School of
Information Systems

SMU
SINGAPORE MANAGEMENT
UNIVERSITY

Basket-Sensitive
Personalized Item Recommendation

Duc-Trong Le, Hady W. Lauw, Yuan Fang

IJCAI 2017
Melbourne, Australia

Outline

- **Motivating examples** and **Models**
  - Factorization Machine (FM)
  - Basket-Sensitive Factorization Machine (BFM)
  - Constrained Basket-Sensitive Factorization Machine (CBFM)
- **Experiments** on BeiRen, Foursquare

Personalized Item Recommendation

Mary

u_i

Adopted Items

$\cdots$

$\{v_*\}$

Candidate Items

$\cdots$

$\{v_j\}$

Learn a real-value function from adoptions to rank candidate items

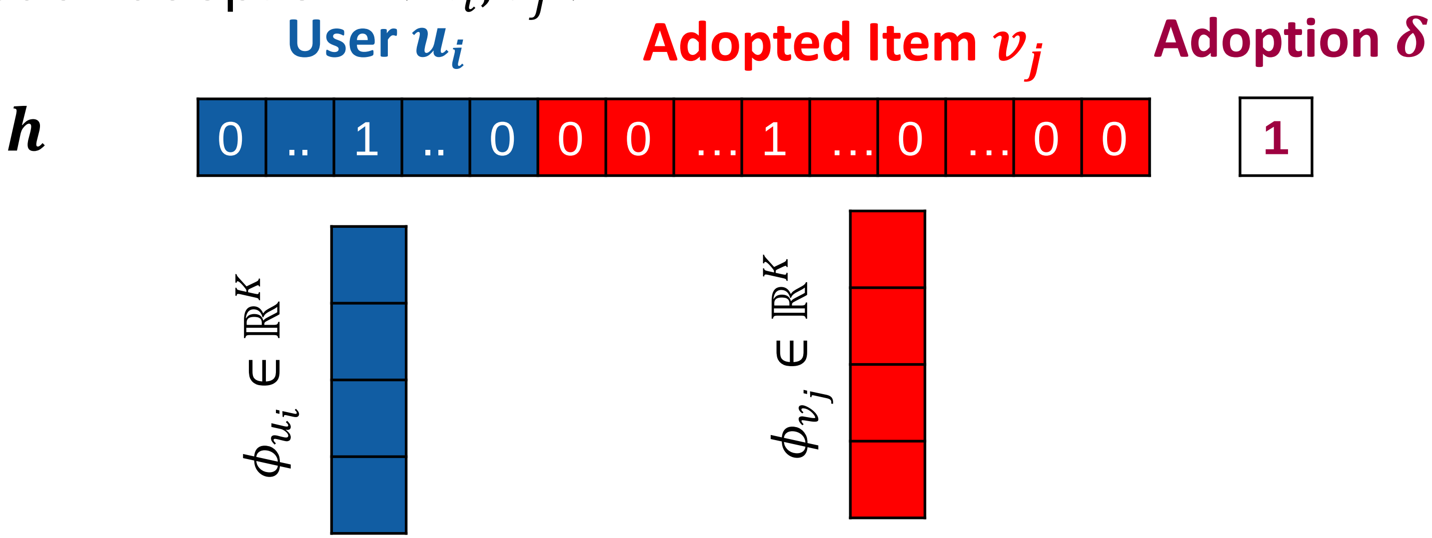
$$F\left(u_i, v_j; \Theta\right)$$

Matrix Factorization:

$$F\left(u_i, v_j; \Theta\right) \propto \phi_i * \phi_j \text{ where } \phi_i, \phi_j \in \mathbb{R}^K$$

Factorization Machine (FM)

For each adoption $< u_i, v_j >$:



$$\underbrace{\mathcal{F}(\boldsymbol{h};\Theta)}_{\text{Adoption Estimation}} = \underbrace{\mu_0 + \sum_{i=1}^{p} \mu_i h_i}_{\text{Global \& Item Bias}} + \underbrace{\sum_{i=1}^{p} \sum_{j=i+1}^{p} h_i h_j \sum_{f}^{K} \boxed{\phi_{i,f}\phi_{j,f}}}_{\text{Interactive Features}}$$

Optimization:

$$OPT\_FM(T) = \text{argmin}_{\Theta} \sum_{\{h\}} l(\mathcal{F}(\boldsymbol{h};\Theta), \delta)$$

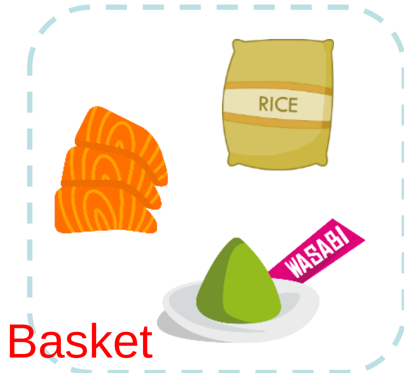
Loss function

The Notion of Basket – Shopping Scenarios

Mary

Shopping cart on …

Monday



Sushi

Thursday

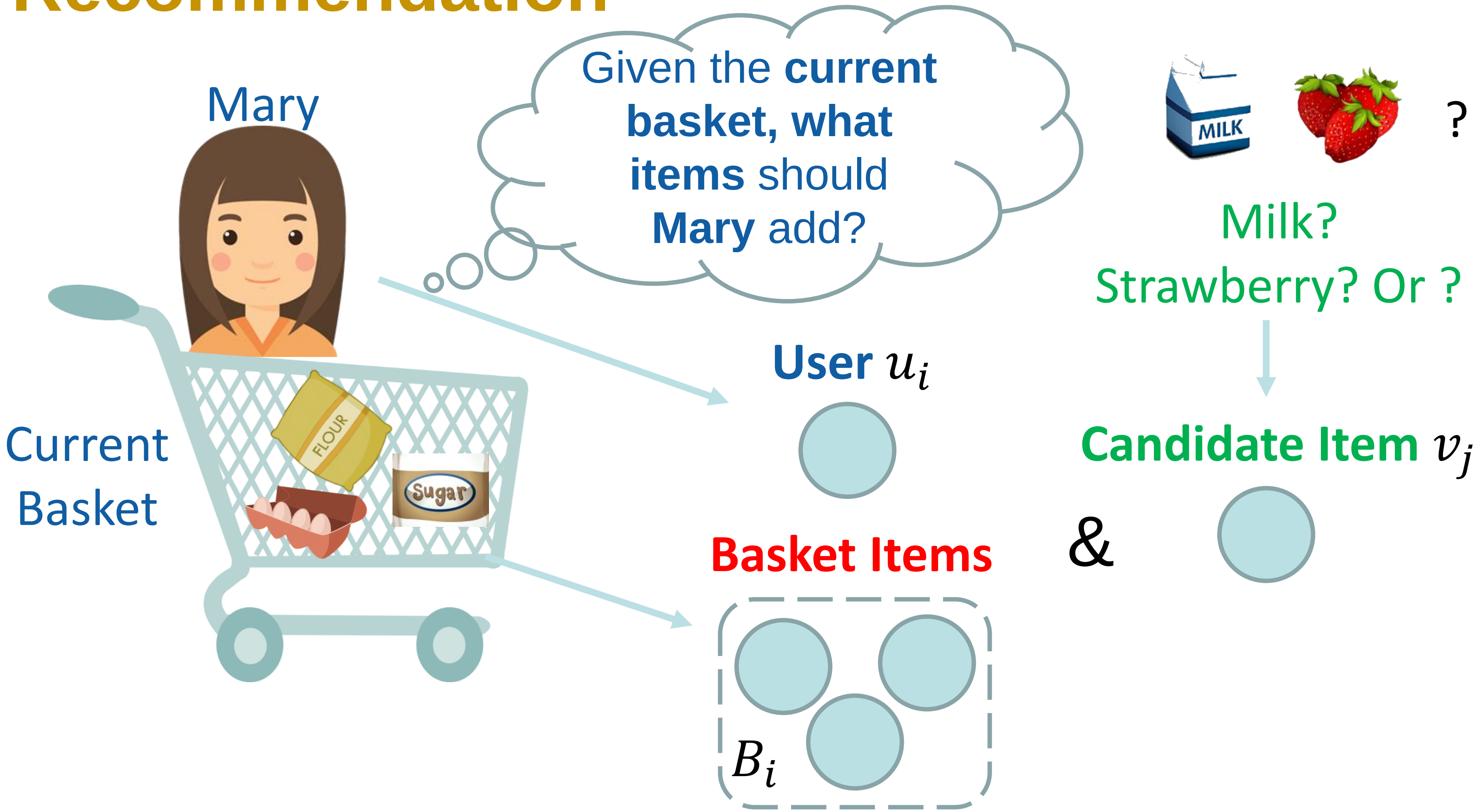
Chili Crab

Saturday

Cake

Items in each basket might be correlated

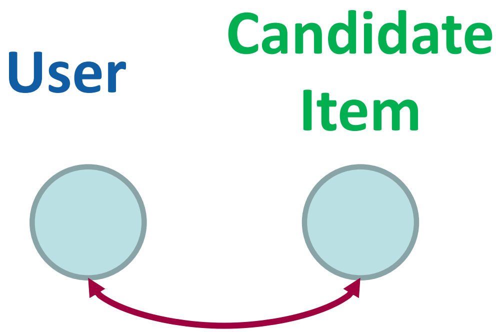
Basket-Sensitive Personalized Item Recommendation



Learn a real-value function to rank candidate items

$$F(u_i, B_i, v_j; \Theta)$$

Modeling Association Types - γ_1



Mary

likes

Milk

Capture personalization

Modeling Association Types - γ_2

Milk

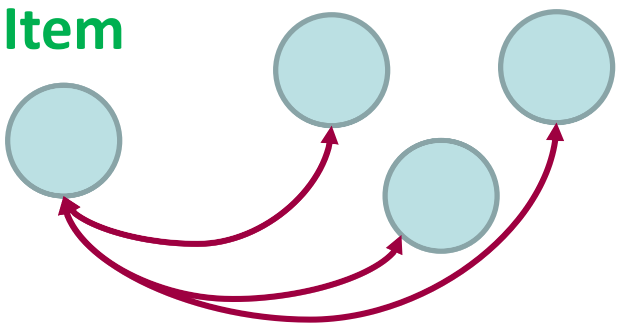
&

Baking Items

Capture correlations

Candidate Item

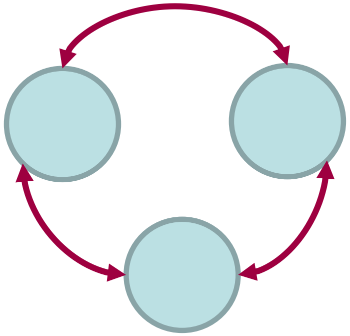
Basket Items



Modeling Association Types - γ_3

Among

Basket Items



Capture correlations in the current basket

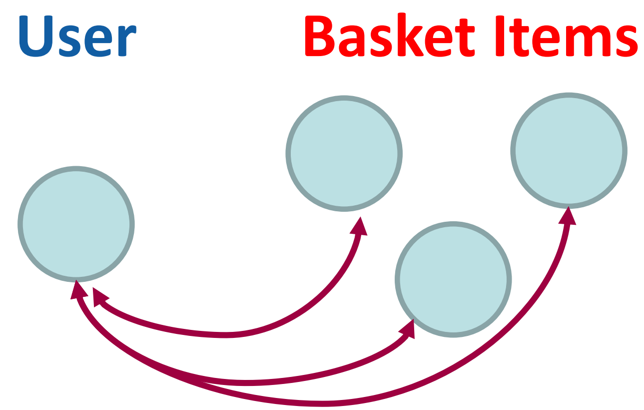
Modeling Association Types - γ_4

Mary

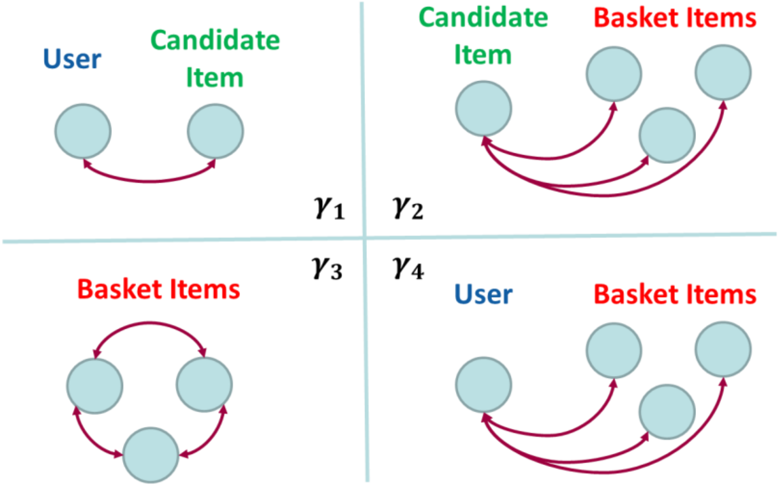
likes

Baking Material

Capture personalization
(Maybe redundant)



Real-Valued Function with Association Types



$$F\left(u_i, B_i, v_j; \Theta\right) \propto \boxed{\gamma_1} . x_i^T y_j + \boxed{\gamma_2} . \sum_{v_k \in B_i} y_j^T z_k$$

Adoption Estimation

User & Candidate Item

Candidate Item & Basket Items

$$+ \boxed{\gamma_3} . \sum_{(v_k \neq v_{k'}) \in B_i} z_k^T z_{k'} + \boxed{\gamma_4} . \sum_{v_k \in B_i} x_i^T z_k$$

Among Basket Items

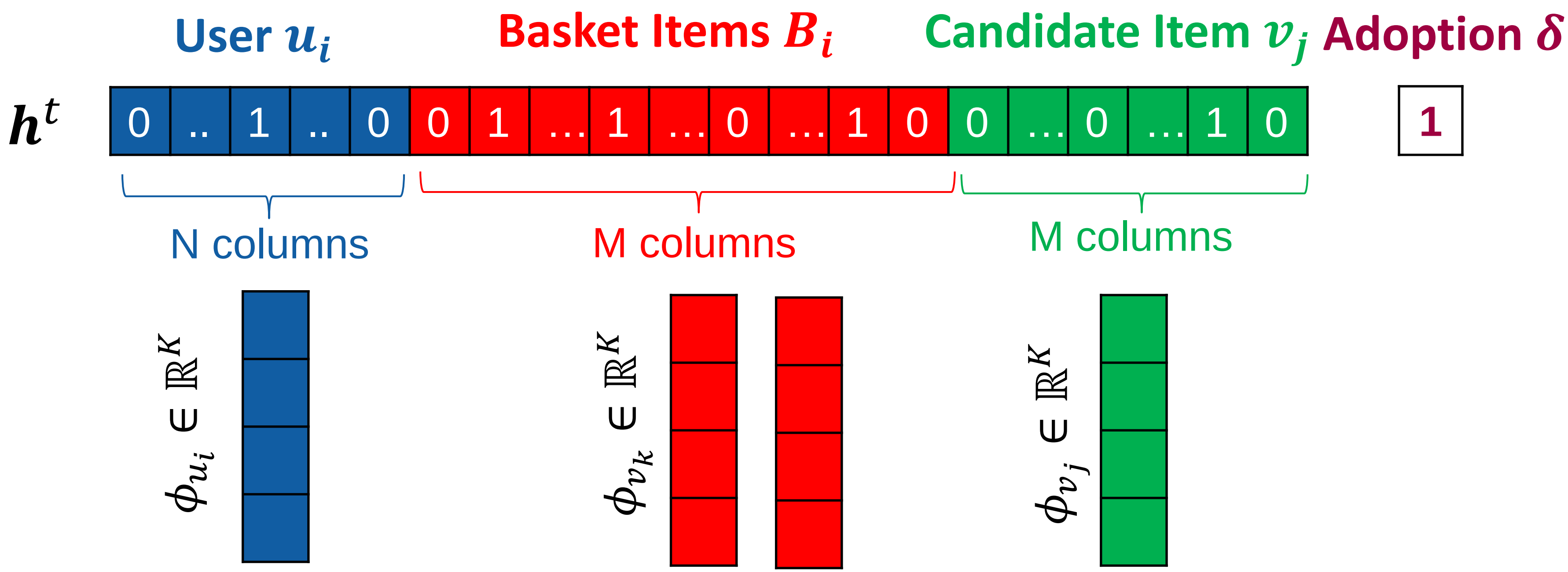
User & Basket Items

where $x_i, y_j, z_k \in \mathbb{R}^K$ are latent vectors; $\gamma_1, \gamma_2, \gamma_3, \gamma_4 \in \{0,1\}$

Basket-Sensitive Factorization Machine (BFM)

Given N users, M items;

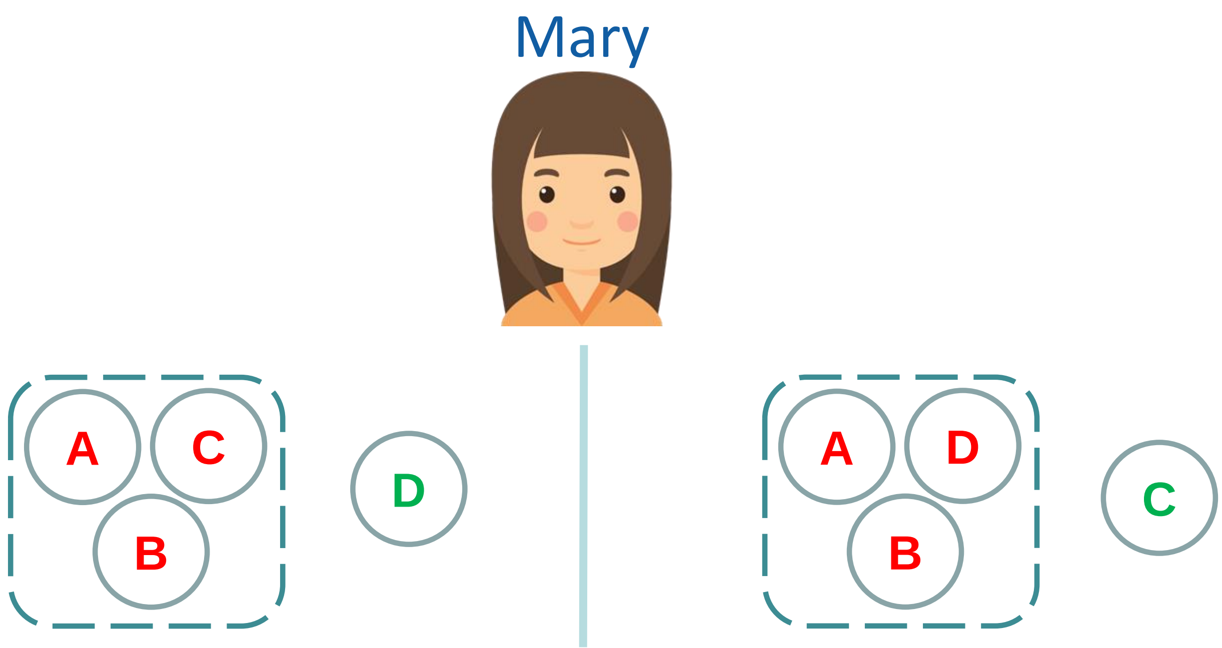
Tuple $t = < u_i, B_i, v_j, \delta > \in T$; $t.\delta \in \{-1, 1\}$;



$$\mathcal{F}(\mathbf{h};\Theta) = \mu_0 + \sum_{i=1}^{p} \mu_i h_i + \boxed{\gamma_1 \sum_{i=1}^{N} \sum_{j=N+1}^{N+M} h_i h_j ({\phi_i}^T \phi_j)} + \boxed{\gamma_2 \sum_{i=N+1}^{N+M} \sum_{j=N+M+1}^{p} h_i h_j ({\phi_i}^T \phi_j)}$$

$$+ \boxed{\gamma_3 \sum_{i=N+M+1}^{p} \sum_{j=i+1}^{p} h_i h_j ({\phi_i}^T \phi_j)} + \boxed{\gamma_4 \sum_{i=1}^{N} \sum_{j=N+M+1}^{p} h_i h_j ({\phi_i}^T \phi_j)}$$

Same-Intent Tuples



$t_1 = < \text{Mary}, \{A, C, B\}, D, 1 >$

$t_2 = < \text{Mary}, \{A, B, D\}, C, 1 >$

Same-Intent Tuples

Constrained BFM (CBFM)

- Given same-intent tuples (t_1, t_2), we expect

$$PMI(t_1, t_2) * \left(\mathcal{F}(\boldsymbol{h}^{t_1}; \Theta) - \mathcal{F}(\boldsymbol{h}^{t_2}; \Theta)\right)^2$$

should be small

Pointwise Mutual Information

Adoption Estimation Difference

- Not all tuple pairs have strong intended effect, only consider the constraint in optimization task:

$$PMI(t, t^m) * \left(\mathcal{F}(\boldsymbol{h}^{t}; \Theta) - \mathcal{F}\left(\boldsymbol{h}^{t^m}; \Theta\right)\right)^2$$

where t, t^m are the current tuple and its same-intent tuple that has the maximum $\mathcal{F}\left(\boldsymbol{h}^{t^m}; \Theta\right)$ score

Optimization

$$\mathrm{OPT\_BFM}(T) = \mathrm{argmin}_{\Theta}\left[\sum_{t\in T} \underbrace{-\ln(\sigma(\mathcal{F}(\mathbf{h}^t;\Theta)\times t.\delta))}_{\text{Log-loss function}} + \underbrace{\sum_{\theta\in\Theta}\lambda_\theta\theta^2}_{\text{Regularization}}\right]$$

$$\mathrm{OPT\_CBFM}(T) = \mathrm{argmin}_{\Theta}\left[\underbrace{\sum_{\theta\in\Theta}\lambda_\theta\theta^2}_{\text{Regularization}} + \sum_{t\in T}\{\underbrace{-\ln(\sigma(\mathcal{F}(\mathbf{h}^t;\Theta)\times t.\delta))}_{\text{Log-loss function}} + \frac{\alpha}{2}\times \underbrace{PMI(t,t^m)\times\left(\mathcal{F}(\mathbf{h}^t;\Theta)-\mathcal{F}(\mathbf{h}^{t^m};\Theta)\right)^2}_{\text{Constraint}}\}\right]$$

Stochastic Gradient Descent with Adaptive Learning Rate

$\alpha, \lambda_\theta \in \mathbb{R}^+$; $\sigma(a) = 1/(1+e^{-a})$

Outline

- **Motivating examples** and **Models**
  - Factorization Machine (FM)
  - Basket-Sensitive Factorization Machine (BFM)
  - Constrained Basket-Sensitive Factorization Machine (CBFM)
- **Experiments** on BeiRen, Foursquare

Experimental Setup

- **Task**: Next adoption recommendation
  - For each testing tuple $t = < u_i, B_i, v_j, \delta >$
  - Hide the observed adoption v_j, require each model to product a ranked list of candidates for u_i based on B_i
- **Datasets**

Dataset	#User	#Item	#Transaction	Avg. #Item / Transaction
BeiRen (Grocery shopping)	9245	5581	87224	6.1
Foursquare (Point-of-Interest)	1548	3619	31377	2.7

Pre-processing: Filter out too few or too popular items; Sample negative tuples ($t.\delta = -1$)

- **Metric**: *Half-life Utility (HLU)*

$$HLU = \frac{1}{|T_{\text{test}}|} \times C \times \sum_{t \in T_{\text{test}}} 2^{\frac{1-r_t}{\beta-1}}; C = 100, \beta = 5$$

School of Information Systems

SMU Singapore Management University

What are effects of the Association Types?

Higher is better

Association				BeiRen	Foursquare
γ_1	γ_2	γ_3	γ_4		
1	0	0	0	1.94	5.45
1	1	0	0	3.35	8.11
1	1	1	0	**3.75***	**8.51***
1	1	0	1	3.59	7.08
1	1	1	1	3.74	8.02

γ_1: User & Candidate Item

γ_2: Candidate & Basket Items

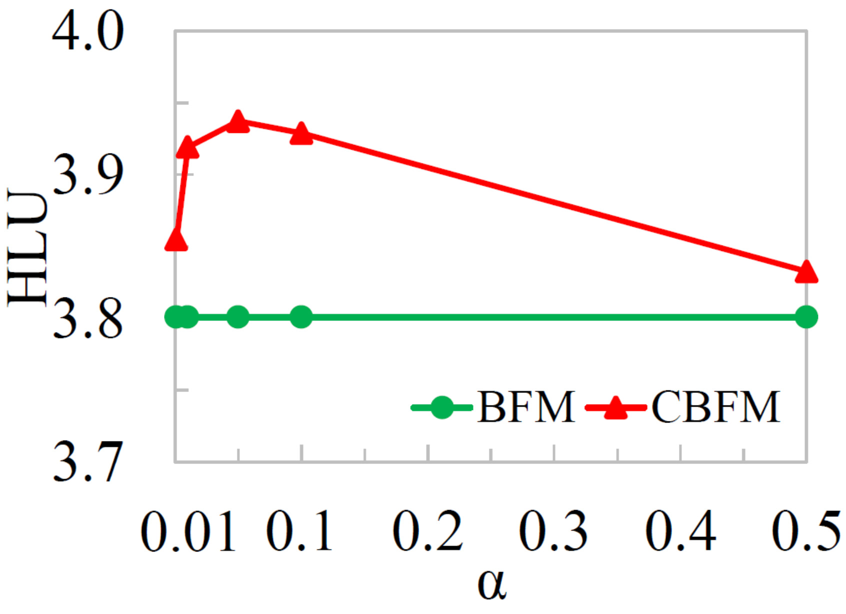
γ_3: Among Basket Items

γ_4: User & Basket Items

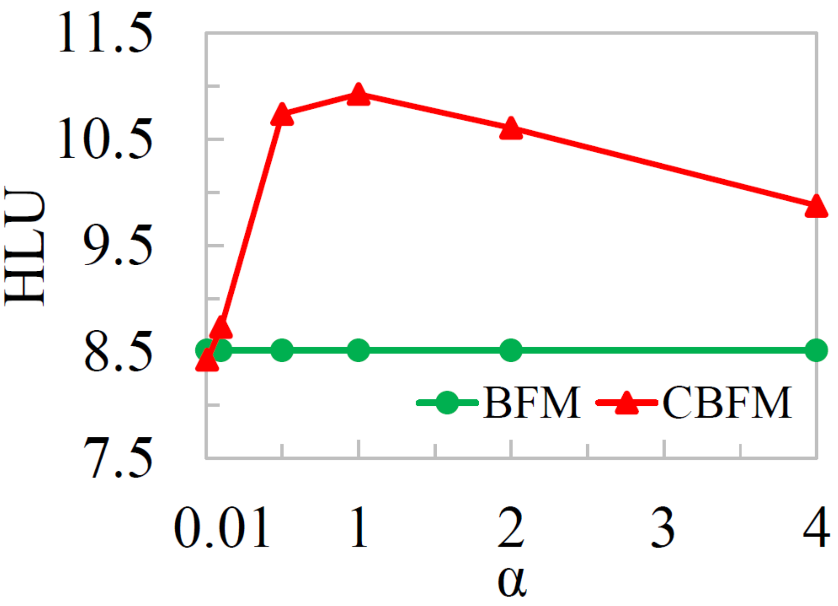
Latent dimension K = 8

How does the constraint influence the prediction performance?

Higher is better



BeiRen

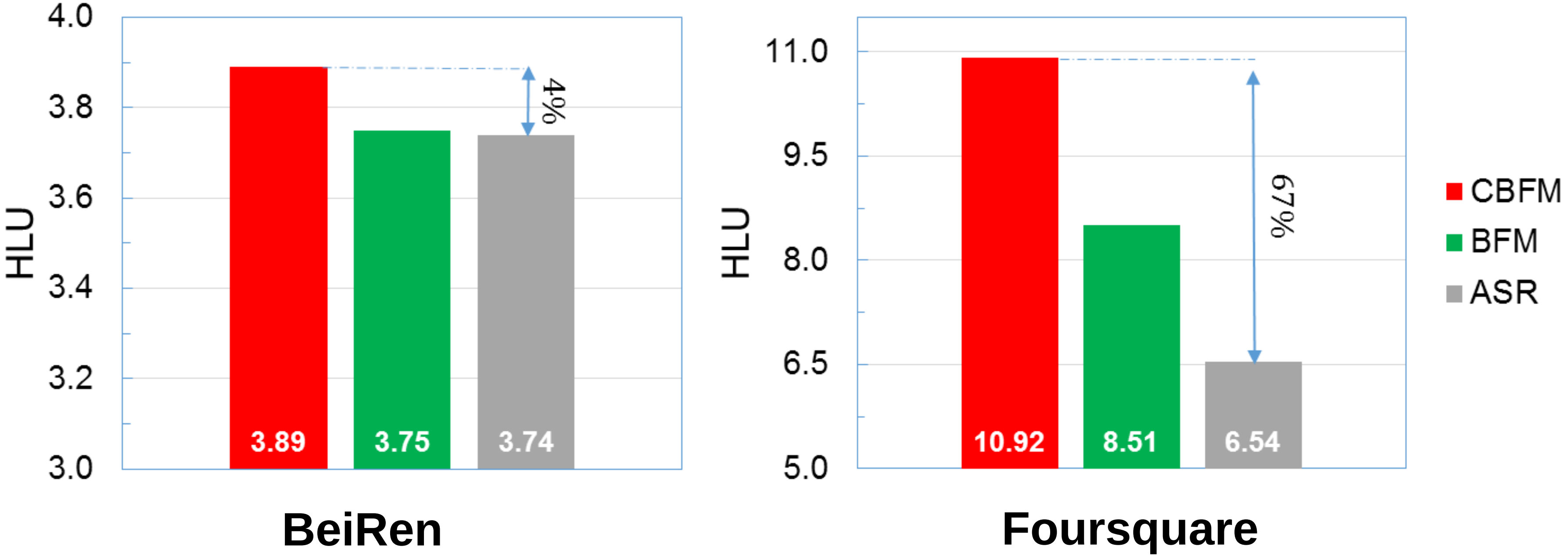


Foursquare

Latent dimension K = 8;

Do CBFM & BFM outperform the Association Rules-based method?

Higher is better



Latent dimension K = 8

Conclusion

- Take into account of a **user's current basket information** in making personalized item recommendations.

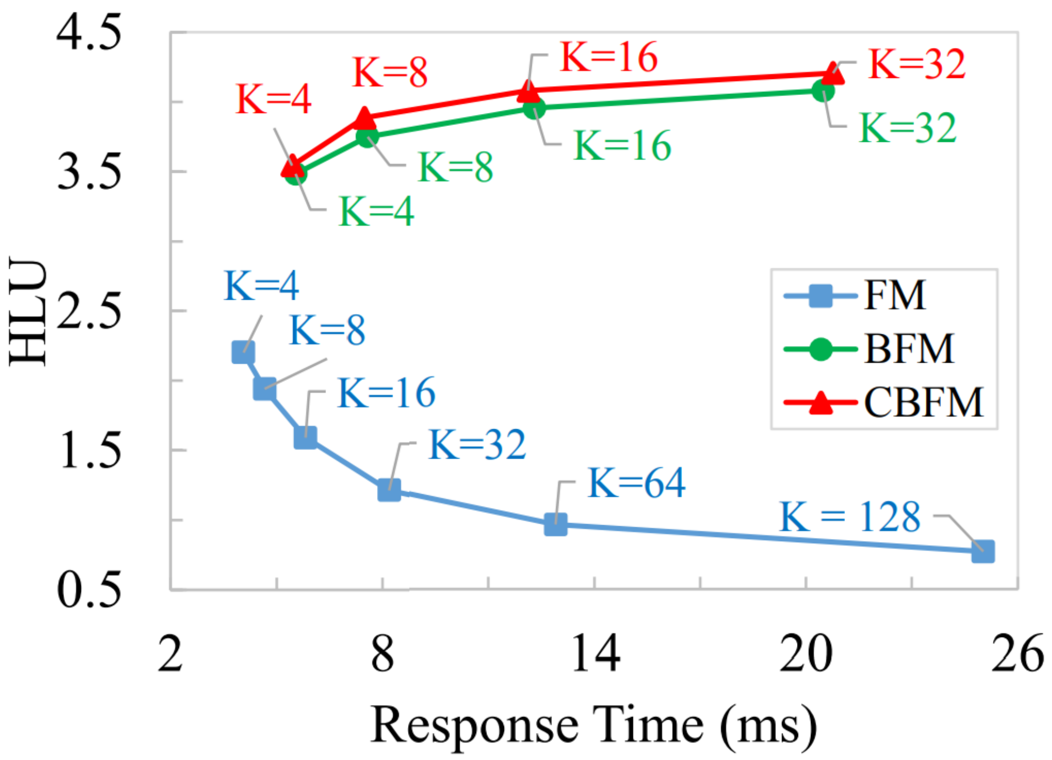
- Propose two models (**BFM** & **CBFM**) contribute **statistically significant** improvements over:
  - **Factorization Machine (FM)**
  - **Association Rule (ASR)**

  in terms of top-K recommendations.

Thank you for listening!
Q&A

Backup slides

What is the relationship between Model Complexity & Response Time?



Half-life Utility and Response Time on **BeiRen**

Negative Tuple Sampling

- For each positive tuple $t = < u_i, B_i, v_j, 1 >$, we sample two negative tuples $t^\neg = < u_i, B_i^\neg, v_j^\neg, -1 >$:
  - As $v_j^\neg$, we pick an item never selected by the user
  - $B_i^\neg$ contains items that never co-occur with either user, $v_j^\neg$, and other current items in $B_i^\neg$