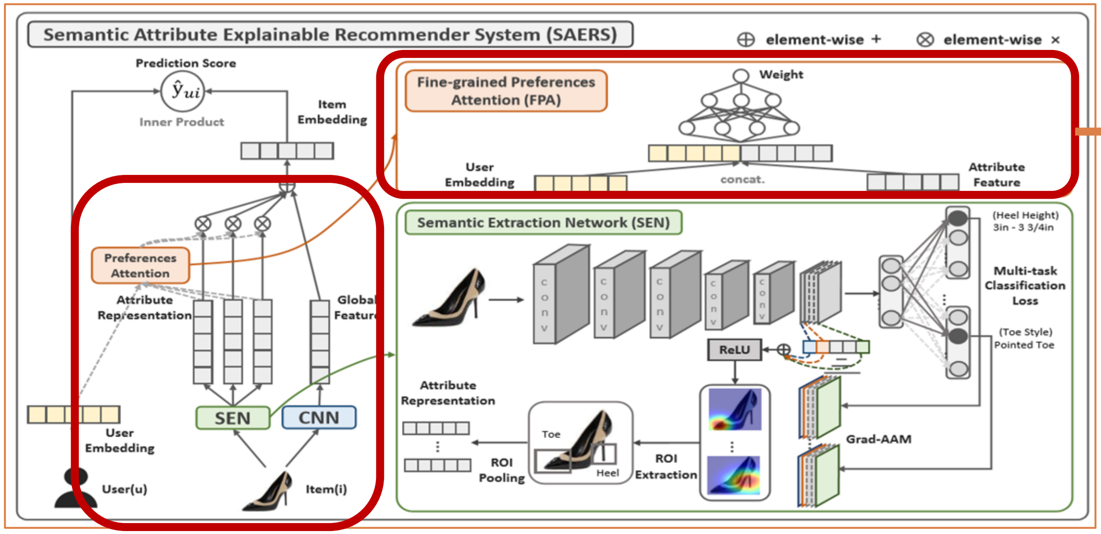
ROI pooling layer

↓

12 attribute representation f_a^k

The Proposed SAERS Framework

2: Projecting User into Semantic Attribute Space



$$f(i) = \sum_{k=1}^{A} \alpha_{ui}^{k} E^k f_a^k(i).$$

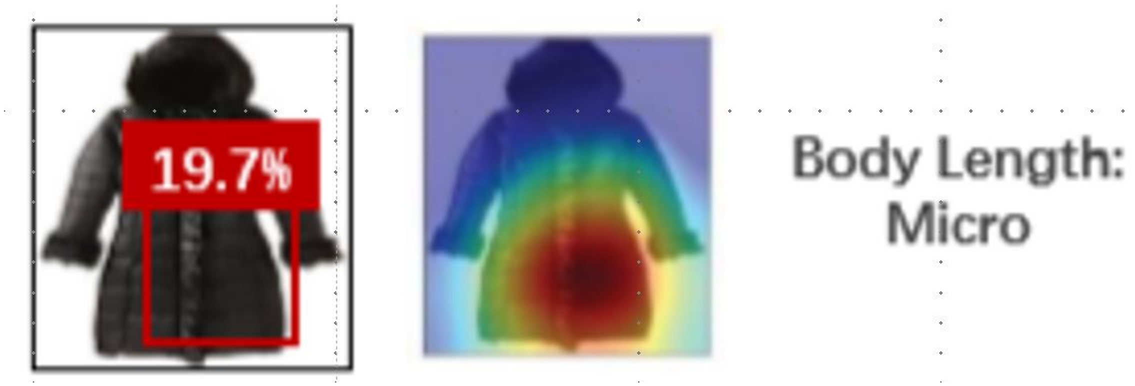
$$\alpha_{ui}^{k} = \mathrm{softmax}\left(\mathcal{D}\left(f(u), E^k f_a^k(i)\right)\right) = \frac{\exp\left(\mathcal{D}\left(f(u), E^k f_a^k(i)\right)\right)}{\sum_{k=1}^{A} \exp\left(\mathcal{D}\left(f(u), E^k f_a^k(i)\right)\right)}$$

$$f(i) = \sum_{k=1}^{A} \alpha_{ui}^{k} E^k f_a^k(i) + f_g(i).$$

Fine-grained Preferences Attention

Model Explanation

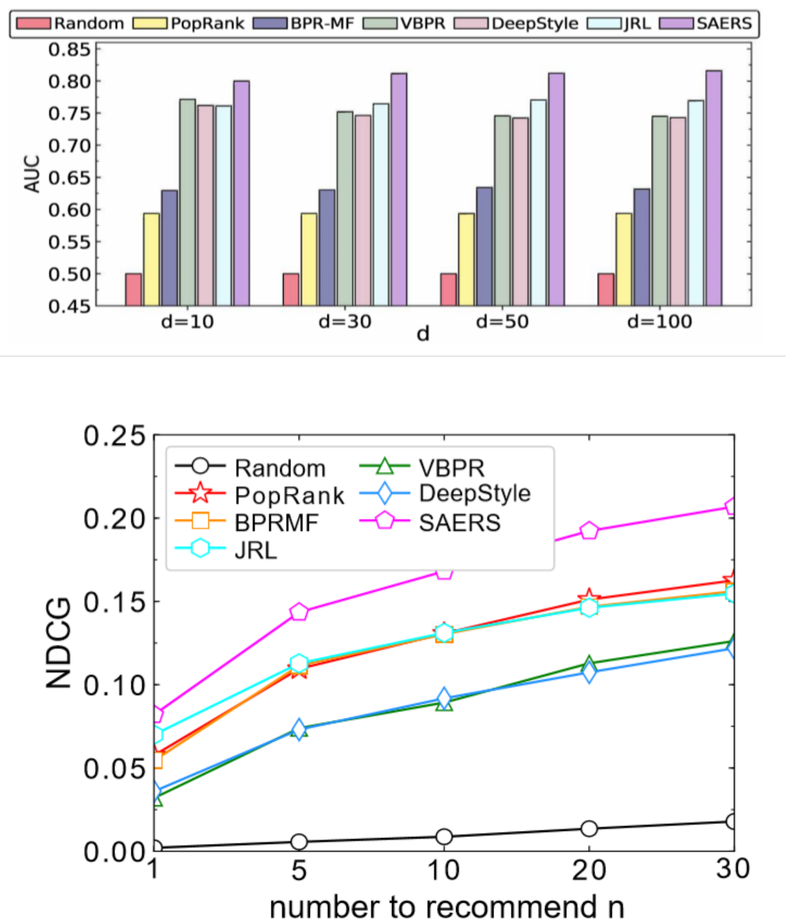
- Personalized explainable recommendations
  - Using a bounding box to highlight which part of the product image the user might like.
  - Providing which semantic attribute the highlighted part belongs to.
  - Providing the possibility that the user likes the semantic attribute



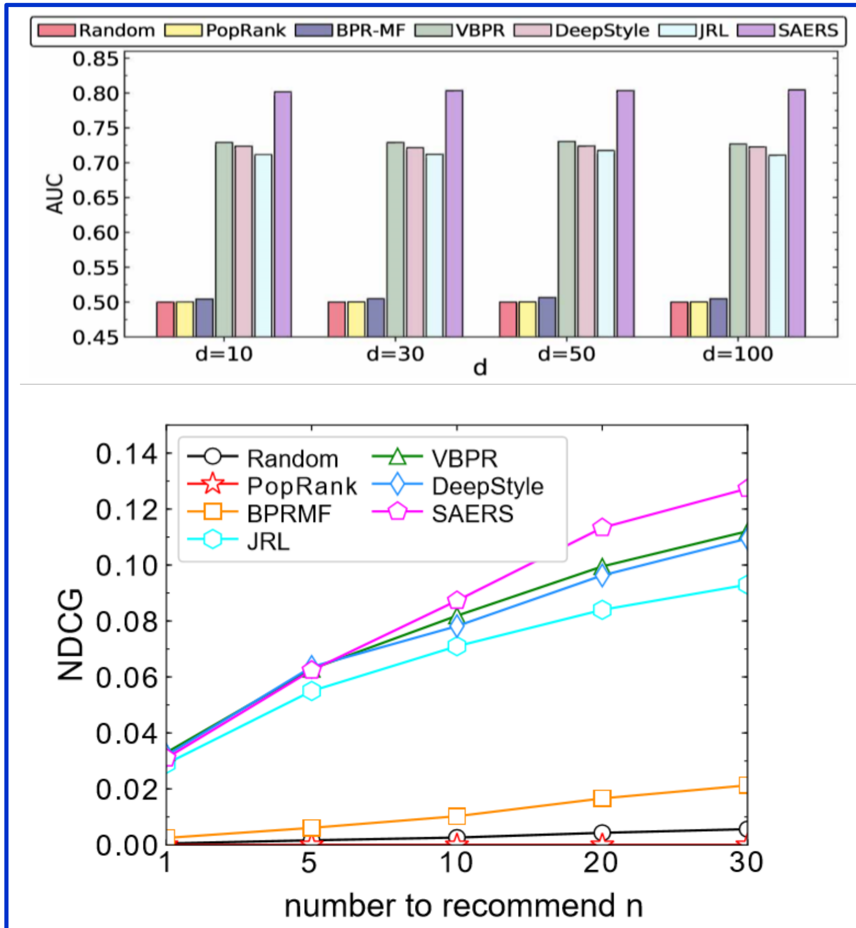
Experiments

- Pretrained **UT-Zap50K & Tianchi Apparel** dataset for semantic attribute learning
  - 50,025 shoes and 180,000 apparels
  - With detailed image-level attribute annotation
- Recommendation task: Amazon Fashion dataset
  - 45,184 users, 166,270 items, and 358,003 records
  - Including men/women's tops, bottoms and shoes
  - Sparsity: **99.9952%**

Experiments: Overall Performance



Overall Recommendation



Cold Recommendation

Experiments

- User attribute visualization

Outline

- Research Background
- Influence Diffusion for Social Recommendation
- Explainable Multimedia based Recommendation
  - A Semantic Attribute Region Guided Approach for Fashion Recommendation
  - Personalized Multimedia Item and Key Frame Recommendation
- A Unified Model for Social Multimedia Recommendation
- Conclusions and Future Work

Le Wu(吴乐), Lei Chen, Yonghui Yang, Richang Hong, Yong Ge, Xing Xie, Meng Wang . Personalized Multimedia Item and Key Frame Recommendation. ***IJCAI***, 2019.

Background

- Visual based content the most eye-catching for users.
- When recommending multimedia items to users, an emerging trend is to present each multimedia item with a display image, e.g., ***a key frame image***.
  - Attract users’ attention to quickly spot the visual content of the item.
  - Enhance recommendation conversation rate with a key frame image.

豆瓣 douban NETFLIX

A movie recommendation page

抖音 快手

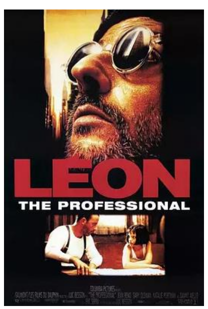
Typical short video recommendation page

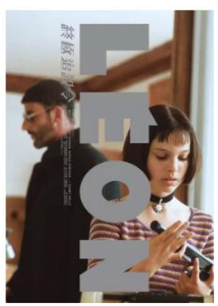
微博 twitter

Image-based advertising

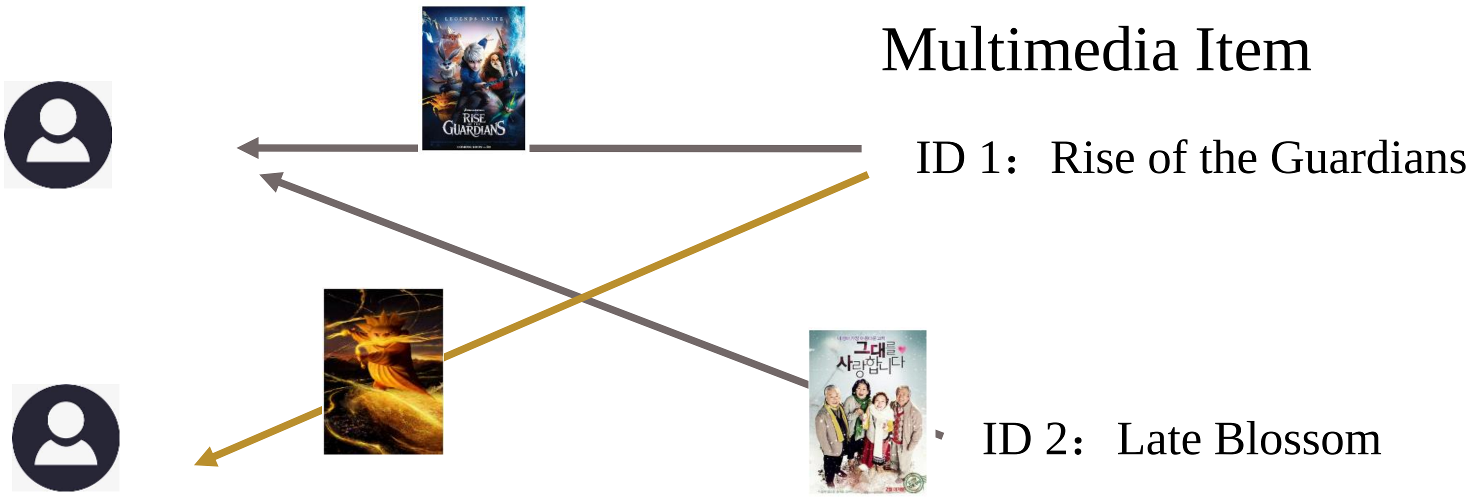
Key Frame Extraction

- A key frame is a brief description of a multimedia item.
- Previous works focus on how to summarize representative content as key frames, which present the same key frame of each multimedia item for all users.
- In the real world, users' visual preferences are not the same but vary from person to person.





Personalized Multimedia and Key Frame Recommendation



Task 1 [Multimedia Item Recommendation]: Predict each user a's unknow preference $\hat{r}_{ai}$ to multimedia item i.

Task 2 [Key Frame Recommendation]: For user a and the recommended item i, predict her unknown fine-grained preference $\hat{l}_{ak}$ to each fame $k(s_{ki} = 1)$.

Challenges

- Discover users' visual preferences **without any user-frame behavior** for key frame recommendation.
  - Nearly all recommendation works need the detailed user-frame interaction behavior to model users' preferences for frames.

- Personalized item recommendation and key frame explanation **at the same time.**

Joint Multimedia Item and Key Frame Recommendation (JIFR)

- Key idea: design a model to discern both the **collaborative and visual dimensions** of users, and model how users make decisive item preferences from these two spaces.

latent collaborative space

the visual space

Input:

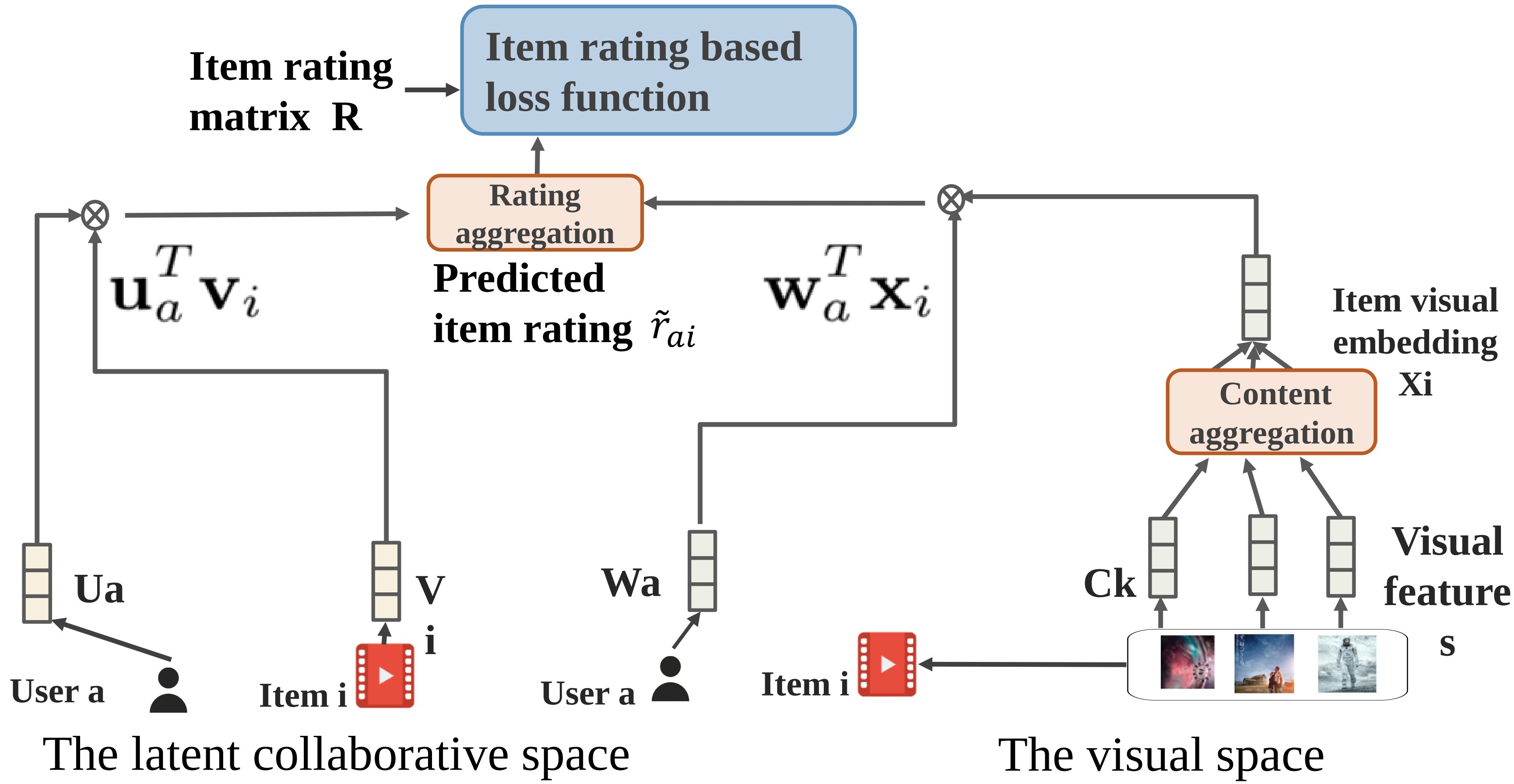
User-Item matrix R

	Item 1	……	Item N
User 1	1		1
……			
User M	0		1

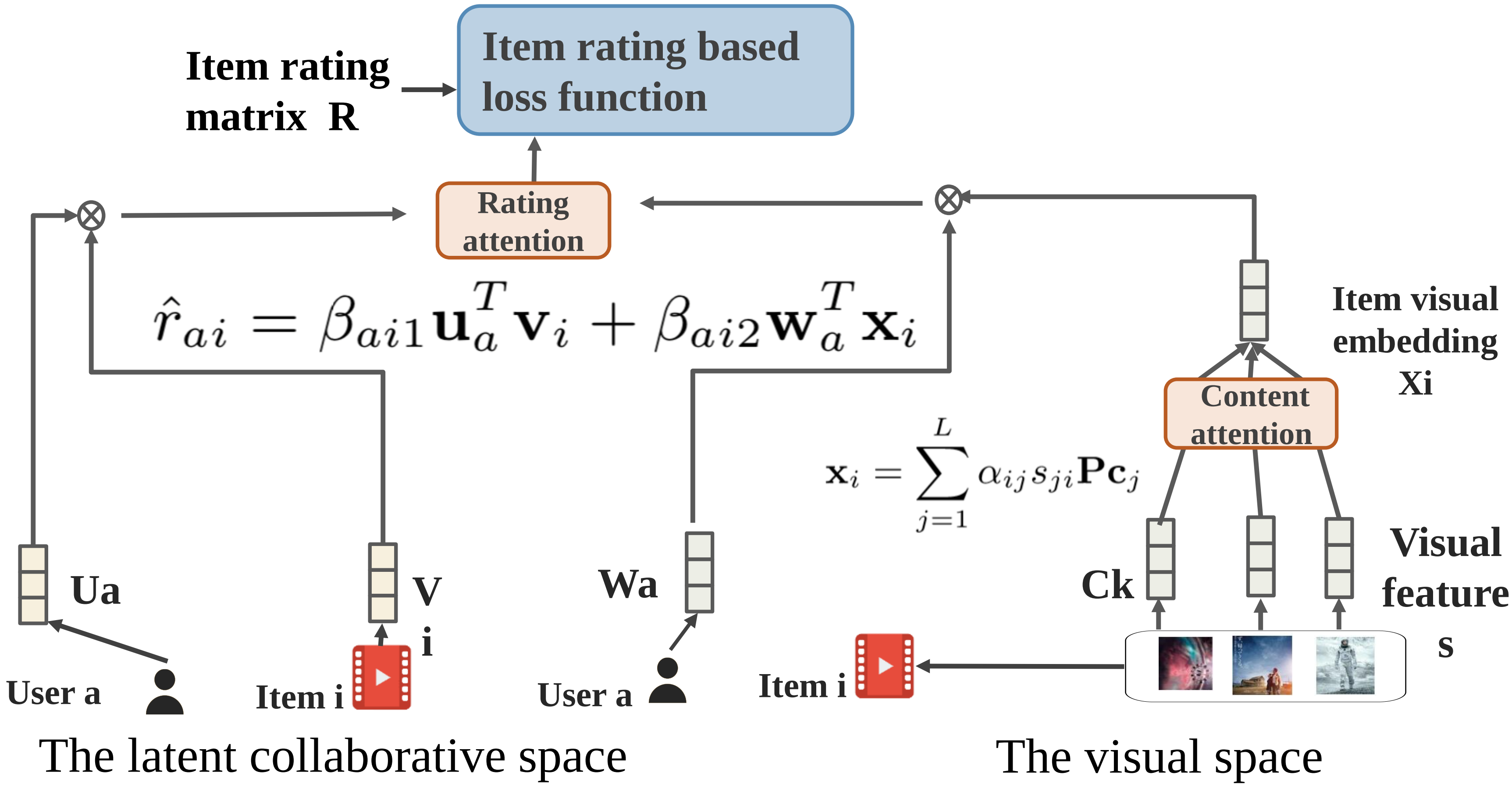
Item-Frame matrix S

	Frame 1	……	Frame L
Item 1	1		0
……			
Item N	0		1

Our Proposed Model



Our Proposed Model



Our Proposed Model

Model Learning

$$\min \mathcal{L}_R = \sum_{a=1}^{M} \sum_{(i,j)\in D_a^R} \sigma(\hat{r}_{ai} - \hat{r}_{aj}) + \lambda_1 ||\Theta_1||_F^2$$

Model Prediction

Multimedia item recommendation

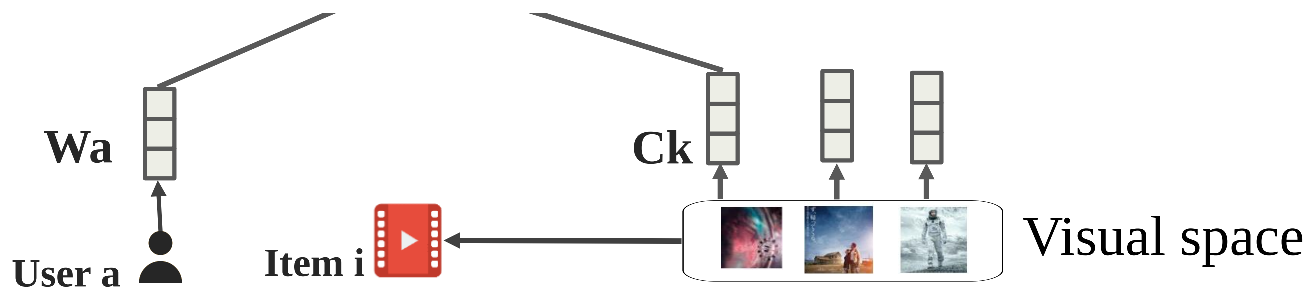
$$\hat{r}_{ai} = \beta_{ai1}\mathbf{u}_a^T\mathbf{v}_i + \beta_{ai2}\mathbf{w}_a^T\mathbf{x}_i$$

Collaborative space

Visual space

Key frame recommendation

$$\hat{l}_{ak} = \mathbf{w}_a^T(\mathbf{P}\mathbf{c}_k)$$



Experiments

Dataset: Douban

#Users	#Movies	#Frames	#Training movie ratings	# Test movie ratings	# Test frames
16,166	12,811	140,916	379,727	98,465	4,760

Movie rating

Frame rating

Users' detailed preferences for frames are only used for test.

Experiments

Baselines:

Item Recommendation:

BPR, **CDL**, **VBPR**, **VPOI**, **ACF**

Key Frame Recommendation:

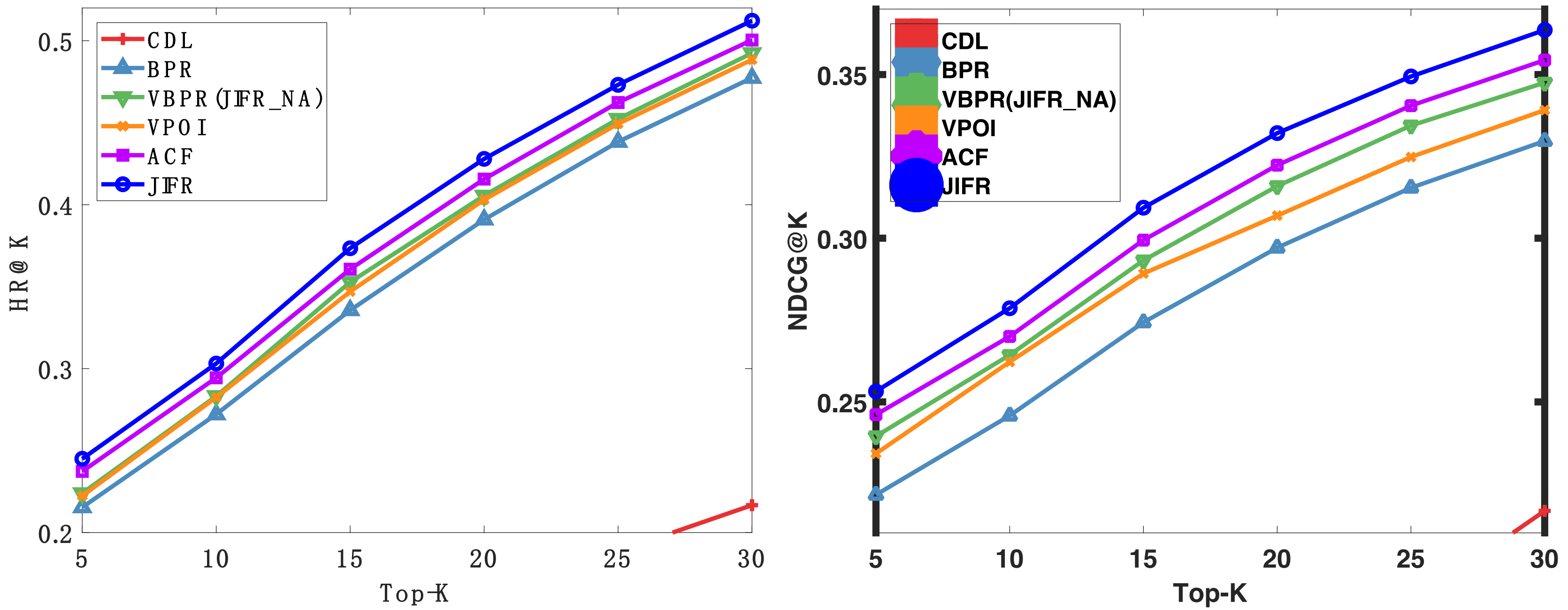
RND, **CDL**, **JIFR_NA**

Model	Input		Task	
	Rating	Image	Item Rec	Frame Rec
BPR [Rendle *et al.*, 2009]	√	×	√	×
CDL [Lei *et al.*, 2016]	√	√	√	√
VBPR [He and McAuley, 2016]	√	√	√	×
VPOI [Wang *et al.*, 2017]	√	√	√	×
ACF [Chen *et al.*, 2017a]	√	√	√	×
JIFR_NA	√	√	√	√
JIFR	√	√	√	√

Evaluation metrics: HR, NDCG

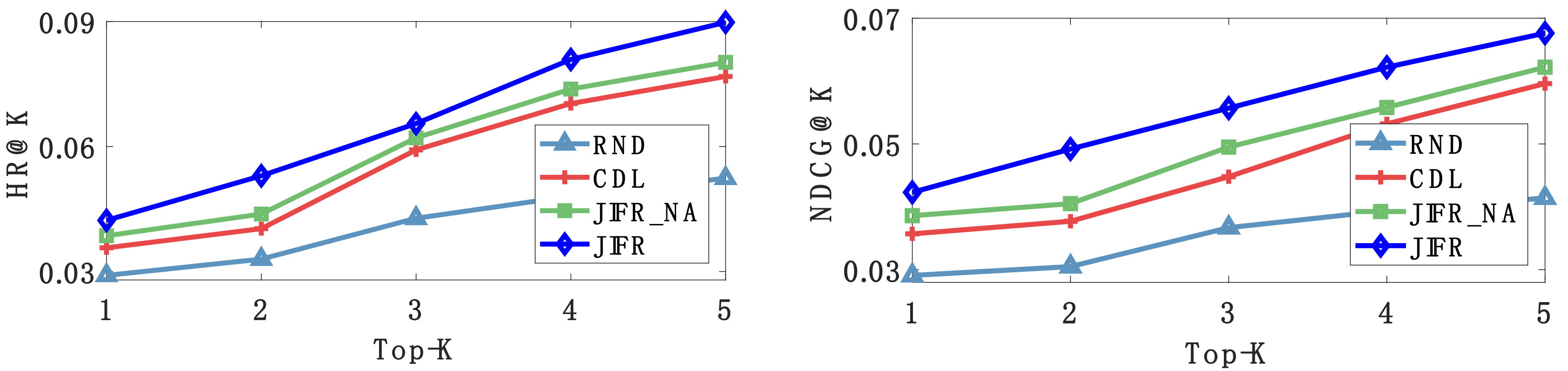
Experiments (1/3)

Multimedia item recommendation performance



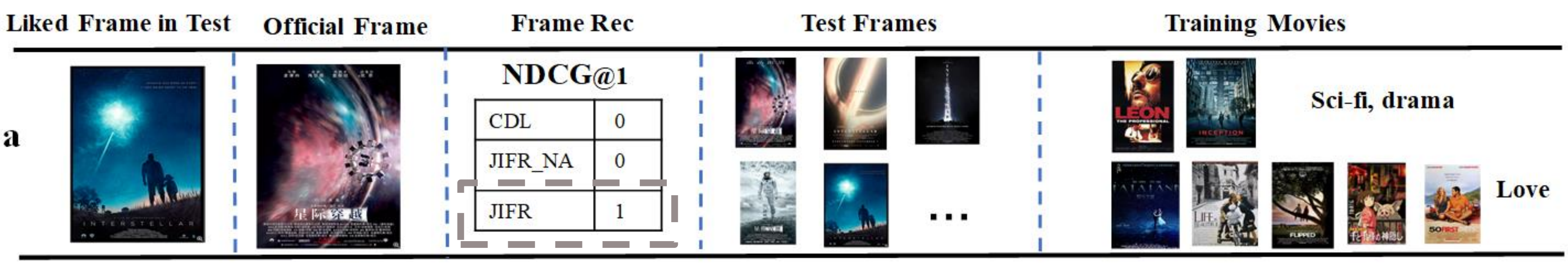
Experiments (2/3)

Key frame recommendation performance



Experiments (3/3)

Case study: personalized key frame recommendation of *Interstellar*



Outline

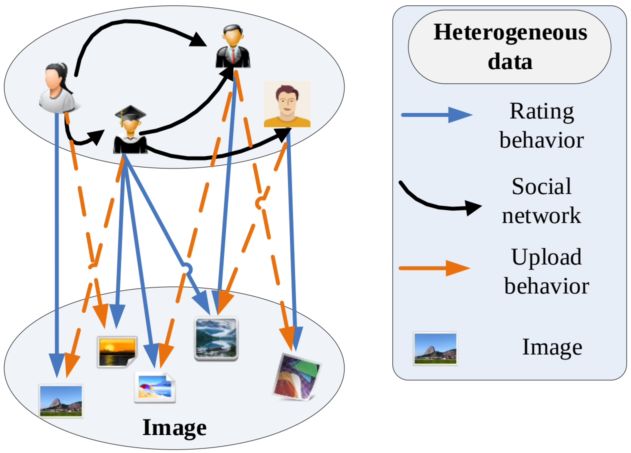
- Research Background
- Influence Diffusion for Social Recommendation
- Explainable Multimedia based Recommendation
- A Unified Model for Social Multimedia Recommendation
- Conclusions and Future Work

Le Wu(吴乐), Lei Chen , Richang Hong, Yanjie Fu, Xing Xie, Meng Wang. A Hierarchical Attention Model for Social Contextual Image Recommendation. ***IEEE TKDE***, 2019.

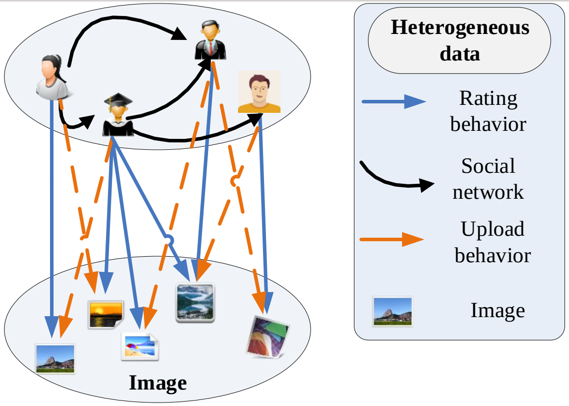
Social Image based Platforms

- Image-based social networking service platforms are very popular in recent years.
  - A picture is worth a thousand words
  - 百闻不如一见

- A typical social image platform
  - User-image interaction behavior
  - User-user social network
  - User-image upload behavior
  - Visual image information



Research Challenges

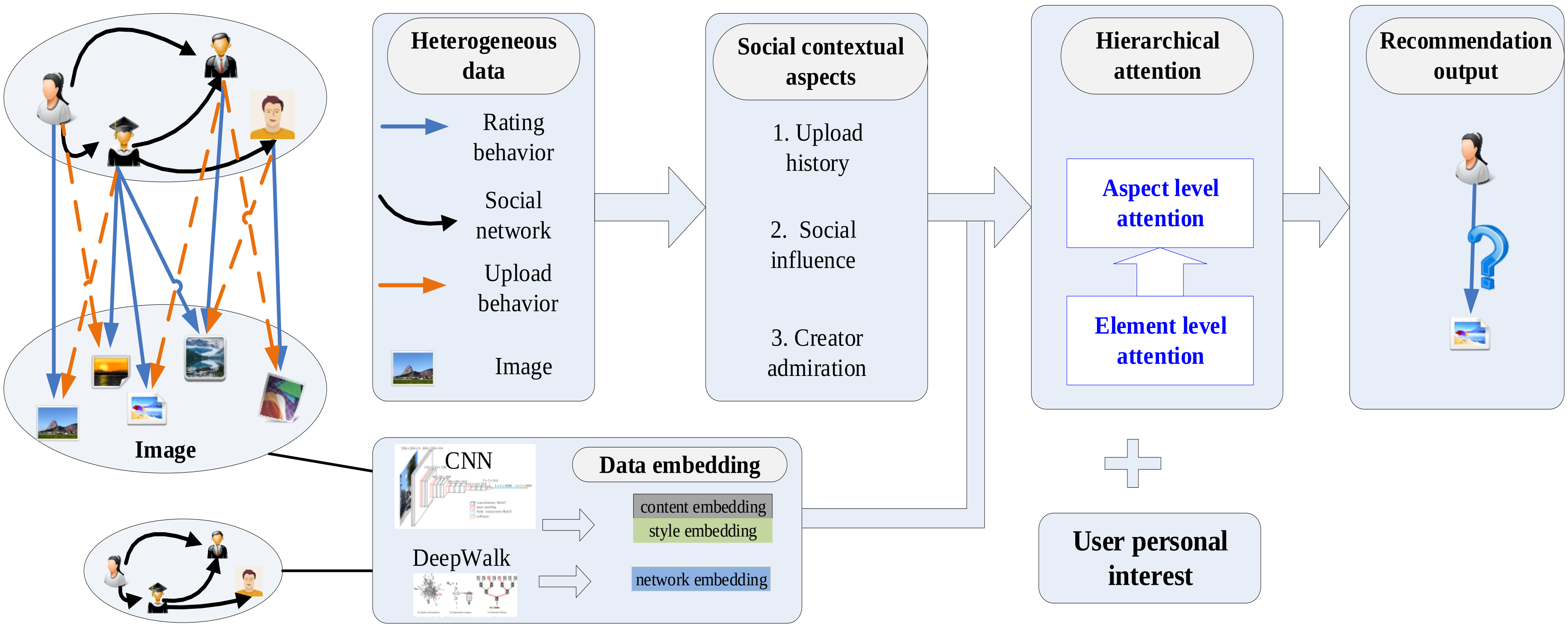


- Previous works focused on either social recommendation or image recommendation

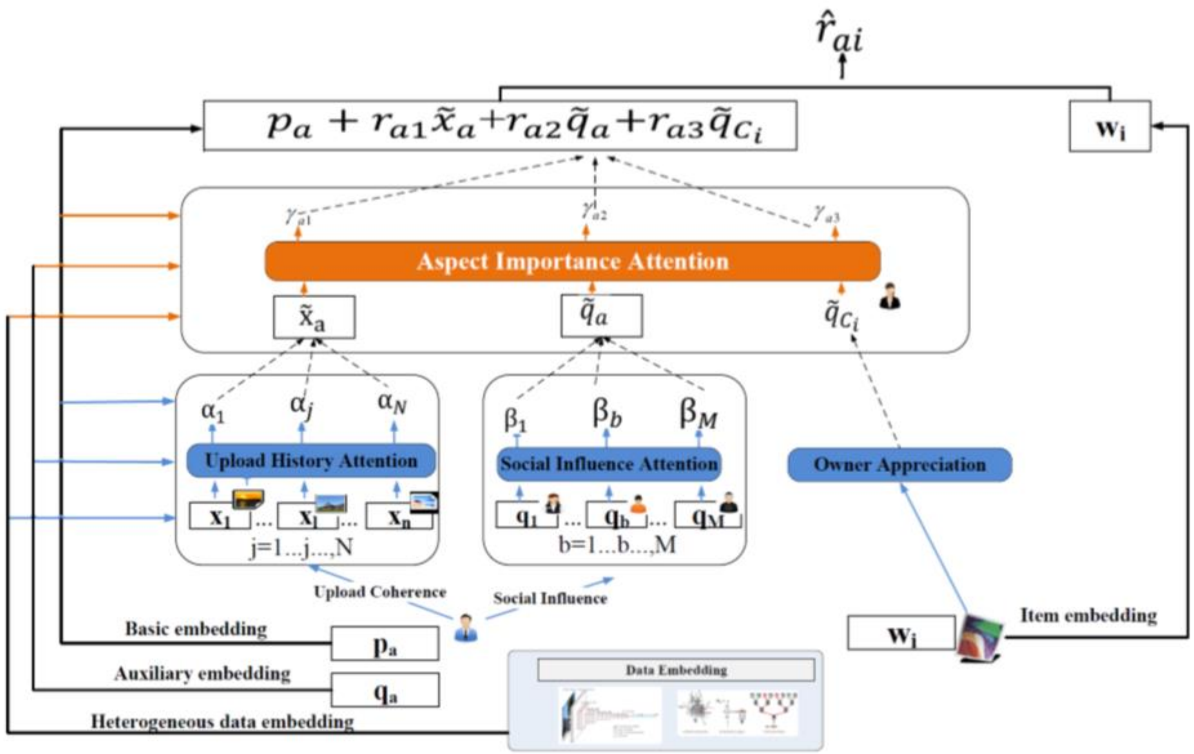
- Given heterogeneous data, how to better summarize the **heterogeneous social contextual aspects** that influence a user's decision in a **holistic** way?

- Furthermore, different users care about different social contextual aspects for their **personalized contextual preference.**

The Overall Framework of the Proposed Model



The Proposed Model



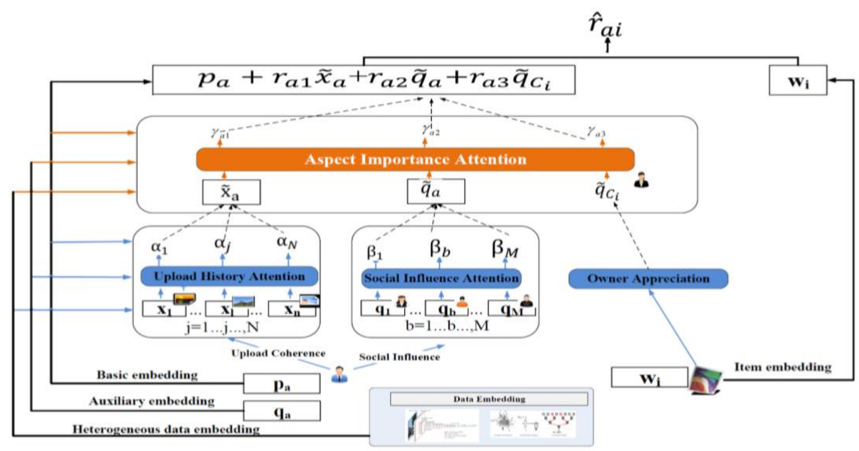
$$\hat{r}_{ai} = \mathbf{w}_i^T(\mathbf{p}_a + \gamma_{a1}\tilde{\mathbf{x}}_a + \gamma_{a2}\tilde{\mathbf{q}}_a + \gamma_{a3}\mathbf{q}_{C_i})$$

where $\tilde{\mathbf{x}}_a = \sum_{j=1}^{N} l_{ja}\alpha_{aj}\mathbf{x}_j, \quad \tilde{\mathbf{q}}_a = \sum_{b=1}^{M} s_{ba}\beta_{ab}\mathbf{q}_b.$

$$\hat{r}_{ai} = \overbrace{\mathbf{p}_a^T\mathbf{w}_i}^{\text{Basic Latent Factor Model}} + \underbrace{\gamma_{a1}\sum_{j=1}^{N}\alpha_{aj}l_{ja}\mathbf{x}_j^T\mathbf{w}_i}_{}^{\text{Item Neighborhood Model}}$$

$$+ \underbrace{\gamma_{a2}\sum_{b=1}^{M}s_{ba}\beta_{ab}\mathbf{q}_b^T\mathbf{w}_i}_{\text{Social Neighborhood Model}} + \underbrace{\gamma_{a3}\mathbf{q}_{C_i}^T\mathbf{w}_i}_{\text{Owner Admiration Bias}},$$

The Proposed Model



- Hierarchical Attentive Social Contextual recommendation
  - **Bottom-layer attention**: the importance of each element within each aspect
    - Upload coherence attention $$\widetilde{x}_a = \sum_{j=1}^{N} l_{ja}\alpha_{aj}\mathbf{x}_j.$$
    - Social influence attention $$\widetilde{q}_a = \sum_{b=1}^{M} s_{ba}\beta_{ab}\mathbf{q}_b.$$
  - **Top-layer attention:** the importance of each aspect

Experiments

- Dataset
  - Start from the NUS-WIDE dataset
  - Crawl the user-related information.

Our proposed HASC model

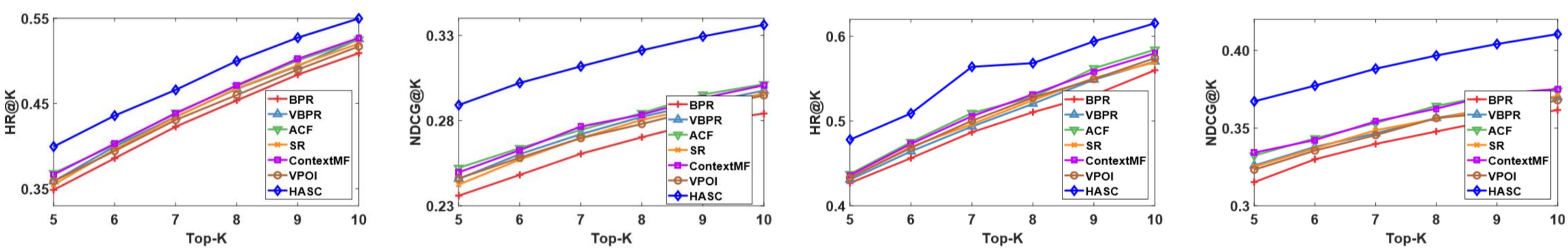
✓ Accuracy

✓ Partially Explanation

The statistics of the two datasets.

Dataset	Users	Images	Ratings	Social Links	Rating Density
F_S	4,418	31,460	761,812	184,991	0.55%
F_L	8,358	105,648	1,323,963,	378,713	0.15%

- Overall performance



(a) HR@K on F_S (b) NDCG@K on F_S (c) HR@K on F_L (d) NDCG@K on F_L

Ablation study

The improvement of using different attention mechanism compared to BPR.

Bottom Layer Attention	Top Layer Attention	F_S		F_L	
		HR	NDCG	HR	NDCG
AVG	AVG	6.44%	10.28%	5.54%	9.02%
MAX	MAX	5.82%	9.55%	4.98%	8.10%
AVG	ATT	7.33%	11.15%	5.95%	9.93%
MAX	ATT	6.84%	10.96%	5.72%	9.55%
ATT	AVG	12.75%	19.23%	8.30%	13.28%
ATT	MAX	12.20%	18.56%	8.02%	12.85%
ATT	ATT	**14.57%**	**22.55%**	**10.67%**	**16.70%**

The improvement of modeling different contextual aspects with our proposed model compared to BPR(U: upload history, S: social influence, C: creator admiration).

Aspects	F_S		F_L	
	HR	NDCG	HR	NDCG
U	8.70%	16.52%	6.44%	11.03%
S	9.63%	16.78%	5.29%	9.65%
C	8.57%	14.53%	4.37%	7.93%
U+S+C	**14.57%**	**22.55%**	**10.67%**	**16.70%**

Outline

- Research Background
- Influence Diffusion for Social Recommendation
- Explainable Multimedia based Recommendation
- A Unified Model for Social Multimedia Recommendation
- Conclusions and Future Work

Conclusions

- Social multimedia recommendation is a popular trend in RS domain
  - Increase model accuracy with auxiliary data.
  - Explainable recommendations with social and multimedia as the explainable components.
- Proposed models
  - A neural influence diffusion model for social recommendation
  - Explainable multimedia based recommendations
  - A hierarchical attention model to tackle the heterogeneous social contextual aspects in social image platforms.

Future Work

- Many recommendation problems could be formulated as the graph form, how to design graph neural network(GNN) based models for recommendation?
  - GNN based recommendation models
  - Robustness and adversarial attacks on GNN based recommendation models.

- Interpretability in recommendation
  - Social path based recommendation
  - Explainable recommendations with language generation techniques

- Emerging applications in social contextual recommendation.
  - Short video recommendation

Thank you!