Theoretically Guaranteed Bidirectional Data Rectification for Robust Sequential Recommendation – Appendix –

This Appendix is divided into three sections. First, Section A empirically validates the feasibility of Assumption 1. Next, in Section B, complete proofs of all the lemmas and theorems are presented. Finally, in Section C, we provide detailed settings of baselines and additional experimental results, including the hyper-parameter analysis and the percentage of rectified data. Our code is available at: `https://github.com/AlchemistYT/BirDRec`.

A The Feasibility of Assumption 1

Assumption 1. *The users' true preference distribution η fulfills the relaxed Multiclass Tsybakov Condition [1] with constants $C > 0$, $\lambda > 0$, and $\alpha_0 \in (0, 1]$, such that for all $\alpha \in (0, \alpha_0]$,*

$$\mathbb{P}\big[\eta_{p_1}(\mathbf{x}_t^u) - \eta_{p_2}(\mathbf{x}_t^u) \leq \alpha\big] \leq C\alpha^{\lambda}. \tag{1}$$

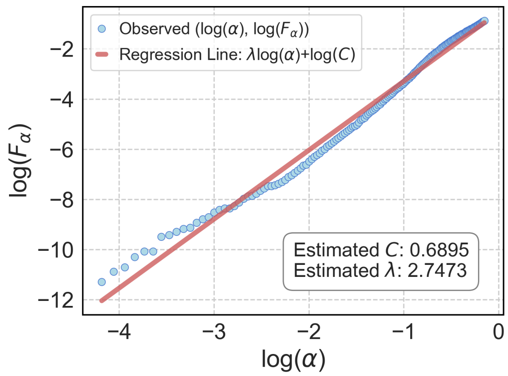
The feasibility of Assumption 1 relies on large C and small λ. In order to estimate the values of C and λ, the initial step is to approximate the true preference distribution η. To achieve this, we first obtain a reliable dataset $\mathcal{D}$ via the heuristic method proposed by [2]. The heuristic method measures the matching degree between the input and target of each instance from two aspects: item co-occurrence and item properties, then those instances with matching degrees lower than a threshold are filtered as unreliable data. We use a large threshold (0.9) to filter the dataset rigorously, ensuring the vast majority of the maintained instances in $\mathcal{D}$ are reliable. Subsequently, as suggested by [3], we train a classic SRS, SASRec [4], on the filtered reliable dataset $\mathcal{D}$ and then use the prediction of SASRec to approximate η. Formally, for each reliable instance $\langle \mathbf{x}_t^u, v_t^u \rangle$ from the reliable dataset $\mathcal{D}$, the SRS prediction score $f_{v_t^u}(\mathbf{x}_t^u)$ is employed as the approximation of $\eta_{v_t^u}(\mathbf{x}_t^u)$.

Next, we densely sample α from 0.05 to 0.9 with step size 0.005 and calculate the corresponding left-hand-side probability of Eq. 1 with the following relative frequency F_α, and collect a series of $\big(\log(\alpha), \log(F_\alpha)\big)$ data points:

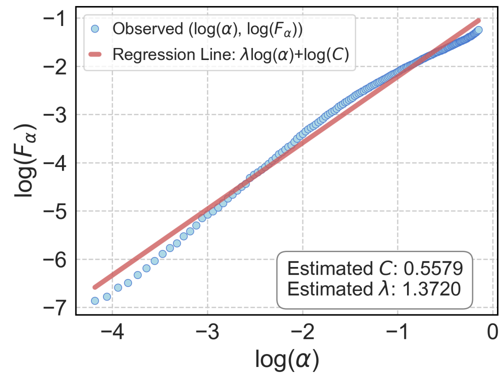
$$F_\alpha = \frac{1}{|\mathcal{D}|} \sum_{\langle \mathbf{x}_t^u, v_t^u \rangle \in \mathcal{D}} \mathbb{I}\big[f_{\hat{p}_1}(\mathbf{x}_t^u) - f_{\hat{p}_2}(\mathbf{x}_t^u) \leq \alpha\big], \tag{2}$$

where $\hat{p}_1$ and $\hat{p}_2$ are respectively the top- and middle-ranked items according to f, namely, $\sum_{v_i \in \mathcal{V}} \mathbb{I}\big[f_{\hat{p}_1}(\mathbf{x}_t^u) \geq f_{v_i}(\mathbf{x}_t^u)\big] = |\mathcal{V}|$, $\sum_{v_i \in \mathcal{V}} \mathbb{I}\big[f_{\hat{p}_2}(\mathbf{x}_t^u) \geq f_{v_i}(\mathbf{x}_t^u)\big] = \lfloor |\mathcal{V}|/2 \rfloor$. We then use $\log(F_\alpha)$ to approximate $\log(C\alpha^{\lambda})$. As shown by the blue dots in Fig. 6, the collected $\big(\log(\alpha), \log(F_\alpha)\big)$ data points are generally linearly distributed, which allows us to estimate C and λ with linear regression according to $\log(C\alpha^{\lambda}) = \log(C) + \lambda \log(\alpha)$. As a result, Fig. 6 shows that the estimated C and λ are respectively restricted in $(0.55, 0.70)$ and $(1.37, 4.01)$ on real-world datasets from various domains, which validates the feasibility of Assumption 1.

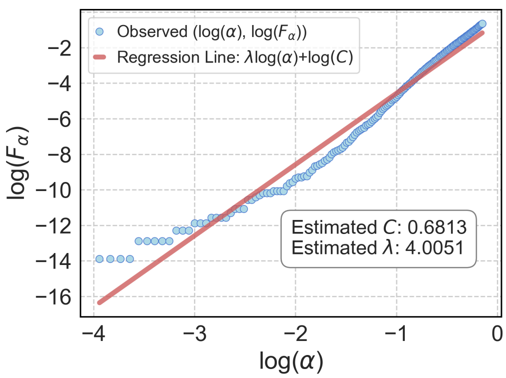
37th Conference on Neural Information Processing Systems (NeurIPS 2023).



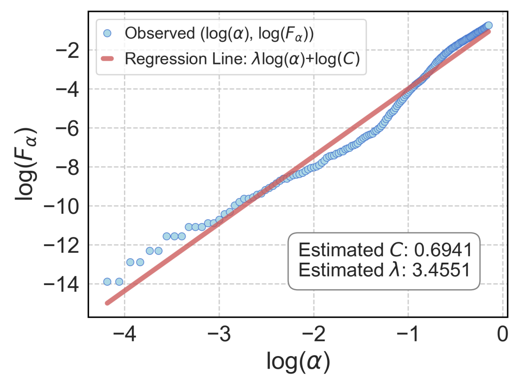
(a) ML-1M



(b) Beauty



(c) Yelp



(d) QK-Vedio

Figure 1: The estimated constants C and λ on various datasets.

B Proofs of Lemmas and Theorems

B.1 The Proof of Theorem 1

Theorem 1. *Given Assumption 1, let* $\{w_h \mid 1 \leq h \leq H, 0 \leq w_h \leq 1, \sum_{h=1}^{H} w_h = 1\}$ *be the weights for averaging prediction scores of different epochs.* $\forall \langle \tilde{\mathbf{x}}_t^u, \tilde{v}_t^u \rangle$, *assume* $\epsilon \leq \alpha_0 \tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)$. *Let* $\gamma = \tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\eta_{p_2}(\tilde{\mathbf{x}}_t^u) + \sum_{v_j \neq \tilde{v}_t^u} \tau_{v_j \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u)$. *We have:* $\mathbb{P}\Big[p_1 = \tilde{v}_t^u, \sum_{h=1}^{H}\big[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\big] \leq \gamma\Big] \leq C(\mathcal{O}(\epsilon))^{\lambda}$.

Proof.

$$
\begin{aligned}
&\mathbb{P}\left[p_1 = \tilde{v}_t^u, \sum_{h=1}^{H}\left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right] \leq \gamma\right] \\
\leq &\mathbb{P}\left[p_1 = \tilde{v}_t^u, \sum_{h=1}^{H} w_h\left[\tilde{\eta}_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) - \epsilon\right] \leq \gamma\right] \\
= &\mathbb{P}\left[p_1 = \tilde{v}_t^u, \eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \sum_{h=1}^{H} w_h\left[\sum_{v_j \in \mathcal{V}} \tau_{v_j \tilde{v}_t^u}\eta_{v_j}(\tilde{\mathbf{x}}_t^u) - \epsilon\right] \leq \gamma\right] \\
\leq &\mathbb{P}\left[p_1 = \tilde{v}_t^u, \eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) + \sum_{v_j \neq \tilde{v}_t^u} \tau_{v_j \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u) \leq \gamma + \epsilon\right] \\
= &\mathbb{P}\left[p_1 = \tilde{v}_t^u, \eta_{p_2}(\tilde{\mathbf{x}}_t^u) \leq \eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) \leq \frac{\gamma + \epsilon - \sum_{v_j \neq \tilde{v}_t^u} \tau_{v_j \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u)}{\tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)}\right]
\end{aligned} \tag{3}
$$

By replacing γ with $\tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\eta_{p_2}(\tilde{\mathbf{x}}_t^u) + \sum_{v_j \neq \tilde{v}_t^u} \tau_{v_j \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u)$, we obtain:

$$\mathbb{P}\left[p_1=\tilde{v}_t^u, \sum_{h=1}^{H}\left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]\le\gamma\right]\le\mathbb{P}\left[\eta_{p_2}(\tilde{\mathbf{x}}_t^u)\le\eta_{p_1}(\tilde{\mathbf{x}}_t^u)\le\eta_{p_2}(\tilde{\mathbf{x}}_t^u)+\frac{\epsilon}{\tau_{\tilde{v}_t^u\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)}\right] \tag{4}$$

Recall that $\epsilon \le \alpha_0\tau_{\tilde{v}_t^u\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)$, which implies $\frac{\epsilon}{\tau_{\tilde{v}_t^u\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)} \le \alpha_0$. Hence, the relaxed Multiclass Tsybakov Condition holds and the probability is bounded by $C\big(\frac{\epsilon}{\tau_{\tilde{v}_t^u\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)}\big)^\lambda$, namely, $C\big(\mathcal{O}(\epsilon)\big)^\lambda$. □

B.2 The Proof of Lemma 1

$$\boldsymbol{E}_{\mathrm{DRUT}}=\underbrace{\mathbb{P}\left[p_1=\tilde{v}_t^u, p_1\neq v_m, \frac{\sum_{h=1}^{H}\left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]}<\beta\right]}_{\text{Case-1}}+\underbrace{\mathbb{P}\left[p_1\neq\tilde{v}_t^u, p_1=v_m, \frac{\sum_{h=1}^{H}\left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]}\ge\beta\right]}_{\text{Case-2}}+\underbrace{\mathbb{P}\left[p_1\neq\tilde{v}_t^u, p_1\neq v_m\right]}_{\text{Case-3}}.$$

Lemma 1. *Given Assumption 1 and the set of weights* $\{w_h \mid 1\le h\le H, 0\le w_h\le 1, \sum_{h=1}^{H}w_h=1\}$, $\forall\langle\tilde{\mathbf{x}}_t^u,\tilde{v}_t^u\rangle$, *assume* $\epsilon\le\min\left[\alpha_0\tau_{\tilde{v}_t^u\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u), \alpha_0\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)\right]$. *Let* $\beta_1=\left[\frac{\tau_{\tilde{v}_t^u\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\eta_{p_2}(\tilde{\mathbf{x}}_t^u)+\sum_{v_j\neq\tilde{v}_t^u}\tau_{v_j\tilde{v}_t^u}\eta_{v_j}(\tilde{\mathbf{x}}_t^u)}{\sum_{h=1}^{H}[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)]}\right]$ *and* $\beta_2=\left[\frac{\sum_{h=1}^{H}[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)]}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)\eta_{p_2}(\tilde{\mathbf{x}}_t^u)+\sum_{v_j\neq v_m}\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u)}\right]$. *We have:* $\beta\le\beta_1$ *guarantees the probability of Case-1 in* $\boldsymbol{E}_{\mathrm{DRUT}}$ *is bounded by* $C\big(\mathcal{O}(\epsilon)\big)^\lambda$, *and* $\beta\ge\beta_2$ *guarantees the probability of Case-2 in* $\boldsymbol{E}_{\mathrm{DRUT}}$ *is bounded by* $C\big(\mathcal{O}(\epsilon)\big)^\lambda$.

Proof. For Case-1 of $\boldsymbol{E}_{\mathrm{DRUT}}$, we have:

$$\begin{aligned}
&\mathbb{P}\left[p_1=\tilde{v}_t^u, p_1\neq v_m, \frac{\sum_{h=1}^{H}\left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]}<\beta_1\right]\\
\le&\mathbb{P}\left[p_1=\tilde{v}_t^u, \frac{\sum_{h=1}^{H}\left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]}<\beta_1\right]\\
\le&\mathbb{P}\left[p_1=\tilde{v}_t^u, \eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\ge\eta_{p_2}(\tilde{\mathbf{x}}_t^u), \sum_{h=1}^{H}w_h\left[\tilde{\eta}_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)-\epsilon\right]<\beta_1\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]\right]\\
=&\mathbb{P}\left[p_1=\tilde{v}_t^u, \eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\ge\eta_{p_2}(\tilde{\mathbf{x}}_t^u), \tilde{\eta}_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)<\beta_1\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]+\epsilon\right]\\
=&\mathbb{P}\left[p_1=\tilde{v}_t^u, \eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\ge\eta_{p_2}(\tilde{\mathbf{x}}_t^u), \sum_{v_j\in\mathcal{V}}\tau_{v_j\tilde{v}_t^u}\eta_{v_j}(\tilde{\mathbf{x}}_t^u)<\beta_1\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]+\epsilon\right]\\
=&\mathbb{P}\left[p_1=\tilde{v}_t^u, \eta_{p_2}(\tilde{\mathbf{x}}_t^u)\le\eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)<\frac{\beta_1\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]+\epsilon-\sum_{v_j\neq\tilde{v}_t^u}\left[\tau_{v_j\tilde{v}_t^u}\eta_{v_j}(\tilde{\mathbf{x}}_t^u)\right]}{\tau_{\tilde{v}_t^u\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)}\right].
\end{aligned} \tag{5}$$

By substituting β_1 with $\frac{\tau_{\tilde{v}_t^u\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\eta_{p_2}(\tilde{\mathbf{x}}_t^u)+\sum_{v_j\neq\tilde{v}_t^u}\tau_{v_j\tilde{v}_t^u}\eta_{v_j}(\tilde{\mathbf{x}}_t^u)}{\sum_{h=1}^{H}[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)]}$ in Eq. 7, we obtain:

$$\begin{aligned}
&\mathbb{P}\left[p_1=\tilde{v}_t^u, p_1\neq v_m, \frac{\sum_{h=1}^{H}\left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]}<\beta_1\right]\\
\le&\mathbb{P}\left[\eta_{p_2}(\tilde{\mathbf{x}}_t^u)\le\eta_{p_1}(\tilde{\mathbf{x}}_t^u)<\eta_{p_2}(\tilde{\mathbf{x}}_t^u)+\frac{\epsilon}{\tau_{\tilde{v}_t^u\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)}\right].
\end{aligned} \tag{6}$$

Recall that $\epsilon \le \alpha_0\tau_{\tilde{v}_t^u\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)$, which implies $\frac{\epsilon}{\tau_{\tilde{v}_t^u\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)} \le \alpha_0$. Hence, the Multiclass Tsybakov Condition holds and the probability of Case-1 is bounded by $C\Big(\frac{\epsilon}{\tau_{\tilde{v}_t^u\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)}\Big)^\lambda$, namely, $C\big(\mathcal{O}(\epsilon)\big)^\lambda$.

Thereafter, for Case-2 of $\boldsymbol{E}_{\text{DRUT}}$, we have:

$$\begin{aligned}
&\mathbb{P}\Bigg[p_1 \neq \tilde{v}_t^u, p_1 = v_m, \frac{\sum_{h=1}^{H}\left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]} \geq \beta_2\Bigg] \\
\leq&\mathbb{P}\Bigg[p_1 = v_m, \frac{\sum_{h=1}^{H}\left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]} \geq \beta_2\Bigg] \\
\leq&\mathbb{P}\Bigg[p_1 = v_m, \eta_m(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \sum_{h=1}^{H} w_h\left[\tilde{\eta}_m(\tilde{\mathbf{x}}_t^u) - \epsilon\right] \leq \frac{\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]}{\beta_2}\Bigg] \\
=&\mathbb{P}\Bigg[p_1 = v_m, \eta_m(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \tilde{\eta}_m(\tilde{\mathbf{x}}_t^u) \leq \frac{\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]}{\beta_2} + \epsilon\Bigg] \\
=&\mathbb{P}\Bigg[p_1 = v_m, \eta_m(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \sum_{v_j \in \mathcal{V}} \tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u) \leq \frac{\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]}{\beta_2} + \epsilon\Bigg] \\
=&\mathbb{P}\Bigg[p_1 = v_m, \eta_{p_2}(\tilde{\mathbf{x}}_t^u) \leq \eta_m(\tilde{\mathbf{x}}_t^u) \leq \frac{\frac{\sum_{h=1}^{H}[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)]}{\beta_2} - \sum_{v_j \neq v_m}\left[\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u)\right]}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)} + \frac{\epsilon}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)}\Bigg].
\end{aligned} \tag{7}$$

We replace β_2 with $\frac{\sum_{h=1}^{H}[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)]}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)\eta_{p_2}(\tilde{\mathbf{x}}_t^u)+\sum_{v_j \neq v_m}\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u)}$ in Eq. 5 and continue the calculation:

$$\begin{aligned}
&\mathbb{P}\Bigg[p_1 \neq \tilde{v}_t^u, p_1 = v_m, \frac{\sum_{h=1}^{H}\left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]} \geq \beta_2\Bigg] \\
\leq&\mathbb{P}\Bigg[\eta_{p_2}(\tilde{\mathbf{x}}_t^u) \leq \eta_m(\tilde{\mathbf{x}}_t^u) \leq \eta_{p_2}(\tilde{\mathbf{x}}_t^u) + \frac{\epsilon}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)}\Bigg].
\end{aligned} \tag{8}$$

Recall that $\epsilon \leq \alpha_0 \tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)$, which implies $\frac{\epsilon}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)} \leq \alpha_0$. Hence, the Multiclass Tsybakov Condition holds and the probabiliy of Case-2 is bounded by $C\big(\frac{\epsilon}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)}\big)^{\lambda}$, namely, $C\big(\mathcal{O}(\epsilon)\big)^{\lambda}$. $\square$

B.3 The Proof of Theorem 2

Theorem 2 (The Upper Bound of $\boldsymbol{E}_{\text{DRUT}}$)**.** *Given Assumption 1 and the set of weights* $\{w_h \mid 1 \leq h \leq H, 0 \leq w_h \leq 1, \sum_{h=1}^{H} w_h = 1\}$, $\forall \langle \tilde{\mathbf{x}}_t^u, \tilde{v}_t^u \rangle$, *let* $\beta_1 = \Big[\frac{\tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\eta_{p_2}(\tilde{\mathbf{x}}_t^u)+\sum_{v_j \neq \tilde{v}_t^u}\tau_{v_j \tilde{v}_t^u}\eta_{v_j}(\tilde{\mathbf{x}}_t^u)}{\sum_{h=1}^{H}[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)]}\Big]$, $\beta_2 = \Big[\frac{\sum_{h=1}^{H}[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)]}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)\eta_{p_2}(\tilde{\mathbf{x}}_t^u)+\sum_{v_j \neq v_m}\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u)}\Big]$, $\xi_1 = |\beta - \beta_1|$, *and* $\xi_2 = |\beta - \beta_2|$. *Assume* $\xi_2 < \beta_2$, $\epsilon \leq \min\Big[\alpha_0 \tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) - \xi_1, \frac{\alpha_0 \tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)\beta_2(\beta_2-\xi_2)-\xi_2}{\beta_2(\beta_2-\xi_2)}, \frac{1}{2}\big[[\tau_{p_1 p_1}(\tilde{\mathbf{x}}_t^u) - \tau_{p_1 v_m}(\tilde{\mathbf{x}}_t^u)][\alpha_0 + \eta_{p_2}(\tilde{\mathbf{x}}_t^u)] - \sum_{v_j \neq p_1}[\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u) - \tau_{v_j p_1}(\tilde{\mathbf{x}}_t^u)]\eta_{v_j}(\tilde{\mathbf{x}}_t^u)\big]\Big]$. *We have:* $\boldsymbol{E}_{\text{DRUT}} \leq C\big(\mathcal{O}(\epsilon+\xi_1)\big)^{\lambda} + C\big(\mathcal{O}(\epsilon+\xi_2)\big)^{\lambda} + C\big(\mathcal{O}(\epsilon)\big)^{\lambda}$.

Proof. For Case-1 of $\boldsymbol{E}_{\mathrm{DRUT}}$, we have:

$$\begin{aligned}
&\mathbb{P}\left[p_1 = \tilde{v}_t^u, p_1 \neq v_m, \frac{\sum_{h=1}^H \left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]} < \beta\right] \\
\leq &\mathbb{P}\left[p_1 = \tilde{v}_t^u, \frac{\sum_{h=1}^H \left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]} < \beta\right] \\
\leq &\mathbb{P}\left[p_1 = \tilde{v}_t^u, \eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \sum_{h=1}^H w_h\left[\tilde{\eta}_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) - \epsilon\right] < \beta \sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]\right] \\
= &\mathbb{P}\left[p_1 = \tilde{v}_t^u, \eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \tilde{\eta}_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) < \beta \sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right] + \epsilon\right] \\
= &\mathbb{P}\left[p_1 = \tilde{v}_t^u, \eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \sum_{v_j \in \mathcal{V}} \tau_{v_j \tilde{v}_t^u} \eta_{v_j}(\tilde{\mathbf{x}}_t^u) < \beta \sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right] + \epsilon\right] \\
\leq &\mathbb{P}\left[p_1 = \tilde{v}_t^u, \eta_{p_2}(\tilde{\mathbf{x}}_t^u) \leq \eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) < \frac{\beta \sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right] + \epsilon - \sum_{v_j \neq \tilde{v}_t^u} \left[\tau_{v_j \tilde{v}_t^u} \eta_{v_j}(\tilde{\mathbf{x}}_t^u)\right]}{\tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)}\right] \\
\leq &\mathbb{P}\left[p_1 = \tilde{v}_t^u, \eta_{p_2}(\tilde{\mathbf{x}}_t^u) \leq \eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) < \frac{(\beta_1 + \xi_1) \sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right] + \epsilon - \sum_{v_j \neq \tilde{v}_t^u} \left[\tau_{v_j \tilde{v}_t^u} \eta_{v_j}(\tilde{\mathbf{x}}_t^u)\right]}{\tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)}\right].
\end{aligned} \tag{9}$$

We substitute β_1 with $\frac{\tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)\eta_{p_2}(\tilde{\mathbf{x}}_t^u) + \sum_{v_j \neq \tilde{v}_t^u} \tau_{v_j \tilde{v}_t^u} \eta_{v_j}(\tilde{\mathbf{x}}_t^u)}{\sum_{h=1}^H [w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)]}$ and obtain:

$$\begin{aligned}
&\mathbb{P}\left[p_1 = \tilde{v}_t^u, p_1 \neq v_m, \frac{\sum_{h=1}^H \left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]} < \beta\right] \\
\leq &\mathbb{P}\left[p_1 = \tilde{v}_t^u, \eta_{p_2}(\tilde{\mathbf{x}}_t^u) \leq \eta_{\tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) < \eta_{p_2}(\tilde{\mathbf{x}}_t^u) + \frac{\epsilon}{\tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)} + \frac{\xi_1 \sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]}{\tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)}\right] \\
\leq &\mathbb{P}\left[\eta_{p_2}(\tilde{\mathbf{x}}_t^u) \leq \eta_{p_1}(\tilde{\mathbf{x}}_t^u) < \eta_{p_2}(\tilde{\mathbf{x}}_t^u) + \frac{\epsilon + \xi_1}{\tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)}\right].
\end{aligned} \tag{10}$$

Recall that $\epsilon \leq \alpha_0 \tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u) - \xi_1$, which implies $\frac{\epsilon + \xi_1}{\tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)} \leq \alpha_0$. Hence, the Multiclass Tsybakov Condition holds and the probabiliy of Case-2 is bounded by $C\left(\frac{\epsilon + \xi_1}{\tau_{\tilde{v}_t^u \tilde{v}_t^u}(\tilde{\mathbf{x}}_t^u)}\right)^\lambda$, namely, $C\big(\mathcal{O}(\epsilon + \xi_1)\big)^\lambda$.

For Case-2 of $\boldsymbol{E}_{\mathrm{DRUT}}$, we have:

$$
\begin{aligned}
&\mathbb{P}\left[p_1 \neq \tilde{v}_t^u, p_1 = v_m, \frac{\sum_{h=1}^H \left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]} \geq \beta\right] \\
\leq&\mathbb{P}\left[p_1 = v_m, \frac{\sum_{h=1}^H \left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]} \geq \beta\right] \\
\leq&\mathbb{P}\left[p_1 = v_m, \eta_m(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \sum_{h=1}^H w_h\left[\tilde{\eta}_m(\tilde{\mathbf{x}}_t^u) - \epsilon\right] \leq \frac{\sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]}{\beta}\right] \\
=&\mathbb{P}\left[p_1 = v_m, \eta_m(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \tilde{\eta}_m(\tilde{\mathbf{x}}_t^u) \leq \frac{\sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]}{\beta} + \epsilon\right] \\
=&\mathbb{P}\left[p_1 = v_m, \eta_m(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \sum_{v_j \in \mathcal{V}} \tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u) \leq \frac{\sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]}{\beta} + \epsilon\right] \\
\leq&\mathbb{P}\left[p_1 = v_m, \eta_{p_2}(\tilde{\mathbf{x}}_t^u) \leq \eta_m(\tilde{\mathbf{x}}_t^u) \leq \frac{\frac{\sum_{h=1}^H [w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)]}{\beta} - \sum_{v_j \neq v_m}\left[\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u)\right]}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)} + \frac{\epsilon}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)}\right] \\
\leq&\mathbb{P}\left[p_1 = v_m, \eta_{p_2}(\tilde{\mathbf{x}}_t^u) \leq \eta_m(\tilde{\mathbf{x}}_t^u) \leq \frac{\frac{\sum_{h=1}^H [w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)]}{\beta_2 - \xi_2} - \sum_{v_j \neq v_m}\left[\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u)\right]}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)} + \frac{\epsilon}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)}\right] \\
=&\mathbb{P}\left[p_1 = v_m, \eta_{p_2}(\tilde{\mathbf{x}}_t^u) \leq \eta_m(\tilde{\mathbf{x}}_t^u) \leq \frac{\frac{\sum_{h=1}^H [w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)]}{\beta_2} - \sum_{v_j \neq v_m}\left[\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u)\right]}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)} + \frac{\epsilon}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)}\right. \\
&\left. + \frac{\xi_2 \sum_{h=1}^H [w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)]}{\beta_2(\beta_2 - \xi_2)\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)}\right]
\end{aligned} \tag{11}
$$

We replace β_2 with $\frac{\sum_{h=1}^H [w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)]}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)\eta_{p_2}(\tilde{\mathbf{x}}_t^u) + \sum_{v_j \neq v_m} \tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u)}$ and continue the calculation:

$$
\begin{aligned}
&\mathbb{P}\left[p_1 \neq \tilde{v}_t^u, p_1 = v_m, \frac{\sum_{h=1}^H \left[w_h f_{\tilde{v}_t^u}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^H \left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]} \geq \beta\right] \\
\leq&\mathbb{P}\left[p_1 = v_m, \eta_{p_2}(\tilde{\mathbf{x}}_t^u) \leq \eta_m(\tilde{\mathbf{x}}_t^u) \leq \eta_{p_2}(\tilde{\mathbf{x}}_t^u) + \frac{\epsilon}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)} + \frac{\xi_2 \sum_{h=1}^H [w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)]}{\beta_2(\beta_2 - \xi_2)\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)}\right] \\
\leq&\mathbb{P}\left[\eta_{p_2}(\tilde{\mathbf{x}}_t^u) \leq \eta_{p_1}(\tilde{\mathbf{x}}_t^u) \leq \eta_{p_2}(\tilde{\mathbf{x}}_t^u) + \frac{\epsilon}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)} + \frac{\xi_2}{\beta_2(\beta_2 - \xi_2)\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)}\right].
\end{aligned} \tag{12}
$$

Recall that $\epsilon \leq \frac{\alpha_0 \tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)\beta_2(\beta_2 - \xi_2) - \xi_2}{\beta_2(\beta_2 - \xi_2)}$, which implies $\frac{\epsilon}{\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)} + \frac{\xi_2}{\beta_2(\beta_2 - \xi_2)\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)} \leq \alpha_0$. Hence, the relaxed Multiclass Tsybakov Condition holds and the probability of Case-2 is bounded by $C\left(\frac{\epsilon\beta_2(\beta_2 - \xi_2) + \xi_2}{\beta_2(\beta_2 - \xi_2)\tau_{v_m v_m}(\tilde{\mathbf{x}}_t^u)}\right)^\lambda$, namely, $C\big(\mathcal{O}(\epsilon + \xi_2)\big)^\lambda$.

Finally, for Case-3 of $\boldsymbol{E}_{\mathrm{DRUT}}$, we have:

$$
\begin{aligned}
&\mathbb{P}\Big[p_1 \neq \tilde{v}_t^u, p_1 \neq v_m\Big] \\
\leq &\mathbb{P}\Big[p_1 \neq v_m\Big] \\
= &\mathbb{P}\left[p_1 \neq v_m, \eta_{p_1}(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \frac{\sum_{h=1}^{H}\left[w_h f_{p_1}^h(\tilde{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]} < 1\right] \\
\leq &\mathbb{P}\left[\eta_{p_1}(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \sum_{h=1}^{H}\left[w_h f_{p_1}^h(\tilde{\mathbf{x}}_t^u)\right] < \sum_{h=1}^{H}\left[w_h f_{v_m}^h(\tilde{\mathbf{x}}_t^u)\right]\right] \\
\leq &\mathbb{P}\left[\eta_{p_1}(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \sum_{h=1}^{H} w_h\left[\tilde{\eta}_{p_1}(\tilde{\mathbf{x}}_t^u) - \epsilon\right] < \sum_{h=1}^{H} w_h\left[\tilde{\eta}_m(\tilde{\mathbf{x}}_t^u) + \epsilon\right]\right] \\
= &\mathbb{P}\left[\eta_{p_1}(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \tilde{\eta}_{p_1}(\tilde{\mathbf{x}}_t^u) - \epsilon < \tilde{\eta}_m(\tilde{\mathbf{x}}_t^u) + \epsilon\right] \\
= &\mathbb{P}\left[\eta_{p_1}(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \sum_{v_j \in \mathcal{V}} \tau_{v_j p_1}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u) < \sum_{v_j \in \mathcal{V}} \tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u) + 2\epsilon\right] \\
= &\mathbb{P}\Bigg[\eta_{p_1}(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \tau_{p_1 p_1}(\tilde{\mathbf{x}}_t^u)\eta_{p_1}(\tilde{\mathbf{x}}_t^u) + \sum_{v_j \neq p_1} \tau_{v_j p_1}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u) < \\
&\quad \tau_{p_1 v_m}(\tilde{\mathbf{x}}_t^u)\eta_{p_1}(\tilde{\mathbf{x}}_t^u) + \sum_{v_j \neq p_1} \tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u)\eta_{v_j}(\tilde{\mathbf{x}}_t^u) + 2\epsilon\Bigg] \\
= &\mathbb{P}\Bigg[\eta_{p_1}(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \eta_{p_1}(\tilde{\mathbf{x}}_t^u)\left[\tau_{p_1 p_1}(\tilde{\mathbf{x}}_t^u) - \tau_{p_1 v_m}(\tilde{\mathbf{x}}_t^u)\right] < \\
&\quad \sum_{v_j \neq p_1}\left[\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u) - \tau_{v_j p_1}(\tilde{\mathbf{x}}_t^u)\right]\eta_{v_j}(\tilde{\mathbf{x}}_t^u) + 2\epsilon\Bigg] \\
= &\mathbb{P}\left[\eta_{p_1}(\tilde{\mathbf{x}}_t^u) \geq \eta_{p_2}(\tilde{\mathbf{x}}_t^u), \eta_{p_1}(\tilde{\mathbf{x}}_t^u) < \frac{\sum_{v_j \neq p_1}\left[\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u) - \tau_{v_j p_1}(\tilde{\mathbf{x}}_t^u)\right]\eta_{v_j}(\tilde{\mathbf{x}}_t^u) + 2\epsilon}{\left[\tau_{p_1 p_1}(\tilde{\mathbf{x}}_t^u) - \tau_{p_1 v_m}(\tilde{\mathbf{x}}_t^u)\right]}\right] \\
= &\mathbb{P}\Bigg[\eta_{p_2}(\tilde{\mathbf{x}}_t^u) \leq \eta_{p_1}(\tilde{\mathbf{x}}_t^u) < \eta_{p_2}(\tilde{\mathbf{x}}_t^u) + \\
&\quad \frac{\sum_{v_j \neq p_1}\left[\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u) - \tau_{v_j p_1}(\tilde{\mathbf{x}}_t^u)\right]\eta_{v_j}(\tilde{\mathbf{x}}_t^u) + 2\epsilon - \left[\tau_{p_1 p_1}(\tilde{\mathbf{x}}_t^u) - \tau_{p_1 v_m}(\tilde{\mathbf{x}}_t^u)\right]\eta_{p_2}(\tilde{\mathbf{x}}_t^u)}{\left[\tau_{p_1 p_1}(\tilde{\mathbf{x}}_t^u) - \tau_{p_1 v_m}(\tilde{\mathbf{x}}_t^u)\right]}\Bigg]
\end{aligned}
\tag{13}
$$

Recall that $\epsilon \leq \frac{1}{2}\Big[\left[\tau_{p_1 p_1}(\tilde{\mathbf{x}}_t^u) - \tau_{p_1 v_m}(\tilde{\mathbf{x}}_t^u)\right]\left[\alpha_0 + \eta_{p_2}(\tilde{\mathbf{x}}_t^u)\right] - \sum_{v_j \neq p_1}\left[\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u) - \tau_{v_j p_1}(\tilde{\mathbf{x}}_t^u)\right]\eta_{v_j}(\tilde{\mathbf{x}}_t^u)\Big]$, which implies $\frac{\sum_{v_j \neq p_1}\left[\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u) - \tau_{v_j p_1}(\tilde{\mathbf{x}}_t^u)\right]\eta_{v_j}(\tilde{\mathbf{x}}_t^u) + 2\epsilon - \left[\tau_{p_1 p_1}(\tilde{\mathbf{x}}_t^u) - \tau_{p_1 v_m}(\tilde{\mathbf{x}}_t^u)\right]\eta_{p_2}(\tilde{\mathbf{x}}_t^u)}{\left[\tau_{p_1 p_1}(\tilde{\mathbf{x}}_t^u) - \tau_{p_1 v_m}(\tilde{\mathbf{x}}_t^u)\right]} \leq \alpha_0$. Hence, the relaxed Multiclass Tsybakov Condition holds and the probability of Case-3 is bounded by $C\left(\frac{\sum_{v_j \neq p_1}\left[\tau_{v_j v_m}(\tilde{\mathbf{x}}_t^u) - \tau_{v_j p_1}(\tilde{\mathbf{x}}_t^u)\right]\eta_{v_j}(\tilde{\mathbf{x}}_t^u) + 2\epsilon - \left[\tau_{p_1 p_1}(\tilde{\mathbf{x}}_t^u) - \tau_{p_1 v_m}(\tilde{\mathbf{x}}_t^u)\right]\eta_{p_2}(\tilde{\mathbf{x}}_t^u)}{\left[\tau_{p_1 p_1}(\tilde{\mathbf{x}}_t^u) - \tau_{p_1 v_m}(\tilde{\mathbf{x}}_t^u)\right]}\right)^{\lambda}$, namely, $C\big(\mathcal{O}(\epsilon)\big)^{\lambda}$.

□

B.4 The Proof of Theorem 3

$$
\boldsymbol{E}_{\mathrm{DDUI}} = \mathbb{P}\left[p_1' = \tilde{v}_{t-l}^u, \frac{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\tilde{v}_{t-l}^u}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\tilde{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]} < \beta'\right] + \mathbb{P}\left[p_1' \neq \tilde{v}_{t-l}^u, \frac{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\tilde{v}_{t-l}^u}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\tilde{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]} \geq \beta'\right]. \tag{14}
$$

Theorem 3 (The Upper Bound of $\boldsymbol{E}_{\mathrm{DDUI}}$). *Given Assumption 1 and the set of weights* $\{w_h \mid 1 \leq h \leq H, 0 \leq w_h \leq 1, \sum_{h=1}^{H} w_h = 1\}$, $\forall \langle \overleftarrow{\mathbf{x}}_t^u, \tilde{v}_{t-l}^u \rangle$, *let* $\beta_1' = \left[\frac{\tau'_{\tilde{v}_{t-l}^u \tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u)\eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u)+\sum_{v_j \neq \tilde{v}_{t-l}^u} \tau'_{v_j \tilde{v}_{t-l}^u} \eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u)}{\sum_{h=1}^{H}[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)]}\right]$, $\beta_2' = \left[\frac{\sum_{h=1}^{H}[w_h \overleftarrow{f}_{\tilde{v}_{t-l}^u}^h(\overleftarrow{\mathbf{x}}_t^u)]}{\tau'_{\overleftarrow{v}_m \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u)+\sum_{v_j \neq \overleftarrow{v}_m} \tau'_{v_j \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u)}\right]$, $\xi_1' = |\beta' - \beta_1'|$, $\xi_2' = |\beta' - \beta_2'|$. *Assume* $\xi_2' < \beta_2'$, $\epsilon' \leq \min\Big[\alpha_0 \tau'_{\tilde{v}_{t-l}^u \tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u) - \xi_1', \frac{\alpha_0 \tau'_{\overleftarrow{v}_m \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\beta_2'(\beta_2'-\xi_2')-\xi_2'}{\beta_2'(\beta_2'-\xi_2')}, \frac{1}{2}\big[[\tau'_{p_1' p_1'}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{p_1' \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)][\alpha_0 + \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u)] - \sum_{v_j \neq p_1'}[\tau'_{v_j \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{v_j p_1'}(\overleftarrow{\mathbf{x}}_t^u)]\eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u)\big]\Big]$. *We have:* $\boldsymbol{E}_{\mathrm{DDUI}} \leq C\big(\mathcal{O}(\epsilon'+\xi_1')\big)^{\lambda} + C\big(\mathcal{O}(\epsilon'+\xi_2')\big)^{\lambda} + C\big(\mathcal{O}(\epsilon')\big)^{\lambda}$.

Proof. For the first term of $\boldsymbol{E}_{\mathrm{DDUI}}$, we have:

$$
\begin{aligned}
&\mathbb{P}\left[p_1' = \tilde{v}_{t-l}^u, \frac{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\tilde{v}_{t-l}^u}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]} < \beta'\right] \\
\leq &\mathbb{P}\left[p_1' = \tilde{v}_{t-l}^u, \eta'_{\tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u) \geq \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u), \sum_{h=1}^{H} w_h\left[\tilde{\eta}'_{\tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u) - \epsilon'\right] < \beta' \sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]\right] \\
= &\mathbb{P}\left[p_1' = \tilde{v}_{t-l}^u, \eta'_{\tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u) \geq \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u), \tilde{\eta}'_{\tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u) < \beta' \sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right] + \epsilon'\right] \\
= &\mathbb{P}\left[p_1' = \tilde{v}_{t-l}^u, \eta'_{\tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u) \geq \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u), \sum_{v_j \in \mathcal{V}} \tau'_{v_j \tilde{v}_{t-l}^u} \eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u) < \beta' \sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right] + \epsilon'\right] \\
\leq &\mathbb{P}\left[p_1' = \tilde{v}_{t-l}^u, \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u) \leq \eta'_{\tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u) < \frac{\beta' \sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right] + \epsilon' - \sum_{v_j \neq \tilde{v}_{t-l}^u}\left[\tau'_{v_j \tilde{v}_{t-l}^u} \eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u)\right]}{\tau'_{\tilde{v}_{t-l}^u \tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u)}\right] \\
\leq &\mathbb{P}\left[p_1' = \tilde{v}_{t-l}^u, \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u) \leq \eta'_{\tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u) < \frac{(\beta_1' + \xi_1') \sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right] + \epsilon' - \sum_{v_j \neq \tilde{v}_{t-l}^u}\left[\tau'_{v_j \tilde{v}_{t-l}^u} \eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u)\right]}{\tau'_{\tilde{v}_{t-l}^u \tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u)}\right].
\end{aligned}
\tag{15}
$$

We substitute β_1' with $\frac{\tau'_{\tilde{v}_{t-l}^u \tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u)\eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u)+\sum_{v_j \neq \tilde{v}_{t-l}^u} \tau'_{v_j \tilde{v}_{t-l}^u} \eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u)}{\sum_{h=1}^{H}[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)]}$ and obtain:

$$
\begin{aligned}
&\mathbb{P}\left[p_1' = \tilde{v}_{t-l}^u, \frac{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\tilde{v}_{t-l}^u}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]} < \beta'\right] \\
\leq &\mathbb{P}\left[p_1' = \tilde{v}_{t-l}^u, \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u) \leq \eta'_{\tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u) < \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u) + \frac{\epsilon'}{\tau'_{\tilde{v}_{t-l}^u \tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u)} + \frac{\xi_1' \sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\tau'_{\tilde{v}_{t-l}^u \tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u)}\right] \\
\leq &\mathbb{P}\left[\eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u) \leq \eta'_{\tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u) < \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u) + \frac{\epsilon' + \xi_1'}{\tau'_{\tilde{v}_{t-l}^u \tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u)}\right].
\end{aligned}
\tag{16}
$$

Recall that $\epsilon' \leq \alpha_0 \tau'_{\tilde{v}_{t-l}^u \tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u) - \xi_1'$, which implies $\frac{\epsilon'+\xi_1'}{\tau'_{\tilde{v}_{t-l}^u \tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u)} \leq \alpha_0$. Hence, the relaxed Multiclass Tsybakov Condition holds, and the probability of the first term of $\boldsymbol{E}_{\mathrm{DDUI}}$ (Eq. 14) is bounded by $C\Big(\frac{\epsilon'+\xi_1'}{\tau'_{\tilde{v}_{t-l}^u \tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u)}\Big)^{\lambda}$, namely, $C\big(\mathcal{O}(\epsilon' + \xi_1')\big)^{\lambda}$.

For the second term of $\boldsymbol{E}_{\mathrm{DDUI}}$, we have:

$$
\begin{aligned}
&\mathbb{P}\left[p_1' \neq \tilde{v}_{t-l}^u, \frac{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\tilde{v}_{t-l}^u}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]} \geq \beta'\right] \\
= &\mathbb{P}\left[p_1' \neq \tilde{v}_{t-l}^u, p_1' = \overleftarrow{v}_m, \frac{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\tilde{v}_{t-l}^u}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]} \geq \beta'\right] \\
&+ \mathbb{P}\left[p_1' \neq \tilde{v}_{t-l}^u, p_1' \neq \overleftarrow{v}_m, \frac{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\tilde{v}_{t-l}^u}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^{H}\left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]} \geq \beta'\right]
\end{aligned}
\tag{17}
$$

For the first term of Eq. 17, we have:

$$
\begin{aligned}
&\mathbb{P}\left[p_1' \neq \tilde{v}_{t-l}^u, p_1' = \overleftarrow{v}_m, \frac{\sum_{h=1}^H \left[w_h \overleftarrow{f}_{\tilde{v}_{t-l}^u}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^H \left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]} \geq \beta'\right] \\
\leq &\mathbb{P}\left[p_1' = \overleftarrow{v}_m, \frac{\sum_{h=1}^H \left[w_h \overleftarrow{f}_{\tilde{v}_{t-l}^u}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^H \left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]} \geq \beta'\right] \\
\leq &\mathbb{P}\left[p_1' = \overleftarrow{v}_m, \eta_{\overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u) \geq \eta_{p_2'}'(\overleftarrow{\mathbf{x}}_t^u), \sum_{h=1}^H w_h \left[\tilde{\eta}_{\overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u) - \epsilon'\right] \leq \frac{\sum_{h=1}^H \left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\beta'}\right] \\
= &\mathbb{P}\left[p_1' = \overleftarrow{v}_m, \eta_{\overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u) \geq \eta_{p_2'}'(\overleftarrow{\mathbf{x}}_t^u), \tilde{\eta}_{\overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u) \leq \frac{\sum_{h=1}^H \left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\beta'} + \epsilon'\right] \\
= &\mathbb{P}\left[p_1' = \overleftarrow{v}_m, \eta_{\overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u) \geq \eta_{p_2'}'(\overleftarrow{\mathbf{x}}_t^u), \sum_{v_j \in \mathcal{V}} \tau_{v_j v_m}'(\overleftarrow{\mathbf{x}}_t^u) \eta_{v_j}'(\overleftarrow{\mathbf{x}}_t^u) \leq \frac{\sum_{h=1}^H \left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\beta'} + \epsilon'\right] \\
\leq &\mathbb{P}\left[p_1' = \overleftarrow{v}_m, \eta_{p_2'}'(\overleftarrow{\mathbf{x}}_t^u) \leq \eta_{\overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u) \leq \frac{\frac{\sum_{h=1}^H [w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)]}{\beta'} - \sum_{v_j \neq v_m} \left[\tau_{v_j v_m}'(\overleftarrow{\mathbf{x}}_t^u) \eta_{v_j}'(\overleftarrow{\mathbf{x}}_t^u)\right]}{\tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)} + \frac{\epsilon'}{\tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)}\right] \\
\leq &\mathbb{P}\left[p_1' = \overleftarrow{v}_m, \eta_{p_2'}'(\overleftarrow{\mathbf{x}}_t^u) \leq \eta_{\overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u) \leq \frac{\frac{\sum_{h=1}^H [w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)]}{\beta_2' - \xi_2'} - \sum_{v_j \neq v_m} \left[\tau_{v_j v_m}'(\overleftarrow{\mathbf{x}}_t^u) \eta_{v_j}'(\overleftarrow{\mathbf{x}}_t^u)\right]}{\tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)} + \frac{\epsilon'}{\tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)}\right] \\
= &\mathbb{P}\left[p_1' = \overleftarrow{v}_m, \eta_{p_2'}'(\overleftarrow{\mathbf{x}}_t^u) \leq \eta_{\overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u) \leq \frac{\frac{\sum_{h=1}^H [w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)]}{\beta_2'} - \sum_{v_j \neq v_m} \left[\tau_{v_j v_m}'(\overleftarrow{\mathbf{x}}_t^u) \eta_{v_j}'(\overleftarrow{\mathbf{x}}_t^u)\right]}{\tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)} + \frac{\epsilon'}{\tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)}\right. \\
&\left. + \frac{\xi_2' \sum_{h=1}^H [w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)]}{\beta_2'(\beta_2' - \xi_2') \tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)}\right]
\end{aligned}
\tag{18}
$$

We replace β_2' with $\frac{\sum_{h=1}^H [w_h \overleftarrow{f}_{\tilde{v}_t^u}^h(\overleftarrow{\mathbf{x}}_t^u)]}{\tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)\eta_{p_2'}'(\overleftarrow{\mathbf{x}}_t^u) + \sum_{v_j \neq v_m} \tau_{v_j v_m}'(\overleftarrow{\mathbf{x}}_t^u)\eta_{v_j}'(\overleftarrow{\mathbf{x}}_t^u)}$ and continue the calculation:

$$
\begin{aligned}
&\mathbb{P}\left[p_1' \neq \tilde{v}_t^u, p_1' = v_m, \frac{\sum_{h=1}^H \left[w_h \overleftarrow{f}_{\tilde{v}_t^u}^h(\overleftarrow{\mathbf{x}}_t^u)\right]}{\sum_{h=1}^H \left[w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)\right]} \geq \beta\right] \\
\leq &\mathbb{P}\left[p_1' = \overleftarrow{v}_m, \eta_{p_2'}'(\overleftarrow{\mathbf{x}}_t^u) \leq \eta_{\overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u) \leq \eta_{p_2'}'(\overleftarrow{\mathbf{x}}_t^u) + \frac{\epsilon'}{\tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)} + \frac{\xi_2' \sum_{h=1}^H [w_h \overleftarrow{f}_{\overleftarrow{v}_m}^h(\overleftarrow{\mathbf{x}}_t^u)]}{\beta_2'(\beta_2' - \xi_2') \tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)}\right] \\
\leq &\mathbb{P}\left[\eta_{p_2'}'(\overleftarrow{\mathbf{x}}_t^u) \leq \eta_{\overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u) \leq \eta_{p_2'}'(\overleftarrow{\mathbf{x}}_t^u) + \frac{\epsilon'}{\tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)} + \frac{\xi_2'}{\beta_2'(\beta_2' - \xi_2') \tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)}\right].
\end{aligned}
\tag{19}
$$

Recall that $\epsilon' \leq \frac{\alpha_0 \tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)\beta_2'(\beta_2' - \xi_2') - \xi_2'}{\beta_2'(\beta_2' - \xi_2')}$, which implies $\frac{\epsilon'}{\tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)} + \frac{\xi_2'}{\beta_2'(\beta_2' - \xi_2')\tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)} \leq \alpha_0$. Hence, the relaxed Multiclass Tsybakov Condition holds and the probability of the first term of Eq. 17 is bounded by $C\left(\frac{\epsilon' \beta_2'(\beta_2' - \xi_2') + \xi_2'}{\beta_2'(\beta_2' - \xi_2')\tau_{\overleftarrow{v}_m \overleftarrow{v}_m}'(\overleftarrow{\mathbf{x}}_t^u)}\right)^\lambda$, namely, $C\big(\mathcal{O}(\epsilon' + \xi_2')\big)^\lambda$.

Finally, for the second term of Eq. 17, we have:

$$
\begin{aligned}
&\mathbb{P}\Bigg[p_1' \neq \tilde{v}_{t-l}^u, p_1' \neq \overleftarrow{v}_m, \frac{\sum_{h=1}^H \big[w_h \overleftarrow{f}^h_{\tilde{v}_{t-l}^u}(\overleftarrow{\mathbf{x}}_t^u)\big]}{\sum_{h=1}^H \big[w_h \overleftarrow{f}^h_{\overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\big]} \geq \beta'\Bigg] \\
=&\mathbb{P}\Bigg[p_1' \neq v_m, \eta'_{p_1'}(\overleftarrow{\mathbf{x}}_t^u) \geq \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u), \frac{\sum_{h=1}^H \big[w_h \overleftarrow{f}^h_{p_1'}(\overleftarrow{\mathbf{x}}_t^u)\big]}{\sum_{h=1}^H \big[w_h \overleftarrow{f}^h_{\overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\big]} < 1\Bigg] \\
\leq&\mathbb{P}\Bigg[\eta'_{p_1'}(\overleftarrow{\mathbf{x}}_t^u) \geq \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u), \sum_{h=1}^H \big[w_h \overleftarrow{f}^h_{p_1'}(\overleftarrow{\mathbf{x}}_t^u)\big] < \sum_{h=1}^H \big[w_h \overleftarrow{f}^h_{\overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\big]\Bigg] \\
\leq&\mathbb{P}\Bigg[\eta'_{p_1'}(\overleftarrow{\mathbf{x}}_t^u) \geq \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u), \sum_{h=1}^H w_h \big[\tilde{\eta}'_{p_1'}(\overleftarrow{\mathbf{x}}_t^u) - \epsilon'\big] < \sum_{h=1}^H w_h \big[\tilde{\eta}'_{\overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u) + \epsilon'\big]\Bigg] \\
=&\mathbb{P}\Bigg[\eta'_{p_1'}(\overleftarrow{\mathbf{x}}_t^u) \geq \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u), \tilde{\eta}'_{p_1'}(\overleftarrow{\mathbf{x}}_t^u) - \epsilon' < \tilde{\eta}'_{\overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u) + \epsilon'\Bigg] \\
=&\mathbb{P}\Bigg[\eta'_{p_1'}(\overleftarrow{\mathbf{x}}_t^u) \geq \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u), \sum_{v_j \in \mathcal{V}} \tau'_{v_j p_1'}(\overleftarrow{\mathbf{x}}_t^u)\eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u) < \sum_{v_j \in \mathcal{V}} \tau'_{\overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u) + 2\epsilon'\Bigg] \\
=&\mathbb{P}\Bigg[\eta'_{p_1'}(\overleftarrow{\mathbf{x}}_t^u) \geq \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u), \tau'_{p_1' p_1'}(\overleftarrow{\mathbf{x}}_t^u)\eta'_{p_1'}(\overleftarrow{\mathbf{x}}_t^u) + \sum_{v_j \neq p_1'} \tau'_{v_j p_1'}(\overleftarrow{\mathbf{x}}_t^u)\eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u) < \\
&\quad \tau'_{p_1' \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\eta'_{p_1'}(\overleftarrow{\mathbf{x}}_t^u) + \sum_{v_j \neq p_1'} \tau'_{\overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u) + 2\epsilon'\Bigg] \\
=&\mathbb{P}\Bigg[\eta'_{p_1'}(\overleftarrow{\mathbf{x}}_t^u) \geq \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u), \eta'_{p_1'}(\overleftarrow{\mathbf{x}}_t^u)\big[\tau'_{p_1' p_1'}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{p_1' \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\big] < \\
&\quad \sum_{v_j \neq p_1'} \big[\tau'_{\overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{v_j p_1'}(\overleftarrow{\mathbf{x}}_t^u)\big]\eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u) + 2\epsilon'\Bigg] \\
=&\mathbb{P}\Bigg[\eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u) \leq \eta'_{p_1'}(\overleftarrow{\mathbf{x}}_t^u) < \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u) + \\
&\quad \frac{\sum_{v_j \neq p_1'} \big[\tau'_{\overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{v_j p_1'}(\overleftarrow{\mathbf{x}}_t^u)\big]\eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u) + 2\epsilon' - \big[\tau'_{p_1' p_1'}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{p_1' \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\big]\eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u)}{\big[\tau'_{p_1' p_1'}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{p_1' \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\big]}\Bigg]
\end{aligned}
\tag{20}
$$

Recall that $\epsilon' \leq \frac{1}{2}\big[[\tau'_{p_1' p_1'}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{p_1' \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)][\alpha_0 + \eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u)] - \sum_{v_j \neq p_1'}[\tau'_{\overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{v_j p_1'}(\overleftarrow{\mathbf{x}}_t^u)]\eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u)\big]$, which implies $\frac{\sum_{v_j \neq p_1'} \big[\tau'_{\overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{v_j p_1'}(\overleftarrow{\mathbf{x}}_t^u)\big]\eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u) + 2\epsilon' - \big[\tau'_{p_1' p_1'}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{p_1' \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\big]\eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u)}{\big[\tau'_{p_1' p_1'}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{p_1' \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\big]} \leq \alpha_0$. Hence, the relaxed Multiclass Tsybakov Condition holds and the probability of Case-3 is bounded by $C\Bigg(\frac{\sum_{v_j \neq p_1'} \big[\tau'_{\overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{v_j p_1'}(\overleftarrow{\mathbf{x}}_t^u)\big]\eta'_{v_j}(\overleftarrow{\mathbf{x}}_t^u) + 2\epsilon' - \big[\tau'_{p_1' p_1'}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{p_1' \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\big]\eta'_{p_2'}(\overleftarrow{\mathbf{x}}_t^u)}{\big[\tau'_{p_1' p_1'}(\overleftarrow{\mathbf{x}}_t^u) - \tau'_{p_1' \overleftarrow{v}_m}(\overleftarrow{\mathbf{x}}_t^u)\big]}\Bigg)^\lambda$, namely, $C\big(\mathcal{O}(\epsilon')\big)^\lambda$. □

B.5 The Proof of Theorem 4

Theorem 4. *Let $\mathcal{R}$ be a list of K items randomly sampled from $\mathcal{V}$ with replacement, $\zeta \in (0, 1)$, $r_H(\mathbf{x}_t^u, v_i) = \sum_{v_j \in \mathcal{V}} \mathbb{I}\big[\sum_{h=1}^H [w_h f^h_{v_j}(\tilde{\mathbf{x}}_t^u)] > \sum_{h=1}^H [w_h f^h_{v_i}(\tilde{\mathbf{x}}_t^u)]\big]$ be the rank of item v_i over the entire item set at the H-th epoch, and $\hat{r}_H(\mathbf{x}_t^u, v_i) = \sum_{v_j \in \mathcal{R}} \mathbb{I}\big[\sum_{h=1}^H [w_h f^h_{v_j}(\tilde{\mathbf{x}}_t^u)] > \sum_{h=1}^H [w_h f^h_{v_i}(\tilde{\mathbf{x}}_t^u)]\big]$ be the rank of v_i over $\mathcal{R}$ at the H-th epoch. We have: $\mathbb{P}\big[\big|\frac{\hat{r}_H(\mathbf{x}_t^u, v_i)}{K} - \frac{r_H(\mathbf{x}_t^u, v_i)}{|\mathcal{V}|}\big| \geq \zeta\big] \leq \exp(-2K\zeta^2)$.*

Proof. $\hat{r}_H(\mathbf{x}_t^u, v_i)$ can be deemed as a random variable that counts the number of items ranked higher than v_i in $\mathcal{R}$. Since the items in $\mathcal{R}$ are randomly sampled from $\mathcal{V}$ with replacement, $\hat{r}_H(\mathbf{x}_t^u, v_i)$ follows a Binomial distribution $\text{Binomial}(K, \frac{r_H(\mathbf{x}_t^u, v_i)}{|\mathcal{V}|})$. Thus, we have:

$$
\mathbb{E}\big[\hat{r}_H(\mathbf{x}_t^u, v_i)\big] = \frac{K \cdot r_H(\mathbf{x}_t^u, v_i)}{|\mathcal{V}|}. \tag{21}
$$

Meanwhile, $\hat{r}_H(\mathbf{x}_t^u, v_i)$ can also be viewed as the sum of K i.i.d. Bernoulli random variables $b_1, b_2, ..., b_k$, where $b_k = 1(1 \leq k \leq K)$ indicates the k-th sampled item for $\mathcal{R}$ is ranked higher than v_i, and $b_k = 0$ otherwise. Hence, by applying Hoeffding's inequality [5], we have:

$$\mathbb{P}\left[\left|\frac{1}{K}\sum_{k=1}^{K} b_k - \mathbb{E}\Big[\frac{1}{K}\sum_{k=1}^{K} b_k\Big]\right| \geq \zeta\right] \leq \exp(-2K\zeta^2). \tag{22}$$

Then by replacing $\sum_{k=1}^{K} b_k$ with $\hat{r}_H(\mathbf{x}_t^u, v_i)$, we obtain:

$$\mathbb{P}\left[\left|\frac{\hat{r}_H(\mathbf{x}_t^u, v_i)}{K} - \frac{\mathbb{E}\big[\hat{r}_H(\mathbf{x}_t^u, v_i)\big]}{K}\right| \geq \zeta\right] \leq \exp(-2K\zeta^2). \tag{23}$$

Next, we replace $\mathbb{E}\big[\hat{r}_H(\mathbf{x}_t^u, v_i)\big]$ with $\frac{K \cdot r_H(\mathbf{x}_t^u, v_i)}{|\mathcal{V}|}$ according to Eq. 21:

$$\mathbb{P}\left[\left|\frac{\hat{r}_H(\mathbf{x}_t^u, v_i)}{K} - \frac{r_H(\mathbf{x}_t^u, v_i)}{|\mathcal{V}|}\right| \geq \zeta\right] \leq \exp(-2K\zeta^2). \tag{24}$$

$\square$

C Additional Experimental Results

C.1 Hyper-parameter Settings for Baselines

For fair comparisons, we implement FPMC with PyTorch. For other baselines, we use the original code provided by the corresponding authors. All the baselines adopt Xavier [6] initializer and Adam [7] optimizer. We empirically find the optimal hyper-parameter setting for each baseline based on the performance on the validation set. The detailed hyper-parameter setting for each baseline is summarized in Table 1.

C.2 More Results on Hyper-parameter Analysis

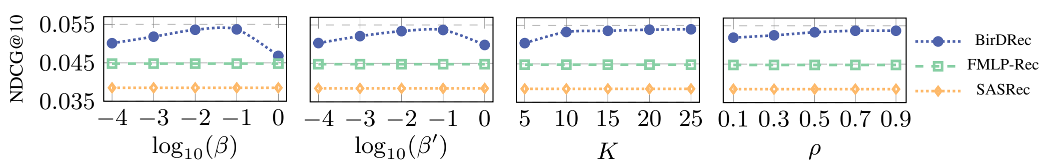


Figure 2: Effect of key hyper-parameters β, β', K, and ρ of BirDRec on Beauty dataset.

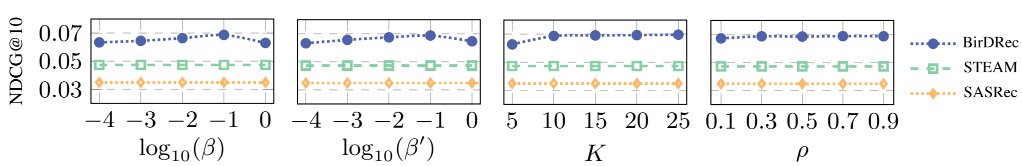


Figure 3: Effect of key hyper-parameters β, β', K, and ρ of BirDRec on Yelp dataset.

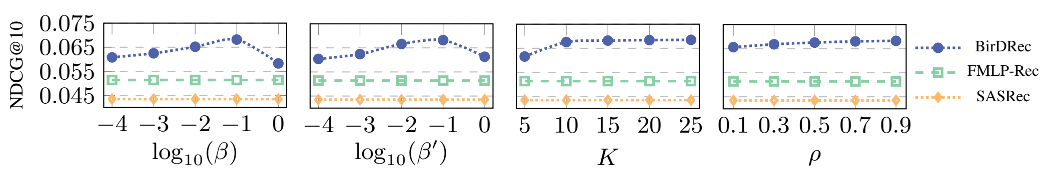
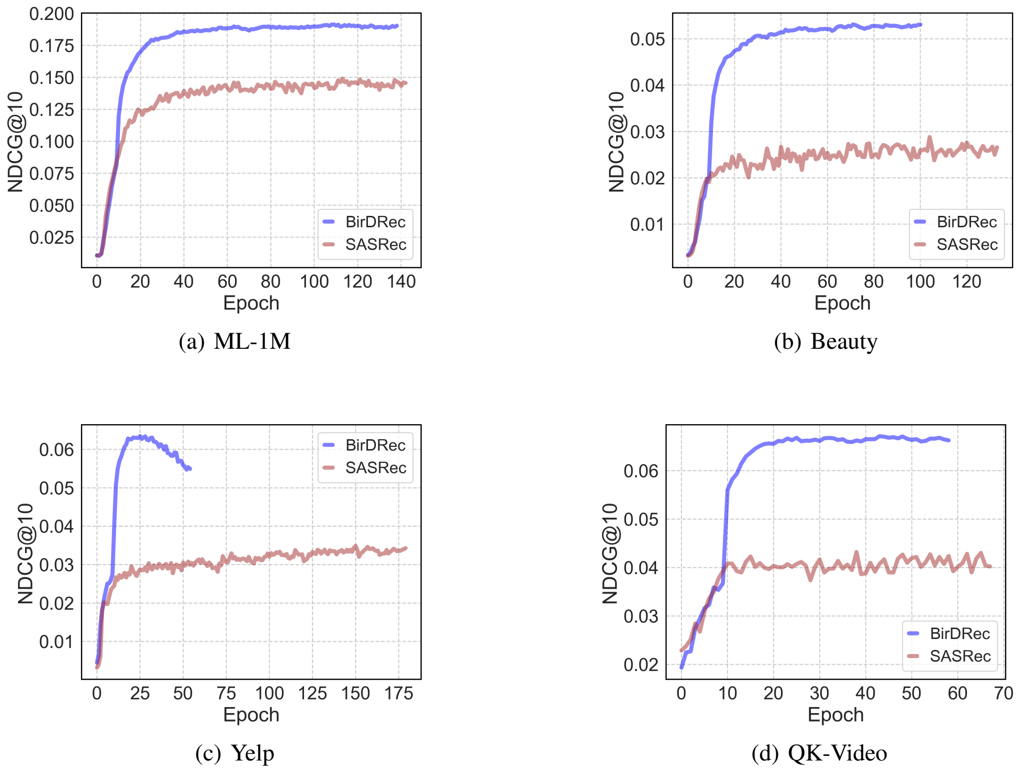


Figure 4: Effect of key hyper-parameters β, β', K, and ρ of BirDRec on QK-Vedio dataset.

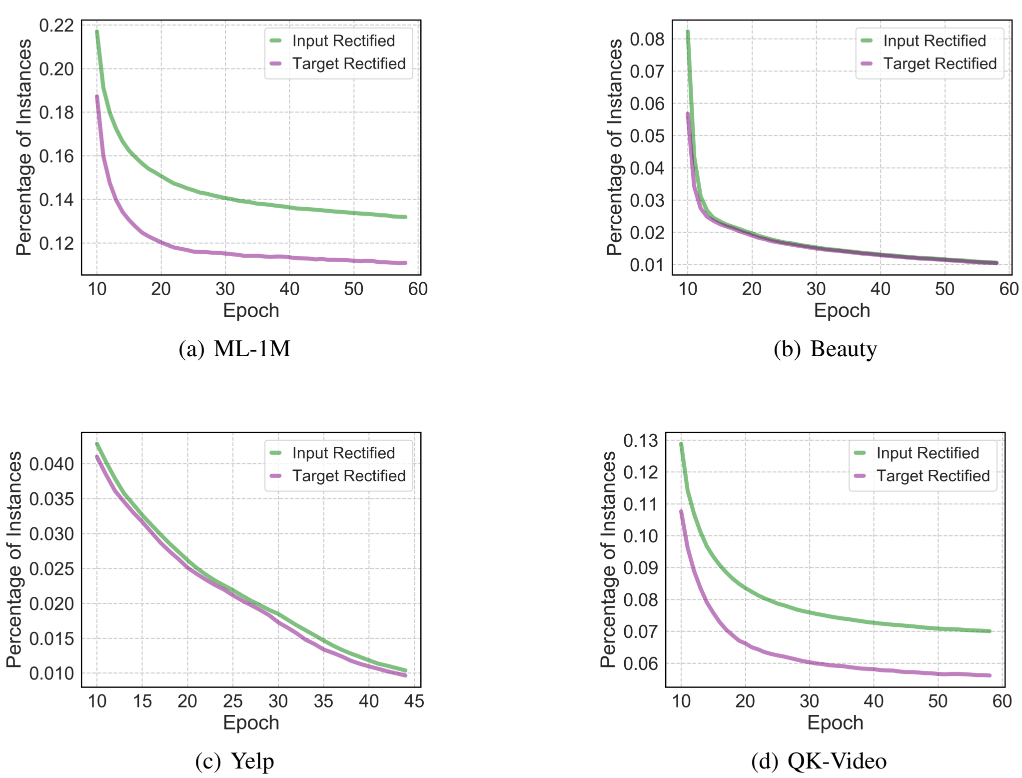
C.3 More Results on the Percentage of Rectified Instances



(a) ML-1M (b) Beauty

(c) Yelp (d) QK-Video

Figure 5: Testing accuracy (NDCG@10) of our BirDRec and SASRec with increasing epochs. The training stops if the best accuracy does not increase in 25 consecutive epochs.



(a) ML-1M (b) Beauty

(c) Yelp (d) QK-Video

Figure 6: The percentage of instances that are rectified with increasing epochs.

Table 1: The search space of hyper-parameters and the optimal settings found by grid search for all baselines on the four real-world datasets.

	Parameter	ML	Be	Ye	QK	Search Space
FPMC	embedding_size	64	64	64	128	$\{16, 32, 64, 128\}$
	learning_rate	10^{-3}	10^{-3}	10^{-2}	10^{-3}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	L2_regularization_coefficient	10^{-2}	10^{-3}	10^{-3}	10^{-2}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	input_length	10	5	5	10	$\{5, 10, 20, 30, 40, 50\}$
	batch_size	1024	512	512	1024	$\{128, 256, 512, 1024\}$
Caser	embedding_size	64	64	64	128	$\{16, 32, 64, 128\}$
	learning_rate	10^{-3}	10^{-3}	10^{-2}	10^{-3}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	L2_regularization_coefficient	10^{-2}	10^{-3}	10^{-3}	10^{-2}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	input_length	5	5	5	5	$\{5, 10, 20, 30, 40, 50\}$
	batch_size	512	256	256	512	$\{128, 256, 512, 1024\}$
	horizontal_filter_num	16	16	16	32	$\{4, 8, 16, 32, 64\}$
	vertical_filter_num	4	4	4	8	$\{1, 2, 4, 8, 16\}$
GRU4Rec	embedding_size	64	64	64	128	$\{16, 32, 64, 128\}$
	learning_rate	10^{-2}	10^{-2}	10^{-2}	10^{-2}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	L2_regularization_coefficient	10^{-2}	10^{-2}	10^{-2}	10^{-2}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	input_length	10	5	5	10	$\{5, 10, 20, 30, 40, 50\}$
	batch_size	1024	512	512	1024	$\{128, 256, 512, 1024\}$
	GRU_unit_number	256	256	256	512	$\{128, 256, 512, 1024\}$
SASRec	embedding_size	64	64	64	128	$\{16, 32, 64, 128\}$
	learning_rate	10^{-2}	10^{-2}	10^{-2}	10^{-3}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	L2_regularization_coefficient	10^{-2}	10^{-2}	10^{-2}	10^{-2}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	input_length	50	20	20	40	$\{5, 10, 20, 30, 40, 50\}$
	batch_size	128	128	512	1024	$\{128, 256, 512, 1024\}$
	self_attention_head_num	2	1	2	4	$\{1, 2, 4, 8\}$
	self_attention_block_num	1	1	1	2	$\{1, 2, 3, 4\}$
BERT4Rec	embedding_size	64	64	64	128	$\{16, 32, 64, 128\}$
	learning_rate	10^{-4}	10^{-4}	10^{-2}	10^{-3}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	L2_regularization_coefficient	10^{-2}	10^{-2}	10^{-2}	10^{-2}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	input_length	50	20	20	40	$\{5, 10, 20, 30, 40, 50\}$
	batch_size	256	256	512	1024	$\{128, 256, 512, 1024\}$
	self_attention_head_num	4	1	2	4	$\{1, 2, 3, 4\}$
	self_attention_block_num	4	1	1	2	$\{1, 2, 3, 4\}$
MAGNN	embedding_size	64	64	64	128	$\{16, 32, 64, 128\}$
	learning_rate	10^{-3}	10^{-3}	10^{-2}	10^{-3}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	L2_regularization_coefficient	10^{-2}	10^{-2}	10^{-2}	10^{-2}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	input_length	5	5	5	5	$\{5, 10, 20, 30, 40, 50\}$
	batch_size	1024	512	512	1024	$\{128, 256, 512, 1024\}$
	GNN_layer_num	2	2	2	2	$\{1, 2, 3, 4\}$
	memory_unit_num	10	10	10	20	$\{5, 10, 15, 20\}$
BERD	embedding_size	64	64	64	128	$\{16, 32, 64, 128\}$
	learning_rate	10^{-3}	10^{-3}	10^{-2}	10^{-3}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	L2_regularization_coefficient	10^{-2}	10^{-2}	10^{-2}	10^{-2}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	input_length	5	5	5	5	$\{5, 10, 20, 30, 40, 50\}$
	batch_size	1024	512	512	1024	$\{128, 256, 512, 1024\}$
	self_attention_head_num	2	2	2	2	$\{1, 2, 3, 4\}$
	self_attention_block_num	1	1	1	2	$\{1, 2, 3, 4\}$
	UGCN_layer_num	2	2	2	2	$\{1, 2, 3, 4\}$
	filter_ratio	0.05	0.05	0.05	0.10	$\{0.05, 0.10, 0.15, 0.20, 0.25\}$
	sample_size of $\mathcal{L}_{sam}$	4	4	4	4	$\{1, 2, 3, 4\}$
FMLP-Rec	embedding_size	64	64	64	128	$\{16, 32, 64, 128\}$
	learning_rate	10^{-3}	10^{-3}	10^{-3}	10^{-3}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	L2_regularization_coefficient	10^{-2}	10^{-3}	10^{-3}	10^{-2}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	input_length	50	20	20	40	$\{5, 10, 20, 30, 40, 50\}$
	batch_size	1024	256	256	1024	$\{128, 256, 512, 1024\}$
	learnable_filter_block_num	2	2	2	2	$\{1, 2, 3, 4\}$
STEAM	embedding_size	64	64	64	128	$\{16, 32, 64, 128\}$
	learning_rate	10^{-3}	10^{-3}	10^{-2}	10^{-3}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	L2_regularization_coefficient	10^{-2}	10^{-3}	10^{-3}	10^{-2}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	input_length	50	20	20	40	$\{5, 10, 20, 30, 40, 50\}$
	batch_size	1024	256	256	1024	$\{128, 256, 512, 1024\}$
	self_attention_head_num	2	1	1	2	$\{1, 2, 3, 4\}$
	self_attention_block_num	1	1	1	2	$\{1, 2, 3, 4\}$
	insertion_probability	0.2	0.4	0.4	0.2	$\{0.1, 0.2, 0.3, 0.4, 0.5\}$
	deletion_probability	0.1	0.1	0.1	0.1	$\{0.1, 0.2, 0.3, 0.4, 0.5\}$
	mask_probability	0.4	0.5	0.5	0.3	$\{0.1, 0.2, 0.3, 0.4, 0.5\}$
BirDRec	embedding_size	64	64	64	64	$\{16, 32, 64, 128\}$
	learning_rate	10^{-3}	10^{-3}	10^{-2}	10^{-3}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	L2_regularization_coefficient	10^{-2}	10^{-2}	10^{-2}	10^{-2}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$
	input_length	5	5	5	5	$\{5, 10, 20, 30, 40, 50\}$
	batch_size	1024	1024	1024	1024	$\{128, 256, 512, 1024\}$
	threshold β in DRUT	10^{-1}	10^{-1}	10^{-1}	10^{-1}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$
	threshold β' in DDUI	10^{-1}	10^{-1}	10^{-1}	10^{-1}	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$
	size K of rectification pools	10	10	10	10	$\{5, 10, 15, 20, 25\}$
	exponential decay rate ρ	0.9	0.9	0.9	0.9	$\{0.1, 0.3, 0.5, 0.7, 0.9\}$

References

[1] Di-Rong Chen and Tao Sun. Consistency of multiclass empirical risk minimization methods based on convex loss. *The Journal of Machine Learning Research*, 7:2435–2447, 2006.

[2] Yatong Sun, Bin Wang, Zhu Sun, and Xiaochun Yang. Does every data instance matter? enhancing sequential recommendation by eliminating unreliable data. In *IJCAI*, pages 1579–1585, 2021.

[3] Songzhu Zheng, Pengxiang Wu, Aman Goswami, Mayank Goswami, Dimitris Metaxas, and Chao Chen. Error-bounded correction of noisy labels. In *ICML*, pages 11447–11457, 2020.

[4] Wang-Cheng Kang and Julian McAuley. Self-attentive sequential recommendation. In *ICDM*, pages 197–206, 2018.

[5] Wassily Hoeffding. *Probability Inequalities for sums of Bounded Random Variables*, pages 409–426. 1994.

[6] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *AISTATS*, pages 249–256, 2010.

[7] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.