School of
Information Systems

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Modeling Sequential Preferences with Dynamic User and Context Factors

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

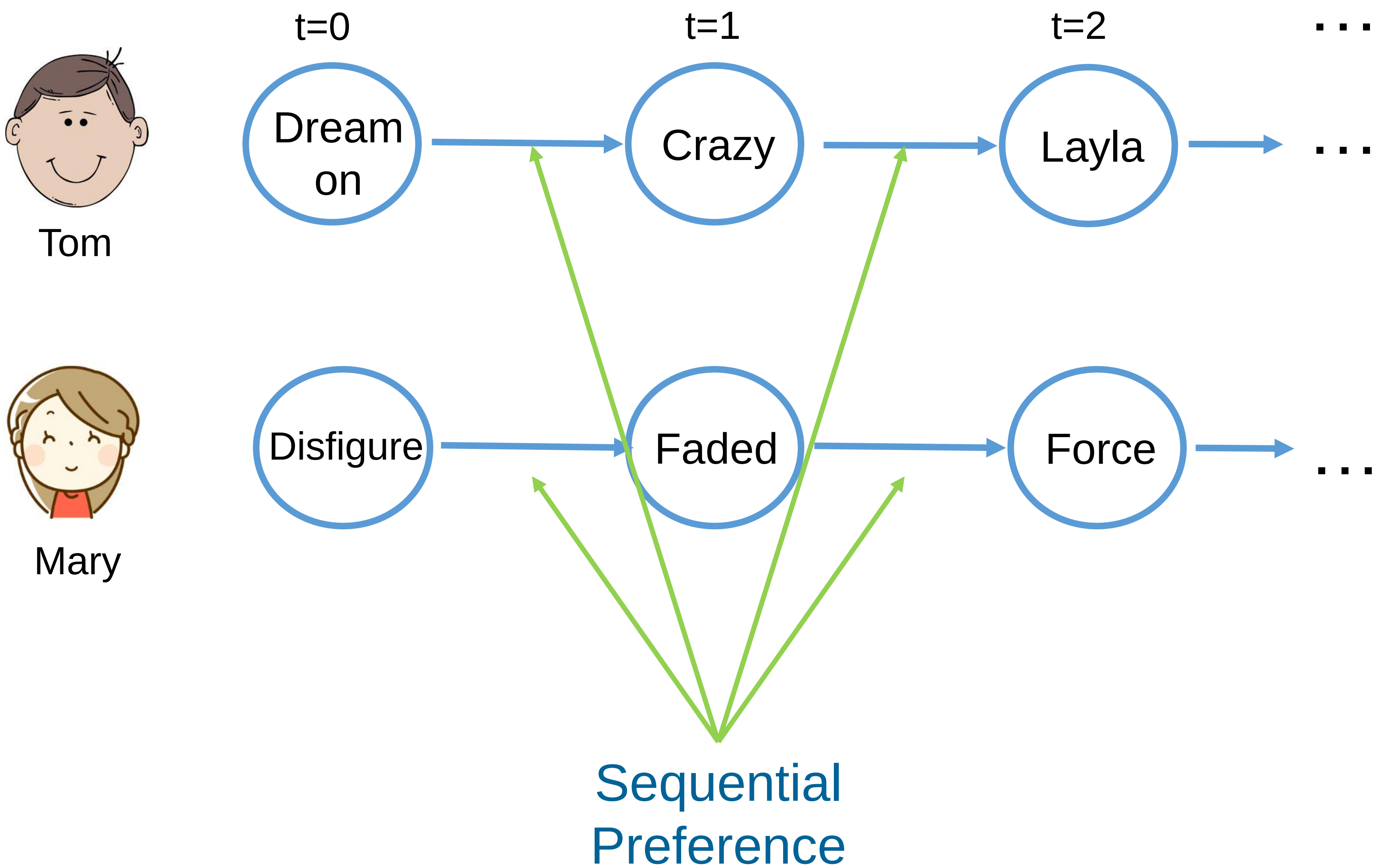
ECML-PKDD 2016

Riva Del Garda, Italy

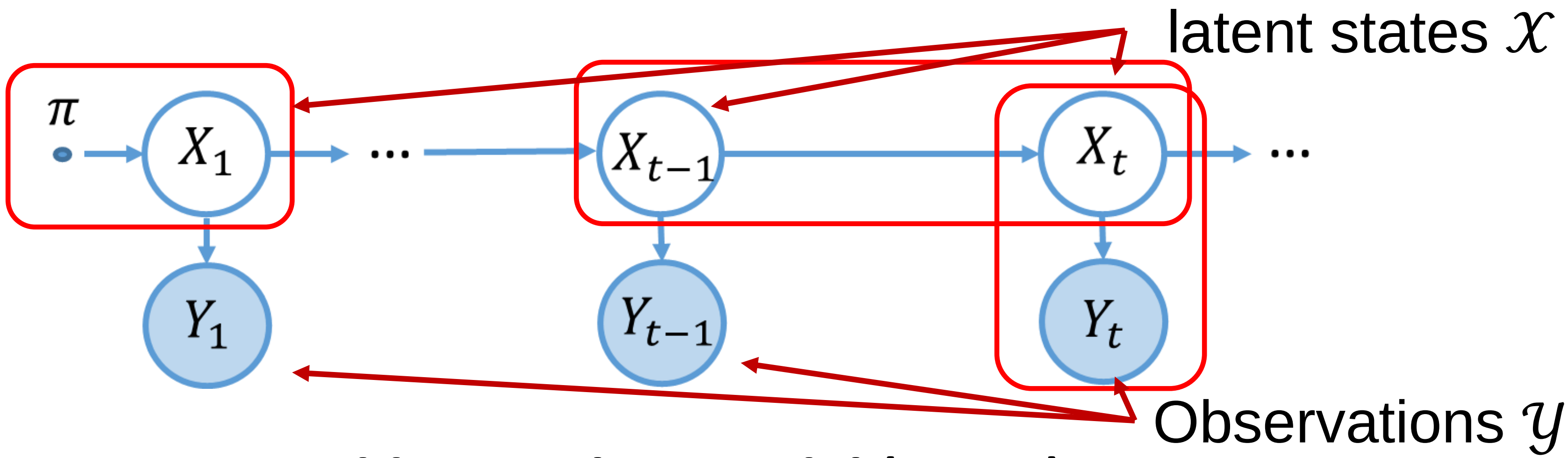
Outline

- **Motivating examples and Models**
  - Modeling sequential preferences (HMM)
  - Modeling Dynamic User-Bias Emissions (SEQ-E)
  - Modeling Dynamic Context-Biased Transitions (SEQ-T)
  - Joint Model (SEQ*)
- **Experiments**
  - Real-life Datasets: Twitter & Yes.com
  - Synthetic Dataset

The notion of Sequence – Song playlists



Modeling Sequential Preferences



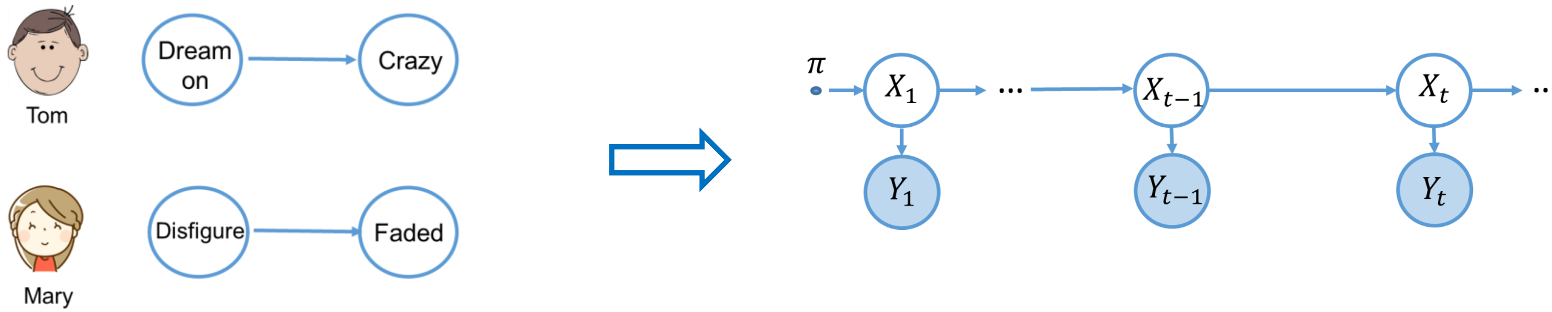
Hidden Markov Model (HMM)

HMM-Formulation: $\theta = (\pi, A, B)$

- π is the initial state distribution: $\pi_x \triangleq P(X_1 = x)$;
- A is the transition matrix: $A_{xu} \triangleq P(X_t = u \mid X_{t-1} = x)$;
- B is the emission matrix: $B_{xy} = P(Y_t = y \mid X_t = x)$;

$$\forall\, x, u \in \mathcal{X};\, y \in \mathcal{Y};\; t \in \{1, 2, \ldots\}$$

Modeling Sequential Preferences



HMM-Example: A HMM model with 2 latent states, 4 items

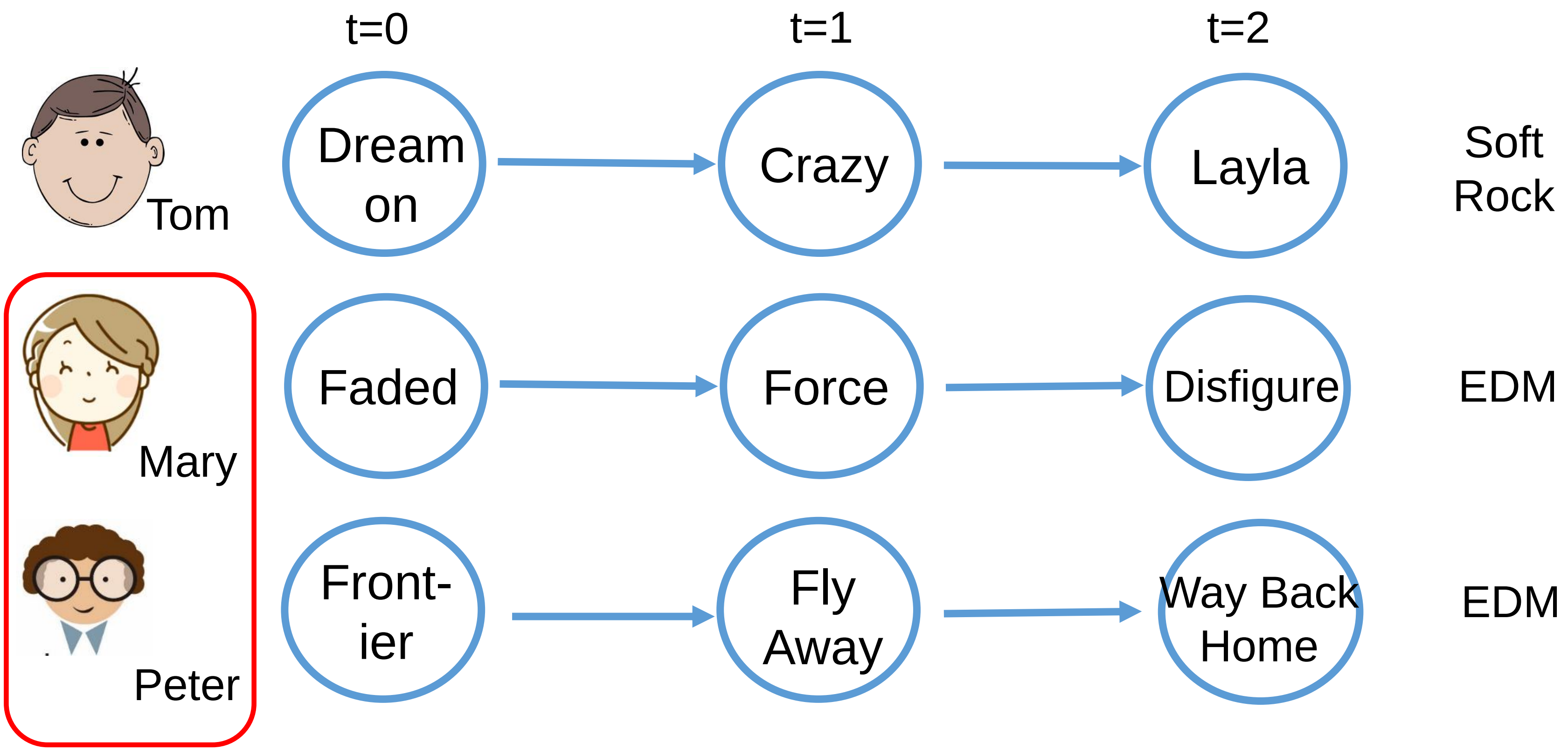
- $\pi = [\pi_0, \pi_1] = [0.8, 0.2]$ $\qquad \pi_0 = P(X_1 = 0) = 0.8$
- $A = \begin{bmatrix} A_{00} & A_{01} \\ A_{10} & A_{11} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}$

  $A_{00} = P(X_t = 0 \mid X_{t-1} = 0) = 0.7$
- $B = \begin{bmatrix} B_{00} & B_{01} & B_{02} & B_{03} \\ B_{10} & B_{11} & B_{12} & B_{13} \end{bmatrix} = \begin{bmatrix} 0.6 & 0.1 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.1 & 0.1 \end{bmatrix}$

  $B_{00} = P(Y_t = \text{"Dream on"} \mid X_t = 0) = 0.6$

  $B_{10} = P(Y_t = \text{"Dream on"} \mid X_t = 1) = 0.3$

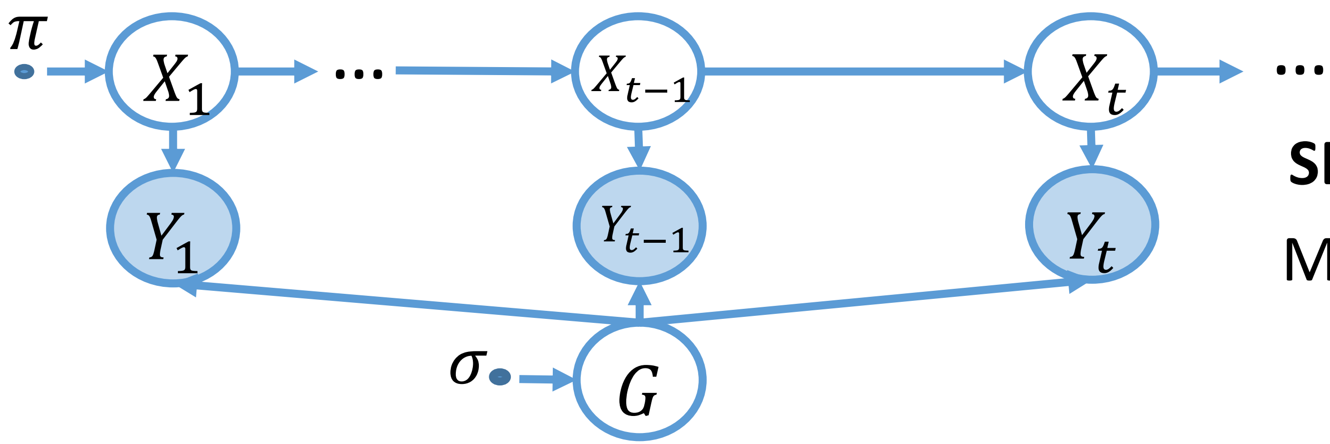
The notion of Group



Hypotheses: There exists **different groups** of users

- Users in the **same group** share the **same emission probabilities**
- Users **across groups** may have **different emission probabilities**.

Modeling Dynamic User-Bias Emissions



SEQ-E
Model

Formulation: $\theta = (\pi, \sigma, A, B)$ with **a set of groups** $\mathcal{G}$

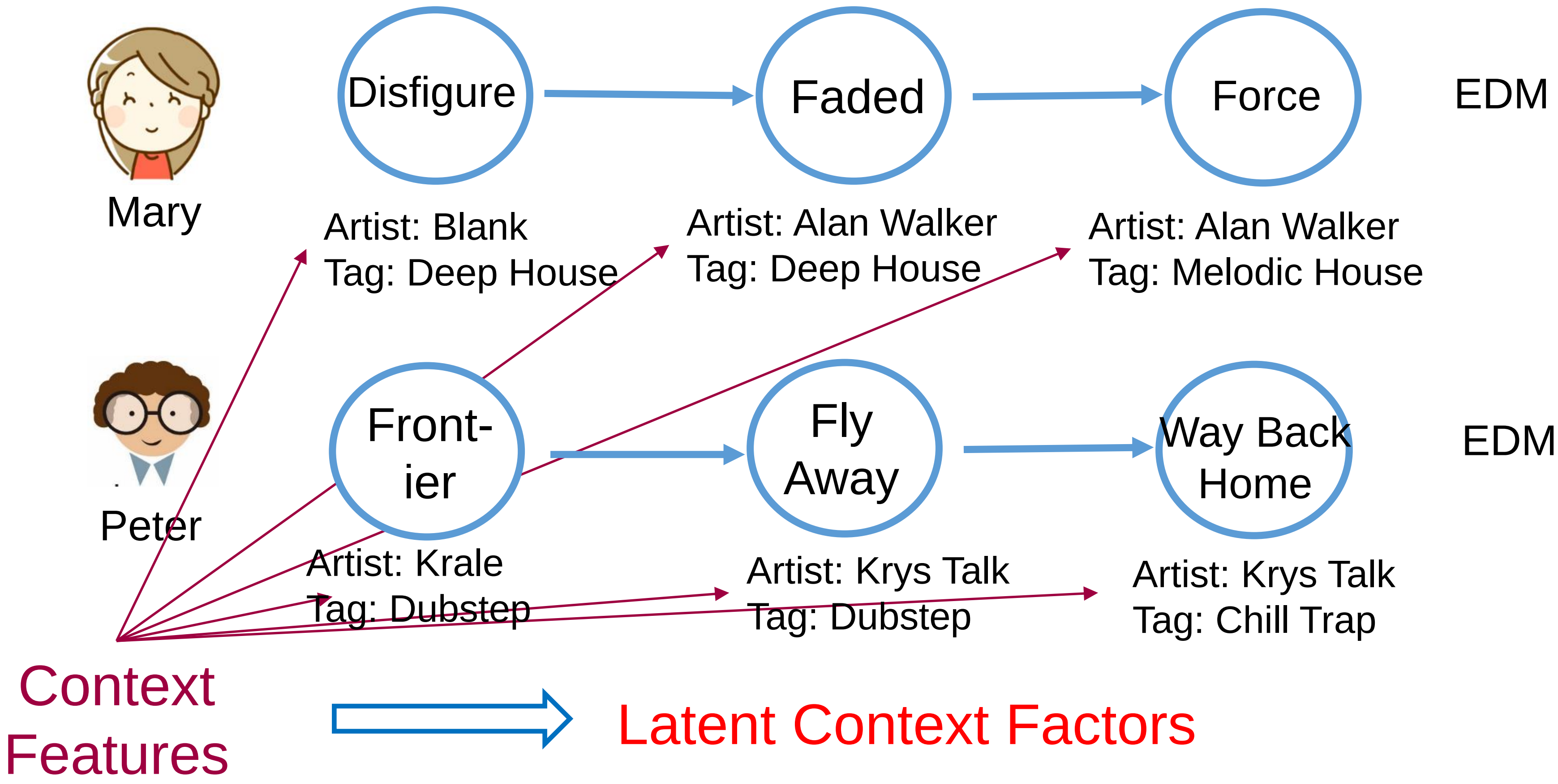
- σ is the **group distribution**: $\sigma_g \triangleq P(G = g)$
- B is the **new emission tensor**: $B_{gxy} \triangleq P(Y_t = y \mid X_t = x, \boldsymbol{G = g})$

$$\forall\, x, u \in \mathcal{X}; y \in \mathcal{Y}; g \in \mathcal{G};\ t \in \{1, 2, \ldots.\}$$

- Example: $B_{000} = P(Y_t = \text{"Dream on"} \mid X_t = 0, \boldsymbol{G = 0})$ = 0.8

$B_{100} = P(Y_t = \text{"Dream on"} \mid X_t = 0, \boldsymbol{G = 1})$ = 0.3

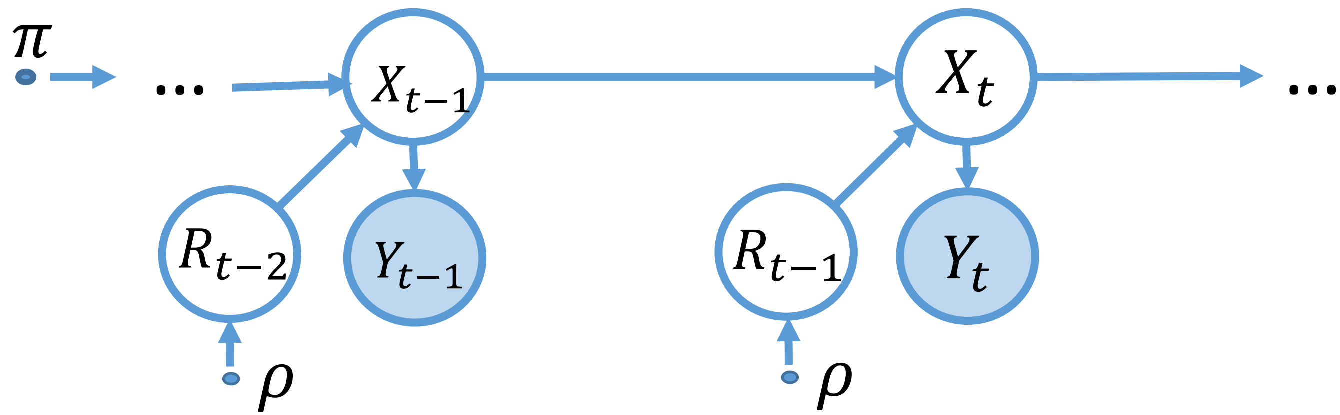
The notion of Context Features, Factors



Hypotheses: There exists **context features** and **factors**

- **Latent context factors** manifest through **context features**
- **Transitions** are **affected** by **latent context factors**.

Modeling Dynamic Context-Biased Transitions



SEQ-T

Model

Hypothesis:

Transitions are **affected** by **latent context factors**

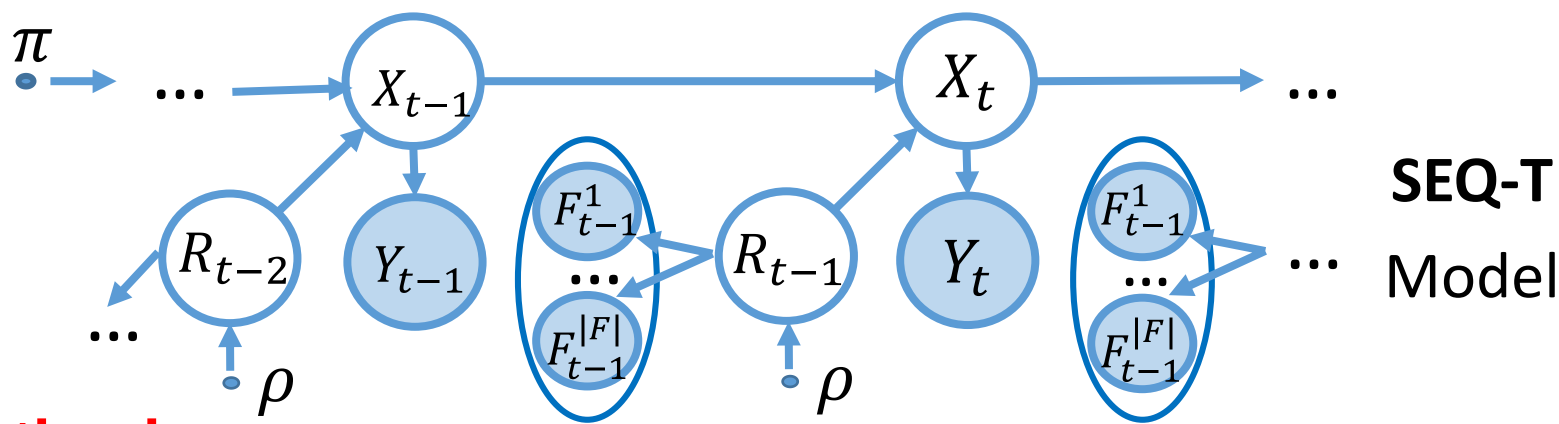
Formulation: $\theta = (\pi, \rho, A, B, C)$ with a context factors set $\mathcal{R}$;

- ρ is the distribution of the **latent context factor**: $\rho_r \triangleq P(R_t = r)$;
- A is the **new transition tensor**: $A_{rxu} \triangleq P(X_t = u | X_{t-1} = x, R_{t-1} = r)$;

  $\forall\, x, u \in \mathcal{X}; r \in \{1, \ldots, |\mathcal{R}|\}$;
- Examples: $A_{100} = P(X_t = 0 | X_{t-1} = 0, \boldsymbol{R_{t-1} = 1})$ = 0.9

  $A_{000} = P(X_t = 0 | X_{t-1} = 0, \boldsymbol{R_{t-1} = 0})$ = 0.4

Modeling Dynamic Context-Biased Transitions



Hypothesis:

Latent context factors manifest through **context features**

Formulation: $\theta = (\pi, \rho, A, B, C)$ with a context features set F = $\{F^1, F^2, \ldots\}$, each F^i takes a values set $\mathcal{F}^i$

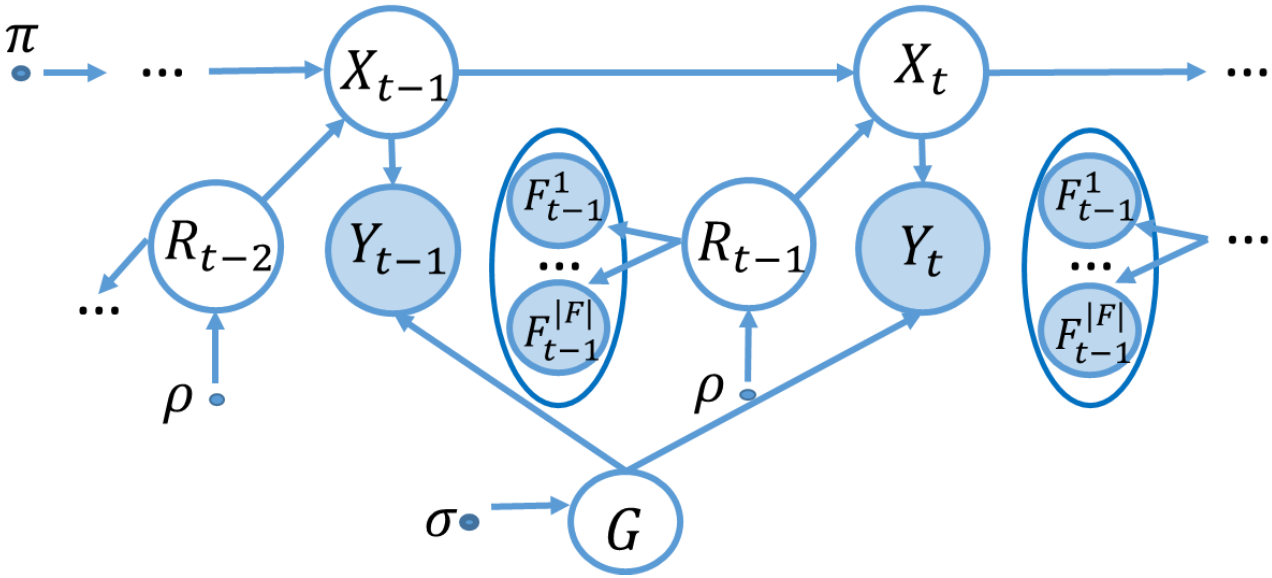
- C is the **feature probability matrix**: $C_{rif} \triangleq P\left(F_t^i = f \,\middle|\, R_t = r\right)$;

  $$r \in \{1, \ldots, |\mathcal{R}|\}; i \in \{1, \ldots, |F|\}; f \in \mathcal{F}^i; \; t \in \{1, 2, \ldots\}$$

- Examples: $C_{101} = P\left(\left(\boldsymbol{F_t^0 = 1}\,\middle|\, R_t = 1\right)\right.$ = 0.99

  $C_{100} = P\left(\left(\boldsymbol{F_t^0 = 0}\,\middle|\, R_t = 1\right)\right.$ = 0.01

Joint Model



SEQ*
Model

<u>**Main idea**</u>: Jointly capture **user** and **context factors** in a single model

<u>**Parameters**</u>: The **six-tuple** $\theta = (\pi, \sigma, \rho, A, B, C)$ as above

<u>**Prediction**</u>: $y^* = \text{argmax}_y P(Y_{T+1} = y | Y_1, .., Y_T, F_1, \ldots, F_T; \theta^*)$

Outline

- **Motivating examples and Models**
  - Modeling sequential preferences (HMM)
  - Modeling Dynamic User-Bias Emissions (SEQ-E)
  - Modeling Dynamic Context-Biased Transitions (SEQ-T)
  - Joint Model (SEQ*)
- **Experiments**
  - Real-life Datasets: Twitter & Yes.com
  - Synthetic Dataset

Experimental Setup – Real-life Datasets

- **Research Question:** Do the latent user and context factors result in significant improvements over HMM ?
- **Datasets**:

Dataset	#Observation	#Sequence	Average Length
Song playlists (Yes.com)	3168	250k	7
Hashtag Sequences (Twitter)	2121	114k	19

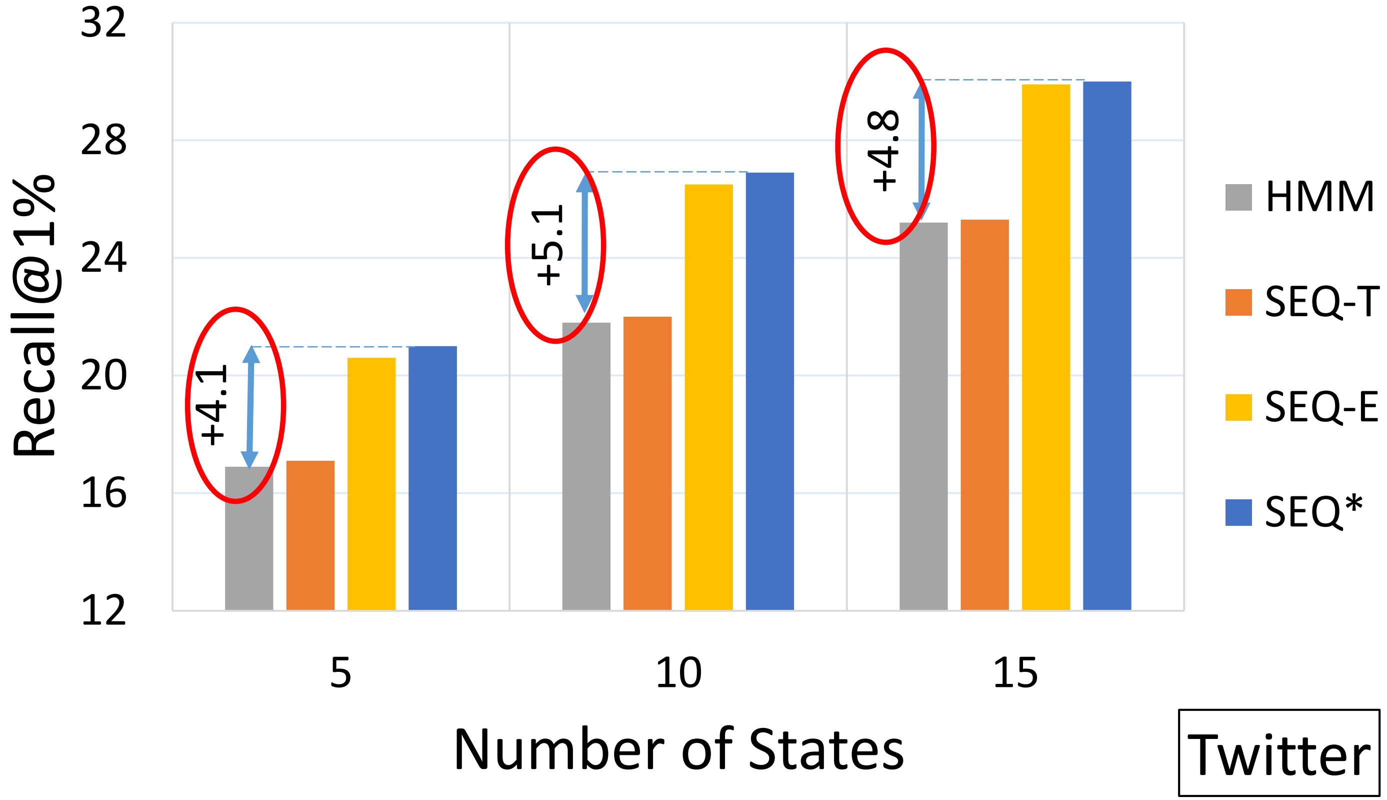
- **Features:**
  - Categories of tags (Yes.com)
  - Tweet information (Twitter): #Retweet, Created Time, etc.

Experimental Setup – Real-life Datasets

- **Task:** Last item prediction
  - For each testing sequence $s = \{Y_1, Y_2, \dots, Y_{T-1}, Y_T\} \quad T \geq 2$
  - Save Y_T as ground-truth target. Predict the last item given the previous items $P\left(Y_{\text{candidate}} \middle| Y_1, Y_2, \dots, Y_{T-1}\right)$
- **Evaluation Metrics**
  - $Recall@K = \frac{\#\,\text{sequences with ground truth in top } K}{\#\,\text{sequences in the testing set}}$

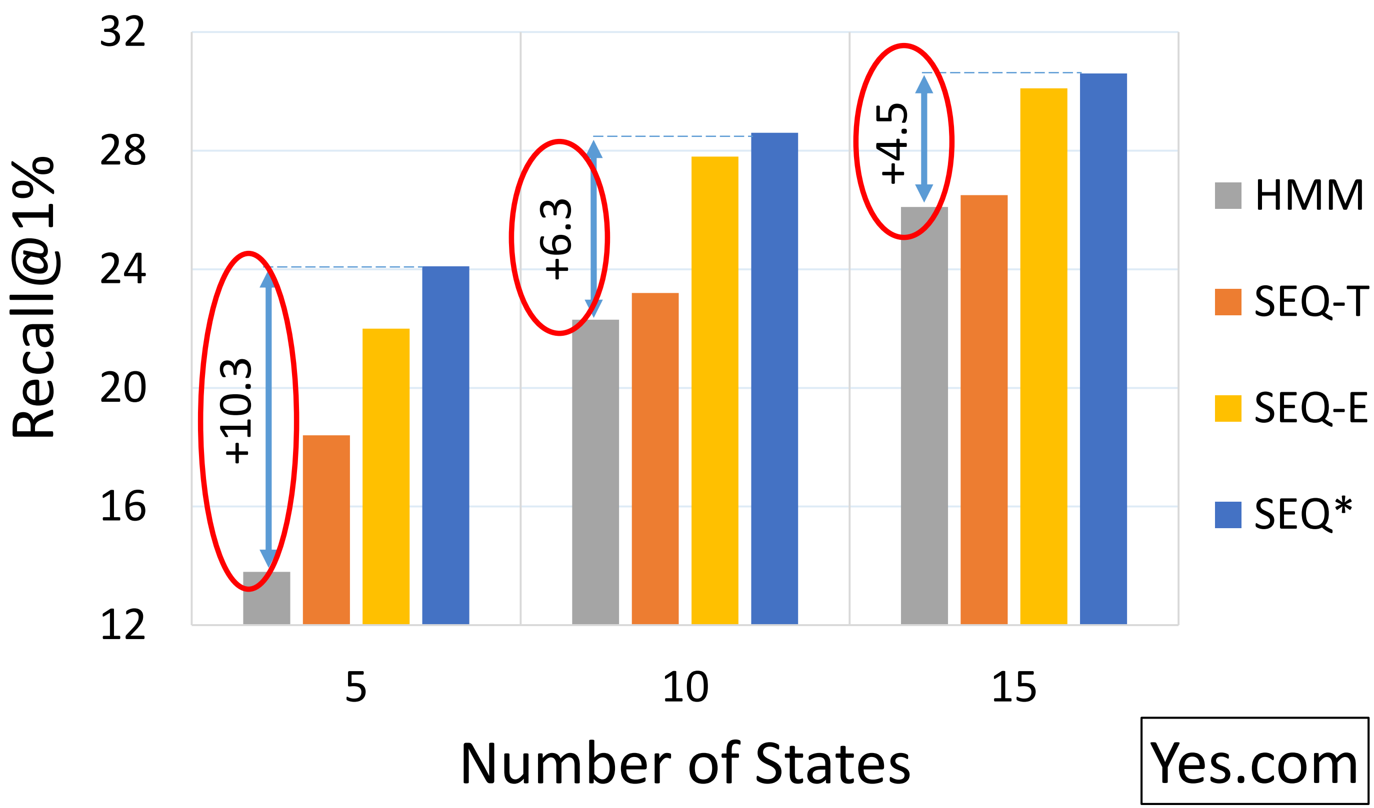
    Example:
    - Given a sequence $s = \{i_{10}, i_2, i_5, i_8\}$;
      $P\left(i_{\text{candidate}} \middle| i_{10}, i_2, i_5\right) \Rightarrow \text{rank}_{\text{i}_8} = 9$
    - $S_{\text{test}} = \{s_1, s_2, s_3\}$ with respective ranks of actual items are 3, 6, 11. Recall@5 = 1/3; Recall@10 = 2/3
  - *Mean Reciprocal Rank (MRR)*

Result – Twitter - Recall@1%



#Group $|\mathcal{G}| = 2$, #Context Factor Level $|\mathcal{R}| = 2$, #Feature $|F| = 7$

Result – Yes.com - Recall@1%



#Group $|\mathcal{G}| = 2$, #Context Factor Level $|\mathcal{R}| = 2$, #Feature $|F| = 11$

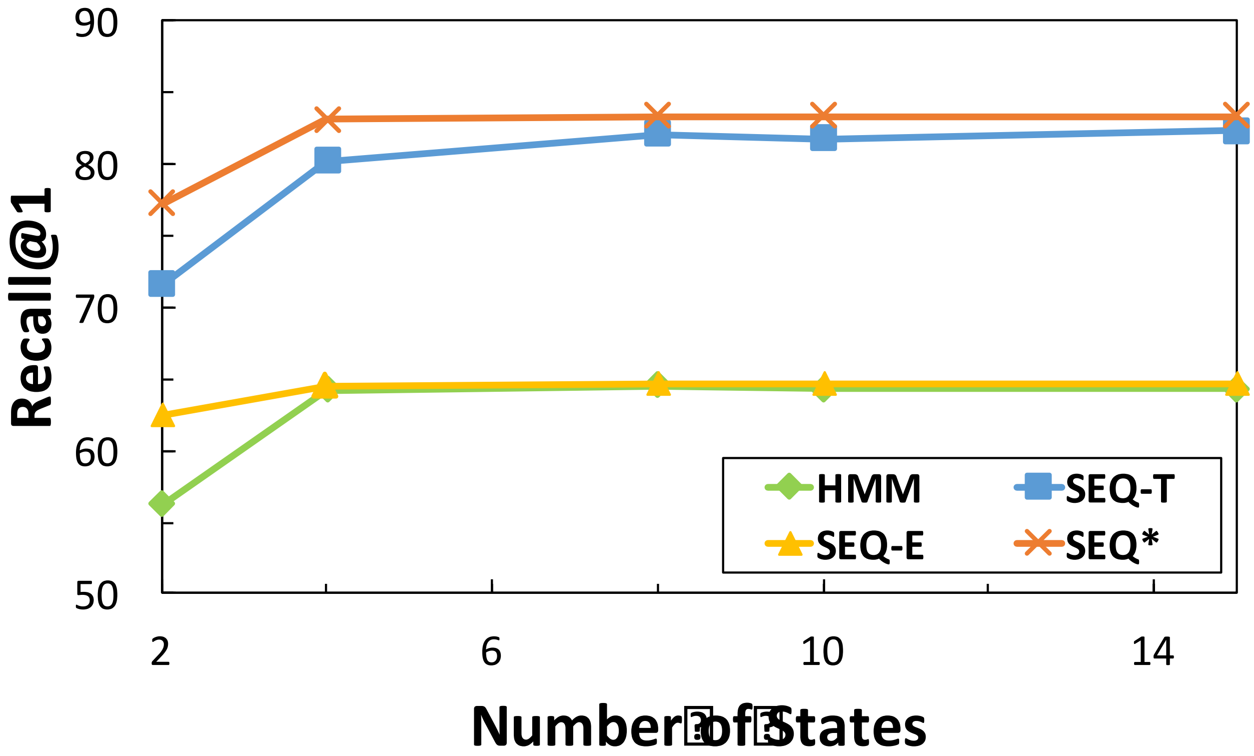
Experimental Setup – Synthetic Dataset

- **Research Questions:**
  - Can the implementation recover parameters from the synthetic dataset?
  - Could the effect of latent user and context factors be simulated by increasing the number of HMM's states?
- **Generative process:**

Dataset	#Observation	#Sequence	Average Length
Synthetic	4	10k	10

- **Task:** Last item prediction
- **Evaluation metrics:** Recall@1, MRR

Result - Synthetic Dataset – Recall@1



#Group $|\mathcal{G}| = 2$, #Context Factor Level $|\mathcal{R}| = 2$, #Feature $|F| = 4$

Conclusion

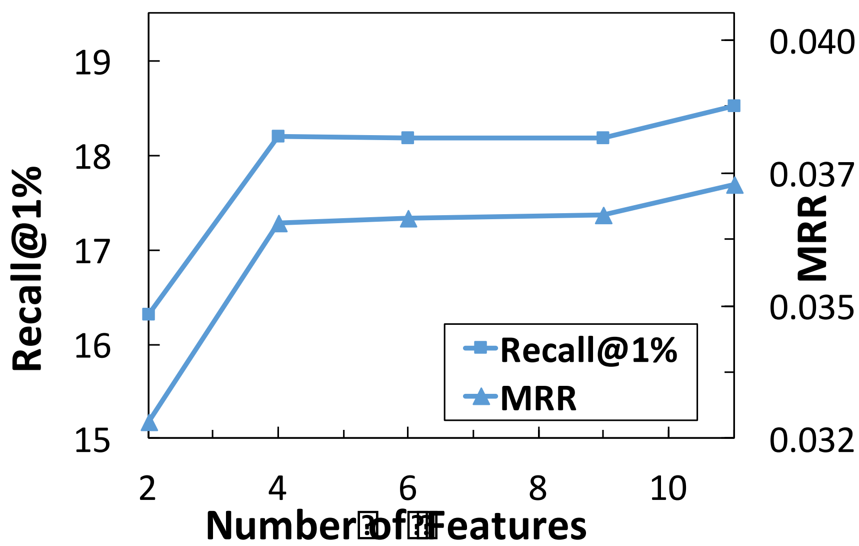
- Introduce and model **dynamic user and context factors** to capture sequential preferences.
- The proposed model contributes **statistically significant** improvement as compared to the baseline HMM in term of top-K recommendations.

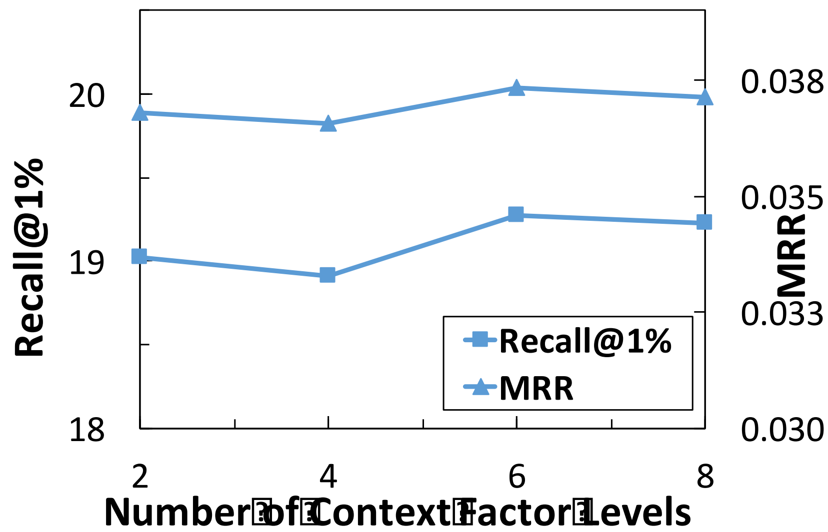
Thank you!
Q&A

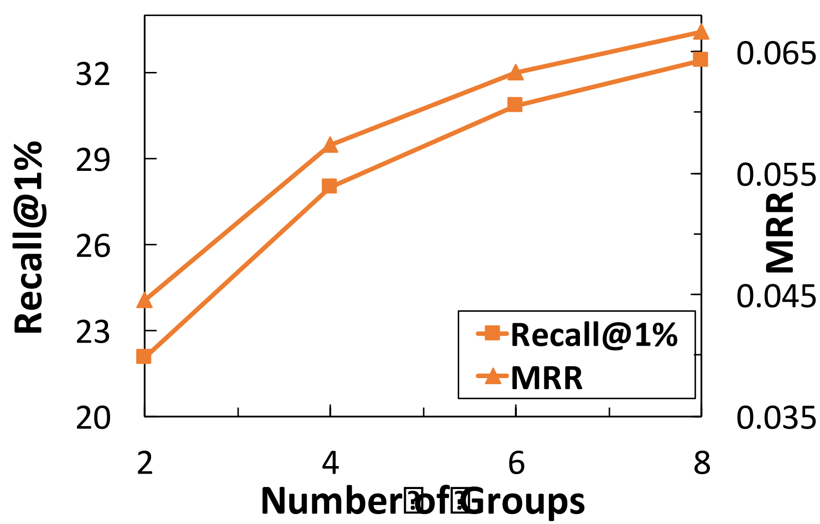
Any further questions, please contact us:
hadywlauw@smu.edu.sg
yfang@i2r.a-star.edu.sg
ductrong.le.2014@smu.edu.sg

Backup Slides

Result – Yes.com – Tuning Parameters







Synthetic Dataset – Generative Process

#Group $\lvert\mathcal{G}\rvert$	2	**#Context Factor Level** $\lvert\mathcal{R}\rvert$	2
#States $\lvert\mathcal{X}\rvert$	2	**#Feature** $\lvert F\rvert$	4
#Observation $\lvert\mathcal{Y}\rvert$	4	**#Feature values** $\lvert\mathcal{F}\rvert$	2

- Initial Probability: $\pi = \{0.8, 0.2\}$
- Latent Context Factor Distribution: $\rho = \{0.3, 0.7\}$
- Group Distribution: $\sigma = \{0.9, 0.1\}$
- Transition Tensor A:
  - The first context factor favors self-transition to the same state
  - The second context factor encourages the state switching.

$$A = [A_0, A_1]; A_1 = \begin{bmatrix} 0.01 & 0.99 \\ 0.70 & 0.30 \end{bmatrix}; A_0 = \begin{bmatrix} 0.99 & 0.01 \\ 0.30 & 0.70 \end{bmatrix}$$

Result – Twitter – Recall@K & MRR

Table 3. Performance of comparative methods on Twitter.com for *Recall@K*

		FREQ	HMM	SEQ-T	SEQ-E	SEQ*	Imp.
5 States	Recall@1%	8.4	16.9	$17.1^{\dagger}$	$20.6^{\S}$	$21.0^{\dagger\S}$	+4.1
	Recall@50	16.1	28.3	$28.6^{\dagger}$	$33.2^{\S}$	$33.7^{\dagger\S}$	+5.4
	Recall@100	25.5	40.6	$40.9^{\dagger}$	$46.0^{\S}$	$46.5^{\dagger\S}$	+5.9
10 States	Recall@1%	8.4	21.8	$22.0^{\dagger}$	$26.5^{\S}$	$26.9^{\dagger\S}$	+5.1
	Recall@50	16.1	34.2	$34.4^{\dagger}$	$39.4^{\S}$	$39.8^{\dagger\S}$	+5.7
	Recall@100	25.5	47.2	$47.4^{\dagger}$	$52.0^{\S}$	$52.4^{\S}$	+5.2
15 States	Recall@1%	8.4	25.2	$25.3^{\dagger}$	$29.9^{\S}$	$30.0^{\dagger\S}$	+4.8
	Recall@50	16.1	38.1	$38.2^{\dagger}$	$43.1^{\S}$	$43.3^{\dagger\S}$	+5.1
	Recall@100	25.5	51.2	$51.3^{\dagger}$	$55.2^{\S}$	$55.3^{\dagger\S}$	+4.1

Table 4. Performance of comparative methods on Twitter.com for *MRR*

	FREQ	HMM	SEQ-T	SEQ-E	SEQ*	Imp.
5 States	0.019	0.045	$0.046^{\dagger}$	$0.062^{\S}$	$0.063^{\dagger\S}$	+0.0183
10 States	0.019	0.063	0.064	$0.084^{\S}$	$0.086^{\dagger\S}$	+0.0227
15 States	0.019	0.076	$0.078^{\dagger}$	$0.100^{\S}$	$0.101^{\dagger\S}$	+0.0246

#Group $|\mathcal{G}|$ = 2, #Context Factor Level $|\mathcal{R}|$ = 2, #Feature $|F|$ = 7

Result – Yes.com – Recall@K & MRR

Table 1. Performance of comparative methods on Yes.com for *Recall@K*

		FREQ	HMM	SEQ-T	SEQ-E	SEQ*	Imp.
5 States	Recall@1%	6.8	13.8	18.4†	22.0§	24.1†§	+10.3
	Recall@50	9.6	19.2	25.1†	29.5§	32.1†§	+13.0
	Recall@100	16.2	29.3	37.0†	42.6§	46.1†§	+16.8
10 States	Recall@1%	6.8	22.3	23.2†	27.8§	28.6†§	+6.3
	Recall@50	9.6	30.0	31.1†	36.9§	38.1†§	+8.1
	Recall@100	16.2	43.4	44.9†	52.1§	53.5†§	+10.2
15 States	Recall@1%	6.8	26.1	26.5†	30.1§	30.6†§	+4.5
	Recall@50	9.6	34.7	35.5†	39.4§	40.2†§	+5.5
	Recall@100	16.2	49.3	50.8†	55.1§	56.3†§	+7.0

Table 2. Performance of comparative methods on Yes.com for *MRR*

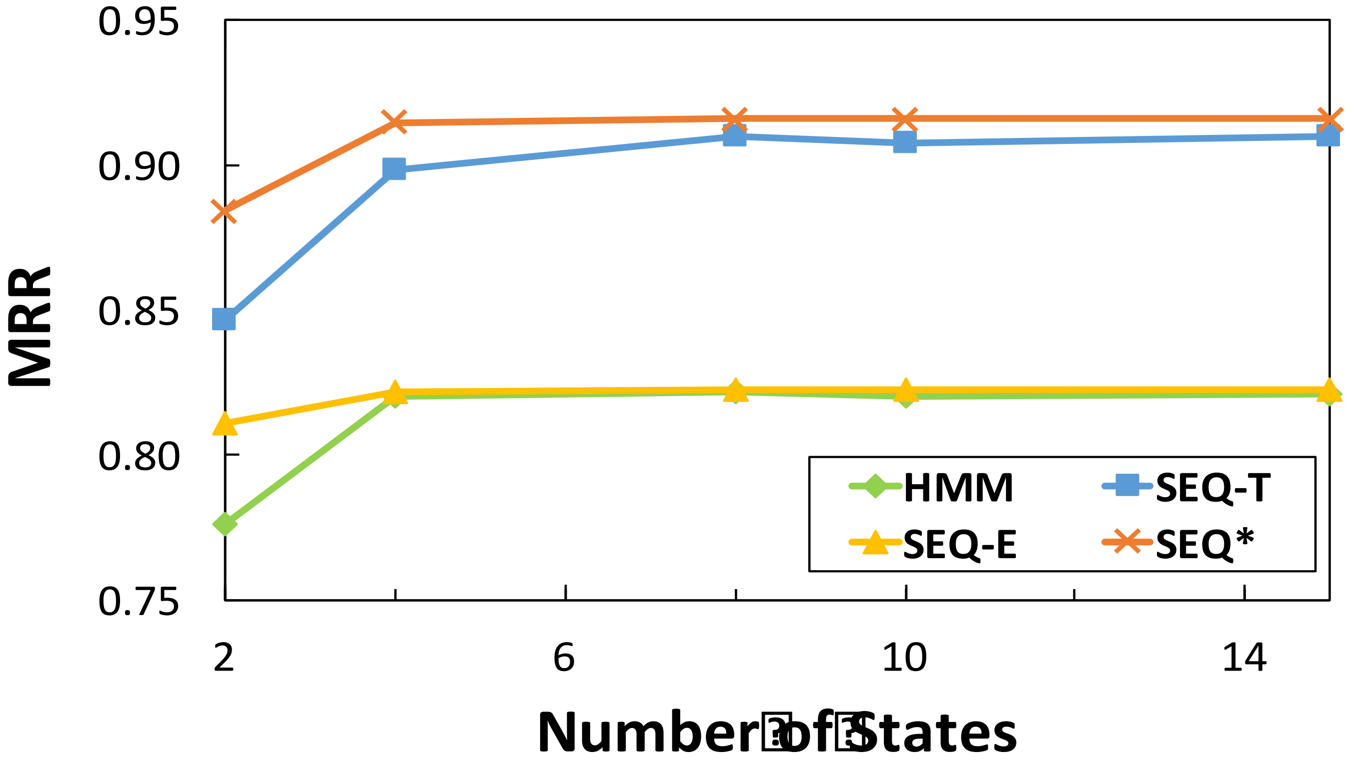
	FREQ	HMM	SEQ-T	SEQ-E	SEQ*	Imp.
5 States	0.014	0.028	0.037†	0.044§	0.049†§	+0.021
10 States	0.014	0.045	0.047†	0.057§	0.059†§	+0.014
15 States	0.014	0.053	0.054†	0.062§	0.063§	+0.009

School of
Information Systems

#Group $|\mathcal{G}|$ = 2, #Context Factor Level $|\mathcal{R}|$ = 2, #Feature $|F|$ = 11

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Result - Synthetic Dataset – MRR



#Group $|\mathcal{G}|$ = 2, #Context Factor Level $|\mathcal{R}|$ = 2, #Feature $|F|$ = 4

Synthetic Dataset – Generative Process

- Emission Tensor B:
  - Each pair of (state, group) favors one of the four items

$$B = [B_0, B_1]; B_0 = \begin{bmatrix} 0.991 & 0.003 \\ 0.003 & 0.003 \\ 0.003 & 0.991 \\ 0.003 & 0.003 \end{bmatrix}; B_1 = \begin{bmatrix} 0.003 & 0.003 \\ 0.991 & 0.003 \\ 0.003 & 0.003 \\ 0.003 & 0.991 \end{bmatrix}$$

- Feature matrix C:
  - Each context factor level is associated with 2 of the 4 features.

$$C = [C_0, C_1]; C_0 = \begin{bmatrix} 0.10 & 0.90 \\ 0.20 & 0.80 \\ 0.90 & 0.10 \\ 0.90 & 0.10 \end{bmatrix}; C_1 = \begin{bmatrix} 0.90 & 0.10 \\ 0.90 & 0.10 \\ 0.10 & 0.90 \\ 0.30 & 0.70 \end{bmatrix}$$