**Algorithm 1:** ProbSparse multi-head self-attention

**Input** : Tensor $\boldsymbol{X} \in \mathbb{R}^{L \times d}$ for representations of behavior-specific sequences

**Output**: Final output representations $\mathrm{PSA}(\boldsymbol{X}) \in \mathbb{R}^{L \times d}$

**foreach** *head* $h = 1$ *to* $H$ **do**

  $\boldsymbol{Q}_h \leftarrow f_{Q_h}(\boldsymbol{X}), \boldsymbol{K}_h \leftarrow f_{K_h}(\boldsymbol{X}), \boldsymbol{V}_h \leftarrow f_{V_h}(\boldsymbol{X})$;

  Randomly select $u$ dot-product pairs from $\boldsymbol{K}_h$ as $\overline{\boldsymbol{K}}_h$, where $u = \alpha \ln L$;

  Calculate $\overline{M}$ through Eq.(15) with $\boldsymbol{Q}_h$ and $\overline{\boldsymbol{K}}_h$;

  Select the top-$u$ queries under $\overline{M}$ as $\overline{\boldsymbol{Q}}_h$;

  Calculate $\boldsymbol{X}_1^{(h)}$ through Eq.(16);

  Set $\boldsymbol{X}_0^{(h)} = \mathsf{mean}(\boldsymbol{V}_h)$;

  $\mathrm{PSA}^{(h)}(\boldsymbol{X}) \leftarrow [\boldsymbol{X}_1^{(h)}, \boldsymbol{X}_0^{(h)}]$, which concatenates along the row axis;

**end**

$\mathrm{PSA}(\boldsymbol{X}) \leftarrow \mathsf{concat}\big(\mathrm{PSA}^{(1)}(\boldsymbol{X}), ..., \mathrm{PSA}^{(H)}(\boldsymbol{X})\big)\boldsymbol{W}_L$;

**return** $\mathrm{PSA}(\boldsymbol{X})$