

---

# Towards a Pairwise Ranking Model with Orderliness and Monotonicity for Label Enhancement

---

## A Proofs of theorems

**Proof of Theorem 2.5.** It is evident that the probabilities  $p(\ell_i \prec \ell_j | z_i, z_j)$ ,  $p(\ell_i \simeq \ell_j | z_i, z_j)$ , and  $p(\ell_i \succ \ell_j | z_i, z_j)$  can all be considered as functions of  $z_i - z_j$ . Therefore,  $\phi(z_i, z_j)$  can be equivalently rewritten as Equation (1).

$$\begin{aligned} \underline{\phi}(\Delta) &= Z^{-1} \exp(a_{\underline{\phi}}\Delta + b_{\underline{\phi}}), \quad \tilde{\phi}(\Delta) = Z^{-1} \exp(a_{\tilde{\phi}}\Delta^2 + b_{\tilde{\phi}}), \quad \bar{\phi}(\Delta) = Z^{-1} \exp(a_{\bar{\phi}}\Delta + b_{\bar{\phi}}), \\ Z &= \exp(a_{\underline{\phi}}\Delta + b_{\underline{\phi}}) + \exp(a_{\tilde{\phi}}\Delta^2 + b_{\tilde{\phi}}) + \exp(a_{\bar{\phi}}\Delta + b_{\bar{\phi}}). \end{aligned} \quad (1)$$

Obviously, any two pairs of label description degrees  $(z_i, z_j)$  and  $(z_m, z_n)$  that satisfy the condition  $z_i - z_j = z_m - z_n$  correspond to the same  $\Delta$  and the same function output.  $\square$

**Proof of Theorem 2.6.** Let  $a_{\bar{\phi}} = -a_{\underline{\phi}} = a$  and  $b_{\bar{\phi}} = b_{\underline{\phi}} = b$ . We have

$$\begin{aligned} \tilde{\phi}(-\Delta) &= \frac{\exp(a_{\tilde{\phi}}(-\Delta)^2 + b_{\tilde{\phi}})}{\exp(a(-\Delta) + b) + \exp(a_{\tilde{\phi}}(-\Delta)^2 + b_{\tilde{\phi}}) + \exp(-a(-\Delta) + b)} = \tilde{\phi}(\Delta), \\ \bar{\phi}(-\Delta) &= \frac{\exp(a(-\Delta) + b)}{\exp(a(-\Delta) + b) + \exp(a_{\tilde{\phi}}(-\Delta)^2 + b_{\tilde{\phi}}) + \exp(-a(-\Delta) + b)} \\ &= \frac{\exp(-a\Delta + b)}{\exp(-a\Delta + b) + \exp(a_{\tilde{\phi}}\Delta^2 + b_{\tilde{\phi}}) + \exp(a\Delta + b)} = \underline{\phi}(\Delta). \end{aligned} \quad (2)$$

$\square$

**Proof of Theorem 2.7.** We first prove the monotonicity of  $\tilde{\phi}$ . The first-order derivative of  $\tilde{\phi}(\Delta)$  w.r.t.  $\Delta$  can be written as follows:

$$C\tilde{\phi}'(\Delta) = (2a_{\tilde{\phi}}\Delta - a_{\bar{\phi}}) \exp(a_{\bar{\phi}}\Delta + b_{\bar{\phi}}) + (2a_{\tilde{\phi}}\Delta - a_{\underline{\phi}}) \exp(a_{\underline{\phi}}\Delta + b_{\underline{\phi}}), \quad C > 0. \quad (3)$$

According to  $a_{\bar{\phi}} = -a_{\underline{\phi}} \exp(b_{\underline{\phi}} - b_{\bar{\phi}})$ , we have

$$a_{\bar{\phi}} = -a_{\underline{\phi}} \exp(b_{\underline{\phi}} - b_{\bar{\phi}}) \iff (C\tilde{\phi}'(\Delta))|_{\Delta=0} = 0. \quad (4)$$

$$\tilde{\phi}''(0)|_{a_{\bar{\phi}} = -a_{\underline{\phi}} \exp(b_{\underline{\phi}} - b_{\bar{\phi}})} = \frac{(2a_{\tilde{\phi}}e^{2b_{\bar{\phi}}} + 2a_{\tilde{\phi}}e^{b_{\bar{\phi}}+b_{\underline{\phi}}} - a_{\underline{\phi}}^2e^{2b_{\underline{\phi}}} - a_{\underline{\phi}}^2e^{b_{\underline{\phi}}+b_{\bar{\phi}}})e^{b_{\bar{\phi}}-b_{\underline{\phi}}}}{e^{2b_{\bar{\phi}}} + e^{2b_{\tilde{\phi}}} + e^{2b_{\underline{\phi}}} + 2e^{b_{\bar{\phi}}+b_{\tilde{\phi}}} + 2e^{b_{\bar{\phi}}+b_{\underline{\phi}}} + 2e^{b_{\tilde{\phi}}+b_{\underline{\phi}}}} < 0. \quad (5)$$

Besides, we have

$$\begin{aligned} C\tilde{\phi}'(\Delta) &= (2a_{\tilde{\phi}}\Delta - a_{\bar{\phi}}) \exp(a_{\bar{\phi}}\Delta + b_{\bar{\phi}}) + (2a_{\tilde{\phi}}\Delta - a_{\underline{\phi}}) \exp(a_{\underline{\phi}}\Delta + b_{\underline{\phi}}) < 0 \\ &\iff (2a_{\tilde{\phi}}\Delta - a_{\bar{\phi}}) \exp(a_{\bar{\phi}}\Delta)(-a_{\underline{\phi}}) + (2a_{\tilde{\phi}}\Delta - a_{\underline{\phi}}) \exp(a_{\underline{\phi}}\Delta)a_{\bar{\phi}} < 0 \\ &\iff (2a_{\tilde{\phi}}\Delta - a_{\bar{\phi}}) \exp(a_{\bar{\phi}}\Delta - a_{\underline{\phi}}\Delta)(-a_{\underline{\phi}}) + (2a_{\tilde{\phi}}\Delta - a_{\underline{\phi}})a_{\bar{\phi}} < 0 \\ &\iff \exp(a_{\bar{\phi}}\Delta - a_{\underline{\phi}}\Delta) > (2a_{\tilde{\phi}}\Delta - a_{\underline{\phi}})a_{\bar{\phi}}(2a_{\tilde{\phi}}\Delta - a_{\bar{\phi}})^{-1}a_{\underline{\phi}}^{-1}. \end{aligned} \quad (6)$$

Since  $\exp(a_{\bar{\phi}}\Delta - a_{\underline{\phi}}\Delta) > 1$  holds for any  $\Delta \in (0, 1)$ , and the first-order derivative of  $(2a_{\tilde{\phi}}\Delta - a_{\underline{\phi}})a_{\bar{\phi}}(2a_{\tilde{\phi}}\Delta - a_{\bar{\phi}})^{-1}a_{\underline{\phi}}^{-1}$  w.r.t.  $\Delta$  is  $-\frac{2a_{\tilde{\phi}}a_{\bar{\phi}}(a_{\bar{\phi}}-a_{\underline{\phi}})}{a_{\underline{\phi}}(2\Delta a_{\tilde{\phi}}-a_{\bar{\phi}})^2} < 0$ . Then

$(2a_{\tilde{\phi}}\Delta - a_{\underline{\phi}})a_{\overline{\phi}}(2a_{\tilde{\phi}}\Delta - a_{\overline{\phi}})^{-1}a_{\underline{\phi}}^{-1} < 1$  holds for any  $\Delta \in (0, 1)$ . Therefore,  $\exp(a_{\overline{\phi}}\Delta - a_{\underline{\phi}}\Delta) > 1 > (2a_{\tilde{\phi}}\Delta - a_{\underline{\phi}})a_{\overline{\phi}}(2a_{\tilde{\phi}}\Delta - a_{\overline{\phi}})^{-1}a_{\underline{\phi}}^{-1}$ , which proves that  $\tilde{\phi}(\Delta)$  is decreasing for  $\Delta \in (0, 1)$ . Similarly, it can also be proven that  $\tilde{\phi}(\Delta)$  is increasing for  $\Delta \in (-1, 0)$ .

Secondly, we prove the monotonicity of  $\overline{\phi}(\Delta)$ . The first-order derivative of  $\overline{\phi}(\Delta)$  w.r.t.  $\Delta$  can be written as follows:

$$\overline{C}\phi'(\Delta) = (a_{\overline{\phi}} - 2a_{\tilde{\phi}}\Delta) \exp(a_{\tilde{\phi}}\Delta^2 + b_{\tilde{\phi}}) + (a_{\overline{\phi}} - a_{\underline{\phi}}) \exp(a_{\underline{\phi}}\Delta + b_{\underline{\phi}}), \quad \overline{C} > 0. \quad (7)$$

$$\begin{aligned} & (a_{\overline{\phi}} - 2a_{\tilde{\phi}}\Delta) \exp(a_{\tilde{\phi}}\Delta^2 + b_{\tilde{\phi}}) + (a_{\overline{\phi}} - a_{\underline{\phi}}) \exp(a_{\underline{\phi}}\Delta + b_{\underline{\phi}}) > 0 \\ \iff & \frac{a_{\overline{\phi}} - 2a_{\tilde{\phi}}\Delta}{a_{\overline{\phi}} - a_{\underline{\phi}}} + \exp(-a_{\tilde{\phi}}\Delta^2 + a_{\underline{\phi}}\Delta + b_{\underline{\phi}} - b_{\tilde{\phi}}) > 0. \end{aligned} \quad (8)$$

It is clear that Equation (8) holds for any  $\Delta \in [-1, 1]$  if  $\Delta > a_{\overline{\phi}}(2a_{\tilde{\phi}})^{-1}$  or  $a_{\overline{\phi}}(2a_{\tilde{\phi}})^{-1} < -1$ . If  $-1 \leq \Delta < a_{\overline{\phi}}(2a_{\tilde{\phi}})^{-1}$  (i.e.,  $\Delta \geq -1$ ,  $2a_{\tilde{\phi}}\Delta - a_{\overline{\phi}} > 0$ , and  $0 < a_{\overline{\phi}} < -2a_{\tilde{\phi}}$ ), Equation (8) can be rewritten as:

$$\frac{a_{\overline{\phi}} - 2a_{\tilde{\phi}}\Delta}{a_{\overline{\phi}} - a_{\underline{\phi}}} + \exp(-a_{\tilde{\phi}}\Delta^2 + a_{\underline{\phi}}\Delta + b_{\underline{\phi}} - b_{\tilde{\phi}}) > 0 \iff -a_{\tilde{\phi}}\Delta^2 + a_{\underline{\phi}}\Delta + b_{\underline{\phi}} - b_{\tilde{\phi}} - \log\left(\frac{2a_{\tilde{\phi}}\Delta - a_{\overline{\phi}}}{a_{\overline{\phi}} - a_{\underline{\phi}}}\right) > 0. \quad (9)$$

Let  $h_{\overline{\phi}}(\Delta) \triangleq -a_{\tilde{\phi}}\Delta^2 + a_{\underline{\phi}}\Delta + b_{\underline{\phi}} - b_{\tilde{\phi}} - \log(2a_{\tilde{\phi}}\Delta - a_{\overline{\phi}}) + \log(a_{\overline{\phi}} - a_{\underline{\phi}})$ . We have

$$\begin{aligned} h'_{\overline{\phi}}(\Delta) &= -2a_{\tilde{\phi}}\Delta + a_{\underline{\phi}} - \frac{2a_{\tilde{\phi}}}{2a_{\tilde{\phi}}\Delta - a_{\overline{\phi}}}, \quad h''_{\overline{\phi}}(\Delta) = \left(\frac{2a_{\tilde{\phi}}}{2a_{\tilde{\phi}}\Delta - a_{\overline{\phi}}}\right)^2 - 2a_{\tilde{\phi}} > 0, \\ h'_{\overline{\phi}}(\overline{\Delta}^*) &= 0 \iff \overline{\Delta}^* = \frac{a_{\overline{\phi}} + a_{\underline{\phi}} + \sqrt{(a_{\overline{\phi}} - a_{\underline{\phi}})^2 - 8a_{\tilde{\phi}}}}{4a_{\tilde{\phi}}} < \frac{a_{\overline{\phi}}}{2a_{\tilde{\phi}}}. \end{aligned} \quad (10)$$

If  $\overline{\Delta}^* \leq -1$ , i.e.,  $a_{\overline{\phi}} \geq 2a_{\tilde{\phi}}(2a_{\tilde{\phi}} + a_{\underline{\phi}} + 1)(-2a_{\tilde{\phi}} - a_{\underline{\phi}})^{-1}$ , then  $h_{\overline{\phi}}(\Delta) \geq h_{\overline{\phi}}(-1)$  holds for any  $\Delta \in [-1, a_{\overline{\phi}}(2a_{\tilde{\phi}})^{-1}]$ , so  $h_{\overline{\phi}}(\Delta) > 0$  if  $h_{\overline{\phi}}(-1) > 0$ . If  $\overline{\Delta}^* > -1$ , i.e.,  $a_{\overline{\phi}} < 2a_{\tilde{\phi}}(2a_{\tilde{\phi}} + a_{\underline{\phi}} + 1)(-2a_{\tilde{\phi}} - a_{\underline{\phi}})^{-1}$ , then  $h_{\overline{\phi}}(\Delta) \geq h_{\overline{\phi}}(\overline{\Delta}^*)$  holds for any  $\Delta \in [-1, a_{\overline{\phi}}(2a_{\tilde{\phi}})^{-1}]$ . In this case,  $h_{\overline{\phi}}(\Delta) > 0$  if  $h_{\overline{\phi}}(\overline{\Delta}^*) > 0$ . Therefore,  $\overline{\phi}(\Delta)$  increases over  $[-1, 1]$  if:

$$(a_{\overline{\phi}}(2a_{\tilde{\phi}})^{-1} < -1) \vee (\overline{\Delta}^* \leq -1 \wedge h_{\overline{\phi}}(-1) > 0) \vee (\overline{\Delta}^* > -1 \wedge h_{\overline{\phi}}(\overline{\Delta}^*) > 0). \quad (11)$$

Finally, we prove that  $\phi(\Delta)$  decreases over  $[-1, 1]$ . First, the first-order derivative can be written as:

$$\underline{C}\phi'(\Delta) = (a_{\underline{\phi}} - 2a_{\tilde{\phi}}\Delta) \exp(a_{\tilde{\phi}}\Delta^2 + b_{\tilde{\phi}}) - (a_{\overline{\phi}} - a_{\underline{\phi}}) \exp(a_{\overline{\phi}}\Delta + b_{\overline{\phi}}), \quad \underline{C} > 0. \quad (12)$$

$$\begin{aligned} & (a_{\underline{\phi}} - 2a_{\tilde{\phi}}\Delta) \exp(a_{\tilde{\phi}}\Delta^2 + b_{\tilde{\phi}}) - (a_{\overline{\phi}} - a_{\underline{\phi}}) \exp(a_{\overline{\phi}}\Delta + b_{\overline{\phi}}) < 0 \\ \iff & \frac{a_{\underline{\phi}} - 2a_{\tilde{\phi}}\Delta}{a_{\underline{\phi}} - a_{\overline{\phi}}} + \exp(-a_{\tilde{\phi}}\Delta^2 + a_{\overline{\phi}}\Delta + b_{\overline{\phi}} - b_{\tilde{\phi}}) > 0. \end{aligned} \quad (13)$$

It is clear that Equation (13) holds if  $a_{\underline{\phi}}(2a_{\tilde{\phi}})^{-1} \geq 1$  or  $\Delta \leq a_{\underline{\phi}}(2a_{\tilde{\phi}})^{-1} < 1$ . If  $a_{\underline{\phi}}(2a_{\tilde{\phi}})^{-1} < \Delta \leq 1$ , Equation (13) can be equivalently rewritten as:

$$\frac{a_{\underline{\phi}} - 2a_{\tilde{\phi}}\Delta}{a_{\underline{\phi}} - a_{\overline{\phi}}} + \exp(-a_{\tilde{\phi}}\Delta^2 + a_{\overline{\phi}}\Delta + b_{\overline{\phi}} - b_{\tilde{\phi}}) > 0 \iff -a_{\tilde{\phi}}\Delta^2 + a_{\overline{\phi}}\Delta + b_{\overline{\phi}} - b_{\tilde{\phi}} - \log\left(\frac{2a_{\tilde{\phi}}\Delta - a_{\underline{\phi}}}{a_{\underline{\phi}} - a_{\overline{\phi}}}\right) > 0. \quad (14)$$

Let  $h_{\underline{\phi}}(\Delta) \triangleq -a_{\tilde{\phi}}\Delta^2 + a_{\overline{\phi}}\Delta + b_{\overline{\phi}} - b_{\tilde{\phi}} - \log(a_{\underline{\phi}} - 2a_{\tilde{\phi}}\Delta) + \log(a_{\overline{\phi}} - a_{\underline{\phi}})$ . We have

$$\begin{aligned} h'_{\underline{\phi}}(\Delta) &= -2a_{\tilde{\phi}}\Delta + a_{\overline{\phi}} - \frac{2a_{\tilde{\phi}}}{2a_{\tilde{\phi}}\Delta - a_{\underline{\phi}}}, \quad h''_{\underline{\phi}}(\Delta) = \left(\frac{2a_{\tilde{\phi}}}{2a_{\tilde{\phi}}\Delta - a_{\underline{\phi}}}\right)^2 - 2a_{\tilde{\phi}} > 0, \\ h'_{\underline{\phi}}(\underline{\Delta}^*) &= 0 \iff \underline{\Delta}^* = \frac{a_{\overline{\phi}} + a_{\underline{\phi}} - \sqrt{(a_{\overline{\phi}} - a_{\underline{\phi}})^2 - 8a_{\tilde{\phi}}}}{4a_{\tilde{\phi}}} > \frac{a_{\underline{\phi}}}{2a_{\tilde{\phi}}}. \end{aligned} \quad (15)$$

If  $\underline{\Delta}^* \geq 1$ , then  $h_{\underline{\phi}}(\Delta) \geq h_{\underline{\phi}}(1)$  holds for any  $\Delta \in [a_{\underline{\phi}}(2a_{\bar{\phi}})^{-1}, 1]$ . In this case,  $h_{\underline{\phi}}(\Delta) > 0$  if  $h_{\underline{\phi}}(1) > 0$ . If  $\underline{\Delta}^* < 1$ , then  $h_{\underline{\phi}}(\Delta) \geq h_{\underline{\phi}}(\underline{\Delta}^*)$  holds for any  $\Delta \in [a_{\bar{\phi}}(2a_{\bar{\phi}})^{-1}, 1]$ . In this case,  $h_{\underline{\phi}}(\Delta) > 0$  if  $h_{\underline{\phi}}(\underline{\Delta}^*) > 0$ . Therefore,  $\underline{\phi}(\Delta)$  decreases over  $[-1, 1]$  if:

$$(a_{\underline{\phi}}(2a_{\bar{\phi}})^{-1} > 1) \vee (\underline{\Delta}^* \geq 1 \wedge h_{\underline{\phi}}(1) > 0) \vee (\underline{\Delta}^* < 1 \wedge h_{\underline{\phi}}(\underline{\Delta}^*) > 0). \quad (16)$$

□

**Proof of Theorem 2.8.** We desire that  $\exists 0 < \Delta^+ < 1$  such that  $\tilde{\phi}(\Delta) > \bar{\phi}(\Delta)$  holds for any  $\Delta < \Delta^+$  and  $\tilde{\phi}(\Delta) < \bar{\phi}(\Delta)$  holds for any  $\Delta > \Delta^+$ , i.e.,

$$\begin{aligned} \forall -1 < \Delta < \Delta^+ \quad \exp(a_{\bar{\phi}}\Delta^2 + b_{\bar{\phi}}) > \exp(a_{\bar{\phi}}\Delta + b_{\bar{\phi}}) &\iff a_{\bar{\phi}}\Delta^2 + b_{\bar{\phi}} > a_{\bar{\phi}}\Delta + b_{\bar{\phi}}, \\ \forall \Delta^+ < \Delta < 1 \quad \exp(a_{\bar{\phi}}\Delta^2 + b_{\bar{\phi}}) < \exp(a_{\bar{\phi}}\Delta + b_{\bar{\phi}}) &\iff a_{\bar{\phi}}\Delta^2 + b_{\bar{\phi}} < a_{\bar{\phi}}\Delta + b_{\bar{\phi}}. \end{aligned} \quad (17)$$

Let  $f_{\bar{\phi}\bar{\phi}}(\Delta) := a_{\bar{\phi}}\Delta^2 + b_{\bar{\phi}} - a_{\bar{\phi}}\Delta - b_{\bar{\phi}}$ . It is obvious that  $f_{\bar{\phi}\bar{\phi}}(\Delta)$  is a quadratic function. Equation (17) holds iff  $f_{\bar{\phi}\bar{\phi}}(-1) > 0$ ,  $f_{\bar{\phi}\bar{\phi}}(0) > 0$ , and  $f_{\bar{\phi}\bar{\phi}}(1) < 0$ , i.e.,

$$f_{\bar{\phi}\bar{\phi}}(-1) > 0 \iff a_{\bar{\phi}} + b_{\bar{\phi}} + a_{\bar{\phi}} - b_{\bar{\phi}} > 0; f_{\bar{\phi}\bar{\phi}}(0) > 0 \iff b_{\bar{\phi}} - b_{\bar{\phi}} > 0; f_{\bar{\phi}\bar{\phi}}(1) < 0 \iff a_{\bar{\phi}} + b_{\bar{\phi}} - a_{\bar{\phi}} - b_{\bar{\phi}} < 0. \quad (18)$$

Similarly, we desire that there exists  $-1 < \Delta^- < 0$  such that  $\tilde{\phi}(\Delta) > \underline{\phi}(\Delta)$  for  $\Delta > \Delta^-$  and  $\tilde{\phi}(\Delta) < \underline{\phi}(\Delta)$  for  $\Delta < \Delta^-$ , i.e.,

$$\forall -1 < \Delta < \Delta^- \quad a_{\bar{\phi}}\Delta^2 + b_{\bar{\phi}} < a_{\underline{\phi}}\Delta + b_{\underline{\phi}}; \quad \forall \Delta^- < \Delta < 1 \quad a_{\bar{\phi}}\Delta^2 + b_{\bar{\phi}} > a_{\underline{\phi}}\Delta + b_{\underline{\phi}}. \quad (19)$$

Let  $f_{\bar{\phi}\underline{\phi}}(\Delta) := a_{\bar{\phi}}\Delta^2 + b_{\bar{\phi}} - a_{\underline{\phi}}\Delta - b_{\underline{\phi}}$ . Equation (19) holds iff  $f_{\bar{\phi}\underline{\phi}}(-1) < 0$ ,  $f_{\bar{\phi}\underline{\phi}}(0) > 0$ , and  $f_{\bar{\phi}\underline{\phi}}(1) > 0$ , i.e.,

$$f_{\bar{\phi}\underline{\phi}}(-1) < 0 \iff a_{\bar{\phi}} + b_{\bar{\phi}} + a_{\underline{\phi}} - b_{\underline{\phi}} < 0; f_{\bar{\phi}\underline{\phi}}(0) > 0 \iff b_{\bar{\phi}} - b_{\underline{\phi}} > 0; f_{\bar{\phi}\underline{\phi}}(1) > 0 \iff a_{\bar{\phi}} + b_{\bar{\phi}} - a_{\underline{\phi}} - b_{\underline{\phi}} > 0. \quad (20)$$

In summary, the probability orderliness assumption holds if

$$\max\{b_{\bar{\phi}} - a_{\bar{\phi}}, b_{\underline{\phi}} + a_{\underline{\phi}}\} < a_{\bar{\phi}} + b_{\bar{\phi}} < \min\{b_{\bar{\phi}} + a_{\bar{\phi}}, b_{\underline{\phi}} - a_{\underline{\phi}}\}, \quad b_{\bar{\phi}} > \max\{b_{\bar{\phi}}, b_{\underline{\phi}}\}. \quad (21)$$

□

## B More Visualization Analysis on PROM

Here, we provide a more comprehensive visualization of the PROM model. In Figure 1, we show the shape of PROM when the parameters  $a$ ,  $b$ , and  $t$  take different values. The parameters of PROM must satisfy the condition (A):  $a > |b - t| \wedge b > 0 \wedge t > 0$ , which serves the probability orderliness of PROM. Further, if the parameters of PROM satisfy the condition (B):  $a \geq 2b$ , PROM can be guaranteed to satisfy the probability monotonicity assumption. Otherwise, if  $a < 2b$  but  $a \geq \sqrt{4b^2 - 2t}$ , the parameters should satisfy condition (C.1):  $t < b + a - \log \frac{2b-a}{2a}$ . If  $a < \sqrt{4b^2 - 2t}$ , the parameters should satisfy condition (C.2):  $t < \frac{2a\sqrt{a^2+2b+a^2}}{4b} - \log \frac{\sqrt{a^2+2b-a}}{2a} + \frac{1}{2}$ . In Figure 2, we show the feasible region of PROM. The parameter  $a$  and the parameter  $b$  take values from  $(0, 10^3)$ , the parameter  $t$  is set to the values 100, 200, 400, and 800, respectively. The blue color represents the regions where conditions (A) and (B) are satisfied; the orange color represents the regions where conditions (A) and condition (C), i.e. the union of conditions (C.1) and (C.2), are satisfied.

$$\begin{aligned} \mathcal{F}_{\text{blue}} &= \{(a, b) \mid (a \geq 2b) \wedge (a > |b - t|) \wedge (b > 0) \wedge (t > 0)\}, \\ \mathcal{F}_{\text{orange}} &= \left\{ (a, b) \mid (a > |b - t| \wedge b > 0 \wedge t > 0) \wedge \right. \\ &\quad \left( (a^2 \geq 4b^2 - 2b \wedge t < b + a - \log \frac{2b-a}{2a}) \vee \right. \\ &\quad \left. \left. \left( t < \frac{2a\sqrt{a^2+2b+a^2}}{4b} - \log \frac{\sqrt{a^2+2b-a}}{2a} + \frac{1}{2} \wedge a^2 < 4b^2 - 2b \right) \right) \right\}. \end{aligned} \quad (22)$$

In Figure 3, we show the marginal effect of different parameters on the shape of PROM. In each sub-figure, a darker color indicates a larger value for the corresponding parameter, and a lighter color indicates a smaller value for the corresponding parameter.

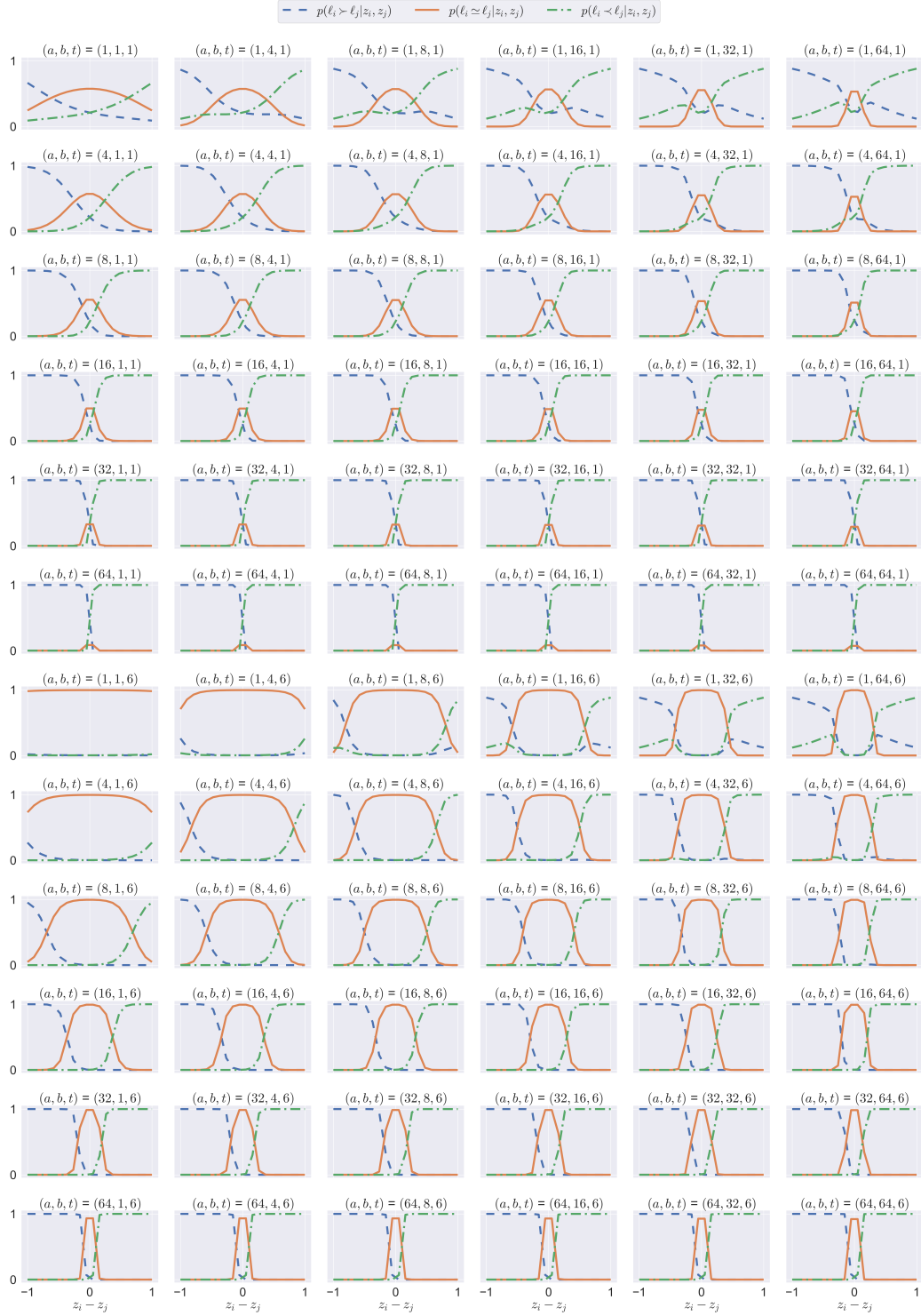


Figure 1: PROM on varying values of parameters. The blue curves correspond to  $p(\ell_i > \ell_j | z_i, z_j)$ , the orange curves correspond to  $p(\ell_i \simeq \ell_j | z_i, z_j)$ , and the green curves correspond to  $p(\ell_i < \ell_j | z_i, z_j)$ . The x-axis denotes  $z_i - z_j$ , y-axis denotes the value of probability.



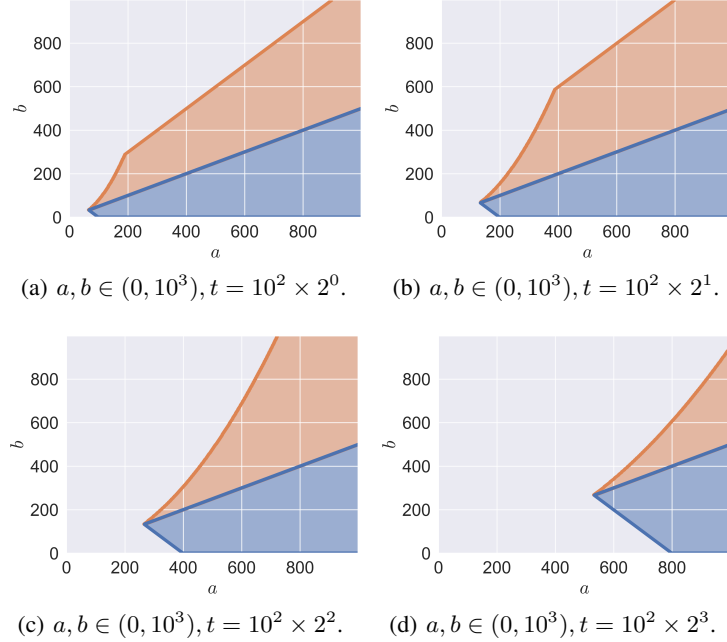


Figure 2: Feasible region of parameters of PROM.  $a$  and  $b$  take values from  $(0, 10^3)$ ,  $t$  takes the values 100, 200, 400 and 800 respectively. The shaded area represents the feasible region.

ID	Dataset	Task	# Instances	# Features	# Labels
1	Painting (Machajdik and Hanbury, 2010)	Image emotion analysis	280	142	8
2	Emotion6 (Peng et al., 2015)	Image emotion analysis	1,980	168	7
3	Music (Lee et al., 2021)	Music emotion analysis	360	5,992	9
4	BU-3DFE (Yin et al., 2006)	Facial expression analysis	2,500	243	6
5	JAFPE (Lyons et al., 1998)	Facial expression analysis	213	243	6
6	Movie (Geng, 2016)	Movie rating prediction	7,755	1,869	5
7	Alpha (Geng, 2016)	Yeast gene expression analysis	2,465	24	18
8	Cdc (Geng, 2016)	Yeast gene expression analysis	2,465	24	15
9	Cold (Geng, 2016)	Yeast gene expression analysis	2,465	24	14
10	Diau (Geng, 2016)	Yeast gene expression analysis	2,465	24	7
11	Dtt (Geng, 2016)	Yeast gene expression analysis	2,465	24	6
12	Elu (Geng, 2016)	Yeast gene expression analysis	2,465	24	6
13	Heat (Geng, 2016)	Yeast gene expression analysis	2,465	24	4
14	Spo (Geng, 2016)	Yeast gene expression analysis	2,465	24	4
15	Spo5 (Geng, 2016)	Yeast gene expression analysis	2,465	24	3
16	Spoem (Geng, 2016)	Yeast gene expression analysis	2,465	24	2

Table 1: Introduction of datasets.

## C Experiments

### C.1 Datasets

“Painting” (Machajdik and Hanbury, 2010) comprises 280 images, where each image is represented by a 142-dimensional feature vector. For the label distributions, about 230 individuals rate each image on eight emotion categories (i.e., entertainment, anger, awe, satisfaction, disgust, excitement, fear, and sadness), and each image has about 14 annotations. “Emotion6” (Peng et al., 2015) comprises 1,980 images, each annotated with voting labels for seven emotion categories (i.e., anger, disgust, happiness, fear, sadness, surprise, and neutral). Following (Ren et al., 2019), we employ principal component analysis to extract a 168-dimensional feature vector for each image. “Movie” dataset (Geng, 2016) aggregates 54,242,292 ratings from 478,656 different users, covering 7,755

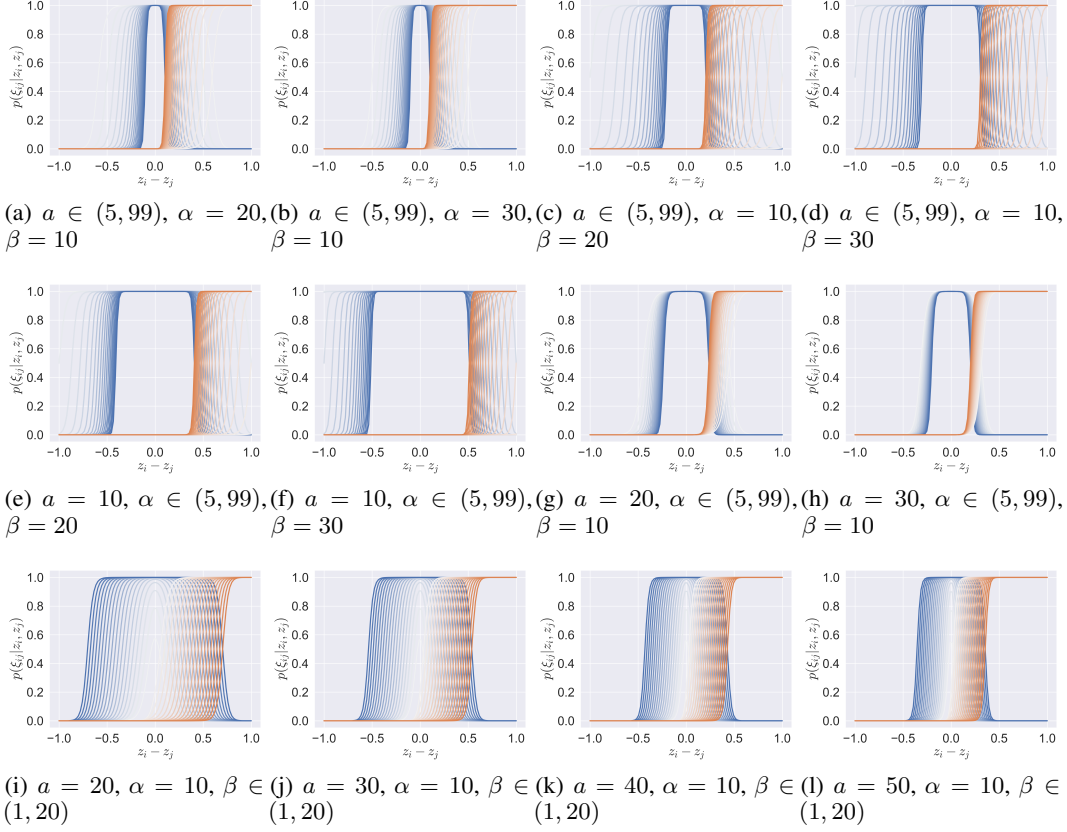


Figure 3: Marginal effect of parameters on the shape of PROM. The x-axis denotes the values of  $z_i - z_j$ , and the y-axis denotes the probability values of  $p(\xi_{ij}|z_i, z_j)$ . The blue curves represent the shape of  $p(\ell_i \simeq \ell_j | z_i, z_j)$ , and the orange curves represent the shape of  $p(\ell_i \prec \ell_j | z_i, z_j)$ . Darker colors indicate larger parameter values, and lighter colors indicate smaller values.

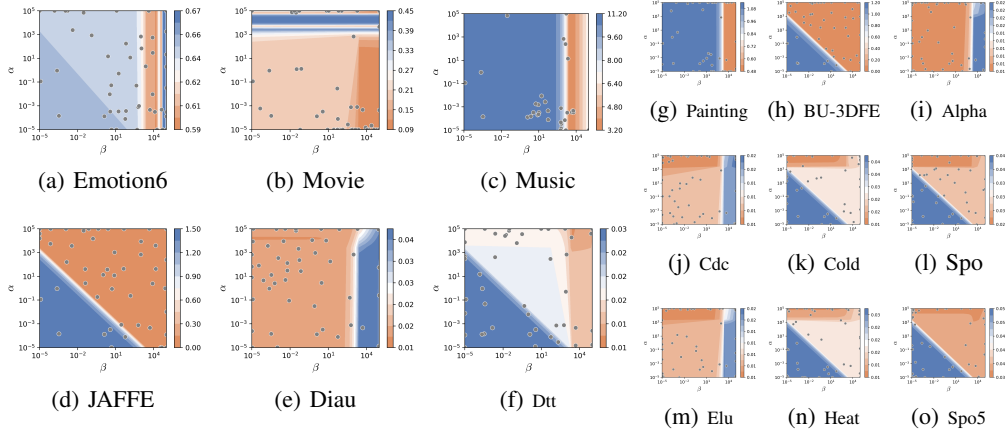


Figure 4: KL-measured prediction performance of LE-PROM under different values of  $\alpha$  and  $\beta$  after averaging out  $\lambda$ .

movies, with ratings ranging from 1 to 5. The features for each movie are extracted based on metadata such as genre, director, and actors, and each movie is represented as a 1,869-dimensional vector. “Music” dataset (Lee et al., 2021) encompasses 360 popular songs from leading music charts in various countries. Each song is represented by a 128-dimensional feature vector. Participants from

the UK, South Korea, and Portugal assess the songs on a 4-level scale according to their perceived moods (i.e., calmness, tension, happiness, sadness, dancability, love, dreaminess, electronic feel, and energy). “JAFPE” dataset (Lyons et al., 1998) comprises 213 gray-scale images of human face, where each image is represented by a 243-dimensional feature vector. Sixty individuals rate each image using a 5-level scale for the six basic emotion labels (i.e., happiness, sadness, surprise, fear, anger, and disgust). “BU-3DFE” dataset (Yin et al., 2006) comprises 2,500 facial expression images, where each image is represented by a 243-dimensional feature vector, and each image is rated by 23 individuals. The ten datasets from “Alpha” to “Spoem” (Geng, 2016) originate from a series of biological experiments conducted on budding yeast, involving 2,465 yeast genes. Each gene is represented by a 24-dimensional phylogenetic profile vector. The label distributions are collected from the gene expression levels at different time points.

## C.2 Evaluation Metrics

We evaluate the prediction performance by Cheb (Chebyshev distance), KL (Kullback–Leibler divergence), Cosine (cosine coefficient), and Intersec (intersection similarity). They are defined by Equation (23).

$$\begin{aligned}
\text{Cheb} \left( \{z_n\}_{n=1}^N, \{\hat{z}_n\}_{n=1}^N \right) &= \frac{1}{N} \sum_{n=1}^N \max \{ |z_n^{\ell_m} - \hat{z}_n^{\ell_m}| \}_{m=1}^M, \\
\text{KL} \left( \{z_n\}_{n=1}^N, \{\hat{z}_n\}_{n=1}^N \right) &= \frac{1}{N} \sum_{n=1}^N \sum_{m=1}^M z_n^{\ell_m} \log \frac{z_n^{\ell_m}}{\hat{z}_n^{\ell_m}}, \\
\text{Cosine} \left( \{z_n\}_{n=1}^N, \{\hat{z}_n\}_{n=1}^N \right) &= \frac{1}{N} \sum_{n=1}^N \frac{\langle z_n, \hat{z}_n \rangle}{\|z_n\|_2 \cdot \|\hat{z}_n\|_2}, \\
\text{Intersec} \left( \{z_n\}_{n=1}^N, \{\hat{z}_n\}_{n=1}^N \right) &= \frac{1}{N} \sum_{n=1}^N \sum_{m=1}^M \min \{ z_n^{\ell_m}, \hat{z}_n^{\ell_m} \},
\end{aligned} \tag{23}$$

where  $z$  denotes the ground-truth label distribution,  $\hat{z}$  denotes the predicted label distribution,  $\langle z_n, \hat{z}_n \rangle$  denotes the inner product of  $z_n$  and  $\hat{z}_n$ ,  $\|z_n\|_2$  denotes the  $L_2$  norm of the vector  $z_n$ .

## C.3 Experiments on Logical Label Enhancement

Here we provide more details of the experiments on logical label enhancement. First, we illustrate how the training set with label distributions is transformed into a training set with logical labels. Second, we show more experimental results.

**Reducing label distributions to logical labels.** The process of reducing the label distribution to a set of logical labels is delineated as follows:

1. Initially, an empty set  $\mathcal{Y}^+$  is established to store the relevant labels (i.e., the labels with logical value of 1).
2. Concurrently, all labels are designated as irrelevant by assigning their logical values to 0.
3. Subsequently, an iterative selection procedure commences.
  - (a) While the cumulative label description degrees  $H$  remains below a predefined threshold  $T$ , the label exhibiting the highest description degree among the unselected labels will be identified and selected.
  - (b) The chosen label will be set as the relevant label, i.e., the logical value of the chosen label will be updated to 1, and incorporated into the set  $\mathcal{Y}^+$ .
  - (c) Then we update the value of  $H$  by adding the description degree of the chosen label.
4. The selection process terminates when  $H$  exceeds the predefined threshold  $T$ .

In our experiments, the predefined threshold  $T$  is set to 0.5.

**More experimental results.** In Table 2 and Table 3, we show the prediction performance of comparison algorithms measured by Chebyshev distance and intersection similarity. In Figure 4, we

Dataset	PROM	CWLD	VIB-ILE	GLLE	KMLE	FCMLE
Painting	(1) 0.259 $\pm$ 0.011	(4) 0.273 $\pm$ 0.006●	(6) 0.761 $\pm$ 0.175●	(1) 0.259 $\pm$ 0.012	(5) 0.311 $\pm$ 0.021●	(4) 0.165 $\pm$ 0.036●
Emotion6	(2) 0.324 $\pm$ 0.006	(5) 0.334 $\pm$ 0.002●	(3) 0.327 $\pm$ 0.007	(4) 0.328 $\pm$ 0.005●	(1) 0.314 $\pm$ 0.009●	(6) 0.344 $\pm$ 0.005●
Movie	(1) 0.127 $\pm$ 0.002	(3) 0.154 $\pm$ 0.002●	(4) 0.159 $\pm$ 0.004●	(2) 0.130 $\pm$ 0.002●	(5) 0.162 $\pm$ 0.004●	(6) 0.172 $\pm$ 0.002●
Music	(1) 0.079 $\pm$ 0.003	(4) 0.086 $\pm$ 0.001●	(6) 0.130 $\pm$ 0.005●	(2) 0.084 $\pm$ 0.002●	(5) 0.102 $\pm$ 0.033●	(2) 0.084 $\pm$ 0.002●
BU-3DFE	(2) 0.126 $\pm$ 0.002	(6) 0.141 $\pm$ 0.002●	(5) 0.139 $\pm$ 0.002●	(1) 0.124 $\pm$ 0.002○	(4) 0.136 $\pm$ 0.003●	(3) 0.132 $\pm$ 0.002●
JAFfE	(1) 0.100 $\pm$ 0.006	(5) 0.134 $\pm$ 0.009●	(4) 0.126 $\pm$ 0.007●	(2) 0.106 $\pm$ 0.007●	(6) 0.242 $\pm$ 0.053●	(3) 0.118 $\pm$ 0.006●
Alpha	(1) 0.014 $\pm$ 0.001	(4) 0.018 $\pm$ 0.000●	(1) 0.014 $\pm$ 0.000●	(4) 0.018 $\pm$ 0.001●	(6) 0.027 $\pm$ 0.002●	(1) 0.014 $\pm$ 0.001
Cdc	(1) 0.017 $\pm$ 0.000	(5) 0.021 $\pm$ 0.000●	(1) 0.017 $\pm$ 0.000●	(4) 0.020 $\pm$ 0.001●	(6) 0.026 $\pm$ 0.001●	(1) 0.017 $\pm$ 0.001
Cold	(1) 0.052 $\pm$ 0.001	(4) 0.060 $\pm$ 0.001●	(1) 0.052 $\pm$ 0.001●	(5) 0.061 $\pm$ 0.005●	(6) 0.086 $\pm$ 0.003●	(3) 0.053 $\pm$ 0.001●
Diau	(1) 0.040 $\pm$ 0.001	(3) 0.046 $\pm$ 0.000●	(4) 0.047 $\pm$ 0.000●	(5) 0.049 $\pm$ 0.001●	(6) 0.085 $\pm$ 0.001●	(2) 0.041 $\pm$ 0.001●
Dtt	(2) 0.038 $\pm$ 0.001	(5) 0.048 $\pm$ 0.001●	(1) 0.036 $\pm$ 0.001○	(4) 0.044 $\pm$ 0.001●	(6) 0.057 $\pm$ 0.001●	(3) 0.040 $\pm$ 0.001○
Elu	(1) 0.016 $\pm$ 0.000	(5) 0.022 $\pm$ 0.000●	(3) 0.017 $\pm$ 0.000●	(4) 0.020 $\pm$ 0.001●	(6) 0.028 $\pm$ 0.001●	(1) 0.016 $\pm$ 0.001●
Heat	(1) 0.043 $\pm$ 0.001	(5) 0.049 $\pm$ 0.000●	(1) 0.043 $\pm$ 0.001●	(4) 0.045 $\pm$ 0.001●	(6) 0.053 $\pm$ 0.001●	(1) 0.043 $\pm$ 0.001●
Spo	(1) 0.059 $\pm$ 0.001	(5) 0.065 $\pm$ 0.001●	(4) 0.061 $\pm$ 0.001●	(2) 0.060 $\pm$ 0.001●	(6) 0.068 $\pm$ 0.002●	(2) 0.060 $\pm$ 0.001●
Spo5	(2) 0.093 $\pm$ 0.001	(6) 0.125 $\pm$ 0.002●	(5) 0.101 $\pm$ 0.002●	(3) 0.099 $\pm$ 0.002●	(4) 0.100 $\pm$ 0.002●	(1) 0.092 $\pm$ 0.001○
Spoem	(1) 0.087 $\pm$ 0.002	(6) 0.132 $\pm$ 0.004●	(2) 0.088 $\pm$ 0.003●	(2) 0.088 $\pm$ 0.003●	(5) 0.094 $\pm$ 0.006●	(2) 0.088 $\pm$ 0.002●

Table 2: Prediction performance ((rank) mean $\pm$ std significance) measured by Chebyshev distance on logical label enhancement.

show the KL-measured prediction performance of LE-PROM under different values of  $\alpha$  and  $\beta$  after averaging out  $\lambda$ . In Figure 5, we show the KL-measured prediction performance of LE-PROM under different values of  $\alpha$  and  $\lambda$  after averaging out  $\beta$ . In Figure 6, we show the KL-measured prediction performance of LE-PROM under different values of  $\beta$  and  $\lambda$  after averaging out  $\alpha$ . Besides, we also empirically study how the parameters  $a$ ,  $b$  and  $t$  affect the prediction performance of LE-PROM. Due to the interdependent value range of these parameters, it is not feasible to present them directly in a grid format as was done for  $\alpha$ ,  $\beta$ , and  $\lambda$ . Therefore, we let  $b = y \cdot a$ , and show the impact of  $a$ ,  $y$ , and  $t$ . We let  $a \in [10, 10^3]$ ,  $y \in (0, 1)$ ,  $t \in (0, 10)$ .

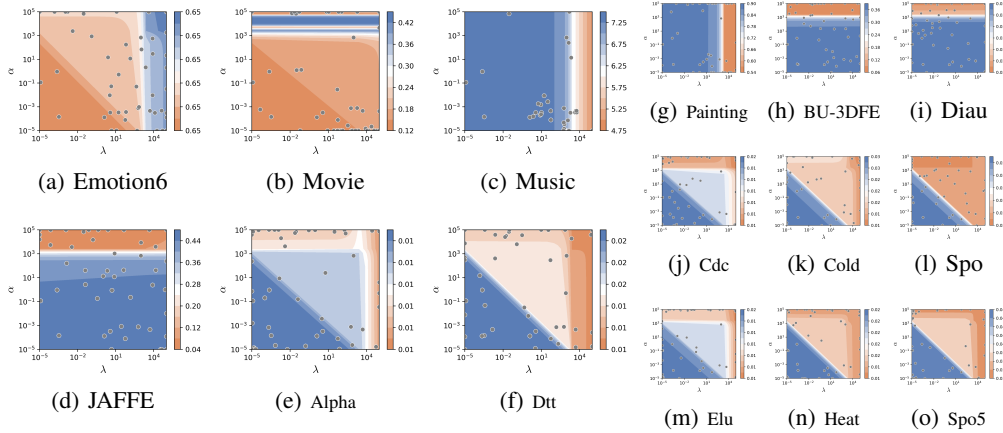


Figure 5: KL-measured prediction performance of LE-PROM under different values of  $\alpha$  and  $\lambda$  after averaging out  $\beta$ .

#### C.4 Experiments on Label Ranking Enhancement

Here we provide the experimental details for label ranking enhancement. First, we illustrate how the performance of different algorithms is evaluated on the label ranking enhancement task. Second, we provide the technical details of the comparison algorithms. Finally, we show more experimental results on label ranking enhancement in Figure 7.

**Experimental method.** Firstly, we partition the dataset with label distributions into a training set and a test set, allocating 70% of the instances to the training set and 30% to the test set. Subsequently,

Dataset	PROM	CWLD	VIB-ILE	GLLE	KMLE	FCMLE
Painting	(1) 0.589 $\pm$ 0.011	(4) 0.544 $\pm$ 0.010●	(6) 0.169 $\pm$ 0.070●	(1) 0.589 $\pm$ 0.011	(5) 0.530 $\pm$ 0.017●	(3) 0.581 $\pm$ 0.011●
Emotion6	(2) 0.561 $\pm$ 0.005	(5) 0.546 $\pm$ 0.004●	(3) 0.556 $\pm$ 0.006●	(3) 0.556 $\pm$ 0.006●	(1) 0.572 $\pm$ 0.008○	(6) 0.536 $\pm$ 0.005●
Movie	(1) 0.823 $\pm$ 0.002	(5) 0.780 $\pm$ 0.003●	(3) 0.790 $\pm$ 0.006●	(2) 0.817 $\pm$ 0.002●	(4) 0.785 $\pm$ 0.005●	(6) 0.755 $\pm$ 0.003●
Music	(1) 0.811 $\pm$ 0.007	(4) 0.788 $\pm$ 0.002●	(6) 0.737 $\pm$ 0.008●	(2) 0.797 $\pm$ 0.007●	(5) 0.778 $\pm$ 0.042●	(3) 0.794 $\pm$ 0.008●
BU-3DFE	(1) 0.852 $\pm$ 0.002	(5) 0.830 $\pm$ 0.002●	(4) 0.839 $\pm$ 0.002●	(1) 0.852 $\pm$ 0.002	(6) 0.824 $\pm$ 0.009●	(3) 0.844 $\pm$ 0.002●
JAFPE	(1) 0.870 $\pm$ 0.007	(5) 0.827 $\pm$ 0.008●	(4) 0.843 $\pm$ 0.005●	(2) 0.860 $\pm$ 0.008●	(6) 0.691 $\pm$ 0.063●	(3) 0.850 $\pm$ 0.006●
Alpha	(1) 0.962 $\pm$ 0.001	(5) 0.944 $\pm$ 0.000●	(3) 0.959 $\pm$ 0.000●	(4) 0.946 $\pm$ 0.001●	(6) 0.922 $\pm$ 0.003●	(2) 0.961 $\pm$ 0.001●
Cdc	(1) 0.958 $\pm$ 0.001	(5) 0.940 $\pm$ 0.000●	(3) 0.954 $\pm$ 0.001●	(4) 0.944 $\pm$ 0.001●	(6) 0.923 $\pm$ 0.003●	(2) 0.957 $\pm$ 0.001●
Cold	(1) 0.940 $\pm$ 0.001	(4) 0.930 $\pm$ 0.001●	(1) 0.940 $\pm$ 0.001	(4) 0.930 $\pm$ 0.002●	(6) 0.905 $\pm$ 0.003●	(3) 0.938 $\pm$ 0.001●
Diau	(1) 0.937 $\pm$ 0.001	(3) 0.924 $\pm$ 0.001●	(4) 0.917 $\pm$ 0.003●	(5) 0.915 $\pm$ 0.003●	(6) 0.841 $\pm$ 0.003●	(2) 0.934 $\pm$ 0.001●
Dtt	(3) 0.957 $\pm$ 0.001	(5) 0.945 $\pm$ 0.001●	(1) 0.958 $\pm$ 0.001○	(4) 0.949 $\pm$ 0.001●	(6) 0.934 $\pm$ 0.004●	(1) 0.958 $\pm$ 0.001○
Elu	(1) 0.958 $\pm$ 0.000	(5) 0.941 $\pm$ 0.001●	(3) 0.957 $\pm$ 0.001●	(4) 0.943 $\pm$ 0.001●	(6) 0.923 $\pm$ 0.002●	(1) 0.958 $\pm$ 0.000
Heat	(1) 0.940 $\pm$ 0.001	(5) 0.928 $\pm$ 0.001●	(1) 0.940 $\pm$ 0.001●	(4) 0.936 $\pm$ 0.001●	(6) 0.927 $\pm$ 0.002●	(1) 0.940 $\pm$ 0.001
Spo	(1) 0.915 $\pm$ 0.001	(5) 0.906 $\pm$ 0.002●	(4) 0.910 $\pm$ 0.001●	(3) 0.912 $\pm$ 0.001●	(6) 0.897 $\pm$ 0.004●	(2) 0.913 $\pm$ 0.002●
Spo5	(1) 0.908 $\pm$ 0.001	(6) 0.875 $\pm$ 0.002●	(4) 0.899 $\pm$ 0.002●	(3) 0.901 $\pm$ 0.002●	(5) 0.890 $\pm$ 0.001●	(1) 0.908 $\pm$ 0.001○
Poem	(1) 0.913 $\pm$ 0.002	(6) 0.868 $\pm$ 0.004●	(2) 0.912 $\pm$ 0.003●	(2) 0.912 $\pm$ 0.003●	(5) 0.906 $\pm$ 0.006●	(2) 0.912 $\pm$ 0.002●

Table 3: Prediction performance ((rank) mean $\pm$ std significance) measured by intersection similarity on logical label enhancement.

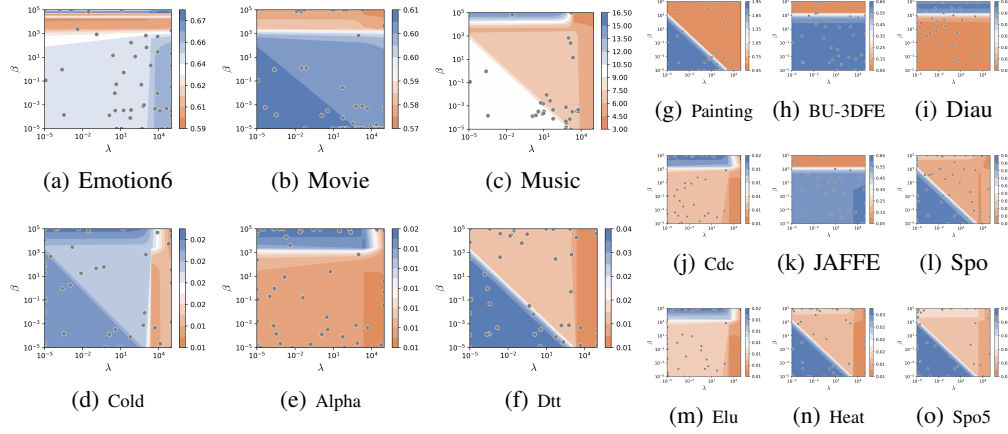


Figure 6: KL-measured prediction performance of LE-PROM under different values of  $\beta$  and  $\lambda$  after averaging out  $\alpha$ .

---

**Algorithm 1** Transforming label rankings to logical labels

---

**Input:** a dataset with label rankings  $\{\mathbf{x}_n, \mathbf{r}_n\}_{n=1}^N$ ;  
Initialize an empty set  $\mathbb{A}$  to store the expanded dataset;  
**for**  $n = 1, 2, \dots, N$  **do**  
     $\mathbf{v} \leftarrow$  Initialize an  $M$ -dimensional all-zero vector;  
     $r^* \leftarrow$  Obtain the maximum value in  $\mathbf{r}_n$ ;  
    **for**  $m = r^*, r^* - 1, \dots, 2, 1$  **do**  
        **for**  $j \in \{j \mid r_{nj} = m\}$  **do**  
             $v_j \leftarrow 1$ ;  
        **end for**  
         $\mathbb{A} \leftarrow \mathbb{A} \cup \{(\mathbf{x}_n, \mathbf{v})\}$ ;  
    **end for**  
**end for**  
**return**  $\mathbb{A}$ .

---

for each instance in the training set, we reduce its label distribution to label rankings.  $z_n^{\ell_i} < z_n^{\ell_j} \implies \ell_i \succ \ell_j$ ,  $z_n^{\ell_i} = z_n^{\ell_j} \implies \ell_i \simeq \ell_j$ ,  $z_n^{\ell_i} > z_n^{\ell_j} \implies \ell_i \prec \ell_j$ . Following this, we train the



Figure 7: Prediction performance on label ranking enhancement.

comparison algorithms (PROM, GLEMR, GLERB, and Bradley-Terry) on the training set with label rankings. Subsequently, we employ the trained models to predict the label distributions for the test instances. Finally, we compute the evaluation metrics between the predicted label distributions and the ground-truth label distributions.

**Comparison algorithms.** We expand the label rankings to multiple logical label vectors since GLERB and GLEMR can only handle the label rankings underlying logical labels. The process of transforming label rankings to logical labels is shown in Algorithm 1. It should be noted that the input label rankings of Algorithm 1 are  $\mathbf{r}$  instead of pairwise label rankings, yet  $\mathbf{r}$  can be obtained from pairwise label rankings deterministically.  $\mathbf{r} \in \{1, 2, \dots, M\}^M$  is a vector of rankings of each label, where the larger value denotes the corresponding label describes the instance to the higher extent. Since Bradley-Terry model cannot be applied to  $\ell_i \simeq \ell_j$ , we randomly convert  $\ell_i \simeq \ell_j$  into  $\ell_i \prec \ell_j$  and  $\ell_i \succ \ell_j$  during the optimization process of comparison algorithms with  $p(\ell_i \prec \ell_j | \ell_i \simeq \ell_j) = p(\ell_i \succ \ell_j | \ell_i \simeq \ell_j) = 0.5$ . Since the comparison algorithms contain modules for capturing other kinds of information, to ensure a fair comparison, we set the hyperparameters  $\alpha, \beta$  of our proposed LE-PROM as 0 and 1, we set the hyperparameters  $K, \lambda_y, \lambda_x$  in GLEMR as 1, we set the hyperparameters  $K, \lambda$  of GLERB as 1 and 50, respectively. Besides, analogous to LE-PROM, we also include  $\mathcal{D}_{\text{KL}}(\hat{p}(\mathbf{u}|\mathbf{x})||q(\mathbf{u}|\mathbf{x}, \mathbf{y}, \boldsymbol{\xi}))$  in the optimization of comparison algorithms to adaptively learn a label distribution predictor during LE process.

## References

- Xin Geng. Label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 28 (7):1734–1748, 2016.
- Harin Lee, Frank Hoeger, Marc Schoenwiesner, Minsu Park, and Nori Jacoby. Cross-cultural mood perception in pop songs and its alignment with mood detection algorithms. In *International Society for Music Information Retrieval Conference*, pages 366–373, 2021.
- Michael Lyons, Shigeru Akamatsu, Miyuki Kamachi, and Jiro Gyoba. Coding facial expressions with Gabor wavelets. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205, 1998.

- Jana Machajdik and Allan Hanbury. Affective image classification using features inspired by psychology and art theory. In *ACM International Conference on Multimedia*, page 83–92, 2010.
- Kuan Chuan Peng, Tsuhan Chen, Amir Sadvnik, and Andrew Gallagher. A mixed bag of emotions: Model, predict, and transfer emotion distributions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 860–868, 2015.
- Tingting Ren, Xiuyi Jia, Weiwei Li, Lei Chen, and Zechao Li. Label distribution learning with label-specific features. In *International Joint Conference on Artificial Intelligence*, pages 3318–3324, 2019.
- Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and M. J. Rosato. A 3D facial expression database for facial behavior research. In *International Conference on Automatic Face and Gesture Recognition*, pages 211–216, 2006.