A Fuzzy Nesterov-Accelerated Nonnegative Latent Factorization of Tensors Model for Efficient Representation to Dynamic Directed Graph Supplementary File

Anonymous

I. Introduction

THIS is the supplementary file for the paper entitled "A Fuzzy Nesterov-Accelerated Nonnegative Latent Factorization of tensors Model for Efficient Representation to Dynamic Directed Graph". We have put the proof of Lemma 1-2 and Theorem 1-2 in Section II, tables and figures of experimental results in Section IV.

II. Model Convergence Analysis

The goal of the FNL model is to represent a HDI tensor $\mathbf{Y}$ via a nonnegative latent factorization. The optimization problem is defined as:

$$\min_{\mathrm{U,W,Z}} \varepsilon\left(\mathrm{U,W,Z}\right) = \sum_{(i,j,k \in \Lambda)} \left(y_{ijk} - \hat{y}_{ijk}\right)^2 + \lambda^t \left(\sigma(u_{ir})^2 + \sigma(w_{jr})^2 + \sigma(z_{kr})^2\right), \tag{S1}$$

where $\hat{y}_{ijk} = \sum_{r=1}^{R} \sigma\left(u_{ir}\right)\sigma\left(w_{jr}\right)\sigma\left(z_{kr}\right)$ and $\sigma\left(*\right)$ represents Hard-Sigmoid function:

$$\sigma\left(\theta\right) = \max(0, \min(1, 0.2\theta + 0.5)), \tag{S2}$$

where Λ and λ_t represent Known set and regularization coefficient on t-th iteration respectively.

For the convenience of analysis, we alternatively fix the counterpart of the active parameter as constant, i.e., we treat w_{jr} and z_{kr} as constant when performing the analyses with u_{ir}. The gradient of u_{ir}:

$$\nabla_{u_{ir}}\varepsilon_{i,j,k} = -\left(y_{ijk} - \sum_{r=1}^{R}\sigma(u_{ir})\sigma(w_{jr})\sigma(z_{kr})\right)\sigma\prime(u_{ir})\sigma(w_{jr})\sigma(z_{kr}) + \lambda^t u_{ir}. \tag{S3}$$

Note that the update of $u_i r$ by NASGD is given as: For the t-th iteration counts.

$$\begin{aligned} t = 1 : u_{ir}^{t+1} &= u_{ir}^{t} - \eta^t\left(1+\gamma^t\right)\nabla_{u_{ir}}\varepsilon_{i,j,k}\left(u_{ir}^{t}\right) \\ t \geq 2 : u_{ir}^{t+1} &= u_{ir}^{t} - \eta^t\nabla_{u_{ir}}\varepsilon_{i,j,k}\left(u_{ir}^{t}\right) \\ &\quad + \gamma^t\left(\left(u_{ir}^{t} - \eta^t\nabla_{u_{ir}}\varepsilon_{i,j,k}\left(u_{ir}^{t}\right)\right) - \left(u_{ir}^{t-1} - \eta^t\nabla_{u_{ir}}\varepsilon_{i,j,k}\left(u_{ir}^{t-1}\right)\right)\right). \end{aligned} \tag{S4}$$

It is worth noting that according to the fuzzy reasoning rule table (Table II), we can get:

$$\begin{aligned} &\eta^t \in [\eta_{min}, \eta_{max}], 0 < \eta_{min} \leq \eta^t \leq \eta_{max}, \\ &\lambda^t \in [\lambda_{min}, \lambda_{max}], 0 < \lambda_{min} \leq \lambda^t \leq \lambda_{max}, \\ &\gamma^t \in [\gamma_{min}, \gamma_{max}], 0 < \gamma_{min} \leq \gamma^t \leq \gamma_{max} < 1. \end{aligned} \tag{S5}$$

We define w_t as:

$$\omega^t = \begin{cases} \frac{\gamma^t}{1-\gamma^t}\left(u_{ir}^t - u_{ir}^{t-1} + \eta^t \nabla_{u_{ir}} \varepsilon_{i,j,k}\left(u_{ir}^{t-1}\right)\right), t \geq 1 \\ 0, t = 0 \end{cases}, \tag{S6}$$

where $\eta^t, \lambda^t, \gamma^t$ are dynamically adjusted according to the fuzzy reasoning $I^{(t)}$ = 0.5(RMSE+MAE), we present the following results.

Lemma 1: $\forall t \geq 0$,the following equation stands:

$$\begin{aligned}
u_{ir}^{t+1} + \omega^{t+1} &= u_{ir}^{t+1} + \frac{\gamma^{t+1}}{1-\gamma^{t+1}}\left(u_{ir}^{t+1} - u_{ir}^t + \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t)\right) \\
&= u_{ir}^t - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \gamma^t \left(\left(u_{ir}^t - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t)\right) - \left(u_{ir}^{t-1} - \eta^{t-1} \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^{t-1})\right)\right) \\
&\quad + \frac{\gamma^{t+1}}{1-\gamma^{t+1}}\left(u_{ir}^{t+1} - u_{ir}^t + \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t)\right) \\
&= u_{ir}^t - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \gamma^t \left(u_{ir}^t - u_{ir}^{t-1} - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \eta^{t-1} \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^{t-1})\right) \\
&\quad + \frac{\gamma^{t+1}}{1-\gamma^{t+1}}\gamma^t \left(u_{ir}^t - u_{ir}^{t-1} - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \eta^{t-1} \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^{t-1})\right) \\
&= u_{ir}^t - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \left(\gamma^t + \frac{\gamma^{t+1}\gamma^t}{1-\gamma^{t+1}}\right)\left(u_{ir}^t - u_{ir}^{t-1} - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \eta^{t-1} \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^{t-1})\right) \\
&= u_{ir}^t - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \gamma^t \left(1 + \frac{\gamma^{t+1}}{1-\gamma^{t+1}}\right)\left(u_{ir}^t - u_{ir}^{t-1} - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \eta^{t-1} \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^{t-1})\right) \\
&= u_{ir}^t - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \gamma^t \frac{1-\gamma^{t+1}+\gamma^{t+1}}{1-\gamma^{t+1}}\left(u_{ir}^t - u_{ir}^{t-1} - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \eta^{t-1} \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^{t-1})\right) \\
&= u_{ir}^t - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \frac{\gamma^t}{1-\gamma^{t+1}}\left(u_{ir}^t - u_{ir}^{t-1} - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \eta^{t-1} \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^{t-1})\right) \\
&= u_{ir}^t - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \frac{\gamma^t}{1-\gamma^t}\left(u_{ir}^t - u_{ir}^{t-1} - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) + \eta^{t-1} \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^{t-1})\right) \\
&= u_{ir}^t + \frac{\gamma^t}{1-\gamma^t}\left(u_{ir}^t - u_{ir}^{t-1} + \eta^{t-1} \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^{t-1})\right) - \eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) - \frac{\gamma^t}{1-\gamma^t}\eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t) \\
&= u_{ir}^t + \omega^t - \frac{\gamma^t}{1-\gamma^t}\eta^t \nabla_{u_{ir}} \varepsilon_{ijk}(u_{ir}^t).
\end{aligned} \tag{S7}$$

Lemma 2: Let $u_{ir}^{-1} = u_{ir}^0$, for any t $t \geq 0, \exists u_{ir}^*$:

$$\begin{aligned}
&\left\|u_{ir}^{t+1} + \omega^{t+1} - u_{ir}^*\right\|^2 \\
=& \left\|u_{ir}^t + \omega^t - \frac{\eta^t}{1-\gamma^t}\nabla_{u_{ir}} \varepsilon_{ijk}\left(u_{ir}^t\right) - u_{ir}^*\right\|^2 \\
=& \left\|\left(u_{ir}^t + \omega^t - u_{ir}^*\right) - \frac{\eta^t}{1-\gamma^t}\nabla_{u_{ir}} \varepsilon_{ijk}\left(u_{ir}^t\right)\right\|^2 \\
=& \left(u_{ir}^t + \omega^t - u_{ir}^*\right)^2 - 2\left\langle\left(u_{ir}^t + \omega^t - u_{ir}^*\right)\frac{\eta^t}{1-\gamma^t}\nabla_{u_{ir}} \varepsilon_{ijk}\left(u_{ir}^t\right)\right\rangle + \left(\frac{\eta^t}{1-\gamma^t}\nabla_{u_{ir}} \varepsilon_{ijk}\left(u_{ir}^t\right)\right)^2 \\
=& \left\|u_{ir}^t + \omega^t - u_{ir}^*\right\|^2 - 2\left\langle\left(u_{ir}^t + \omega^t - u_{ir}^*\right)\frac{\eta^t}{1-\gamma^t}\nabla_{u_{ir}} \varepsilon_{ijk}\left(u_{ir}^t\right)\right\rangle + \left(\frac{\eta^t}{1-\gamma^t}\right)^2\left\|\nabla_{u_{ir}} \varepsilon_{ijk}\left(u_{ir}^t\right)\right\|^2.
\end{aligned} \tag{S8}$$

Note that (S7) can yield the appearance of u_{ir} when $t = 0$. By setting $u_{ir}^{-1} = u_{ir}^0$ the above inequality still holds. Hence, Lemma 2 stands.

Based on Lemmas 1-2, we present the following important result.

Theorem 1 (Convergence of NASGD):for any $t \geq 0$, let $\|\nabla u_{ir} \varepsilon_{ijk}\left(u_{ir}^t\right)\| \leq G, \left|u_{ir}^{t+1} - u_{ir}^t\right| \leq A$, where G and A are both positive constants, by setting $B = A + 2\eta_{max}G$ as B be a positive constant, when $t \in \{0, ..., T\}$, based on Lemma 2, we have:

$$
\begin{aligned}
\left\|u_{ir}^{t+1}+\omega^{t+1}-u_{ir}^{*}\right\|^{2}=&\left\|u_{ir}^{t}+\omega^{t}-u_{ir}^{*}\right\|^{2}-2\left\langle\left(u_{ir}^{t}+\omega^{t}-u_{ir}^{*}\right)\frac{\eta^{t}}{1-\gamma^{t}}\nabla_{u_{ir}}\varepsilon_{ijk}\left(u_{ir}^{t}\right)\right\rangle \\
&+\left(\frac{\eta^{t}}{1-\gamma^{t}}\right)^{2}\left\|\nabla_{u_{ir}}\varepsilon_{ijk}\left(u_{ir}^{t}\right)\right\|^{2}.
\end{aligned} \tag{S9}
$$

First of all, function $\varepsilon_{ijk}\left(u_{ir}\right)$ is convex, therefore, about $\left(u_{ir}^{t}+\omega^{t}-u_{ir}^{*}\right)\nabla_{u_{ir}}\varepsilon_{ijk}\left(u_{ir}^{t}\right)$, we have:

$$
\begin{aligned}
&\left(u_{ir}^{t}+\omega^{t}-u_{ir}^{*}\right)\nabla_{u_{ir}}\varepsilon_{ijk}\left(u_{ir}^{t}\right)\\
&=\left(u_{ir}^{t}-u_{ir}^{*}\right)\nabla_{u_{ir}}\varepsilon_{ijk}\left(u_{ir}^{t}\right)+\omega^{t}\nabla_{u_{ir}}\varepsilon_{ijk}\left(u_{ir}^{t}\right)\\
&\geq\varepsilon_{ijk}\left(u_{ir}^{t}\right)-\varepsilon_{ijk}\left(u_{ir}^{*}\right)+\frac{\gamma^{t}}{1-\gamma^{t}}\left(u_{ir}^{t}-u_{ir}^{t-1}+\eta^{t-1}\nabla_{u_{ir}}\varepsilon_{ijk}\left(u_{ir}^{t-1}\right)\right)\nabla_{u_{ir}}\varepsilon_{ijk}\left(u_{ir}^{t}\right).
\end{aligned} \tag{S10}
$$

Secondly, since y_{ijk}, $\sigma\left(*\right)$, λ_{t} are bounded, therefore, setting $\left\|\nabla u_{ir}\varepsilon_{ijk}\left(u_{ir}^{t}\right)\right\|\leq G$, we have:

$$
\begin{aligned}
&\left|\omega^{t}\nabla_{u_{ir}}\varepsilon_{ijk}\left(u_{ir}^{t}\right)\right|\\
&\leq\frac{\gamma^{t}}{1-\gamma^{t}}G\left|u_{ir}^{t}-u_{ir}^{t-1}+\eta^{t-1}\nabla_{u_{ir}}\varepsilon_{ijk}\left(u_{ir}^{t-1}\right)\right|\\
&\leq\frac{\gamma^{t}}{1-\gamma^{t}}G\gamma^{t-1}\left|u_{ir}^{t-1}-u_{ir}^{t-2}+\eta^{t-2}\nabla_{u_{ir}}\varepsilon_{ijk}\left(u_{ir}^{t-2}\right)-\eta^{t-1}\nabla_{u_{ir}}\varepsilon_{ijk}\left(u_{ir}^{t-1}\right)\right|\\
&\leq\frac{\gamma_{\max}}{1-\gamma_{\max}}G\left(A+2\eta_{\max}G\right).
\end{aligned} \tag{S11}
$$

Finally, substitute S(10) and S(11) into S(9) to get:

$$
\begin{aligned}
\left\|u_{ir}^{t+1}+\omega^{t+1}-u_{ir}^{*}\right\|^{2}\leq&\left\|u_{ir}^{t}+\omega^{t}-u_{ir}^{*}\right\|^{2}-2\frac{\eta^{t}}{1-\gamma^{t}}\left(\varepsilon_{ijk}\left(u_{ir}^{t}\right)-\varepsilon_{ijk}\left(u_{ir}^{*}\right)\right)\\
&+2\frac{\eta^{t}}{1-\gamma^{t}}\frac{\gamma_{\max}}{1-\gamma_{\max}}GB+\left(\frac{\eta^{t}}{1-\gamma^{t}}\right)^{2}G^{2}.
\end{aligned} \tag{S12}
$$

Accumulate from $t=0$ to T:

$$
\begin{aligned}
\left\|u_{ir}^{T+1}+\omega^{T+1}-u_{ir}^{*}\right\|^{2}\leq&\left\|u_{ir}^{0}+\omega^{0}-u_{ir}^{*}\right\|^{2}-2\sum_{t=0}^{T}\left(\frac{\eta^{t}}{1-\gamma^{t}}\left(\varepsilon_{ijk}\left(u_{ir}^{t}\right)-\varepsilon_{ijk}\left(u_{ir}^{*}\right)\right)\right)\\
&+2\frac{\gamma_{\max}}{1-\gamma_{\max}}GB\sum_{t=0}^{T}\left(\frac{\eta^{t}}{1-\gamma^{t}}\right)+G^{2}\sum_{t=0}^{T}\left(\left(\frac{\eta^{t}}{1-\gamma^{t}}\right)^{2}\right).
\end{aligned} \tag{S13}
$$

According to $\omega^{0}=0,\left\|u_{ir}^{t+1}+\omega_{ir}^{t+1}-u_{ir}^{*}\right\|^{2}\geq 0:$

$$
\begin{aligned}
&\sum_{t=0}^{T}\left(\frac{\eta^{t}}{1-\gamma^{t}}\left(\varepsilon_{ijk}\left(u_{ir}^{t}\right)-\varepsilon_{ijk}\left(u_{ir}^{*}\right)\right)\right)\\
&\leq\frac{\left\|u_{ir}^{0}-u_{ir}^{*}\right\|^{2}}{2}+\frac{\gamma_{\max}}{1-\gamma_{\max}}GB\sum_{t=0}^{T}\left(\frac{\eta^{t}}{1-\gamma^{t}}\right)+\frac{1}{2}G^{2}\sum_{t=0}^{T}\left(\left(\frac{\eta^{t}}{1-\gamma^{t}}\right)^{2}\right)\\
&\leq\frac{\left\|u_{ir}^{0}-u_{ir}^{*}\right\|^{2}}{2}+\frac{\eta_{\max}}{1-\gamma_{\max}}\frac{\gamma_{\max}}{1-\gamma_{\max}}GB\left(T+1\right)+\frac{1}{2}G^{2}\frac{\eta_{\max}}{1-\gamma_{\max}}\left(T+1\right).
\end{aligned} \tag{S14}
$$

And then,

$$
\begin{aligned}
&\frac{1}{T+1}\sum_{t=0}^{T}\left(\frac{\eta^{t}}{1-\gamma^{t}}\left(\varepsilon_{ijk}\left(u_{ir}^{t}\right)-\varepsilon_{ijk}\left(u_{ir}^{*}\right)\right)\right)\\
&\leq\frac{\sum\limits_{t=0}^{T}\left(\frac{\eta^{t}}{1-\gamma^{t}}\left(\varepsilon_{ijk}\left(u_{ir}^{t}\right)-\varepsilon_{ijk}\left(u_{ir}^{*}\right)\right)\right)}{\frac{\eta_{\min}}{1-\gamma_{\max}}(T+1)}\\
&\leq\frac{\left\|u_{ir}^{0}-u_{ir}^{*}\right\|^{2}}{2\frac{\eta_{\min}}{1-\gamma_{\max}}(T+1)}+\frac{\eta_{\max}}{1-\gamma_{\min}}\frac{\gamma_{\max}}{1-\gamma_{\max}}GB+\frac{1}{2}G^{2}\frac{\eta_{\max}}{1-\gamma_{\max}}.
\end{aligned}
\tag{S15}
$$

Let $\bar{u}_{ir}=\sum_{t=0}^{T}u_{ir}^{t}/(T+1)$, from the convexity of the function, we can get:

$$
\begin{aligned}
&\varepsilon_{ijk}\left(\bar{u}_{ir}^{t}\right)-\varepsilon_{ijk}\left(u_{ir}^{*}\right)\\
&\leq\frac{1}{T+1}\sum_{t=0}^{T}\left(\frac{\eta^{t}}{1-\gamma^{t}}\left(\varepsilon_{ijk}\left(u_{ir}^{t}\right)-\varepsilon_{ijk}\left(u_{ir}^{*}\right)\right)\right)\\
&\leq\frac{\left\|u_{ir}^{0}-u_{ir}^{*}\right\|^{2}}{2\frac{\eta_{\min}}{1-\gamma_{\max}}(T+1)}+\frac{\eta_{\max}}{1-\gamma_{\min}}\frac{\gamma_{\max}}{1-\gamma_{\max}}GB+\frac{1}{2}G^{2}\frac{\eta_{\max}}{1-\gamma_{\max}}.
\end{aligned}
\tag{S16}
$$

The proof of Theorem 1 can be done by fuzzy reasoning to ensure that η_t, λ_t, γ_t are obtained in a limited range. According to the same principle, $\varepsilon_{i,j,k}\left(w_{jr}\right)$ converges by training w jr by fixing u_{ir} and z_{kr} as a constant, and $\varepsilon_{ijk}\left(Z_{kr}\right)$ converges by training z_{kr} by fixing u_{ir} and w_{jr} as a constant. Therefore, Theorem 1 stands.

III. EXPERIMENTAL RESULTS

Here are some supplementary tables and figures used in the experiment.

TABLE S1
THE PERFORMANCE OF FNL WITH MANUALLY-TUNED AND SELF-ADAPTED HYPER-PARAMETERS.

Datasets		**Manual-tuning**	**Self-adaption**	**Gap**
D1	RMSE	0.3390$_{\pm 4.32E-04}$	**0.3387**$_{\pm \mathbf{6.80E\text{-}04}}$	0.08%
	*Per	5802.11$_{\pm 48.73}$	**37.48**$_{\pm \mathbf{0.66}}$	99.35%
	MAE	0.2452$_{\pm 2.77E-04}$	**0.2450**$_{\pm \mathbf{1.81E\text{-}04}}$	0.08%
	**Per	8123.96$_{\pm 29.15}$	**39.82**$_{\pm \mathbf{0.55}}$	99.50%
D2	RMSE	0.3469$_{\pm 5.29E-05}$	**0.3466**$_{\pm \mathbf{1.61E\text{-}05}}$	0.08%
	*Per	2613.86$_{\pm 25.76}$	**22.74**$_{\pm \mathbf{0.35}}$	99.12%
	MAE	0.2477$_{\pm 3.64E-05}$	**0.2474**$_{\pm \mathbf{1.36E\text{-}05}}$	0.12%
	**Per	3267.33$_{\pm 17.63}$	**24.17**$_{\pm \mathbf{2.35}}$	99.26%
D3	RMSE	**0.3189**$_{\pm \mathbf{7.19E\text{-}05}}$	0.3192$_{\pm 3.93E-06}$	0.09%
	*Per	1058.04$_{\pm 34.17}$	**4.99**$_{\pm \mathbf{0.43}}$	99.52%
	MAE	**0.2184**$_{\pm \mathbf{3.80E\text{-}06}}$	0.2187$_{\pm 2.12E-06}$	0.09%
	**Per	1143.65$_{\pm 56.56}$	**3.23**$_{\pm \mathbf{0.32}}$	99.71%
D4	RMSE	0.3164$_{\pm 4.51E-06}$	**0.3161**$_{\pm \mathbf{6.49E\text{-}06}}$	0.09%
	*Per	719.59$_{\pm 9.77}$	**8.03**$_{\pm \mathbf{0.55}}$	98.88%
	MAE	0.2167$_{\pm 3.33E-06}$	**0.2166**$_{\pm \mathbf{2.78E\text{-}06}}$	0.09%
	**Per	770.99$_{\pm 5.61}$	**4.59**$_{\pm \mathbf{0.66}}$	99.41%
D5	RMSE	**0.2900**$_{\pm \mathbf{5.29E\text{-}04}}$	0.2904$_{\pm 2.18E-04}$	0.13%
	*Per	1010.39$_{\pm 11.8}$	**8.88**$_{\pm \mathbf{0.66}}$	99.12%
	MAE	0.1998$_{\pm 1.50E-05}$	**0.1996**$_{\pm \mathbf{1.81E\text{-}04}}$	0.10%
	**Per	1309.77$_{\pm 18.74}$	**9.24**$_{\pm \mathbf{0.43}}$	99.29%
D6	RMSE	0.3430$_{\pm 2.14E-04}$	**0.3426**$_{\pm \mathbf{5.61E\text{-}06}}$	0.14%
	*Per	2280.92$_{\pm 32.15}$	**5.33**$_{\pm \mathbf{0.04}}$	99.76%
	MAE	**0.2488**$_{\pm \mathbf{5.37E\text{-}04}}$	0.2493$_{\pm 2.98E-06}$	0.16%
	**Per	2348.05$_{\pm 18.99}$	**5.33**$_{\pm \mathbf{0.04}}$	99.77%
D7	RMSE	0.3250$_{\pm 3.21E-04}$	**0.3245**$_{\pm \mathbf{2.22E\text{-}04}}$	0.15%
	*Per	590361$_{\pm \mathbf{3761}}$	**4722.89**$_{\pm 129.8}$	99.20%
	MAE	0.2298$_{\pm \mathbf{5.01E\text{-}04}}$	**0.2294**$_{\pm \mathbf{3.85E\text{-}04}}$	0.17%
	**Per	678780$_{\pm \mathbf{1120}}$	**5028.77**$_{\pm 425.6}$	99.26%
D8	RMSE	0.3199$_{\pm 5.01E-04}$	**0.3117**$_{\pm \mathbf{4.00E\text{-}04}}$	0.06%
	*Per	121296$_{\pm 407}$	**1083.52**$_{\pm \mathbf{110}}$	99.10%
	MAE	**0.2162**$_{\pm \mathbf{3.10E\text{-}04}}$	0.2164$_{\pm 2.98E-04}$	0.09%
	**Per	140988$_{\pm 356}$	**1137.18**$_{\pm \mathbf{129}}$	99.19%

***/**Time cost per training in RMSE/MAE (Seconds).**

TABLE S2
HYPER-PARAMETER SETTINGS OF ALL MODELS

Datasets	Hyper-parameter Setting						
D1	M2: $R=10$, $\lambda=10^{-5}$, $\eta=0.001$	M3: $R=10$, $\lambda=10^{-5}$, $\eta=0.001$	M4: $R=10$, $\lambda=4\times10^{-8}$, $\gamma=480$	M5: $R=10$, $\lambda_1=0.04$, $\lambda_2=0.2$	M6: $R=10$, $\lambda_1=0.5$, $\lambda_2=0.0625$	M7: $R=10$, $\eta=0.1$, $\lambda=0.1$	M8: $R=10$, $\eta=0.1$, $\lambda=0.001$
D2	M2: $R=10$, $\lambda=10^{-5}$, $\eta=0.001$	M3: $R=10$, $\lambda=10^{-5}$, $\eta=0.002$	M4: $R=10$, $\lambda=10^{-8}$, $\gamma=500$	M5: $R=10$, $\lambda_1=0.04$, $\lambda_2=0.2$	M6: $R=10$, $\lambda_1=0.125$, $\lambda_2=0.25$	M7: $R=10$, $\eta=0.2$, $\lambda=0.1$	M8: $R=10$, $\eta=0.1$, $\lambda=0.0001$
D3	M2: $R=10$, $\lambda=10^{-5}$, $\eta=0.001$	M3: $R=10$, $\lambda=10^{-5}$, $\eta=0.004$	M4: $R=10$, $\lambda=10^{-8}$, $\gamma=450$	M5: $R=10$, $\lambda_1=0.005$, $\lambda_2=0.4$	M6: $R=10$, $\lambda_1=0.25$, $\lambda_2=0.0625$	M7: $R=10$, $\eta=0.1$, $\lambda=0.2$	M8: $R=10$, $\eta=0.2$, $\lambda=0.01$
D4	M2: $R=10$, $\lambda=10^{-5}$, $\eta=0.002$	M3: $R=10$, $\lambda=10^{-5}$, $\eta=0.005$	M4: $R=10$, $\lambda=4\times10^{-8}$, $\gamma=500$	M5: $R=10$, $\lambda_1=0.007$, $\lambda_2=0.4$	M6: $R=10$, $\lambda_1=0.5$, $\lambda_2=0.0625$	M7: $R=10$, $\eta=0.1$, $\lambda=0.1$	M8: $R=10$, $\eta=0.2$, $\lambda=0.001$
D5	M2: $R=10$, $\lambda=10^{-4}$, $\eta=0.001$	M3: $R=10$, $\lambda=10^{-4}$, $\eta=0.008$	M4: $R=10$, $\lambda=10^{-7}$, $\gamma=480$	M5: $R=10$, $\lambda_1=0.07$, $\lambda_2=0.3$	M6: $R=10$, $\lambda_1=0.5$, $\lambda_2=0.0625$	M7: $R=10$, $\eta=0.2$, $\lambda=0.2$	M8: $R=10$, $\eta=0.1$, $\lambda=0.002$
D6	M2: $R=10$, $\lambda=10^{-5}$, $\eta=0.001$	M3: $R=10$, $\lambda=10^{-5}$, $\eta=0.004$	M4: $R=10$, $\lambda=10^{-8}$, $\gamma=450$	M5: $R=10$, $\lambda_1=0.006$, $\lambda_2=0.5$	M6: $R=10$, $\lambda_1=0.25$, $\lambda_2=0.03125$	M7: $R=10$, $\eta=0.1$, $\lambda=0.1$	M8: $R=10$, $\eta=0.2$, $\lambda=0.01$
D7	M2: $R=10$, $\lambda=10^{-5}$, $\eta=0.01$	M3: $R=10$, $\lambda=10^{-5}$, $\eta=0.004$	M4: $R=10$, $\lambda=10^{-8}$, $\gamma=480$	M5: $R=10$, $\lambda_1=0.005$, $\lambda_2=0.4$	M6: $R=10$, $\lambda_1=0.25$, $\lambda_2=0.0625$	M7: $R=10$, $\eta=0.1$, $\lambda=0.2$	M8: $R=10$, $\eta=0.2$, $\lambda=0.01$
D8	M2: $R=10$, $\lambda=10^{-4}$, $\eta=0.01$	M3: $R=10$, $\lambda=10^{-5}$, $\eta=0.004$	M4: $R=10$, $\lambda=10^{-8}$, $\gamma=500$	M5: $R=10$, $\lambda_1=0.004$, $\lambda_2=0.5$	M6: $R=10$, $\lambda_1=0.5$, $\lambda_2=0.03125$	M7: $R=10$, $\eta=0.1$, $\lambda=0.2$	M8: $R=10$, $\eta=0.1$, $\lambda=0.01$

TABLE S3
RMSE, MAE, WIN/LOSS COUNTS OF M1-8 ON D1-8.

Datasets		M1	M2	M3	M4	M5	M6	M7	M8
D1	**RMSE**	**0.3387**$_{\pm 6.8E-4}$	0.3605$_{\pm 8.3E-4}$	0.3431$_{\pm 3.5E-4}$	0.3647$_{\pm 7.1E-4}$	0.3498$_{\pm 4.8E-5}$	0.3531$_{\pm 4.5E-5}$	0.5871$_{\pm 6.1E-5}$	0.5869$_{\pm 4.2E-6}$
	MAE	**0.2450**$_{\pm 1.8E-4}$	0.2498$_{\pm 6.3E-4}$	0.2492$_{\pm 2.1E-4}$	0.2557$_{\pm 4.8E-4}$	0.2512$_{\pm 7.4E-5}$	0.2508$_{\pm 4.5E-5}$	0.4126$_{\pm 7.6E-5}$	0.4111$_{\pm 2.7E-6}$
D2	**RMSE**	**0.3466**$_{\pm 1.6E-5}$	0.3697$_{\pm 1.0E-3}$	0.3485$_{\pm 5.2E-4}$	0.3675$_{\pm 4.5E-4}$	0.3535$_{\pm 7.4E-5}$	0.3534$_{\pm 8.1E-5}$	0.6027$_{\pm 1.4E-5}$	0.6024$_{\pm 8.6E-6}$
	MAE	**0.2474**$_{\pm 1.3E-5}$	0.2951$_{\pm 6.4E-2}$	0.2552$_{\pm 3.6E-4}$	0.2619$_{\pm 4.5E-3}$	0.2556$_{\pm 5.8E-5}$	0.2560$_{\pm 4.9E-5}$	0.4259$_{\pm 1.7E-5}$	0.4246$_{\pm 1.1E-5}$
D3	**RMSE**	**0.3192**$_{\pm 3.9E-6}$	0.3296$_{\pm 2.2E-4}$	0.3252$_{\pm 2.1E-0}$	0.3671$_{\pm 2.4E-0}$	0.3344$_{\pm 2.7E-5}$	0.3325$_{\pm 1.8E-4}$	0.4758$_{\pm 1.9E-6}$	0.4759$_{\pm 3.9E-6}$
	MAE	**0.2187**$_{\pm 2.1E-6}$	0.2234$_{\pm 3.1E-4}$	0.2273$_{\pm 2.0E-4}$	0.2471$_{\pm 1.3E-3}$	0.2219$_{\pm 8.1E-5}$	0.2282$_{\pm 1.2E-4}$	0.3125$_{\pm 4.3E-6}$	0.3123$_{\pm 1.4E-6}$
D4	**RMSE**	**0.3161**$_{\pm 6.4E-6}$	0.3359$_{\pm 2.7E-4}$	0.3172$_{\pm 1.0E-4}$	0.3855$_{\pm 5.3E-4}$	0.3425$_{\pm 7.5E-5}$	0.3489$_{\pm 8.6E-5}$	0.4856$_{\pm 9.0E-7}$	0.4857$_{\pm 3.8E-6}$
	MAE	**0.2166**$_{\pm 2.7E-6}$	0.2216$_{\pm 5.6E-4}$	0.2186$_{\pm 3.7E-5}$	0.2770$_{\pm 3.2E-4}$	0.2314$_{\pm 7.8E-5}$	0.2377$_{\pm 1.6E-5}$	0.3242$_{\pm 5.6E-6}$	0.3239$_{\pm 1.5E-6}$
D5	**RMSE**	**0.2904**$_{\pm 2.1E-4}$	0.3021$_{\pm 5.5E-4}$	0.2939$_{\pm 3.4E-4}$	0.2962$_{\pm 8.5E-3}$	0.2985$_{\pm 6.1E-5}$	0.3050$_{\pm 9.6E-5}$	0.4794$_{\pm 1.6E-6}$	0.4793$_{\pm 1.8E-7}$
	MAE	**0.1996**$_{\pm 1.8E-4}$	0.2033$_{\pm 1.0E-3}$	0.2039$_{\pm 2.8E-3}$	0.2029$_{\pm 8.9E-3}$	0.2089$_{\pm 2.0E-4}$	0.2161$_{\pm 1.2E-5}$	0.3353$_{\pm 4.7E-5}$	0.3336$_{\pm 3.1E-6}$
D6	**RMSE**	**0.3426**$_{\pm 5.6E-6}$	0.3717$_{\pm 7.0E-4}$	0.3477$_{\pm 2.6E-4}$	0.3489$_{\pm 3.3E-4}$	0.3436$_{\pm 5.4E-2}$	0.3440$_{\pm 1.0E-5}$	0.5881$_{\pm 5.0E-6}$	0.5880$_{\pm 3.5E-6}$
	MAE	**0.2493**$_{\pm 2.9E-6}$	0.2570$_{\pm 2.4E-4}$	0.2555$_{\pm 1.1E-4}$	0.2583$_{\pm 2.1E-5}$	0.2599$_{\pm 3.9E-4}$	0.2593$_{\pm 5.1E-5}$	0.4295$_{\pm 2.6E-5}$	0.4287$_{\pm 7.0E-6}$
D7	**RMSE**	**0.3245**$_{\pm 2.2E-4}$	0.3440$_{\pm 8.3E-4}$	0.3294$_{\pm 3.6E-4}$	0.3488$_{\pm 7.3E-4}$	0.3359$_{\pm 5.1E-5}$	0.3375$_{\pm 6.4E-5}$	-	-
	MAE	**0.2294**$_{\pm 3.8E-4}$	0.2420$_{\pm 6.8E-4}$	0.2351$_{\pm 4.4E-4}$	0.2478$_{\pm 6.6E-4}$	0.2386$_{\pm 5.8E-5}$	0.2397$_{\pm 6.1E-5}$	-	-
D8	**RMSE**	**0.3117**$_{\pm 4.0E-4}$	0.3304$_{\pm 7.8E-4}$	0.3164$_{\pm 4.2E-4}$	0.3251$_{\pm 8.0E-4}$	0.3226$_{\pm 6.2E-5}$	0.3242$_{\pm 6.9E-5}$	-	-
	MAE	**0.2164**$_{\pm 3.0E-4}$	0.2283$_{\pm 6.0E-4}$	0.2283$_{\pm 5.9E-4}$	0.2337$_{\pm 6.2E-4}$	0.2251$_{\pm 5.4E-5}$	0.2261$_{\pm 5.6E-5}$	-	-
Win/Loss		-	16/0	16/0	16/0	16/0	16/0	16/0	16/0
F-rank		1.00	4.28	2.59	4.88	3.63	4.31	7.75	7.31

TABLE S4
SINGLE TRAINING TIME COST IN RMSE/MAE(SECONDS), WIN/LOSS COUNTS OF M1-8 ON D1-8

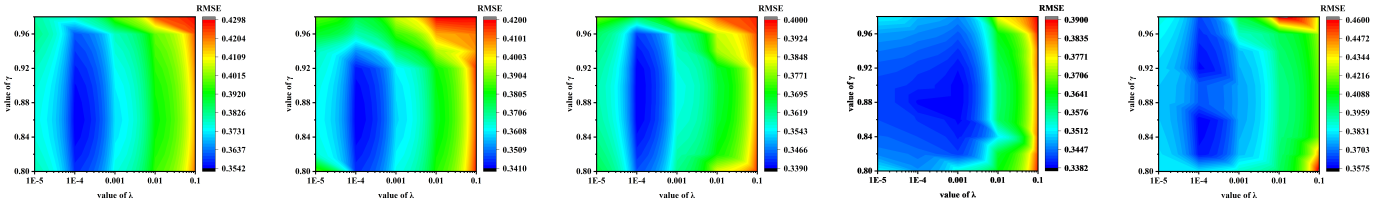
Datasets		M1	M2	M3	M4	M5	M6	M7	M8
D1	RMSE	**37.48$_{\pm 0.66}$**	1222.25$_{\pm 2428.2}$	1358.66$_{\pm 61.69}$	57.01$_{\pm 5.65}$	180.48$_{\pm 7.25}$	358.29$_{\pm 6.03}$	1419.7$_{\pm 113.98}$	522.58$_{\pm 10.01}$
	MAE	**39.82$_{\pm 0.55}$**	1318.37$_{\pm 468.16}$	1357.62$_{\pm 61.76}$	55.97$_{\pm 3.42}$	92.55$_{\pm 0.77}$	206.14$_{\pm 2.85}$	1812.03$_{\pm 132.77}$	1097.77$_{\pm 54.89}$
D2	RMSE	**22.74$_{\pm 0.35}$**	693.79$_{\pm 40.64}$	700.9$_{\pm 51.93}$	38.83$_{\pm 2.82}$	79.12$_{\pm 5.57}$	96.65$_{\pm 3.98}$	986.4$_{\pm 119.81}$	378.21$_{\pm 3.17}$
	MAE	**24.17$_{\pm 2.35}$**	740.65$_{\pm 41.59}$	689.27$_{\pm 44.32}$	38.83$_{\pm 1.73}$	61.89$_{\pm 3.34}$	54.67$_{\pm 3.91}$	1222.48$_{\pm 132.77}$	857.7$_{\pm 2.31}$
D3	RMSE	**4.99$_{\pm 0.43}$**	270.32$_{\pm 95.46}$	122.41$_{\pm 3.54}$	31.32$_{\pm 3.03}$	25.05$_{\pm 0.98}$	61.74$_{\pm 4.18}$	394.79$_{\pm 11.95}$	175.21$_{\pm 2.63}$
	MAE	**3.23$_{\pm 0.32}$**	272.08$_{\pm 96.06}$	119.19$_{\pm 4.27}$	30.5$_{\pm 4.79}$	15.03$_{\pm 1.11}$	51.38$_{\pm 2.95}$	482.28$_{\pm 7.46}$	342.87$_{\pm 6.23}$
D4	RMSE	**8.03$_{\pm 0.55}$**	171.97$_{\pm 60.97}$	129.93$_{\pm 0.28}$	22.63$_{\pm 0.65}$	17.02$_{\pm 0.63}$	34.27$_{\pm 1.92}$	219.58$_{\pm 5.08}$	158.55$_{\pm 51.02}$
	MAE	**4.59$_{\pm 0.66}$**	182.39$_{\pm 69.95}$	128.18$_{\pm 1.32}$	23.18$_{\pm 0.62}$	12.07$_{\pm 1.68}$	40.09$_{\pm 1.72}$	275.74$_{\pm 10.48}$	296.24$_{\pm 96.35}$
D5	RMSE	**8.88$_{\pm 0.66}$**	138.95$_{\pm 9.59}$	128.5$_{\pm 3.71}$	11.71$_{\pm 0.94}$	26.98$_{\pm 0.41}$	42.2$_{\pm 3.53}$	390.7$_{\pm 14.97}$	272$_{\pm 16.51}$
	MAE	**9.24$_{\pm 0.43}$**	141.5$_{\pm 8.23}$	127.79$_{\pm 4.95}$	12.66$_{\pm 0.52}$	14.68$_{\pm 0.61}$	59.06$_{\pm 2.34}$	509.75$_{\pm 23.17}$	901.16$_{\pm 11.82}$
D6	RMSE	**5.33$_{\pm 0.04}$**	214.96$_{\pm 3.58}$	258.28$_{\pm 4.71}$	19.57$_{\pm 2.82}$	35.77$_{\pm 1.1}$	65.38$_{\pm 1.91}$	799.05$_{\pm 9.98}$	146.72$_{\pm 26.42}$
	MAE	**5.33$_{\pm 0.04}$**	192.14$_{\pm 4.36}$	220.03$_{\pm 7.75}$	6.34$_{\pm 0.93}$	8.82$_{\pm 0.65}$	116.52$_{\pm 3.96}$	1125.55$_{\pm 5.81}$	575.61$_{\pm 93.9}$
D7	RMSE	**4722.89$_{\pm 129.8}$**	153965$_{\pm 13366}$	171167$_{\pm 3378}$	7175$_{\pm 311}$	22549$_{\pm 398}$	44765$_{\pm 671}$	-	-
	MAE	**5028.77$_{\pm 425.6}$**	171479$_{\pm 23392}$	171614$_{\pm 11662}$	7040$_{\pm 233}$	14130$_{\pm 926.34}$	25803$_{\pm 2207}$	-	-
D8	RMSE	**2083.52$_{\pm 110}$**	135316$_{\pm 11325.60}$	139236$_{\pm 2862}$	4646$_{\pm 263}$	15175$_{\pm 337}$	210273$_{\pm 568}$	-	-
	MAE	**3137.18$_{\pm 129}$**	135923$_{\pm 13284}$	140292$_{\pm 3358}$	5592$_{\pm 308}$	13190$_{\pm 395}$	25836$_{\pm 667}$	-	-
Win/Loss		-	16/0	16/0	16/0	16/0	16/0	16/0	16/0
F-rank		1.00	6.00	5.38	2.25	2.81	3.94	7.75	6.25

TABLE S5
WILCOXON SIGNED-RANKS TEST FOR TABLE S3

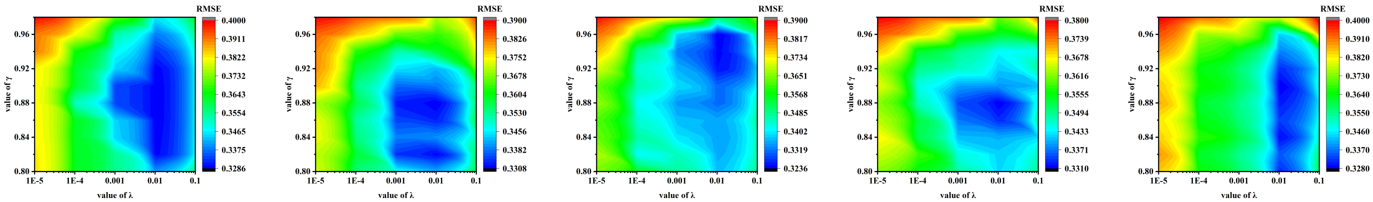
Comparison	*R+*	*R-*	*p-value*
M1 vs M2	136	0	$1.53{\times}10^{-5}$
M1 vs M3	136	0	$1.53{\times}10^{-5}$
M1 vs M4	136	0	$1.53{\times}10^{-5}$
M1 vs M5	136	0	$1.53{\times}10^{-5}$
M1 vs M6	136	0	$1.53{\times}10^{-5}$
M1 vs M7	136	0	$1.53{\times}10^{-5}$
M1 vs M8	136	0	$1.53{\times}10^{-5}$

TABLE S6
WILCOXON SIGNED-RANKS TEST FOR TABLE S4

Comparison	*R+*	*R-*	*p-value*
M1 vs M2	136	0	$1.53{\times}10^{-5}$
M1 vs M3	136	0	$1.53{\times}10^{-5}$
M1 vs M4	136	0	$1.53{\times}10^{-5}$
M1 vs M5	136	0	$1.53{\times}10^{-5}$
M1 vs M6	136	0	$1.53{\times}10^{-5}$
M1 vs M7	136	0	$1.53{\times}10^{-5}$
M1 vs M8	136	0	$1.53{\times}10^{-5}$

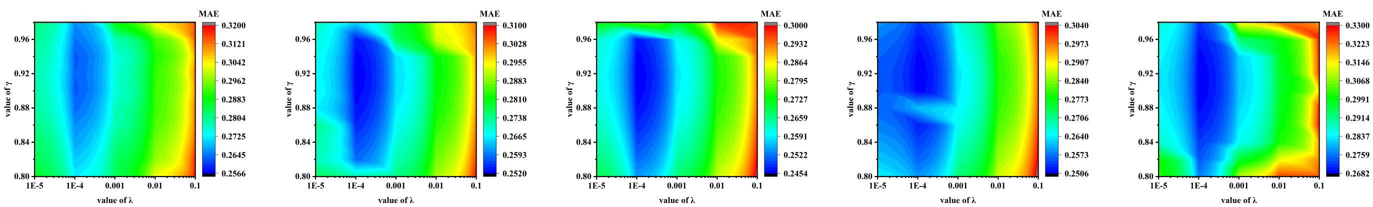


(a) RMSE on D1($\eta = 2^{-5}$) (b) RMSE on D1($\eta = 2^{-4}$) (c) RMSE on D1($\eta = 2^{-3}$) (d) RMSE on D1($\eta = 2^{-2}$) (e) RMSE on D1($\eta = 2^{-1}$)

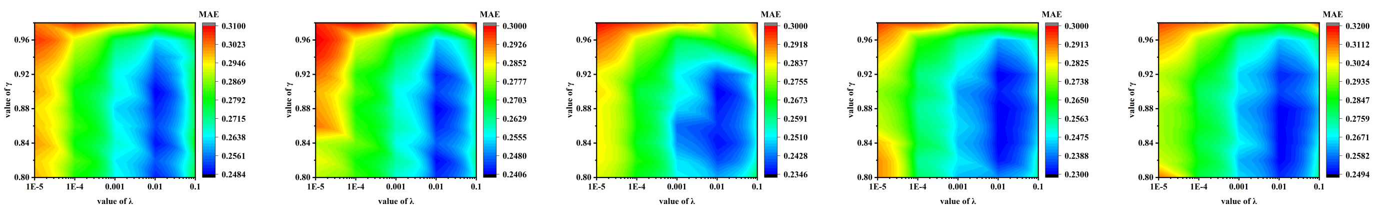


(f) RMSE on D7($\eta = 2^{-5}$) (g) RMSE on D7($\eta = 2^{-4}$) (h) RMSE on D7($\eta = 2^{-3}$) (i) RMSE on D7($\eta = 2^{-2}$) (j) RMSE on D7($\eta = 2^{-1}$)

Fig. S1. RMSE of FNL as η, λ and γ varies on D1 and D7.

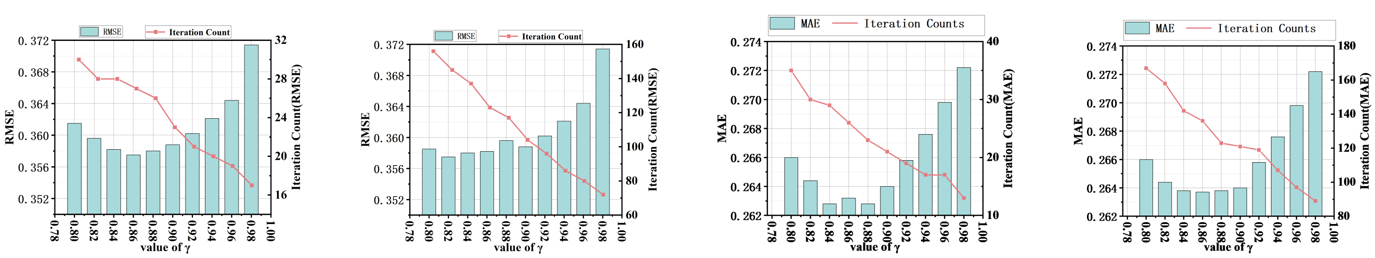


(a) MAE on D1($\eta = 2^{-5}$) (b) MAE on D1($\eta = 2^{-4}$) (c) MAE on D1($\eta = 2^{-3}$) (d) MAE on D1($\eta = 2^{-2}$) (e) MAE on D1($\eta = 2^{-1}$)

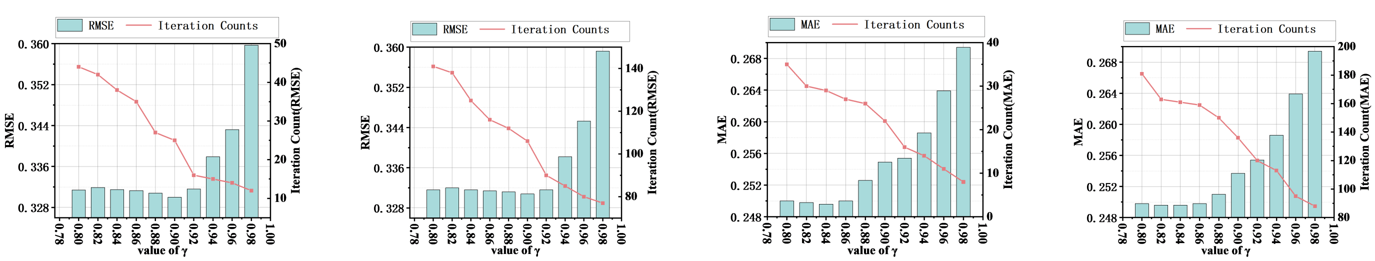


(f) MAE on D7($\eta = 2^{-5}$) (g) MAE on D7($\eta = 2^{-4}$) (h) MAE on D7($\eta = 2^{-3}$) (i) MAE on D7($\eta = 2^{-2}$) (j) MAE on D7($\eta = 2^{-1}$)

Fig. S2. MAE of FNL as η, λ and γ varies on D1 and D7.

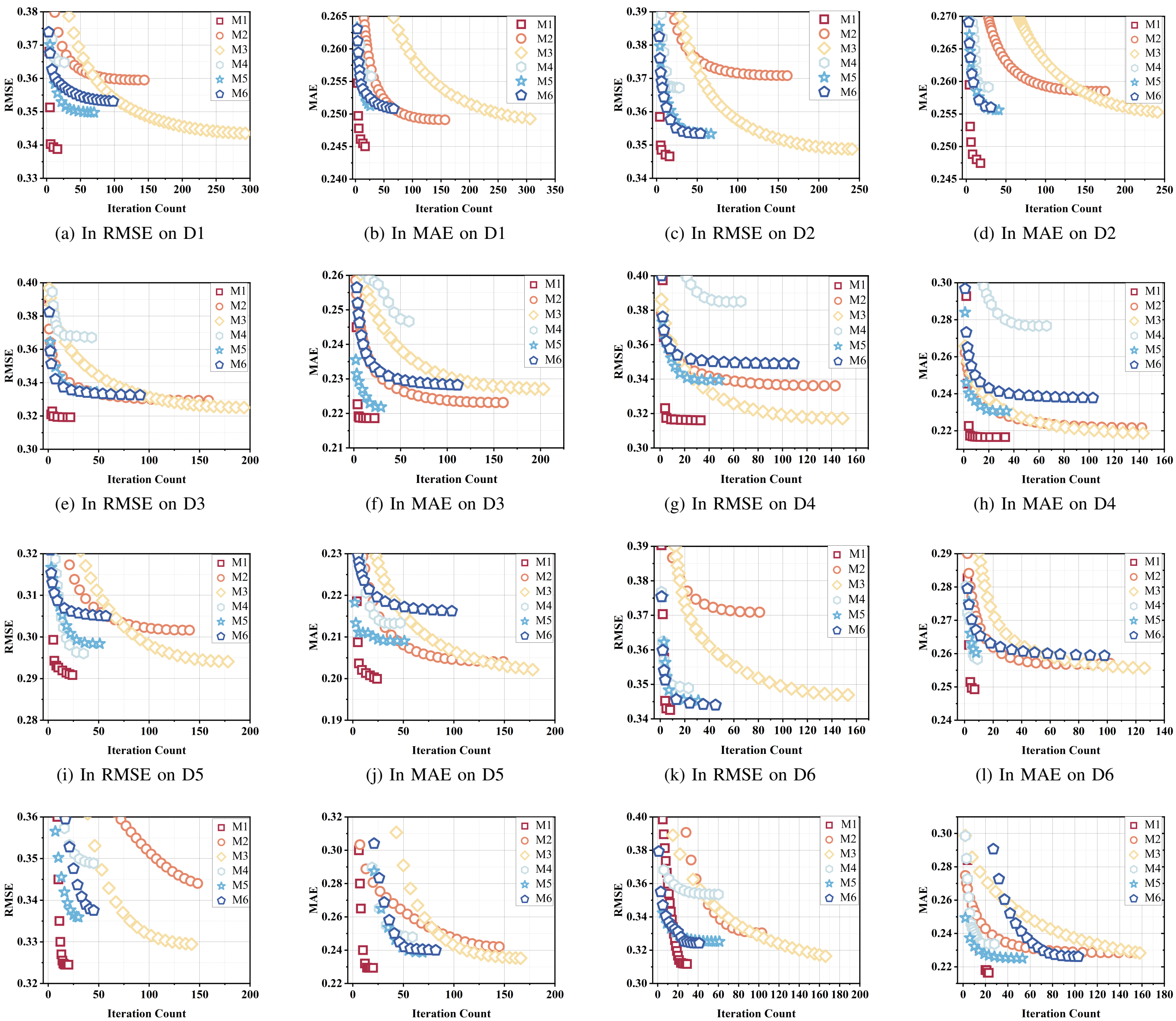


(a) RMSE on D1(η=2^{-1}, λ=10^{-4}) (b) RMSE on D1(η=2^{-5}, λ=10^{-4}) (c) MAE on D1(η=2^{-1}, λ=10^{-4}) (d) MAE on D1(η=2^{-5}, λ=10^{-4})



(e) RMSE on D7(η=2^{-1}, λ=10^{-2}) (f) RMSE on D7(η=2^{-5}, λ=10^{-2}) (g) MAE on D7(η=2^{-1}, λ=10^{-2}) (h) MAE on D7(η=2^{-5}, λ=10^{-2})

Fig. S3. RMSE/MAE and Iteration of FNL as η, λ and γ varies on D1 and D7.



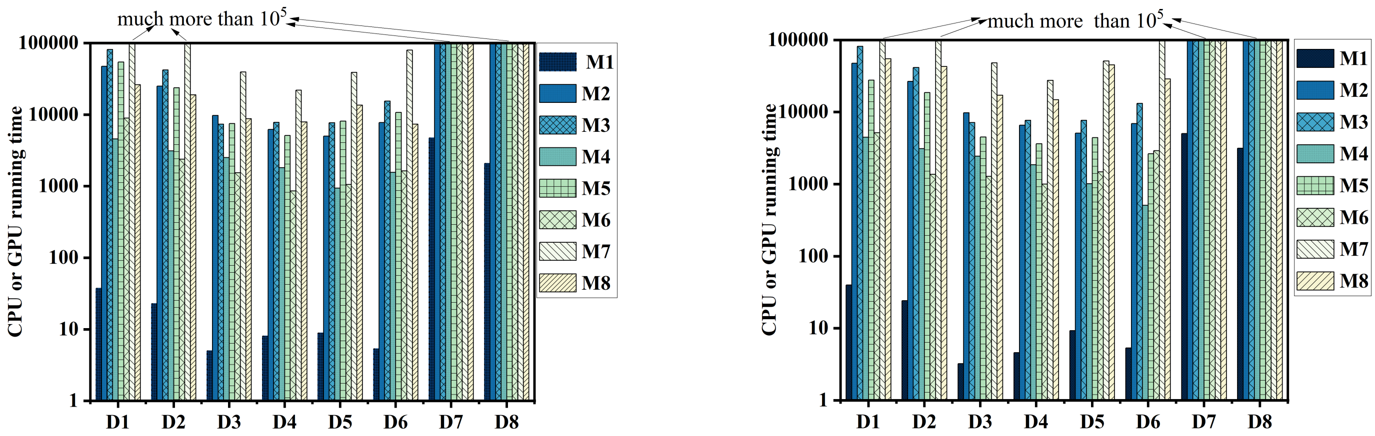
(a) In RMSE on D1 (b) In MAE on D1 (c) In RMSE on D2 (d) In MAE on D2

(e) In RMSE on D3 (f) In MAE on D3 (g) In RMSE on D4 (h) In MAE on D4

(i) In RMSE on D5 (j) In MAE on D5 (k) In RMSE on D6 (l) In MAE on D6

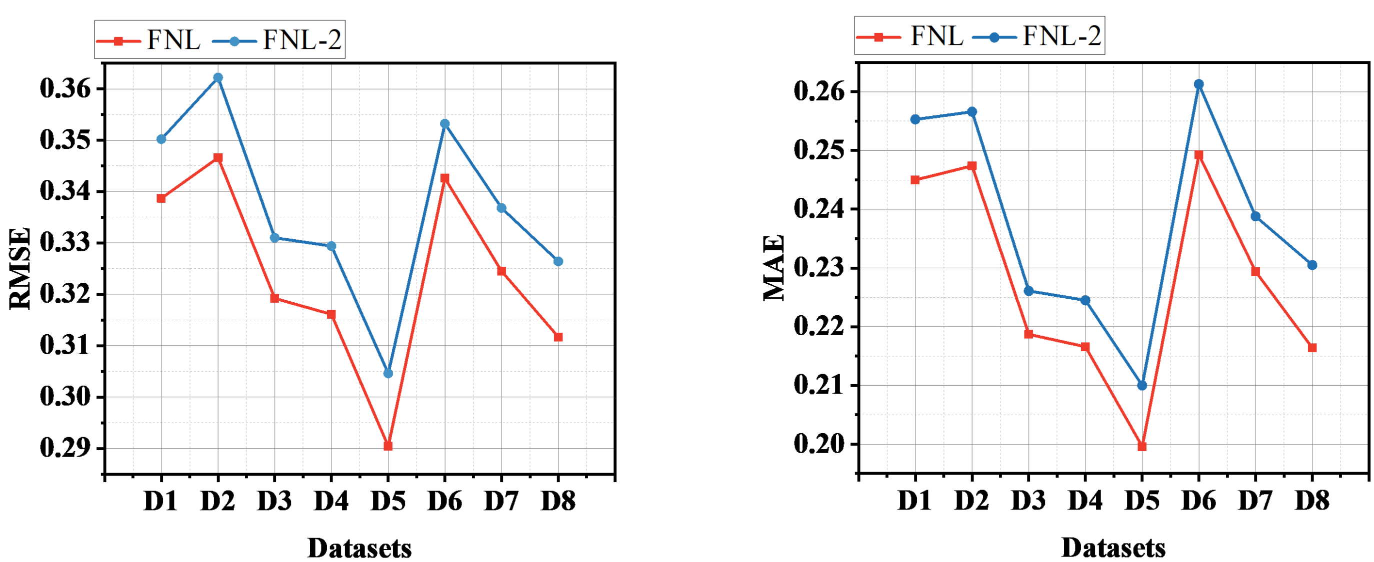
(m) In RMSE on D7 (n) In MAE on D7 (o) In RMSE on D8 (p) In MAE on D8

Fig. S4. Training curves of M1-6 in RMSE and MAE on D1-8.



(a) RMSE (b) MAE

Fig. S5. The total time cost of M1-8 for obtaining the lowest RMSE and MAE on D1-8.



(a) RMSE of FNL and FNL-2 on D1-D8 (b) MAE of FNL and FNL-2 on D1-D8

Fig. S6. Ablation study with nonlinear activation function on D1-D8.