Mutual Information Learning for CTR Prediction (Supplementary Material)

1 PROOF OF THE INFORMATION FEATURE SELECTION EMBEDDING LAYER

THEOREM 1.1. *Under the CTRLID assumption of orders k_1 and k_2, the high-order conditional mutual information is equivalent to:*

$$I(X_m; L|S) = \frac{1}{\binom{2}{k_1}\binom{|S|}{k_2}} \sum_{L^i_{k_1} \in L_{k_1}} \sum_{S^j_{k_2} \in S_{k_2}} \left(I(X_m; L^i_{k_1}|S^j_{k_2}) + \sum_{l_t \in L - L^i_{k_1}} I(X_m; l_t|L^i_{k_1}, S^j_{k_2}) \right) \tag{1}$$

Proof. First, we have the CTRLID assumption:

$$p(L|X_m, S) = \prod_{L^i_{k_1} \in L_{k_1}} \prod_{S^j_{k_2} \in S_{k_2}} \left[p(L^i_{k_1}|X_m, S^j_{k_2}) \prod_{l_t \in L - L^i_{k_1}} p(l_t|X_m, S^j_{k_2}, L^i_{k_1}) \right]^{\frac{1}{\binom{2}{k_1}\binom{|S|}{k_2}}} \tag{2}$$

$$= \prod_{L^i_{k_1} \in L_{k_1}} \prod_{S^j_{k_2} \in S_{k_2}} \left[p(L^i_{k_1}|X_m, S^j_{k_2}) \tag{3}$$

$$\prod_{l_t \in L - L^i_{k_1}} p(l_t|X_m, S^j_{k_2}, L^i_{k_1}) \right]^{\frac{1}{\binom{2}{k_1}\binom{|S|}{k_2}}} \tag{4}$$

Now, we can express the mutual information as follows:

$$I(X_m; L|S) = \frac{1}{\binom{2}{k_1}\binom{|S|}{k_2}} \sum_{L^i_{k_1} \in L_{k_1}} \sum_{S^j_{k_2} \in S_{k_2}} \left\{ H(L^i_{k_1}|X_m, S^j_{k_2}) + \sum_{l_t \in L - L^{k_1}_i} H(l_t|X_m, L^i_{k_1}, S^j_{k_2}) \right\} \tag{5}$$

According to the definition of conditional mutual information in Formula 5 we have:

$$H(L^i_{k_1}|X_m, S^j_{k_2}) = H(L^i_{k_1}|S^j_{k_2}) - I(X_m; L^i_{k_1}|S^j_{k_2}) \tag{6}$$

$$H(l_t|X_m, L^i_{k_1}, S^j_{k_2}) = H(l_t|L^i_{k_1}, S^j_{k_2}) - I(X_m; l_t|L^i_{k_1}, S^j_{k_2}) \tag{7}$$

Let $\theta_1 = H(L^i_{k_1}|S^j_{k_2})$ and $\theta_2 = H(l_t|L^i_{k_1}, S^j_{k_2})$, which are two constants. Thus,

$$I(X_m; L|S) = H(L|S) - H(L|X_m, S)$$

$$= H(L|S) - \frac{1}{\binom{2}{k_1}\binom{|S|}{k_2}} \sum_{L^i_{k_1} \in L_{k_1}} \sum_{S^j_{k_2} \in S_{k_2}} \left\{ H(L^i_{k_1}|X_m, S^j_{k_2}) + \sum_{l_t \in L - L_{k_1}{}^i} H(l_t|X_m, L^i_{k_1}, S^j_{k_2}) \right\} \tag{8}$$

$$= H(L|S) - \frac{1}{\binom{2}{k_1}\binom{|S|}{k_2}} \sum_{L^i_{k_1} \in L_{k_1}} \sum_{S^j_{k_2} \in S_{k_2}} \left\{ \theta_1 - I(X_m; L^i_{k_1}|S^j_{k_2}) + \sum_{l_t \in L - L^i_{k_1}} (\theta_2 - I(X_m; l_t|L^i_{k_1}, S^j_{k_2})) \right\} \tag{9}$$

$$= \frac{1}{\binom{2}{k_1}\binom{|S|}{k_2}} \sum_{L^i_{k_1} \in L_{k_1}} \sum_{S^j_{k_2} \in S_{k_2}} \left\{ I(X_m; L^i_{k_1}|S^j_{k_2}) + \sum_{l_t \in L - L^i_{k_1}} I(X_m; l_t|L^i_{k_1}, S^j_{k_2}) \right\} + H(L|S) - \theta_1 - \theta_2 \tag{10}$$

$$\propto \frac{1}{\binom{2}{k_1}\binom{|S|}{k_2}} \sum_{L^i_{k_1} \in L_{k_1}} \sum_{S^j_{k_2} \in S_{k_2}} \left\{ I(X_m; L^i_{k_1}|S^j_{k_2}) + \sum_{l_t \in L - L^i_{k_1}} I(X_m; l_t|L^i_{k_1}, S^j_{k_2}) \right\} \tag{11}$$

where $\theta_1 - \theta_2$ is a constant.

THEOREM 1.2. *Under the CCTRFD assumption of orders k_1 and k_2, the high-order conditional mutual information is equivalent to:*

$$I(X_m; L|S) = \frac{1}{\binom{2}{k_1}\binom{|S|}{k_2}} \sum_{L^i_{k_1} \in L_{k_1}} \sum_{S^j_{k_2} \in S_{k_2}} \{ I(X_m; L^i_{k_1}, S^j_{k_2}) \} \tag{12}$$

$$- \frac{1}{\binom{|S|}{k_2}} \sum_{S^j_{k_2} \in S_{k_2}} I(X_m; S^j_{k_2}) \tag{13}$$

Proof. According to the definition, we have:

$$p(X_m|L,S) = \prod_{L^i_{k_1} \in L_{k_1}} \prod_{S^j_{k_2} \in S_{k_2}} [p(X_m|L^i_{k_1}, S^j_{k_2})]^{\frac{1}{\binom{2}{k_1}\cdot\binom{|S|}{k_2}}} \quad (14)$$

$$p(X_m|S) = \prod_{S^j_{k_2} \in S_{k_2}} [p(X_m|S^j_{k_2})]^{\frac{1}{\binom{|S|}{k_2}}} \quad (15)$$

$$H(X_m|S) = \frac{1}{\binom{|S|}{\mathrm{k}_2}} \sum_{S^j_{\mathrm{k}_2} \in S_{\mathrm{k}_2}} \{H(X_m|S^j_{\mathrm{k}_2})\} \quad (16)$$

$$H(X_m|L,S) = \frac{1}{\binom{2}{\mathrm{k}_1}\binom{|S|}{\mathrm{k}_2}} \sum_{L^i_{\mathrm{k}_1} \in L_{\mathrm{k}_1}} \sum_{S^j_{\mathrm{k}_2} \in S_{\mathrm{k}_2}} \{H(X_m|L^i_{\mathrm{k}_1}, S^j_{\mathrm{k}_2})\} \quad (17)$$

$$\begin{aligned}
I(X_m;L|S) &= H(X_m|S) - H(X_m|L,S) \\
&= \frac{1}{\binom{|S|}{\mathrm{k}_2}} \sum_{S^j_{\mathrm{k}_2} \in S_{\mathrm{k}_2}} \{H(X_m|S^j_{\mathrm{k}_2})\} \\
&\quad - \frac{1}{\binom{2}{\mathrm{k}_1}\binom{|S|}{\mathrm{k}_2}} \sum_{L^i_{\mathrm{k}_1} \in L_{\mathrm{k}_1}} \sum_{S^j_{\mathrm{k}_2} \in S_{\mathrm{k}_2}} \{H(X_m|L^i_{\mathrm{k}_1}, S^j_{\mathrm{k}_2})\} \\
&= \frac{1}{|S|_{\mathrm{k}_2}} \sum_{S^j_{\mathrm{k}_2} \in S_{\mathrm{k}_2}} \{H(X_m) - I(X_m;S^j_{\mathrm{k}_2})\} \\
&\quad - \frac{1}{\binom{2}{\mathrm{k}_1}\binom{|S|}{\mathrm{k}_2}} \sum_{L^i_{\mathrm{k}_1} \in L_{\mathrm{k}_1}} \sum_{S^j_{\mathrm{k}_2} \in S_{\mathrm{k}_2}} \{H(X_m) - I(X_m;L^i_{\mathrm{k}_1}, S_{\mathrm{k}_2})^j\} \\
&= H(X_m) - \frac{1}{\binom{|S|}{\mathrm{k}_2}} \sum_{S^j_{\mathrm{k}_2} \in S_{\mathrm{k}_2}} \{I(X_m;S^j_{\mathrm{k}_2})\} - H(X_m) \\
&\quad + \frac{1}{\binom{2}{\mathrm{k}_1}\binom{|S|}{\mathrm{k}_2}} \sum_{L^i_{\mathrm{k}_1} \in L_{\mathrm{k}_1}} \sum_{S^j_{\mathrm{k}_2} \in S_{\mathrm{k}_2}} \{I(X_m;L^i_{\mathrm{k}_1}, S^j_{\mathrm{k}_2})\} \\
&= \frac{1}{\binom{2}{\mathrm{k}_1}\binom{|S|}{\mathrm{k}_2}} \sum_{L^i_{\mathrm{k}_1} \in L_{\mathrm{k}_1}} \sum_{S^j_{\mathrm{k}_2} \in S_{\mathrm{k}_2}} \{I(X_m;L^i_{\mathrm{k}_1}, S^j_{\mathrm{k}_2})\} \\
&\quad - \frac{1}{\binom{|S|}{\mathrm{k}_2}} \sum_{S^j_{\mathrm{k}_2} \in S_{\mathrm{k}_2}} \{I(X_m;S^j_{\mathrm{k}_2})\}
\end{aligned}$$

COROLLARY 1

Under the CCTRFD assumption of order k_1 and k_2, the high-order conditional mutual information is equivalent to:

$$I(X_m;L|S) = \frac{1}{\binom{2}{k_1}\cdot\binom{|S|}{k_2}} \sum_{L^i_{k_1} \in L_{k_1}} \sum_{S^j_{k_2} \in S_{k_2}} \{I(X_m;L^i_{k_1}|S^j_{k_2})\} \quad (18)$$

Proof. According to the definition of joint mutual information, we obtain:

$$I(X_m;L^i_{k_1}, S^j_{k_2}) = I(X_m;S^j_{k_2}) + I(X_m;L^i_{k_1}|S^j_{k_2}) \quad (19)$$

Then,

$$I(X_m;L|S) = \frac{1}{\binom{2}{k_1}\cdot\binom{|S|}{k_2}} \sum_{L^i_{k_1} \in L_{k_1}} \sum_{S^j_{k_2} \in S_{k_2}} \{I(X_m;L^i_{k_1}|S^j_{k_2})\} \quad (20)$$

When $k_1 = 1$, $k_2 = 1$:

$$I(X_m;L|S) = \frac{1}{2\cdot|S|} \sum_{l_i \in L} \sum_{X_j \in S} \{I(X_m;l_i|X_j)\} \quad (21)$$

COROLLARY 2

Under the CCTRFD assumption of order k_1 and k_2, the high-order conditional mutual information is equivalent to:

$$\begin{aligned}
I(X_m;L|S) = \frac{1}{\binom{2}{k_1}} \sum_{L^i_{k_1} \in L_{k_1}} I(X_m;L^i_{k_1}) + \frac{1}{\binom{2}{k_1}\cdot\binom{|S|}{k_2}} \sum_{L^{k_1}_i \in L_{k_1}} \sum_{S^{k_2}_j \in S_{k_2}} \\
\{I(X_m;S^j_{k_2}|L^i_{k_1})\} - \frac{1}{\binom{|S|}{k_2}} \sum_{S^j_{k_2} \in S_{k_2}} \{I(X_m;S^j_{k_2})\} \quad (22)
\end{aligned}$$

Proof. According to the definition of joint mutual information, we obtain:

$$I(X_m;L^i_{k_1}, S^j_{k_2}) = I(X_m;L^i_{k_1}) + I(X_m;S^j_{k_2}|L^i_{k_1}) \quad (23)$$

Then,

$$\begin{aligned}
I(X_m;L|S) &= \frac{1}{\binom{2}{k_1}} \sum_{L^i_{k_1} \in L_{k_1}} I(X_m;L^i_{k_1}) \\
&\quad + \frac{1}{\binom{2}{k_1}\cdot\binom{|S|}{k_2}} \sum_{L^i_{k_1} \in L_{k_1}} \sum_{S^j_{k_2} \in S_{k_2}} \{I(X_m;S^j_{k_2}|L^i_{k_1})\} \\
&\quad - \frac{1}{\binom{|S|}{k_2}} \sum_{S^{k_2}_j \in S_{k_2}} \{I(X_m;S^j_{k_2})\} \quad (24)
\end{aligned}$$

When $k_1 = 1$, $k_2 = 1$:

$$\begin{aligned}
I(X_m;L|S) &= \frac{1}{2} \sum_{l_i \in L} I(X_m;li) + \frac{1}{2\cdot|S|} \sum_{l_i \in L} \sum_{X_j \in S} \{I(X_m;X_j|l_i)\} \\
&\quad - \frac{1}{|S|} \sum_{X_j \in S} \{I(X_m;X_j)\}
\end{aligned}$$

REFERENCES