A Personalized Interest-Forgetting Markov Model for Recommendations

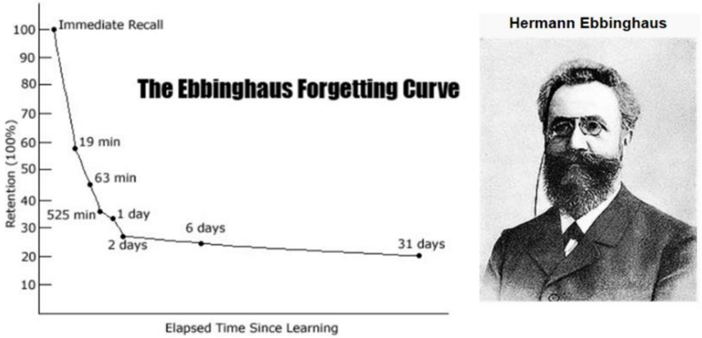
Jun Chen, Chaokun Wang, Jianmin Wang

Tsinghua University, China

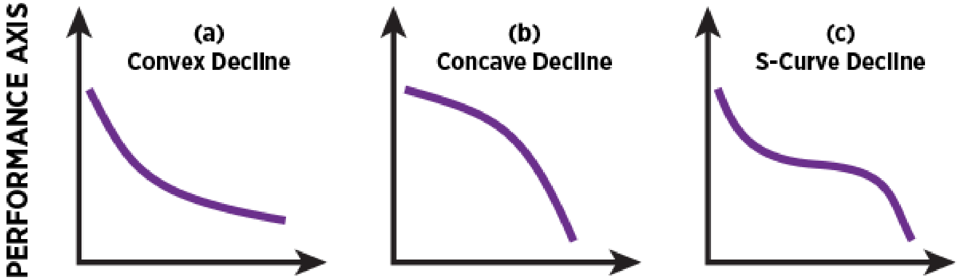
chenjun14@mails.tsinghua.edu.cn, {chaokun, jimwang}@tsinghua.edu.cn

Review on Forgetting Curve (FC)

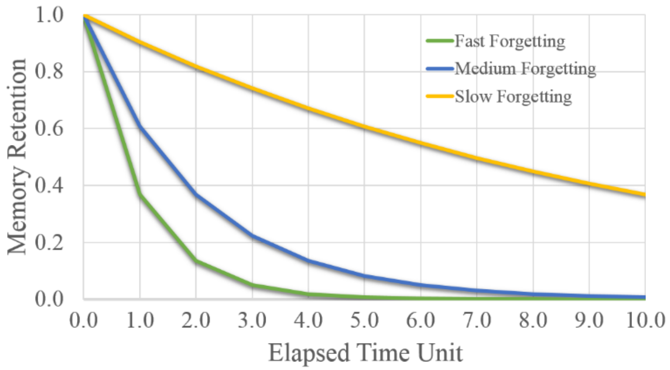
Memory Forgetting



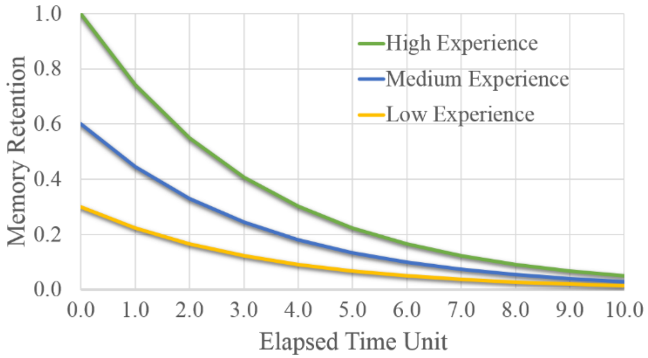
Ebbinghaus FC



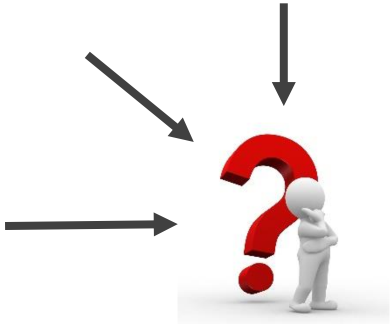
More FCs



Forgetting speeds



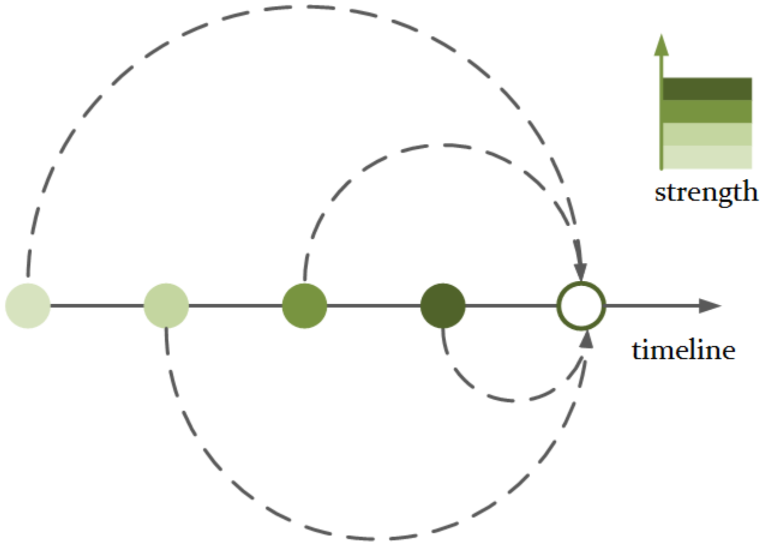
Starting experience

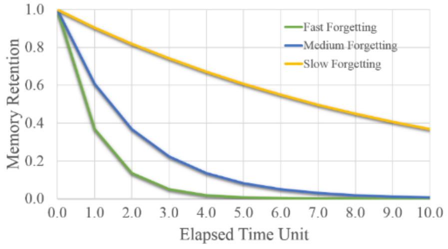


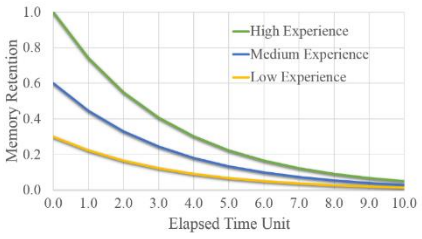
An intelligent recommender system?

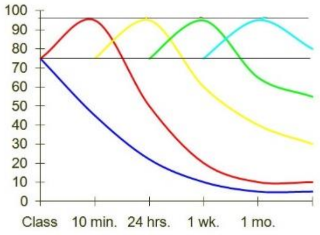
Forgetting of User Interests

- **Interest-Forgetting**
  - User u's interest upon item x loses as time elapses after the consumption.
  - Importance to influence current user interest.
- **Some issues**
  - Modeling of interest-forgetting.
  - Personalization
    - Forgetting speeds
    - Starting experience
    - Re-learning/Reconsumption









Major Contributions

1. We considered the interest-forgetting in recommendations towards a "human-minded" recommender system.
2. A personalized framework for interest-forgetting Markov model with multiple implementations on experience and interest retention.
3. An effective solution to item recommendation problem compared with the state-of-the-art.

Related Works

- **Markov model & Recommendation**
  - First-order Markov chain model[Rendle et al 2010, Cheng et al 2013].
  - High-order Markov model[Raftery 1985]
  - Variable-order Markov models[Begleiter et al 2004, Dimitrakakis 2010].
- **Memory Forgetting & Learning**
  - Forgetting models[Ebbinghaus 1885, Nembhard et al 2001, Averell et al 2011].
  - Learning models[Jaber et al 1997, Anzanello et al 2011]
  - Data filtering & updating[Packer et al 2011, Freedman et al 2011]

Problem Formulation

- **Variable-Order Markov (VOM) model based Recommendation**
  - Given an item trace $\mathcal{X}^{u,t} = \{x_1^u, x_2^u, \ldots, x_t^u\}$ of user u, recommend Top-N unseen items with the largest transition probability:

$$P(x|\mathcal{X}^{u,t}) = P(\mathrm{X}_{t+1}^u = x|\mathrm{X}_t^u = x_t^u, \ldots, \mathrm{X}_1^u = x_1^u)$$

  - Exponential expansion on the number of states ☹
- **λ-VOM**
  - *Step-wise weighted* *first-order* Markov model

$$P(x|\mathcal{X}^{u,t}) = \sum_{j=1}^{t} \lambda_j^{u,t} \, P\left(\mathrm{X}_{t+1}^u = x \middle| \mathrm{X}_{t+1-j}^u = x_{t+1-j}^u\right)$$

Framework

- **λ-VOM**

$$P(x|\mathcal{X}^{u,t}) = \sum_{j=1}^{t} \lambda_j^{u,t} P\left(x|x_{t+1-j}^{u}\right)$$

$P\left(x|x_{t+1-j}^{u}\right)$: one-step transition probability.

$\lambda_j^{u,t}$: personalized interest-forgetting component.

$$\lambda_j^{u,t} = \Upsilon_{x_{t+1-j}^{u}}^{u,t} \Phi_{x_{t+1-j}^{u}}^{u,t}$$

- Starting Experience: $\Upsilon_x^{u,t} \propto f(x,u,t)$, monotonically increasing with frequency.
- Interest Retention: $\Phi_x^{u,t} \propto 1/j$, monotonically decreasing with elapsed time steps.

IFMM Framework

- **Objective**
  - Minimize the negative log-likelihood of the probabilities to recommend the last item in each training trace.

$$\Theta^* = \operatorname*{argmin}_{\Theta} \mathcal{L} = -\sum_{u \in \mathcal{U}} \sum_{\mathcal{X}^{u,t} \in \mathcal{X}^u} \ln(P(x|\mathcal{X}^{u,t}))$$

$$= -\sum_{u \in \mathcal{U}} \sum_{\mathcal{X}^{u,t} \in \mathcal{X}^u} \ln(\sum_{j=1}^{t} \Upsilon^{u,t}_{x^u_{t+1-j}} \Phi^{u,t}_{x^u_{t+1-j}} P(x|x^u_{t+1-j})),$$

$$s.t. \quad 1 \le \Upsilon^{u,t}_x \le 2, 0 \le \Phi^{u,t}_x \le 1, 0 \le P(x|x^u_{t+1-j}) \le 1.$$

  - Parameters Θ^* could be learned via stochastic gradient descent method.
  - One-step transition probability can be directly computed.

Framework Specifications

- **One-Step Transition Probability**
  - Conditional probability of observing x_i *after* x_j.

$$P(x_i|x_j) = \frac{\sum_{u \in \mathcal{U}} \sum_{\mathcal{X}^{u,t} \in \mathcal{X}^u} \mathbb{1}_{\{x_j, x_i\} \subseteq \mathcal{X}^{u,t}}}{\sum_{u \in \mathcal{U}} \sum_{\mathcal{X}^{u,t} \in \mathcal{X}^u} \mathbb{1}_{x_j \in \mathcal{X}^{u,t}}}$$

Framework Specifications

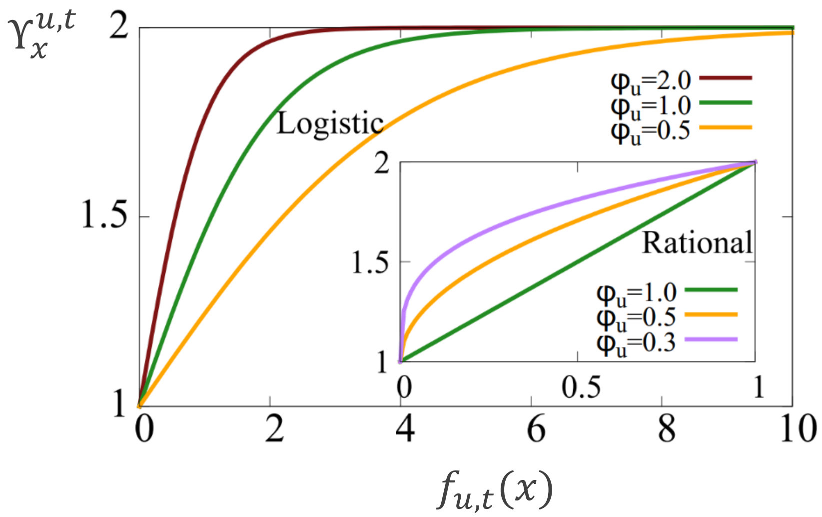
- **Starting Experience**
  - Logistic function

$$\Upsilon_x^{u,t} = \frac{2}{1+e^{-\phi_u f_{u,t}(x)}} \qquad (f_{u,t}(x) \geq 0, \phi_u \geq 0)$$

  - Rational function (normalized frequencies)

$$\Upsilon_x^{u,t} = 1 + (f_{u,t}(x))^{\phi_u} \quad (0 \leq f_{u,t}(x) \leq 1, 0 \leq \phi_u \leq 1)$$

Starting Experience: $\Upsilon_x^{u,t} \propto f(x,u,t)$, monotonically increasing with frequency.



Experience Curves

Starting experience measures the personalized accumulative interest a user has upon a certain item before forgetting.

Framework Specifications

- **Interest Retention**
  - Log-Linear function[Wright 1936]

    $$\Phi^{u,t}_{x^u_{t+1-j}} = C_u j^{-\alpha_u},$$

    $1 \le j \le t; 0 \le \alpha_u \le 1; 0 < C_u \le 1$
  - Exponential function[Knecht 1974]

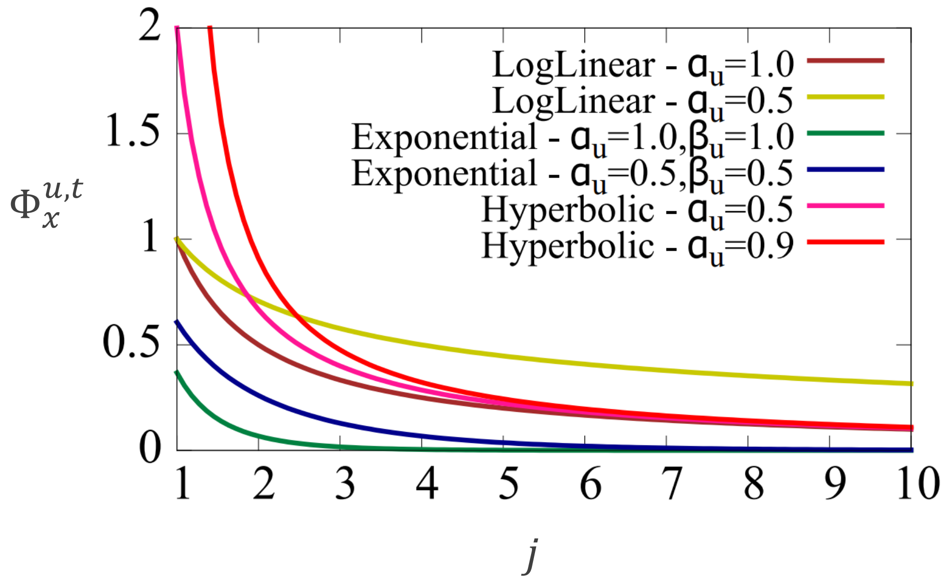
    $$\Phi^{u,t}_{x^u_{t+1-j}} = C_u j^{-\alpha_u} e^{-\beta_u j},$$

    $1 \le j \le t; 0 \le \alpha_u, \beta_u \le 1; 0 < C_u \le 1$
  - Hypobolic function[Mazur and Hastie 1978]

    $$\Phi^{u,t}_{x^u_{t+1-j}} = \frac{C_u}{j - \alpha_u}, 0 \le \alpha_u < 1, 0 < C_u \le 1$$

Interest retention measures the personalized residual interest of a user upon a certain item after forgetting.

Interest Retention: $\Phi^{u,t}_x \propto 1/j$, monotonically decreasing with elapsed time.



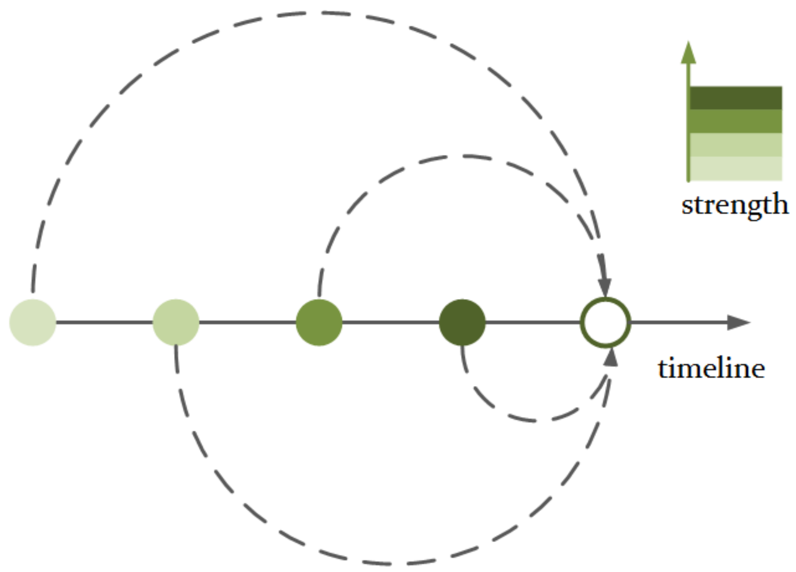
Interest Retention Curves

Personalized Recommendation

- **IFMM Framework**

$$P(x|\mathcal{X}^{u,t}) = \sum_{j=1}^{t} \Upsilon^{u,t}_{x^u_{t+1-j}} \Phi^{u,t}_{x^u_{t+1-j}} P\left(x \middle| x^u_{t+1-j}\right)$$

  - Forgetting speeds
  - Starting experience
  - Re-learning/Reconsumption



Top-N item recommendation with the largest values of $P(x|\mathcal{X}^{u,t})$.

Experiments

- **Data Set**
  - **Last.fm** music listening data set.
  - 992 users, 964,464 songs, 16,986,614 listening records.
  - Partition each user's listening history with a time shreshold, e.g. 1 hour.
  - Remove listening records whose duration is less than 30 secs.
  - 80% traces for training, 20% traces for test.

- **Comparative Methods**
  - Markov model based
    - Factorizing Personalized Markov Chain (FPMC) [Rendle 2010, Cheng 2013]
    - Topic Sensitive PageRank (TSPR) [Haveliwala 2002]
  - Graph-based preference fusion (STG) [Xiang 2010]
  - Sequential pattern based (SEQ) [Hariri et al 2012]

Experiments

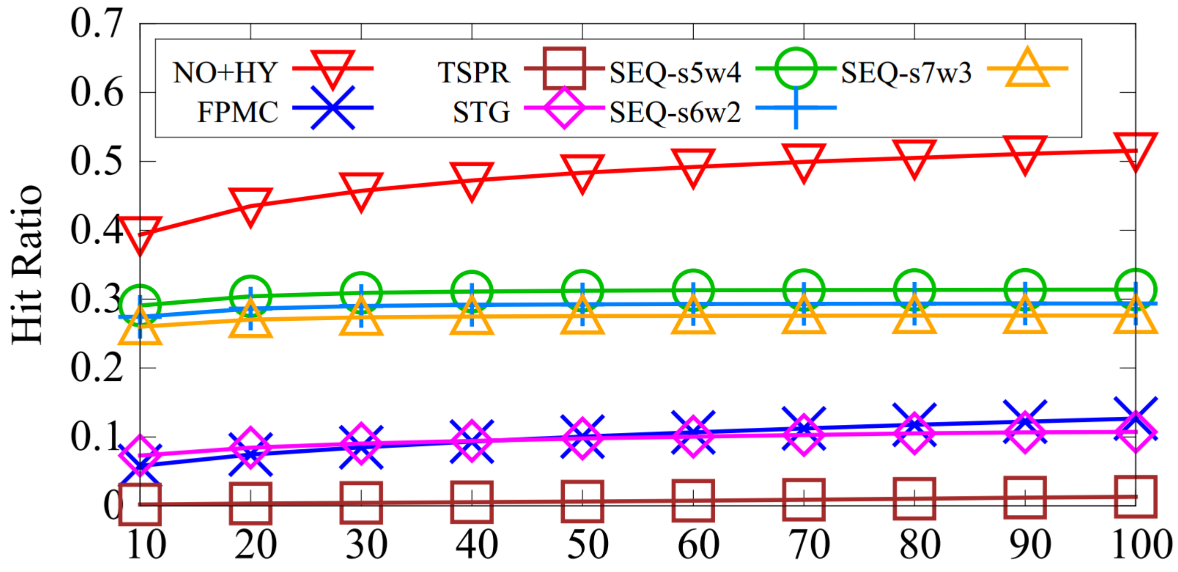
- **Accuracy of the proposed methods**
  - Starting Experience
    - NM: rational function
    - NO: logistic function
  - Interest Retention
    - LL: log-linear
    - EX: exponential
    - HY: hypobolic

Method	Top10	Top30	Top50	Top70	Top90	Top100
NM+LL	0.3288	0.4163	0.4536	0.4756	0.4919	0.4983
NM+EX	0.3335	0.4189	0.4553	0.4766	0.4924	0.4983
NM+HY	**0.3954**	**0.4573**	0.4816	0.4962	0.5068	0.5113
NO+LL	0.3289	0.4164	0.4537	0.4757	0.4921	0.4984
NO+EX	0.3345	0.4198	0.4557	0.4769	0.4927	0.4987
NO+HY	0.3935	**0.4573**	**0.4834**	**0.4991**	**0.5109**	**0.5154**

NO+HY performs the best, and is selected as the representative.

Experiments

- **Accuracy Comparisons**
  - NO+HY
  - SEQ
    - -s5w4: sup 5, winsize 4
    - -s7w3: sup 7, winsize 3
    - -s6w2: sup 6, winsize 2
  - FPMC
  - STG
  - TSPR



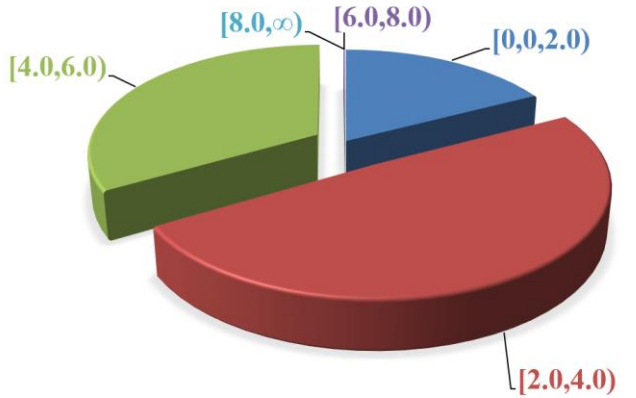
NO+HY improves 10%-20% in recommendation accuracy compared with the best of the reference methods.

Experiments

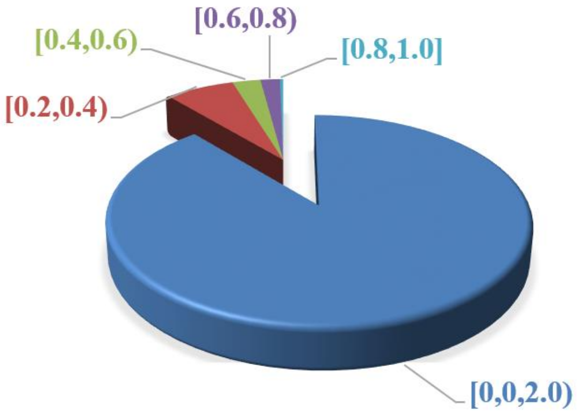
- **Personalized parameters distribution**
  - NO+HY

$$\Upsilon_x^{u,t} = \frac{2}{1+e^{-\phi_u f_{u,t}(x)}} \qquad (f_{u,t}(x) \geq 0, \phi_u \geq 0)$$

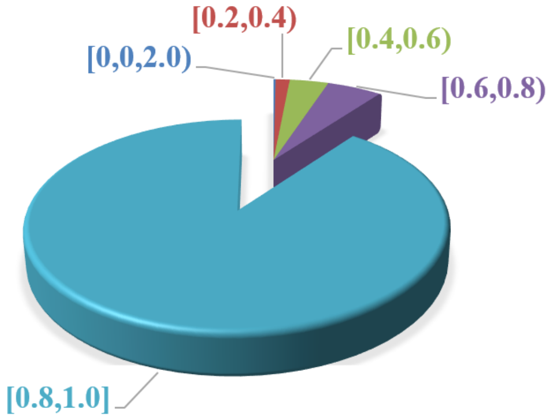
$$\Phi_{x_{t+1-j}^u}^{u,t} = \frac{C_u}{j-\alpha_u}, 0 \leq \alpha_u < 1, 0 < C_u \leq 1$$



ϕ_u



α_u



C_u

Conclusions

- **Forgetting** is an intrinsic feature of human beings, and should be taken into account in recommender systems.
- We proposed **λ-VOM** to simplify the computation of variable-order Markov model.
- We brought forward a **personalized framework** which integrates interest-forgetting and Markov model.
- Multiple **forgetting curve models** and **experience models** have been evaluated under our framework to find an optimal solution.
- IFMM provides various strategies for **personalization**.
- The experimental results proved the **effectiveness** of our method in recommendation tasks.

Thank You ~
Any Question?

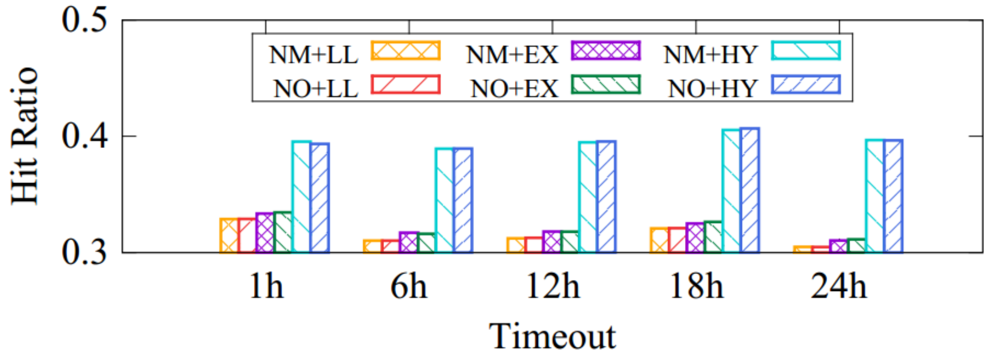
Jun Chen, Chaokun Wang, Jianmin Wang

Tsinghua University, China
chenjun14@mails.tsinghua.edu.cn, {chaokun, jimwang}@tsinghua.edu.cn

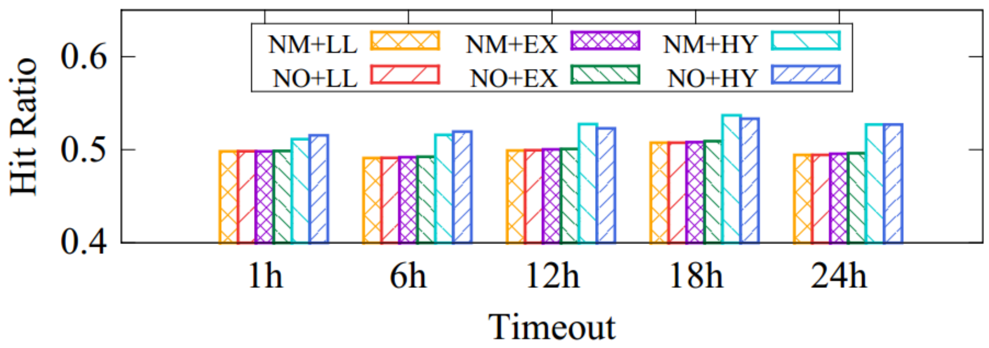
Experiments

- **Timeout Threshold**
  - Influence general length of traces.
  - Larger value, longer traces.

Very slight impact upon the recommendation accuracy



(a) Top-10 recommendations.



(b) Top-100 recommendations.