An Adaptive Alternating-direction-method of Multipliers-Incorporated Approach to Nonnegative Latent Factor Analysis: Supplementary File

This is the supplementary file for the paper entitled *An Adaptive Alternating-direction-method of Multipliers-incorporated Approach to Non-negative Latent Factor Analysis*. Detailed convergence proof of A^{2}NLF, additional experimental results and several supplementary files are presented here.

I. Convergence of A^{2}NLF

A. Proof of Lemma 1

Note that A^{2}NLF's learning objective is non-convex. According to [17], any of its limit points where the gradient becomes zero can be a local/global optimum, or saddle point. Hence, such a limit point can be treated as a solution. Supposing that the optimal solution to $a_{u,k(q)}$ by (4c) is $a^t_{u,k(q)}$. Thus, it fulfills the following condition:

$$\lambda^t_{(q)}\left|\Lambda(u)\right|\left(p^t_{u,k(q)}-a^t_{u,k(q)}+\frac{h^{t-1}_{u,k(q)}}{\lambda^t_{(q)}\left|\Lambda(u)\right|}\right)=0. \tag{S1}$$

Following (4e) and (5c), by applying the update rule of $h_{u,k(q)}$ to (S1), we have:

$$h^t_{u,k(q)}=\left(\eta^t_{(q)}-1\right)\lambda^t_{(q)}\left|\Lambda(u)\right|\left(p^t_{u,k(q)}-a^t_{u,k(q)}\right). \tag{S2}$$

Then (17) stands based on (S2). Following the same principle, we can derive the optimality condition of (5*b*) related to $x_{i,k(q)}$:

$$\lambda^t_{(q)}\left|\Lambda(i)\right|\left(z^t_{i,k(q)}-x^t_{i,k(q)}+\frac{w^{t-1}_{i,k(q)}}{\lambda^t_{(q)}\left|\Lambda(i)\right|}\right)=0,\ \Rightarrow w^t_{i,k(q)}=\left(\eta^t_{(q)}-1\right)\lambda^t_{(q)}\left|\Lambda(i)\right|\left(z^t_{i,k(q)}-x^t_{i,k(q)}\right). \tag{S3}$$

Then (18) holds based on (S3). Hence, *Lemma* 1 holds, and **Step 1** is implemented.□

B. Proof of Lemma 2

Considering the difference between $g\left(p^t_{u,k(q)},z^t_{i,k(q)},a^{t-1}_{u,k(q)},x^{t-1}_{i,k(q)},s^{t-1}_{2(q)}\right)$ and $g\left(s^{t-1}_{1(q)},s^{t-1}_{2(q)}\right)$, we have:

$$\begin{aligned}
&g\left(p^t_{u,k(q)},z^t_{i,k(q)},a^{t-1}_{u,k(q)},x^{t-1}_{i,k(q)},s^{t-1}_{2(q)}\right)-g\left(s^{t-1}_{1(q)},s^{t-1}_{2(q)}\right)\\
&\overset{(*)}{=}\left(\sum_{i\in\Lambda(u)}z^{t-1}_{i,k(q)}\left(y_{u,i}-\sum_{l_1=1}^{k-1}p^t_{u,l_1(q)}z^t_{i,l_1(q)}-p^t_{u,k(q)}z^{t-1}_{i,k(q)}-\sum_{l_2=k+1}^{d}p^{t-1}_{u,l_2(q)}z^{t-1}_{i,l_2(q)}\right)-\lambda^t_{(q)}\left|\Lambda(u)\right|\left(p^t_{u,k(q)}-a^{t-1}_{u,k(q)}+\frac{h^{t-1}_{u,k(q)}}{\lambda^t_{(q)}\left|\Lambda(u)\right|}\right)\right)\left(p^{t-1}_{u,k(q)}-p^t_{u,k(q)}\right)\\
&\quad+\left(\sum_{u\in\Lambda(i)}p^{t-1}_{u,k(q)}\left(y_{u,i}-\sum_{l_1=1}^{k-1}p^t_{u,l_1(q)}z^t_{i,l_1(q)}-p^{t-1}_{u,k(q)}z^t_{i,k(q)}-\sum_{l_2=k+1}^{d}p^{t-1}_{u,l_2(q)}z^{t-1}_{i,l_2(q)}\right)-\lambda^t_{(q)}\left|\Lambda(i)\right|\left(z^t_{i,k(q)}-x^{t-1}_{i,k(q)}+\frac{w^{t-1}_{i,k(q)}}{\lambda^t_{(q)}\left|\Lambda(i)\right|}\right)\right)\left(z^{t-1}_{i,k(q)}-z^t_{i,k(q)}\right)\\
&\quad-\frac{1}{2}\left(\sum_{i\in\Lambda(u)}\left(z^{t-1}_{i,k(q)}\right)^2+\lambda^t_{(q)}\left|\Lambda(u)\right|\right)\left(p^t_{u,k(q)}-p^{t-1}_{u,k(q)}\right)^2-\frac{1}{2}\left(\sum_{u\in\Lambda(i)}\left(p^{t-1}_{u,k(q)}\right)^2+\lambda^t_{(q)}\left|\Lambda(i)\right|\right)\left(z^t_{i,k(q)}-z^{t-1}_{i,k(q)}\right)^2.
\end{aligned} \tag{S4}$$

Considering (5a)'s optimality condition, (S4) is transformed as:

$$
\begin{aligned}
& g\left(p_{u,k(q)}^{t}, z_{i,k(q)}^{t}, a_{u,k(q)}^{t-1}, x_{i,k(q)}^{t-1}, s_{2(q)}^{t-1}\right) - g\left(s_{1(q)}^{t-1}, s_{2(q)}^{t-1}\right) \\
& = -\frac{1}{2}\left(\sum_{i\in\Lambda(u)}\left(z_{i,k(q)}^{t-1}\right)^2 + \lambda_{(q)}^{t}\left|\Lambda(u)\right|\right)\left(p_{u,k(q)}^{t} - p_{u,k(q)}^{t-1}\right)^2 - \frac{1}{2}\left(\sum_{u\in\Lambda(i)}\left(p_{u,k(q)}^{t-1}\right)^2 + \lambda_{(q)}^{t}\left|\Lambda(i)\right|\right)\left(z_{i,k(q)}^{t} - z_{i,k(q)}^{t-1}\right)^2.
\end{aligned}
\tag{S5}
$$

Thus, the difference between $g\left(s_{1(q)}^{t}, s_{2(q)}^{t-1}\right)$ and $g\left(p_{u,k(q)}^{t}, z_{i,k(q)}^{t}, a_{u,k(q)}^{t-1}, x_{i,k(q)}^{t-1}, s_{2(q)}^{t-1}\right)$ is:

$$
g\left(s_{1(q)}^{t}, s_{2(q)}^{t-1}\right) - g\left(p_{u,k(q)}^{t}, z_{i,k(q)}^{t}, a_{u,k(q)}^{t-1}, x_{i,k(q)}^{t-1}, s_{2(q)}^{t-1}\right) = -\left(\lambda_{(q)}^{t}\left|\Lambda(u)\right|/2\right)\left(a_{u,k(q)}^{t} - a_{u,k(q)}^{t-1}\right)^2 - \left(\lambda_{(q)}^{t}\left|\Lambda(i)\right|/2\right)\left(x_{i,k(q)}^{t} - x_{i,k(q)}^{t-1}\right)^2.
\tag{S6}
$$

Moreover, $g\left(s_{1(q)}^{t}, s_{2(q)}^{t}\right)$ and $g\left(s_{1(q)}^{t}, s_{2(q)}^{t-1}\right)$ yields:

$$
\begin{aligned}
& g\left(s_{1(q)}^{t}, s_{2(q)}^{t}\right) - g\left(s_{1(q)}^{t}, s_{2(q)}^{t-1}\right) \\
& = \left(p_{u,k(q)}^{t} - a_{u,k(q)}^{t}\right)\left(h_{u,k(q)}^{t} - h_{u,k(q)}^{t-1}\right) + \left(z_{i,k(q)}^{t} - x_{i,k(q)}^{t}\right)\left(w_{i,k(q)}^{t} - w_{i,k(q)}^{t-1}\right) \\
& \overset{\text{(I)}}{=} \frac{1}{\eta_{(q)}^{t}\lambda_{(q)}^{t}\left|\Lambda(u)\right|}\left(h_{u,k(q)}^{t} - h_{u,k(q)}^{t-1}\right)^2 + \frac{1}{\eta_{(q)}^{t}\lambda_{(q)}^{t}\left|\Lambda(i)\right|}\left(w_{i,k(q)}^{t} - w_{i,k(q)}^{t-1}\right)^2 \\
& \overset{\text{(II)}}{\le} \frac{2\left|\Lambda(u)\right|}{\eta_{(q)}^{t}\lambda_{(q)}^{t}}\left(\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}p_{u,k(q)}^{t} - \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}p_{u,k(q)}^{t-1}\right)^2 + \left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}a_{u,k(q)}^{t} - \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}a_{u,k(q)}^{t-1}\right)^2\right) \\
& + \frac{2\left|\Lambda(i)\right|}{\eta_{(q)}^{t}\lambda_{(q)}^{t}}\left(\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}z_{i,k(q)}^{t} - \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}z_{i,k(q)}^{t-1}\right)^2 + \left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}x_{i,k(q)}^{t} - \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}x_{i,k(q)}^{t-1}\right)^2\right),
\end{aligned}
\tag{S7}
$$

where (I) is based on the update rules of ($h_{u,k(q)}$, $w_{i,k(q)}$) given in (4e), (4f) and (5c), and (II) is achieved with *Lemma* 1.With (S5)-(S7), we have the following deduction:

$$
\begin{aligned}
& g\left(s_{1(q)}^{t}, s_{2(q)}^{t}\right) - g\left(s_{1(q)}^{t-1}, s_{2(q)}^{t-1}\right) \\
& \le -\frac{1}{2}\left(\sum_{i\in\Lambda(u)}\left(z_{i,k(q)}^{t-1}\right)^2 + \lambda_{(q)}^{t}\left|\Lambda(u)\right|\right)\left(p_{u,k(q)}^{t} - p_{u,k(q)}^{t-1}\right)^2 - \frac{\lambda_{(q)}^{t}\left|\Lambda(u)\right|}{2}\left(a_{u,k(q)}^{t} - a_{u,k(q)}^{t-1}\right)^2 \\
& -\frac{1}{2}\left(\sum_{u\in\Lambda(i)}\left(p_{u,k(q)}^{t-1}\right)^2 + \lambda_{(q)}^{t}\left|\Lambda(i)\right|\right)\left(z_{i,k(q)}^{t} - z_{i,k(q)}^{t-1}\right)^2 - \frac{\lambda_{(q)}^{t}\left|\Lambda(i)\right|}{2}\left(x_{i,k(q)}^{t} - x_{i,k(q)}^{t-1}\right)^2 \\
& + \frac{2\left|\Lambda(u)\right|}{\eta_{(q)}^{t}\lambda_{(q)}^{t}}\left(\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}p_{u,k(q)}^{t} - \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}p_{u,k(q)}^{t-1}\right)^2 + \left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}a_{u,k(q)}^{t} - \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}a_{u,k(q)}^{t-1}\right)^2\right) \\
& + \frac{2\left|\Lambda(i)\right|}{\eta_{(q)}^{t}\lambda_{(q)}^{t}}\left(\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}z_{i,k(q)}^{t} - \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}z_{i,k(q)}^{t-1}\right)^2 + \left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}x_{i,k(q)}^{t} - \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}x_{i,k(q)}^{t-1}\right)^2\right).
\end{aligned}
\tag{S8}
$$

Owing to (20a), (21a) stands, which indicates that the augmented Lagrangian function (3) related to the q-th particle is non-increasing as $a_{u,k(q)}^{t}>0$ and $x_{i,k(q)}^{t}>0$. Then after the t-th iteration, the partial objective from (3) related to the q-th particle is formulated as:

$$
\begin{aligned}
& g\left(s_{1(q)}^{t}, s_{2(q)}^{t}\right) \\
& = \frac{1}{2}\sum_{y_{u,i}\in\Lambda}\left(y_{u,i} - \sum_{l_1=1}^{k}p_{u,l_1(q)}^{t}z_{i,l_1(q)}^{t} - \sum_{l_2=k+1}^{d}p_{u,l_2(q)}^{t-1}z_{i,l_2(q)}^{t-1}\right)^2 \\
& + \sum_{u}\left(\left(\sum_{l_1=1}^{k}h_{u,l_1(q)}^{t}\left(p_{u,l_1(q)}^{t} - a_{u,l_1(q)}^{t}\right)\right) + \left(\sum_{l_2=k+1}^{d}h_{u,l_2(q)}^{t-1}\left(p_{u,l_2(q)}^{t-1} - a_{u,l_2(q)}^{t-1}\right)\right)\right) + \sum_{i}\left(\left(\sum_{l_1=1}^{k}w_{i,l_1(q)}^{t}\left(z_{i,l_1(q)}^{t} - x_{i,l_1(q)}^{t}\right)\right) + \left(\sum_{l_2=k+1}^{d}w_{i,l_2(q)}^{t-1}\left(z_{i,l_2(q)}^{t-1} - x_{i,l_2(q)}^{t-1}\right)\right)\right) \\
& + \sum_{u}\frac{\lambda_{(q)}^{t}\left|\Lambda(u)\right|}{2}\left(\sum_{l_1=1}^{k}\left(p_{u,l_1(q)}^{t} - a_{u,l_1(q)}^{t}\right)^2 + \sum_{l_2=k+1}^{d}\left(p_{u,l_2(q)}^{t-1} - a_{u,l_2(q)}^{t-1}\right)^2\right) + \sum_{i}\frac{\lambda_{(q)}^{t}\left|\Lambda(i)\right|}{2}\left(\sum_{l_1=1}^{k}\left(z_{i,l_1(q)}^{t} - x_{i,l_1(q)}^{t}\right)^2 + \sum_{l_2=k+1}^{d}\left(z_{i,l_2(q)}^{t-1} - x_{i,l_2(q)}^{t-1}\right)^2\right).
\end{aligned}
\tag{S9}
$$

By substituting (S2) and (S3) into (S9), we have:

$$\begin{aligned}
&g\left(s_{1(q)}^{t}, s_{2(q)}^{t}\right) \\
&=\frac{1}{2}\sum_{y_{u,i}\in\Lambda}\left(y_{u,i}-\sum_{l_1=1}^{k}p_{u,l_1(q)}^{t}z_{i,l_1(q)}^{t}-\sum_{l_2=k+1}^{d}p_{u,l_2(q)}^{t-1}z_{i,l_2(q)}^{t-1}\right)^2 \\
&+\sum_{u}\left|\Lambda(u)\right|\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}\sum_{l_1=1}^{k}\left(p_{u,l_1(q)}^{t}-a_{u,l_1(q)}^{t}\right)^2+\left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}\sum_{l_2=k+1}^{d}\left(p_{u,l_2(q)}^{t-1}-a_{u,l_2(q)}^{t-1}\right)^2\right) \\
&+\sum_{i}\left|\Lambda(i)\right|\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}\sum_{l_1=1}^{k}\left(z_{i,l_1(q)}^{t}-x_{i,l_1(q)}^{t}\right)^2+\left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}\sum_{l_2=k+1}^{d}w_{i,l_2(q)}^{t}\left(z_{i,l_2(q)}^{t-1}-x_{i,l_2(q)}^{t-1}\right)^2\right) \\
&+\sum_{u}\frac{\lambda_{(q)}^{t}\left|\Lambda(u)\right|}{2}\left(\sum_{l_1=1}^{k}\left(p_{u,l_1(q)}^{t}-a_{u,l_1(q)}^{t}\right)^2+\sum_{l_2=k+1}^{d}\left(p_{u,l_2(q)}^{t-1}-a_{u,l_2(q)}^{t-1}\right)^2\right)+\sum_{i}\frac{\lambda_{(q)}^{t}\left|\Lambda(i)\right|}{2}\left(\sum_{l_1=1}^{k}\left(z_{i,l_1(q)}^{t}-x_{i,l_1(q)}^{t}\right)^2+\sum_{l_2=k+1}^{d}\left(z_{i,l_2(q)}^{t-1}-x_{i,l_2(q)}^{t-1}\right)^2\right).
\end{aligned} \tag{S10}$$

(S10) indicates that if (20b) is fulfilled, (21b) holds, thereby making (3) related to the q-th particle lower-bounded as $a_{u,k(q)}^{t}$>0 and $x_{i,k(q)}^{t}$>0. Based on the above inferences, *Lemma* 2 stands, and **Step 2** is implemented.□

C. Proof of Theorem 1

Part a. Following *Lemma* 2, $g\left(s_{1(q)}^{t}, s_{2(q)}^{t}\right)$ converges as $t\rightarrow\infty$, indicating that:

$$\lim_{t\to\infty} g\left(s_{1(q)}^{t}, s_{2(q)}^{t}\right)-g\left(s_{1(q)}^{t-1}, s_{2(q)}^{t-1}\right)\rightarrow 0. \tag{S11}$$

With (20), when (S1) is fulfilled, the upper-bound of $g\left(s_{1(q)}^{t}, s_{2(q)}^{t}\right)-g\left(s_{1(q)}^{t-1}, s_{2(q)}^{t-1}\right)$ is zero as $t\rightarrow\infty$, thereby achieving (22). Following (S8) and (22), we have [24]:

$$\lim_{t\to\infty}\left(p_{u,k(q)}^{t}-p_{u,k(q)}^{t-1}\right)\rightarrow 0, \tag{S12a}$$

$$\lim_{t\to\infty}\left(z_{i,k(q)}^{t}-z_{i,k(q)}^{t-1}\right)\rightarrow 0, \tag{S12b}$$

$$\lim_{t\to\infty}\left(a_{u,k(q)}^{t}-a_{u,k(q)}^{t-1}\right)\rightarrow 0, \tag{S12c}$$

$$\lim_{t\to\infty}\left(x_{i,k(q)}^{t}-x_{i,k(q)}^{t-1}\right)\rightarrow 0, \tag{S12d}$$

$$\lim_{t\to\infty}\left(\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}p_{u,k(q)}^{t}-\left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}p_{u,k(q)}^{t-1}\right)^2+\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}a_{u,k(q)}^{t}-\left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}a_{u,k(q)}^{t-1}\right)^2\right)\rightarrow 0, \tag{S12e}$$

$$\lim_{t\to\infty}\left(\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}z_{i,k(q)}^{t}-\left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}z_{i,k(q)}^{t-1}\right)^2+\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}x_{i,k(q)}^{t}-\left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}x_{i,k(q)}^{t-1}\right)^2\right)\rightarrow 0. \tag{S12f}$$

Based on (17), (18) and (S12), we have the following inferences:

$$\lim_{t\to\infty}\left(h_{u,k(q)}^{t}-h_{u,k(q)}^{t-1}\right)\rightarrow 0, \tag{S13a}$$

$$\lim_{t\to\infty}\left(w_{i,k(q)}^{t}-w_{i,k(q)}^{t-1}\right)\rightarrow 0. \tag{S13b}$$

Based on (4e), (4f) and (S13), we conclude that (23) is fulfilled.

Part b. Firstly, following (4a), (4b) and (5a), the update rules of ($p_{u,k(q)}$, $z_{i,k(q)}$) can be rearranged as:

$$\left(p_{u,k(q)}^{t-1}-p_{u,k(q)}^{t}\right)\left(\sum_{i\in\Lambda(u)}\left(z_{i,k(q)}^{t-1}\right)^2+\lambda_{(q)}^{t}\left|\Lambda(u)\right|\right)$$
$$=\sum_{i\in\Lambda(u)}z_{i,k(q)}^{t-1}\left(\sum_{l_1=1}^{k}p_{u,l_1(q)}^{t}z_{i,l_1(q)}^{t}+\sum_{l_2=k+1}^{d}p_{u,l_2(q)}^{t-1}z_{i,l_2(q)}^{t-1}-y_{u,i}\right)+\lambda_{(q)}^{t}\left|\Lambda(u)\right|\left(p_{u,k(q)}^{t-1}-a_{u,k(q)}^{t-1}\right)+h_{u,k(q)}^{t-1}, \tag{S14a}$$

$$\left(z_{i,k(q)}^{t-1}-z_{i,k(q)}^{t}\right)\left(\sum_{u\in\Lambda(i)}\left(p_{i,k(q)}^{t-1}\right)^2+\lambda_{(q)}^{t}\left|\Lambda(i)\right|\right)$$
$$=\sum_{u\in\Lambda(i)}p_{u,k(q)}^{t-1}\left(\sum_{l_1=1}^{k}p_{u,l_1(q)}^{t}z_{i,l_1(q)}^{t}+\sum_{l_2=k+1}^{d}p_{u,l_2(q)}^{t-1}z_{i,l_2(q)}^{t-1}-y_{u,i}\right)+\lambda_{(q)}^{t}\left|\Lambda(i)\right|\left(z_{i,k(q)}^{t-1}-x_{i,k(q)}^{t-1}\right)+w_{i,k(q)}^{t-1}. \tag{S14b}$$

Then by substituting (23) and (S12) into (S14), we have:

$$\sum_{i\in\Lambda(u)}z_{i,k(q)}^{t-1}\left(\sum_{l_1=1}^{k}p_{u,l_1(q)}^{t}z_{i,l_1(q)}^{t}+\sum_{l_2=k+1}^{d}p_{u,l_2(q)}^{t-1}z_{i,l_2(q)}^{t-1}-y_{u,i}\right)+h_{u,k(q)}^{t-1}\to 0, \tag{S15a}$$

$$\sum_{u\in\Lambda(i)}p_{u,k(q)}^{t-1}\left(\sum_{l_1=1}^{k}p_{u,l_1(q)}^{t}z_{i,l_1(q)}^{t}+\sum_{l_2=k+1}^{d}p_{u,l_2(q)}^{t-1}z_{i,l_2(q)}^{t-1}-y_{u,i}\right)+w_{i,k(q)}^{t-1}\to 0. \tag{S15b}$$

Hence, considering a limit point $\left\{s_{1(q)}^{*},s_{2(q)}^{*}\right\}$ of a sequence $\left\{s_{1(q)}^{t},s_{2(q)}^{t}\right\}$ generated by the update rules of $\{s_{1(q)}, s_{2(q)}\}$ based on (4) and (5), the following KKT conditions are satisfied with (23) and (S15):

$$\sum_{i\in\Lambda(u)}z_{i,k(q)}^{*}\left(\sum_{k=1}^{d}p_{u,k(q)}^{*}z_{i,k(q)}^{*}-y_{u,i}\right)+h_{u,k(q)}^{*}\to 0, \tag{S16a}$$

$$\sum_{u\in\Lambda(i)}p_{u,k(q)}^{*}\left(\sum_{k=1}^{d}p_{u,k(q)}^{*}z_{i,k(q)}^{*}-y_{u,i}\right)+w_{i,k(q)}^{*}\to 0, \tag{S16b}$$

$$p_{u,k}^{*}-a_{u,k}^{*}\to 0, \tag{S16c}$$

$$z_{i,k}^{*}-x_{i,k}^{*}\to 0. \tag{S16d}$$

Afterwards, considering the remaining KKT conditions regarding constraints $a_{u,k(q)}>0$ and $x_{i,k(q)}>0$, we extend the original augmented Lagrangian g:

$$g_{(q)}^{\#}=g_{(q)}-Tr\left(M_{(q)}\left(A_{(q)}\right)^{T}\right)-Tr\left(N_{(q)}\left(X_{(q)}\right)^{T}\right)=g_{(q)}-\sum_{(u,k)}m_{u,k(q)}a_{u,k(q)}-\sum_{(i,k)}n_{i,k(q)}x_{i,k(q)}, \tag{S17}$$

where the operator $Tr(\cdot)$ computes the trace of an enclosed matrix, and the definition of $g_{(q)}$ is given by:

$$g_{(q)}=\frac{1}{2}\sum_{y_{u,i}\in\Lambda}\left(y_{u,i}-\sum_{k=1}^{d}p_{u,k(q)}z_{i,k(q)}\right)^2+\sum_{(u,k)}h_{u,k(q)}\left(p_{u,k(q)}-a_{u,k(q)}\right)+\sum_{(u,k)}\frac{\lambda_{(q)}\left|\Lambda(u)\right|}{2}\left(p_{u,k(q)}-a_{u,k(q)}\right)^2$$
$$+\sum_{(i,k)}w_{i,k(q)}\left(z_{i,k(q)}-x_{i,k(q)}\right)+\sum_{(i,k)}\frac{\lambda_{(q)}\left|\Lambda(i)\right|}{2}\left(z_{i,k(q)}-x_{i,k(q)}\right)^2. \tag{S18}$$

For the partial derivatives of $g_{(q)}^{\#}$ with $a_{u,k(q)}$ and $x_{i,k(q)}$, we have:

$$\begin{cases} \dfrac{\partial g_{(q)}^{\#}}{\partial a_{u,k}} = -\lambda_{(q)} |\Lambda(u)| \left(p_{u,k(q)} - a_{u,k(q)} + \dfrac{h_{u,k(q)}}{\lambda_{(q)} |\Lambda(u)|} \right) - m_{u,k(q)} = 0, \\ \dfrac{\partial g_{(q)}^{\#}}{\partial x_{i,k}} = -\lambda_{(q)} |\Lambda(i)| \left(z_{i,k(q)} - x_{i,k(q)} + \dfrac{w_{i,k(q)}}{\lambda_{(q)} |\Lambda(i)|} \right) - n_{i,k(q)} = 0, \end{cases} \Rightarrow \begin{cases} m_{u,k} = -\lambda_{(q)} |\Lambda(u)| \left(p_{u,k(q)} - a_{u,k(q)} + \dfrac{h_{u,k(q)}}{\lambda_{(q)} |\Lambda(u)|} \right), \\ n_{i,k} = -\lambda_{(q)} |\Lambda(i)| \left(z_{i,k(q)} - x_{i,k(q)} + \dfrac{w_{i,k(q)}}{\lambda_{(q)} |\Lambda(i)|} \right). \end{cases} \tag{S19}$$

Then, with the KKT conditions of $\forall m_{u,k(q)}$, $a_{u,k(q)}$: $m_{u,k(q)}a_{u,k(q)}=0$ and $\forall n_{i,k(q)}$, $x_{i,k(q)}$: $n_{i,k(q)}x_{i,k(q)}=0$ for (S17), we achieve the following equations based on (S19) [21, 24]:

$$\begin{cases} a_{u,k(q)} \left(-\lambda_{(q)} |\Lambda(u)| \left(p_{u,k(q)} - a_{u,k(q)} + \dfrac{h_{u,k(q)}}{\lambda_{(q)} |\Lambda(u)|} \right) \right) = 0, \\ x_{i,k(q)} \left(-\lambda_{(q)} |\Lambda(i)| \left(z_{i,k(q)} - x_{i,k(q)} + \dfrac{w_{i,k(q)}}{\lambda_{(q)} |\Lambda(i)|} \right) \right) = 0, \end{cases} \Rightarrow \begin{cases} a_{u,k(q)} = p_{u,k(q)} + \dfrac{h_{u,k(q)}}{\lambda_{(q)} |\Lambda(u)|}, \\ x_{i,k(q)} = z_{i,k(q)} + \dfrac{w_{i,k(q)}}{\lambda_{(q)} |\Lambda(i)|}. \end{cases} \tag{S20}$$

To satisfy the nonnegativity of output LFs $a_{u,k(q)}$ and $x_{i,k(q)}$, (S20) can be rewritten as:

$$\begin{cases} a_{u,k(q)} = \max\left(0, p_{u,k(q)} + \dfrac{h_{u,k(q)}}{\lambda_{(q)} |\Lambda(u)|} \right), \\ x_{i,k(q)} = \max\left(0, z_{i,k(q)} + \dfrac{w_{i,k(q)}}{\lambda_{(q)} |\Lambda(i)|} \right). \end{cases} \tag{S21}$$

Note that (S21) is consistent with the update rules of $a_{u,k(q)}$ and $x_{i,k(q)}$ based on (4c) and (4d). Therefore, (S17)-(S21) show that an A^{2}NLF model's learning rules are closely connected with the KKT conditions of its learning objective.

Then considering the KKT conditions related to $a_{u,k(q)}$:

$$\left. \frac{\partial g_{(q)}^{\#}}{\partial a_{u,k(q)}} \right|_{a_{u,k(q)} = a_{u,k(q)}^{*}} = -\lambda_{(q)}^{*} |\Lambda(u)| \left(p_{u,k(q)}^{*} - a_{u,k(q)}^{*} + \frac{h_{u,k(q)}^{*}}{\lambda_{(q)}^{*} |\Lambda(u)|} \right) - m_{u,k(q)}^{*} = 0, \tag{S22a}$$

$$m_{u,k(q)}^{*} a_{u,k(q)}^{*} = 0, \tag{S22b}$$

$$a_{u,k(q)}^{*} \geq 0, \tag{S22c}$$

$$m_{u,k(q)}^{*} \geq 0, \tag{S22d}$$

where $a_{u,k(q)}^{*}$ is a KKT stationary point of $a_{u,k(q)}$, and $m_{u,k(q)}^{*}$ is a limit point of the sequence $\{m_{u,k(q)}^{t}\}$ generated by the update rules of $m_{u,k}$ based on (S19). According to (S17)-(S21) and $a_{u,k(q)}^{t}=0$, conditions (S22a)-(S22c) are satisfied. Thus, we have:

$$m_{u,k(q)}^{*} = -\lambda_{(q)}^{*} |\Lambda(u)| \left(p_{u,k(q)}^{*} - a_{u,k(q)}^{*} + \frac{h_{u,k(q)}^{*}}{\lambda_{(q)}^{*} |\Lambda(u)|} \right). \tag{S23}$$

Thus, we focus on condition (S22d). Since $a_{u,k(q)}^{t}>0$ in this case, the update rule for $a_{u,k(q)}$ is given as:

$$a_{u,k(q)}^{*} \leftarrow p_{u,k(q)}^{*} + \frac{h_{u,k(q)}^{*}}{\lambda_{(q)}^{*} |\Lambda(u)|}. \tag{S24}$$

By substituting (S24) into (S23), we have $m_{u,k(q)}^{*}=0$. Hence, conditions (S22c) and (S22d) are fulfilled. Note that as $x_{i,k(q)}^{t}>0$ in this case, the proof regarding the KKT conditions of $x_{u,k(q)}$ can be achieved similarly. *Theorem* 1 stands, and **Step 3** is implemented. □

D. Proof of Lemma 3

The difference between $g\left(p_{u,k(q)}^{t}, z_{i,k(q)}^{t}, a_{u,k(q)}^{t-1}, x_{i,k(q)}^{t-1}, s_{2(q)}^{t-1}\right)$ and $g\left(s_{1(q)}^{t-1}, s_{2(q)}^{t-1}\right)$ in this case is also given by (S5). Considering the fact of $a_{u,k(q)}^{t}$=0 and, $x_{i,k(q)}^{t}$>0 the difference between $g\left(s_{1(q)}^{t}, s_{2(q)}^{t-1}\right)$ and $g\left(p_{u,k(q)}^{t}, z_{i,k(q)}^{t}, a_{u,k(q)}^{t-1}, x_{i,k(q)}^{t-1}, s_{2(q)}^{t-1}\right)$ is:

$$g\left(s_{1(q)}^{t}, s_{2(q)}^{t-1}\right) - g\left(p_{u,k(q)}^{t}, z_{i,k(q)}^{t}, a_{u,k(q)}^{t-1}, x_{i,k(q)}^{t-1}, s_{2(q)}^{t-1}\right) = -\frac{\lambda_{(q)}^{t}\left|\Lambda(u)\right|}{2}\left(a_{u,k(q)}^{t-1}\right)^{2} - \frac{\lambda_{(q)}^{t}\left|\Lambda(i)\right|}{2}\left(x_{i,k(q)}^{t} - x_{i,k(q)}^{t-1}\right)^{2}. \tag{S25}$$

Moreover, $g\left(s_{1(q)}^{t}, s_{2(q)}^{t}\right)$ and $g\left(s_{1(q)}^{t}, s_{2(q)}^{t-1}\right)$ yields:

$$\begin{aligned}
&g\left(s_{1(q)}^{t}, s_{2(q)}^{t}\right) - g\left(s_{1(q)}^{t}, s_{2(q)}^{t-1}\right) \\
&= \left(p_{u,k(q)}^{t} - a_{u,k(q)}^{t}\right)\left(h_{u,k(q)}^{t} - h_{u,k(q)}^{t-1}\right) + \left(z_{i,k(q)}^{t} - x_{i,k(q)}^{t}\right)\left(w_{i,k(q)}^{t} - w_{i,k(q)}^{t-1}\right) \\
&\overset{\text{(I)}}{=} \eta_{(q)}^{t}\lambda_{(q)}^{t}\left|\Lambda(u)\right|\left(p_{u,k(q)}^{t} - a_{u,k(q)}^{t}\right)^{2} + \frac{1}{\eta_{(q)}^{t}\lambda_{(q)}^{t}\left|\Lambda(i)\right|}\left(w_{i,k(q)}^{t} - w_{i,k(q)}^{t-1}\right)^{2} \\
&\overset{\text{(II)}}{\le} \eta_{(q)}^{t}\lambda_{(q)}^{t}\left|\Lambda(u)\right|\left(p_{u,k(q)}^{t}\right)^{2} + \frac{2\left|\Lambda(i)\right|}{\eta_{(q)}^{t}\lambda_{(q)}^{t}}\left(\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t} z_{i,k(q)}^{t} - \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1} z_{i,k(q)}^{t-1}\right)^{2} + \left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t} x_{i,k(q)}^{t} - \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1} x_{i,k(q)}^{t-1}\right)^{2}\right),
\end{aligned} \tag{S26}$$

where (I) is based on the update rules of ($h_{u,k(q)}$, $w_{i,k(q)}$) given in (4e), (4f) and (5c), and (II) is achieved with (18) and $a_{u,k(q)}^{t}$=0. With (S5), (S25) and (S26), we have the following deduction:

$$\begin{aligned}
&g\left(s_{1(q)}^{t}, s_{2(q)}^{t}\right) - g\left(s_{1(q)}^{t-1}, s_{2(q)}^{t-1}\right) \\
&\le -\frac{1}{2}\left(\sum_{i\in\Lambda(u)}\left(z_{i,k(q)}^{t-1}\right)^{2} + \lambda_{(q)}^{t}\left|\Lambda(u)\right|\right)\left(p_{u,k(q)}^{t} - p_{u,k(q)}^{t-1}\right)^{2} - \frac{\lambda_{(q)}^{t}\left|\Lambda(u)\right|}{2}\left(a_{u,k(q)}^{t-1}\right)^{2} \\
&\quad -\frac{1}{2}\left(\sum_{u\in\Lambda(i)}\left(p_{u,k(q)}^{t-1}\right)^{2} + \lambda_{(q)}^{t}\left|\Lambda(i)\right|\right)\left(z_{i,k(q)}^{t} - z_{i,k(q)}^{t-1}\right)^{2} - \frac{\lambda_{(q)}^{t}\left|\Lambda(i)\right|}{2}\left(x_{i,k(q)}^{t} - x_{i,k(q)}^{t-1}\right)^{2} \\
&\quad +\eta_{(q)}^{t}\lambda_{(q)}^{t}\left|\Lambda(u)\right|\left(p_{u,k(q)}^{t}\right)^{2} + \frac{2\left|\Lambda(i)\right|}{\eta_{(q)}^{t}\lambda_{(q)}^{t}}\left(\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t} z_{i,k(q)}^{t} - \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1} z_{i,k(q)}^{t-1}\right)^{2} + \left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t} x_{i,k(q)}^{t} - \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1} x_{i,k(q)}^{t-1}\right)^{2}\right).
\end{aligned} \tag{S27}$$

Owing to (25), (21a) stands, which indicates that the augmented Lagrangian function (3) related to the q-th particle is non-increasing as $a_{u,k(q)}^{t}$=0 and $x_{i,k(q)}^{t}$>0 in this case. Then after the t-th iteration, we substitute $a_{u,k(q)}^{t}$=0 into (S10):

$$\begin{aligned}
g\left(s_{1(q)}^{t}, s_{2(q)}^{t}\right) = &\frac{1}{2}\sum_{y_{u,i}\in\Lambda}\left(y_{u,i} - \sum_{l_1=1}^{k} p_{u,l_1(q)}^{t} z_{i,l_1(q)}^{t} - \sum_{l_2=k+1}^{d} p_{u,l_2(q)}^{t-1} z_{i,l_2(q)}^{t-1}\right)^{2} \\
&+\sum_{u}\left|\Lambda(u)\right|\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}\left(\sum_{l_1=1}^{k-1}\left(p_{u,l_1(q)}^{t} - a_{u,l_1(q)}^{t}\right)^{2} + \left(p_{u,k(q)}^{t}\right)^{2}\right) + \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}\sum_{l_2=k+1}^{d}\left(p_{u,l_2(q)}^{t-1} - a_{u,l_2(q)}^{t-1}\right)^{2}\right) \\
&+\sum_{i}\left|\Lambda(i)\right|\left(\left(\eta_{(q)}^{t}-1\right)\lambda_{(q)}^{t}\sum_{l_1=1}^{k}\left(z_{i,l_1(q)}^{t} - x_{i,l_1(q)}^{t}\right)^{2} + \left(\eta_{(q)}^{t-1}-1\right)\lambda_{(q)}^{t-1}\sum_{l_2=k+1}^{d} w_{i,l_2(q)}^{t}\left(z_{i,l_2(q)}^{t-1} - x_{i,l_2(q)}^{t-1}\right)^{2}\right) \\
&+\sum_{u}\frac{\lambda_{(q)}^{t}\left|\Lambda(u)\right|}{2}\left(\sum_{l_1=1}^{k-1}\left(p_{u,l_1(q)}^{t} - a_{u,l_1(q)}^{t}\right)^{2} + \left(p_{u,k(q)}^{t}\right)^{2} + \sum_{l_2=k+1}^{d}\left(p_{u,l_2(q)}^{t-1} - a_{u,l_2(q)}^{t-1}\right)^{2}\right) \\
&+\sum_{i}\frac{\lambda_{(q)}^{t}\left|\Lambda(i)\right|}{2}\left(\sum_{l_1=1}^{k}\left(z_{i,l_1(q)}^{t} - x_{i,l_1(q)}^{t}\right)^{2} + \sum_{l_2=k+1}^{d}\left(z_{i,l_2(q)}^{t-1} - x_{i,l_2(q)}^{t-1}\right)^{2}\right).
\end{aligned} \tag{S28}$$

(S28) indicates that if (20b) is fulfilled, (21b) holds, thereby making (3) related to the q-th particle lower-bounded as $a_{u,k(q)}^{t}$=0, and $x_{i,k(q)}^{t}$>0 in this case. Based on the above inferences, *Lemma* 2 stands, and **Step 4** is implemented.□

E. Proof of Theorem 2

Part a. Following *Lemma* 3, $g\left(s_{1(q)}^{t}, s_{2(q)}^{t}\right)$ converges as $t \to \infty$, indicating that (S11) is fulfilled. With (22), (25) and (S27), we have (S12a), (S12b), (S12d), (S12f) and the following inferences:

$$\lim_{t\to\infty} a_{u,k(q)}^{t-1} \to 0, \tag{S29a}$$

$$\lim_{t\to\infty} p_{u,k(q)}^{t} \to 0. \tag{S29b}$$

Then according to (S12f), (S13b) is fulfilled. Hence, based on (S13b), (S29b) and $a_{u,k(q)}^{t}$=0, (23) is fulfilled.

Part b. Firstly, considering a limit point $\left\{s_{1(q)}^{*}, s_{2(q)}^{*}\right\}$ of a sequence $\left\{s_{1(q)}^{t}, s_{2(q)}^{t}\right\}$ generated by the update rules of $\{s_{1(q)}, s_{2(q)}\}$ based on (4) and (5), according to (23) and (S15), (S16) holds when $a_{u,k(q)}^{t}$=0, and $x_{i,k(q)}^{t}$>0 in this case. Then considering the KKT conditions related to $a_{u,k(q)}$, i.e., (S22).

According to (S17)-(S21) and with $a_{u,k(q)}^{t}$=0, conditions (S22a)-(S22c) are naturally satisfied. Thus, we focus on analyzing condition (S22d). Since we have $a_{u,k(q)}^{t}$=0 in this case, the following inequality holds:

$$p_{u,k(q)}^{*} + \frac{h_{u,k(q)}^{*}}{\lambda_{(q)}^{*}\left|\Lambda(u)\right|} \le 0. \tag{S30}$$

Note that (S30) indicates that:

$$m_{u,k(q)}^{*} = -\lambda_{(q)}^{*}\left|\Lambda(u)\right|\left(p_{u,k(q)}^{*} + \frac{h_{u,k(q)}^{*}}{\lambda_{(q)}^{*}\left|\Lambda(u)\right|}\right) \ge 0. \tag{S31}$$

Thus, condition (S22) are all fulfilled in this case. Note that as $x_{i,k(q)}^{t}$>0 in this case, the proof regarding the KKT conditions of $x_{u,k(q)}$ can be achieved similarly. *Theorem* 2 stands, and **Step 5** is implemented. ■

II. SUPPLEMENTARY TABLES

TABLE S1. PSO Parameters Settings in (7).

w	c_1	c_2	r_1, r_2
0.729	2	2	uniform random numbers∈[0,1].

TABLE S2 PSO PARAMETERS SETTINGS IN (13)

$\breve{\lambda}$	$\hat{\lambda}$	$\breve{\eta}$	$\hat{\eta}$	$\breve{v}_{\lambda}$	$\hat{v}_{\lambda}$	$\breve{v}_{\eta}$	$\hat{v}_{\eta}$
0.2	2	1	2	$-\hat{v}_{\lambda}$	$0.2\times(\hat{\lambda}-\breve{\lambda})$	$-\hat{v}_{\eta}$	$0.2\times(\hat{\eta}-\breve{\eta})$

TABLE S3. Grid-search Range of Hyper-parameters in M1-6 on D1-4.

No.	Hyper-parameters	Grid-searching Range
M1	Augmentation coefficient λ	2^{-6}, 2^{-4}, 2^{-2}, 1, 2^{2}, 2^{4}, 2^{6}, 2^{8}
	Learning rate η	8, 4, 2, 1, 2^{-1}, 2^{-2}, 2^{-3}
M2	Penalty coefficient λ	2, 1.8, 1.6, 1.4, 1.2, 1.0, 0.8
M3	Regularization coefficient μ	8, 4, 2, 1, 2^{-1}, 2^{-2}, 2^{-3}
	L_p coefficient p	1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0
M4	Penalty coefficient α	1, 10, 10^{2}, 10^{3}, 10^{4}, 10^{5}
	Penalty coefficient β	10, 10^{2}, 10^{3}, 10^{4}, 10^{5}, 10^{6}
M5	Penalty coefficient ρ	2^{-3}, 2^{-5}, 2^{-7}, 2^{-9}, 2^{-11}, 2^{-13}, 2^{-15}, 2^{-17}
M6	Regularization coefficient λ	10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.1, 1
	Learning rate η	10^{-5}, 5×10^{-5}, 10^{-4}, 5×10^{-4}, 10^{-3}, 5×10^{-3}, 10^{-2}
	Batch size bs	32, 64, 128, 256, 512, 1024
M7	Regularization coefficient λ	10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.1, 1
	Learning rate η	10^{-5}, 5×10^{-5}, 10^{-4}, 5×10^{-4}, 10^{-3}, 5×10^{-3}, 10^{-2}
	Batch size bs	32, 64, 128, 256, 512, 1024

TABLE S4. Optimal Hyper-parameters during M1's ten times' training process on D1-4

No.	**Type**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
D1	**RMSE**	$\lambda=2^{-4}$, $\eta=2$	$\lambda=2^{-4}$, $\eta=2$	$\lambda=2^{-4}$, $\eta=2$	$\lambda=2^{-2}$, $\eta=1$	$\lambda=2^{-4}$, $\eta=2$	$\lambda=2^{-4}$, $\eta=2$	$\lambda=2^{-4}$, $\eta=2$	$\lambda=2^{-4}$, $\eta=2$	$\lambda=2^{-2}$, $\eta=1$	$\lambda=2^{-4}$, $\eta=2$
	MAE	$\lambda=2^{-2}$, $\eta=1$	$\lambda=2^{-4}$, $\eta=1$	$\lambda=2^{-4}$, $\eta=1$	$\lambda=2^{-2}$, $\eta=1$	$\lambda=2^{-4}$, $\eta=1$	$\lambda=2^{-4}$, $\eta=1$	$\lambda=2^{-4}$, $\eta=1$	$\lambda=2^{-2}$, $\eta=1$	$\lambda=2^{-4}$, $\eta=1$	$\lambda=2^{-4}$, $\eta=1$
D2	**RMSE**	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2$
	MAE	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=2$
D3	**RMSE**	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2$
	MAE	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=2^{-2}$	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2^{-2}$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=2$
D4	**RMSE**	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=1$
	MAE	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=2$	$\lambda=1$, $\eta=1$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=2^{-1}$	$\lambda=1$, $\eta=2^{-1}$

TABLE S5. Optimal Hyper-parameters during M2's ten times' training process on D1-4

No.	**Type**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
D1	**RMSE**	$\lambda=1.6$	$\lambda=1.6$	$\lambda=1.6$	$\lambda=1.8$	$\lambda=1.6$	$\lambda=1.8$	$\lambda=1.6$	$\lambda=1.8$	$\lambda=1.6$	$\lambda=1.6$
	MAE	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.2$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.2$
D2	**RMSE**	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$
	MAE	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$
D3	**RMSE**	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$
	MAE	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$	$\lambda=1.8$
D4	**RMSE**	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure
	MAE	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure

[1] Note that M2 fails to achieve the final results on D4: their memory requirements are too large to meet on our experimental environment as shown in Table V.

TABLE S6. Optimal Hyper-parameters during M3's ten times' training process on D1-4

No.	**Type**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
D1	**RMSE**	$\mu=2^{-1}$, $p=1.9$	$\mu=2^{-1}$, $p=1.8$	$\mu=2^{-1}$, $p=1.7$	$\mu=2^{-1}$, $p=1.8$	$\mu=2^{-1}$, $p=1.8$	$\mu=2^{-1}$, $p=1.7$	$\mu=2^{-1}$, $p=1.8$	$\mu=2^{-1}$, $p=1.9$	$\mu=2^{-1}$, $p=1.8$	$\mu=2^{-1}$, $p=1.8$
	MAE	$\mu=2^{-1}$, $p=1.8$	$\mu=2^{-1}$, $p=1.9$	$\mu=2^{-1}$, $p=1.8$	$\mu=2^{-1}$, $p=1.7$	$\mu=2^{-1}$, $p=1.8$	$\mu=2^{-1}$, $p=1.7$	$\mu=2^{-1}$, $p=1.8$	$\mu=2^{-1}$, $p=1.8$	$\mu=2^{-1}$, $p=1.9$	$\mu=2^{-1}$, $p=1.9$
D2	**RMSE**	$\mu=2$, $p=1.9$	$\mu=4$, $p=1.9$	$\mu=4$ $p=1.9$	$\mu=2^{-1}$, $p=1.8$	$\mu=2^{-1}$, $p=1.9$	$\mu=2$, $p=1.9$	$\mu=2^{-1}$, $p=1.8$	$\mu=2$, $p=1.9$	$\mu=2$, $p=1.8$	$\mu=2$, $p=1.8$
	MAE	$\mu=4$, $p=1.9$	$\mu=2$, $p=1.9$	$\mu=2$, $p=1.8$	$\mu=4$, $p=1.9$	$\mu=4$, $p=1.9$	$\mu=2$, $p=1.9$	$\mu=1$, $p=1.9$	$\mu=2$, $p=1.8$	$\mu=2$, $p=1.8$	$\mu=2$, $p=1.9$
D3	**RMSE**	$\mu=4$, $p=1.9$	$\mu=4$, $p=1.9$	$\mu=2$, $p=1.8$	$\mu=2$, $p=1.8$	$\mu=4$, $p=1.9$	$\mu=4$, $p=1.9$	$\mu=2$, $p=1.9$	$\mu=2$, $p=1.9$	$\mu=2$, $p=1.9$	$\mu=2$, $p=1.9$
	MAE	$\mu=2$, $p=1.9$	$\mu=2$, $p=1.8$	$\mu=2$, $p=1.9$	$\mu=2$, $p=1.8$	$\mu=4$, $p=1.9$	$\mu=4$, $p=1.9$	$\mu=4$, $p=1.9$	$\mu=2$, $p=1.8$	$\mu=2$, $p=1.9$	$\mu=2$, $p=1.9$
D4	**RMSE**	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure
	MAE	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure

[1] Note that M3 fails to achieve the final results on D4: their memory requirements are too large to meet on our experimental environment as shown in Table V.

TABLE S7. Optimal Hyper-parameters during M4's ten times' training process on D1-4

No.	**Type**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
D1	**RMSE**	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$

	MAE	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^3$, $\beta=10^4$
D2	**RMSE**	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^4$, $\beta=10^2$	$\alpha=10^3$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^3$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^2$	$\alpha=10^3$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^4$, $\beta=10^2$	$\alpha=10^3$, $\beta=10^3$
	MAE	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^3$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^3$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^3$, $\beta=10^3$
D3	**RMSE**	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^4$, $\beta=10^3$
	MAE	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^3$, $\beta=10^4$	$\alpha=10^4$, $\beta=10^3$	$\alpha=10^3$, $\beta=10^4$
D4	**RMSE**	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure
	MAE	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure

[1] Note that M4 fails to achieve the final results on D4: their memory requirements are too large to meet on our experimental environment as shown in Table V.

TABLE S8. Optimal Hyper-parameters during M5's ten times' training process on D1-4

No.	**Type**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
D1	**RMSE**	$\rho=2^{-5}$	$\rho=2^{-5}$	$\rho=2^{-5}$	$\rho=2^{-5}$	$\rho=2^{-5}$	$\rho=2^{-5}$	$\rho=2^{-5}$	$\rho=2^{-5}$	$\rho=2^{-5}$	$\rho=2^{-5}$
	MAE	$\rho=2^{-7}$	$\rho=2^{-5}$	$\rho=2^{-7}$	$\rho=2^{-5}$	$\rho=2^{-5}$	$\rho=2^{-7}$	$\rho=2^{-5}$	$\rho=2^{-5}$	$\rho=2^{-7}$	$\rho=2^{-5}$
D2	**RMSE**	$\rho=2^{-15}$	$\rho=2^{-13}$	$\rho=2^{-15}$	$\rho=2^{-13}$	$\rho=2^{-15}$	$\rho=2^{-13}$	$\rho=2^{-15}$	$\rho=2^{-15}$	$\rho=2^{-15}$	$\rho=2^{-13}$
	MAE	$\rho=2^{-13}$	$\rho=2^{-15}$	$\rho=2^{-15}$	$\rho=2^{-15}$	$\rho=2^{-13}$	$\rho=2^{-15}$	$\rho=2^{-13}$	$\rho=2^{-13}$	$\rho=2^{-15}$	$\rho=2^{-15}$
D3	**RMSE**	$\rho=2^{-13}$	$\rho=2^{-15}$	$\rho=2^{-13}$	$\rho=2^{-15}$	$\rho=2^{-13}$	$\rho=2^{-15}$	$\rho=2^{-13}$	$\rho=2^{-13}$	$\rho=2^{-15}$	$\rho=2^{-13}$
	MAE	$\rho=2^{-13}$	$\rho=2^{-15}$	$\rho=2^{-13}$	$\rho=2^{-13}$	$\rho=2^{-15}$	$\rho=2^{-13}$	$\rho=2^{-13}$	$\rho=2^{-15}$	$\rho=2^{-15}$	$\rho=2^{-13}$
D4	**RMSE**	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure
	MAE	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure	[1]Failure

[1] Note that M5 fails to achieve the final results on D4: their memory requirements are too large to meet on our experimental environment as shown in Table V.

TABLE S9. Optimal Hyper-parameters during M6's ten times' training process on D1-4

No.	**Type**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
D1	**RMSE**	$\lambda=10^{-4}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-4}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-4}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=512
	MAE	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512
D2	**RMSE**	$\lambda=10^{-4}$, $\eta=10^{-3}$, bs=64	$\lambda=10^{-4}$, $\eta=10^{-3}$, bs=64	$\lambda=10^{-1}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-1}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-2}$, $\eta=10^{-3}$, bs=64	$\lambda=10^{-1}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-2}$, $\eta=10^{-3}$, bs=64	$\lambda=10^{-2}$, $\eta=10^{-3}$, bs=64	$\lambda=10^{-1}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-4}$, $\eta=10^{-3}$, bs=64
	MAE	$\lambda=10^{-4}$, $\eta=10^{-2}$, bs=64	$\lambda=10^{-3}$, $\eta=10^{-2}$, bs=64	$\lambda=10^{-4}$, $\eta=10^{-2}$, bs=64	$\lambda=10^{-3}$, $\eta=10^{-2}$, bs=64	$\lambda=10^{-4}$, $\eta=10^{-2}$, bs=64	$\lambda=10^{-2}$, $\eta=10^{-3}$, bs=64	$\lambda=10^{-4}$, $\eta=10^{-2}$, bs=64	$\lambda=10^{-2}$, $\eta=10^{-3}$, bs=64	$\lambda=10^{-2}$, $\eta=10^{-3}$, bs=64	$\lambda=10^{-2}$, $\eta=10^{-3}$, bs=64
D3	**RMSE**	$\lambda=10^{-4}$, $\eta=5\times10^{-4}$, bs=128	$\lambda=10^{-4}$, $\eta=5\times10^{-4}$, bs=128	$\lambda=10^{-4}$, $\eta=5\times10^{-4}$, bs=128	$\lambda=10^{-4}$, $\eta=5\times10^{-4}$, bs=128	$\lambda=10^{-4}$, $\eta=5\times10^{-4}$, bs=128	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=128	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=128	$\lambda=10^{-4}$, $\eta=5\times10^{-4}$, bs=128	$\lambda=10^{-4}$, $\eta=5\times10^{-4}$, bs=128	$\lambda=10^{-4}$, $\eta=5\times10^{-4}$, bs=128

	MAE	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=128	$\lambda=10^{-4}$, $\eta=10^{-3}$, bs=128	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=128	$\lambda=10^{-5}$, $\eta=10^{-3}$, bs=128	$\lambda=10^{-4}$, $\eta=10^{-3}$, bs=128	$\lambda=10^{-4}$, $\eta=10^{-3}$, bs=128	$\lambda=10^{-4}$, $\eta=10^{-3}$, bs=128	$\lambda=10^{-4}$, $\eta=10^{-3}$, bs=128	$\lambda=10^{-4}$, $\eta=10^{-3}$, bs=128	$\lambda=10^{-4}$, $\eta=10^{-3}$, bs=128
D4	**RMSE**	$\lambda=10^{-3}$, $\eta=10^{-4}$, bs=512	$\lambda=10^{-4}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-5}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-4}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-4}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-3}$, $\eta=10^{-4}$, bs=512	$\lambda=10^{-5}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-3}$, $\eta=10^{-4}$, bs=512	$\lambda=10^{-5}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-3}$, $\eta=10^{-4}$, bs=512
	MAE	$\lambda=10^{-2}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-5}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-5}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512

bs denotes batch size adopted by M6 on an HDI matrix;

TABLE S10. Optimal Hyper-parameters during M7's ten times' training process on D1-4

No.	**Type**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
D1	**RMSE**	$\lambda=10^{-2}$, $\eta=10^{-4}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=10^{-4}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-4}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-4}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-4}$, bs=512	$\lambda=10^{-2}$, $\eta=10^{-4}$, bs=512
	MAE	$\lambda=10^{-2}$, $\eta=10^{-4}$, bs=512	$\lambda=10^{-2}$, $\eta=10^{-4}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=5\times10^{-3}$, bs=512	$\lambda=10^{-2}$, $\eta=10^{-4}$, bs=512	$\lambda=10^{-2}$, $\eta=10^{-4}$, bs=512	$\lambda=10^{-2}$, $\eta=10^{-4}$, bs=512	$\lambda=10^{-2}$, $\eta=10^{-4}$, bs=512	$\lambda=10^{-2}$, $\eta=10^{-4}$, bs=512
D2	**RMSE**	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=10^{-4}$, bs=64	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=10^{-4}$, bs=64	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=10^{-4}$, bs=64
	MAE	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=64	$\lambda=10^{-3}$, $\eta=5\times10^{-3}$, bs=64
D3	**RMSE**	$\lambda=0.1$, $\eta=10^{-3}$, bs=128	$\lambda=0.1$, $\eta=5\times10^{-3}$, bs=128	$\lambda=0.1$, $\eta=10^{-3}$, bs=128	$\lambda=0.1$, $\eta=10^{-3}$, bs=128	$\lambda=0.1$, $\eta=10^{-3}$, bs=128	$\lambda=0.1$, $\eta=5\times10^{-3}$, bs=128	$\lambda=0.1$, $\eta=10^{-3}$, bs=128	$\lambda=0.1$, $\eta=10^{-3}$, bs=128	$\lambda=0.1$, $\eta=10^{-3}$, bs=128	$\lambda=0.1$, $\eta=10^{-3}$, bs=128
	MAE	$\lambda=0.1$, $\eta=10^{-3}$, bs=128	$\lambda=0.1$, $\eta=10^{-4}$, bs=128	$\lambda=0.1$, $\eta=10^{-4}$, bs=128	$\lambda=0.1$, $\eta=5\times10^{-3}$, bs=128	$\lambda=0.1$, $\eta=5\times10^{-3}$, bs=128	$\lambda=0.1$, $\eta=10^{-3}$, bs=128	$\lambda=0.1$, $\eta=10^{-4}$, bs=128	$\lambda=0.1$, $\eta=10^{-3}$, bs=128	$\lambda=0.1$, $\eta=10^{-3}$, bs=128	$\lambda=0.1$, $\eta=10^{-3}$, bs=128
D4	**RMSE**	$\lambda=0.1$, $\eta=10^{-3}$, bs=512	$\lambda=0.1$, $\eta=5\times10^{-3}$, bs=512	$\lambda=0.1$, $\eta=5\times10^{-4}$, bs=512	-	-	-	-	-	-	-
	MAE	$\lambda=0.1$, $\eta=10^{-3}$, bs=512	$\lambda=0.1$, $\eta=10^{-3}$, bs=512	$\lambda=0.1$, $\eta=5\times10^{-3}$, bs=512	-	-	-	-	-	-	-

Note that M7 need to consume about 18 days for manually implementing the tuning process of its hyper-parameters on D4. Hence, only third times validation process for M7 on D4 is completed; bs denotes batch size adopted by M7 on an HDI matrix.

TABLE S11. RMSE and Time Cost of M1 and M8 on D1-4.

Dataset	Model	Prediction Accuracy		Tuning Time Cost (Secs)*	Testing Time Cost (Secs)**	Total Time Cost (Secs)
D1	M1	RMSE	0.2373±2.2E-6	428±22.7	6±2.4	434±25
		MAE	0.1815±1.1E-6	439±25.4	6±2.8	445±28
	M8	RMSE	**0.2339±2.7E-4**	-	-	**46±4**
		MAE	**0.1792±3.3E-4**	-	-	**60±5**
D2	M1	RMSE	1.0187±1.1E-6	271±45.4	4±1.7	275±47
		MAE	0.8079±2.5E-6	305±38.3	3±0.9	308±39
	M8	RMSE	**1.0172±7.4E-4**	-	-	**25±5**
		MAE	**0.7856±7.9E-4**	-	-	**29±8**
D3	M1	RMSE	**0.8665±7.8E-4**	739±72.3	21±14.7	759±87
		MAE	0.6829±1.7E-6	756±53.8	3±0.5	763±54
	M8	RMSE	0.8675±8.2E-4	-	-	**26±6**
		MAE	**0.6787±7.3E-4**	-	-	**30±9**
D4	M1	RMSE	0.8096±2.9E-6	2,934±353.8	38±24.6	2,972±378
		MAE	0.6221±7.9E-7	9,203±828.9	33±20.4	9,236±849
	M8	RMSE	**0.8089±9.4E-4**	-	-	**334±46**
		MAE	**0.6193±5.1E-4**	-	-	**358±31**

* Time cost consumed by M1 for manually grid-searching optimal hyper-parameters;

** Time cost consumed by M1 with achieved hyper-parameters.

TABLE S12. Storage Complexity and Memory Requirements of M1-8 on D1-4.

No.	Storage Complexity	Memory Requirements (GB)			
		D1	D2	D3	D4
M1	$\Theta((\lvert U\rvert+\lvert I\rvert)\times d+\lvert\Lambda\rvert)$	**0.7**	**0.6**	**0.5**	**1.7**
M2	$\Theta(\lvert U\rvert\times\lvert I\rvert)$	17.6	1.8	9.3	>256
M3	$\Theta(\lvert U\rvert\times\lvert I\rvert)$	15.4	1.9	8.6	>256
M4	$\Theta(\lvert U\rvert\times\lvert I\rvert)$	21.9	3.5	11.5	>256
M5	$\Theta(\lvert U\rvert\times\lvert I\rvert)$	26.5	3.9	12.7	>256
M6	$\Theta(\mathrm{bs}^{*}\times\lvert U\rvert+\lvert\Lambda\rvert)$	2.9	1.7	1.7	3.1
M7	$\Theta((\lvert U\rvert+\lvert I\rvert)\times d+\mathrm{bs}\times 1^{\mathrm{st}}\mathrm{LN}^{**}+\lvert\Lambda\rvert)$	2.1	1.2	1.8	6.5
M8	$\Theta((\lvert U\rvert+\lvert I\rvert)\times d+\lvert\Lambda\rvert)$	**0.7**	**0.6**	**0.5**	**1.7**

* bs denotes batch size adopted by M6 and M7 on an HDI matrix;

** 1^{st}LN denotes the number of neuron in the 1^{st} layer and is set at $2\times d$ according to [10].

TABLE S13. RMSE/MAE of M1-8 on D1-4, including Win/Loss counts and Friedman Rank, where ● indicates that M8 has higher RMSE/MAE than the rival models

No.	CASE	M1	M2	M3	M4	M5	M6	M7	M8
D1	RMSE	0.2373$_{\pm 2.2E-6}$	0.3058$_{\pm 5.0E-4}$	0.3047$_{\pm 4.3E-5}$	0.2384$_{\pm 1.0E-4}$	0.4913$_{\pm 1.0E-4}$	●**0.2302$_{\pm 2.6E-3}$**	0.2354$_{\pm 2.3E-3}$	0.2339$_{\pm 2.7E-4}$
	MAE	0.1815$_{\pm 1.1E-6}$	0.2439$_{\pm 4.4E-5}$	0.2422$_{\pm 3.2E-5}$	0.1832$_{\pm 4.8E-5}$	0.4111$_{\pm 1.6E-2}$	●**0.1792$_{\pm 7.9E-5}$**	0.1842$_{\pm 2.2E-4}$	0.1793$_{\pm 3.3E-4}$
D2	RMSE	1.0187$_{\pm 1.1E-6}$	1.1281$_{\pm 7.2E-3}$	1.1257$_{\pm 1.8E-4}$	1.0787$_{\pm 8.6E-7}$	1.8808$_{\pm 2.8E-2}$	1.1494$_{\pm 4.1E-5}$	1.0425$_{\pm 3.9E-4}$	**1.0172$_{\pm 7.4E-4}$**
	MAE	0.8079$_{\pm 2.5E-6}$	0.9256$_{\pm 1.5E-3}$	0.9229$_{\pm 1.9E-4}$	0.8580$_{\pm 6.2E-6}$	1.5403$_{\pm 2.6E-2}$	0.9257$_{\pm 2.1E-4}$	0.8014$_{\pm 4.9E-2}$	**0.7856$_{\pm 7.9E-4}$**
D3	RMSE	●0.8665$_{\pm 7.8E-4}$	1.0336$_{\pm 5.4E-4}$	1.0848$_{\pm 5.3E-5}$	0.8713$_{\pm 3.7E-4}$	2.2963$_{\pm 9.8E-3}$	0.8845$_{\pm 1.1E-3}$	0.8982$_{\pm 4.4E-4}$	0.8675$_{\pm 8.2E-4}$
	MAE	0.6829$_{\pm 1.7E-6}$	0.8832$_{\pm 4.0E-4}$	0.9021$_{\pm 4.2E-5}$	0.6802$_{\pm 3.3E-4}$	1.9190$_{\pm 9.6E-3}$	0.7021$_{\pm 5.5E-3}$	0.7067$_{\pm 4.1E-4}$	**0.6787$_{\pm 7.3E-4}$**
D4	RMSE	0.8096$_{\pm 2.9E-6}$	[1]Failure	[1]Failure	[1]Failure	[1]Failure	0.8436$_{\pm 1.2E-3}$	[a]0.8657$_{\pm 3.5E-4}$	**0.8089$_{\pm 9.4E-4}$**
	MAE	0.6221$_{\pm 7.9E-7}$	[1]Failure	[1]Failure	[1]Failure	[1]Failure	0.6530$_{\pm 4.9E-3}$	[a]0.6650$_{\pm 2.9E-4}$	**0.6193$_{\pm 5.1E-4}$**
Win/Loss		7/1	8/0	8/0	8/0	8/0	6/2	8/0	-
Friedman Rank		2.5	6.38	6.13	4.38	7.63	3.75	3.88	**1.38**

[1] Note that M2-M5 fails to achieve the final results on D4: their memory requirements are too large to meet on our experimental environment as shown in Table S12.

[a] Note that M7 need to consume about 18 days for manually implementing the tuning process of its hyper-parameters on D4. Hence, only third times validation process for M7 on D4 is completed, and the corresponding performance is marked as '[a]' on D4.

TABLE S14. Total time cost of M1-8 in RMSE/MAE on D1-4 (Secs), including Win/Loss counts and Friedman Rank

No.	CASE	M1	M2	M3	M4	M5	M6	M7	M8
D1	RMSE	434$_{\pm 25}$	38,499$_{\pm 2,332}$	43,271$_{\pm 2,962}$	209,670$_{\pm 36,625}$	3,856$_{\pm 578}$	10,804$_{\pm 832}$	87,425$_{\pm 9,048}$	**46$_{\pm 4}$**
	MAE	445$_{\pm 28}$	51,443$_{\pm 2,643}$	189,090$_{\pm 29,135}$	210,745$_{\pm 27,190}$	3,912$_{\pm 103}$	13,432$_{\pm 2,219}$	78,395$_{\pm 8,698}$	**60$_{\pm 5}$**
D2	RMSE	275$_{\pm 47}$	236$_{\pm 35}$	128$_{\pm 21}$	138$_{\pm 39}$	55$_{\pm 7}$	2,109$_{\pm 295}$	1,264$_{\pm 99}$	**25$_{\pm 5}$**
	MAE	308$_{\pm 39}$	32$_{\pm 3}$	128$_{\pm 27}$	766$_{\pm 64}$	56$_{\pm 6}$	2,256$_{\pm 633}$	1,156$_{\pm 286}$	**29$_{\pm 8}$**
D3	RMSE	759$_{\pm 87}$	42,578$_{\pm 3,258}$	71,927$_{\pm 2,718}$	45,849$_{\pm 2,479}$	717$_{\pm 22}$	8,359$_{\pm 532}$	9,921$_{\pm 2,586}$	**26$_{\pm 6}$**
	MAE	763$_{\pm 54}$	1,906$_{\pm 291}$	12,294$_{\pm 2,478}$	48,821$_{\pm 2,363}$	716$_{\pm 53}$	7,700$_{\pm 219}$	8,998$_{\pm 1,446}$	**30$_{\pm 9}$**
D4	RMSE	2,972$_{\pm 378}$	[1]Failure	[1]Failure	[1]Failure	[1]Failure	266,885$_{\pm 12,775}$	[a]1,644,502$_{\pm 98,039}$	**334$_{\pm 46}$**
	MAE	9,236$_{\pm 849}$	[1]Failure	[1]Failure	[1]Failure	[1]Failure	232,974$_{\pm 13,157}$	[a]997,259$_{\pm 84,311}$	**358$_{\pm 31}$**
Win/Loss		8/0	8/0	8/0	8/0	8/0	8/0	8/0	8/0
Friedman Rank		3.13	5	6	6.75	3.5	4.88	5.75	**1**

[1] It means the same with that in Table S13;

[a] It means the same with that in Table S13.

TABLE S15. Results of the Wilcoxon signed-ranks test on RMSE/MAE of Table S13.

Comparison	*R+*	*R-*	*p-value**
M8 vs M1	34	2	**0.0117**
M8 vs M2	36	0	**0.0039**
M8 vs M3	36	0	**0.0039**
M8 vs M4	36	0	**0.0039**
M8 vs M5	36	0	**0.0039**
M8 vs M6	33	3	**0.0195**
M8 vs M7	36	0	**0.0039**

***With a significance level of 0.1, the accepted hypotheses are highlighted.

TABLE S16. Results of the Wilcoxon signed-ranks test on time cost of Table S14.

Comparison	*R+*	*R-*	*p-value**
M8 vs M1	36	0	**0.0039**
M8 vs M2	36	0	**0.0039**
M8 vs M3	36	0	**0.0039**
M8 vs M4	36	0	**0.0039**
M8 vs M5	36	0	**0.0039**
M8 vs M6	36	0	**0.0039**
M8 vs M7	36	0	**0.0039**

***With a significance level of 0.1, the accepted hypotheses are highlighted.