Table 3: Some notations and their corresponding explanations used in the paper.

$\mathcal{U} = \{u\}$	the set of users
$\mathcal{V} = \{v\}$	the set of items
$\mathcal{B}$	the set of user behaviors
k	behavior types for the enumeration or illustration
$v_u^\ell \in \mathcal{V}$	the item interacted with by user u at the ℓ-th time step
$\mathbf{b}_u^\ell \in \mathbb{R}^{\lvert\mathcal{B}\rvert}$	the behavior set (multi-hot vector) involved in the interaction term of user u at the ℓ-th time step
$b_{u,k}^\ell \in \{0,1\}$	$b_{u,k}^\ell = 1$ if user u has an interaction on item v_u^ℓ with the k-th behavior at the ℓ-th time step, otherwise $b_{u,k}^\ell = 0$
$\mathcal{S}_u = \{(v_u^\ell, \mathbf{b}_u^\ell)\}$	the interaction sequence with behavior set of user u
$d \in \mathbb{R}$	latent dimension of embeddings
$L \in \mathbb{R}$	length of the user sequence
$\boldsymbol{I} \in \mathbb{R}^{\lvert\mathcal{V}\rvert \times d}$	the embedding look-up table for all the items
$\boldsymbol{E} \in \mathbb{R}^{L \times d}$	the matrix form of item embeddings for the whole user sequence
$\boldsymbol{P} \in \mathbb{R}^{L \times d}$	position embedding matrix
$\boldsymbol{G} \in \mathbb{R}^{\lvert\mathcal{B}\rvert \times d}$	global behavior embedding matrix
$\boldsymbol{\mathcal{F}} \in \mathbb{R}^{\lvert\mathcal{U}\rvert \times \lvert\mathcal{B}\rvert}$	user behavioral preference factor matrix
$\mathbf{e}_{v_u^\ell} \in \mathbb{R}^{1 \times d}$	item embedding of item v_u^ℓ
$\mathbf{p}_\ell \in \mathbb{R}^{1 \times d}$	position embedding at the ℓ-th time step
$\boldsymbol{\beta}_u^\ell \in \mathbb{R}^d$	embedding of the behavior set corresponding to the interaction term of user sequence $\mathcal{S}_u$ at the ℓ-th time step
$\boldsymbol{B} \in \mathbb{R}^{L \times d}$	the matrix form of behavior set embeddings for the whole user sequence
$\mathfrak{B} \in \mathbb{R}^{L \times \lvert\mathcal{B}\rvert}$	the matrix form of behavior set multi-hot vectors for the whole user sequence
α	the sampling constant in PSA
C	the total number of blocks of L-MSAB
H	the total number of heads in PSA mechanism
$\star \in \{(ib), (pb)\}$	used to distinguish representations from different perspectives
$\widetilde{\boldsymbol{X}} \in \mathbb{R}^{L \times d}$	the output representations from the DCBA layer
$\boldsymbol{H} \in \mathbb{R}^{L \times L}$	Hamming distance attention weight matrix
$\boldsymbol{M}^s \in \mathbb{R}^{L \times \lvert\mathcal{B}\rvert}$	scaling mask in UB-FEEL
$\boldsymbol{M}^e \in \mathbb{R}^{L \times \lvert\mathcal{B}\rvert}$	embedding enhancement mask in UB-FEEL
$\boldsymbol{M}_\Delta \in \mathbb{R}^{u \times L}$	causality mask corresponding to the indices of the top-u queries in PSA
$\overline{\boldsymbol{X}}_k^{(h)} \in \mathbb{R}^{u \times \frac{d}{H}}$	the output representations with top-u dominant queries under the k-th behavior for head h in PSA
$\widehat{\boldsymbol{X}}_k \in \mathbb{R}^{L \times d}$	the output representations of L-MSAB under the k-th behavior
$\boldsymbol{\gamma}^{\ell+1} \in \mathbb{R}^{1 \times d}$	non-user-personalized behavior set embedding at the $(\ell+1)$-th time step
$\boldsymbol{X}^{\text{EMP}} \in \mathbb{R}^{L \times d}$	the output representations along EMP
$\boldsymbol{X}^{\text{IMP}} \in \mathbb{R}^{L \times d}$	the output representations along IMP