Modeling Sequential Preferences with Dynamic User and Context Factors (Supplementary Material)

Duc-Trong Le[1], Yuan Fang[2], and Hady W. Lauw[1]

[1] School of Information Systems, Singapore Management University, Singapore
ductrong.le.2014@phdis.smu.edu.sg,hadywlauw@smu.edu.sg
[2] Institute for Infocomm Research, Singapore
yfang@i2r.a-star.edu.sg

1 Synthetic Dataset

Here, we first describe the original parameters for the synthetic data generation, and then show the parameters learned by each method.

Original Parameter Values. To generate a synthetic dataset, we construct a simple model with 2 latent groups ($|\mathcal{G}| = 2$), 2 latent states ($|\mathcal{X}| = 2$), 2 levels of context factor ($|\mathcal{R}| = 2$), 4 items ($|\mathcal{Y}| = 4$), 4 features ($|F| = 4$) each with 2 binary feature values.

The six-tuple parameter $\theta = (\pi, \sigma, \rho, A, B, C)$ is specified such that:

- Each (group, state) combination predominately generates 1 of the 4 items.
- Each context factor level is characterized by a pair of context features.
- One context factor level predominately supports self-transition to the same state. The other level predominately supports switching to the other state.

The original parameter values used during synthetic data generation are shown in the *Original Values* column of Table 1.

Learned Parameter Values. We run each comparative model on the generated synthetic dataset with 10 thousand sequences, each of length 10. Table 1 also shows the parameter values learned by each model.

HMM**:** It seems to take the advantage of the grouping probability to create hidden states and aggregate the emission probabilities of the two groups. Its transition probability also favors that of the *majority* context factor level.

SEQ-E**:** It learns the group distribution σ well, but the initial distribution of hidden states π is a bit off. This affects the emission probability. The transition probability amounts to an aggregation of the two context factor levels' transition matrices.

SEQ-T**:** It can recover most of the parameters, except the emission probability due to its not taking into account the user bias.

SEQ***:** Importantly, SEQ* is the only model that can virtually recover all of the original parameter values with very light noises.

Table 1. Synthetic Parameters: Original Values vs. Learned Values by Various Models

Parameter	Original Values	Learned Values			
		HMM	SEQ-E	SEQ-T	SEQ*
$\pi = [\pi_0, \pi_1]$	[0.80, 0.20]	[0.90, 0.10]	[0.75, 0.25]	[0.80, 0.20]	[0.80, 0.20]
$\sigma = [\sigma_0, \sigma_1]$	[0.90, 0.10]	N.A.	[0.90, 0.10]	N.A.	[0.90, 0.10]
$\rho = [\rho_0, \rho_1]$	[0.30, 0.70]	N.A.	N.A.	[0.30, 0.70]	[0.30, 0.70]
$A = [A_0, A_1]$ $A_0 = \begin{bmatrix} A_{000} & A_{001} \\ A_{010} & A_{011} \end{bmatrix}$ $A_1 = \begin{bmatrix} A_{100} & A_{101} \\ A_{110} & A_{111} \end{bmatrix}$	$A_0 = \begin{bmatrix} 0.010 & 0.990 \\ 0.700 & 0.300 \end{bmatrix}$ $A_1 = \begin{bmatrix} 0.990 & 0.010 \\ 0.300 & 0.700 \end{bmatrix}$	$A_0 = \begin{bmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{bmatrix}$	$A_0 = \begin{bmatrix} 0.670 & 0.330 \\ 0.280 & 0.720 \end{bmatrix}$	$A_0 = \begin{bmatrix} 0.008 & 0.992 \\ 0.703 & 0.297 \end{bmatrix}$ $A_1 = \begin{bmatrix} 0.994 & 0.006 \\ 0.292 & 0.708 \end{bmatrix}$	$A_0 = \begin{bmatrix} 0.004 & 0.996 \\ 0.696 & 0.304 \end{bmatrix}$ $A_1 = \begin{bmatrix} 0.993 & 0.007 \\ 0.293 & 0.707 \end{bmatrix}$
$B = [B_0, B_1]$ $B_0 = \begin{bmatrix} B_{000} & B_{001} \\ B_{002} & B_{003} \\ B_{010} & B_{011} \\ B_{012} & B_{013} \end{bmatrix}$ $B_1 = \begin{bmatrix} B_{100} & B_{101} \\ B_{102} & B_{103} \\ B_{110} & B_{111} \\ B_{112} & B_{113} \end{bmatrix}$	$B_0 = \begin{bmatrix} 0.991 & 0.003 \\ 0.003 & 0.003 \\ 0.003 & 0.991 \\ 0.003 & 0.003 \end{bmatrix}$ $B_1 = \begin{bmatrix} 0.003 & 0.003 \\ 0.991 & 0.003 \\ 0.003 & 0.003 \\ 0.003 & 0.991 \end{bmatrix}$	$B_0 = \begin{bmatrix} 0.605 & 0.389 \\ 0.003 & 0.003 \\ 0.002 & 0.003 \\ 0.605 & 0.390 \end{bmatrix}$	$B_0 = \begin{bmatrix} 0.989 & 0.006 \\ 0.003 & 0.002 \\ 0.202 & 0.790 \\ 0.005 & 0.003 \end{bmatrix}$ $B_1 = \begin{bmatrix} 0.002 & 0.002 \\ 0.980 & 0.018 \\ 0.003 & 0.005 \\ 0.028 & 0.784 \end{bmatrix}$	$B_0 = \begin{bmatrix} 0.891 & 0.004 \\ 0.101 & 0.004 \\ 0.008 & 0.890 \\ 0.003 & 0.099 \end{bmatrix}$	$B_0 = \begin{bmatrix} 0.991 & 0.002 \\ 0.003 & 0.004 \\ 0.010 & 0.984 \\ 0.003 & 0.003 \end{bmatrix}$ $B_1 = \begin{bmatrix} 0.002 & 0.003 \\ 0.990 & 0.005 \\ 0.002 & 0.004 \\ 0.014 & 0.980 \end{bmatrix}$
$C = [C_0, C_1]$ $C_0 = \begin{bmatrix} C_{000} & C_{001} \\ C_{010} & C_{011} \\ C_{020} & C_{021} \\ C_{030} & C_{031} \end{bmatrix}$ $C_1 = \begin{bmatrix} C_{100} & C_{101} \\ C_{110} & C_{111} \\ C_{120} & C_{121} \\ C_{130} & C_{131} \end{bmatrix}$	$C_0 = \begin{bmatrix} 0.10 & 0.90 \\ 0.20 & 0.80 \\ 0.90 & 0.10 \\ 0.90 & 0.10 \end{bmatrix}$ $C_1 = \begin{bmatrix} 0.90 & 0.10 \\ 0.90 & 0.10 \\ 0.10 & 0.90 \\ 0.30 & 0.70 \end{bmatrix}$	N.A.	N.A.	$C_0 = \begin{bmatrix} 0.09 & 0.91 \\ 0.19 & 0.81 \\ 0.90 & 0.10 \\ 0.90 & 0.10 \end{bmatrix}$ $C_1 = \begin{bmatrix} 0.90 & 0.10 \\ 0.90 & 0.10 \\ 0.08 & 0.92 \\ 0.28 & 0.72 \end{bmatrix}$	$C_0 = \begin{bmatrix} 0.09 & 0.91 \\ 0.19 & 0.81 \\ 0.90 & 0.10 \\ 0.90 & 0.10 \end{bmatrix}$ $C_1 = \begin{bmatrix} 0.90 & 0.10 \\ 0.90 & 0.10 \\ 0.08 & 0.92 \\ 0.28 & 0.72 \end{bmatrix}$