A Appendix

A.1 Proof of Theorem 1

According to Definition 3.3, the boundary ranking error is characterized as the proportion of examples that were initially ranked accurately but became misranked because of their proximity to the ranking model's boundary. In order to mitigate this boundary ranking error, we propose the incorporation of a trainable upper bound as an alternative. This measure guarantees an improved robustness of the ranking model, effectively tackling vulnerabilities associated with the boundary ranking error.

To derive a trainable upper bound, we employ the neighborhood of $(\boldsymbol{q}, \boldsymbol{d}_i)$, denoted as $\pi_n(\boldsymbol{q}, \boldsymbol{d}_i) = \pi_y(\boldsymbol{q}, \boldsymbol{d}_i) - 1$. This serves as a bridge to determine the objective for unsupervised documents in relation to the perturbed document $\boldsymbol{d}_i$. Specifically, the boundary ranking error is calculated by,

$$\begin{aligned}
&\mathcal{R}_{\text{bdy}}(f) \\
&= \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon), \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) = \pi_y(\boldsymbol{q}, \boldsymbol{d}_i)], \\
&\leq \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon)], \\
&= \mathbb{E}_{\boldsymbol{d}_i \sim \mathcal{D}} \max_{\boldsymbol{d_i}' \in \mathbb{B}(\boldsymbol{d_i}, \epsilon)} \mathbb{I}\{[\pi_f(\boldsymbol{q}, \boldsymbol{d}_i) - \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)] \\
&\qquad \cdot [\pi_f(\boldsymbol{q}, \boldsymbol{d}_i') - \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)] \leq 0\}, \\
&= \mathbb{E}_{\boldsymbol{d}_i \sim \mathcal{D}} \max_{\boldsymbol{d_i}' \in \mathbb{B}(\boldsymbol{d_i}, \epsilon)} \mathbb{I}\{[\pi_f(\boldsymbol{q}, \boldsymbol{d}_i) - \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)] \\
&\qquad \neq [\pi_f(\boldsymbol{q}, \boldsymbol{d}_i') - \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)]\}, \\
&= \mathbb{E}_{\boldsymbol{d}_i \sim \mathcal{D}} \max_{\boldsymbol{d}_i' \in \mathbb{B}(\boldsymbol{d}_i, \epsilon)} \mathbb{I}\left\{\pi_f(\boldsymbol{q}, \boldsymbol{d}_i) \neq \pi_f(\boldsymbol{q}, \boldsymbol{d}_i')\right\}.
\end{aligned} \tag{18}$$

This completes the proof of Theorem 1. This showcases the possibility of refining the upper bound for the boundary ranking error without necessitating dependence on the ground-truth ranking list π_y. Instead, we have devised an objective aimed at promoting the invariance of the ranked list against perturbations, thereby giving rise to the loss function presented in Eq. 11.

A.2 Tightness of Theorem 1

Whether the upper bound in Theorem 1 is sufficiently tight is of great importance. In the following, we provide the necessary proof to state its tightness. First, we present the gap between the upper bound of the boundary ranking error and the error itself. Then, we prove that this gap has an upper limit, which is closely related to the natural ranking error. Finally, by demonstrating that the upper limit of the gap can be effectively optimized, we prove that the upper bound of the boundary ranking error in Theorem 1 is sufficiently tight.

The gap between the boundary ranking error and its upper bound. Since the upper bound of the boundary ranking error is always larger than itself, the gap $\mathcal{G}$ between the boundary ranking error $\mathcal{R}_{\text{bdy}}(f)$ and its upper bound $\mathcal{R}^*_{\text{bdy}}(f)$ is denoted as,

$$\begin{aligned}
&\mathcal{G}\left(\mathcal{R}^*_{\text{bdy}}(f), \mathcal{R}_{\text{bdy}}(f)\right) \\
&= \left|\mathcal{R}^*_{\text{bdy}}(f) - \mathcal{R}_{\text{bdy}}(f)\right| \\
&= \mathcal{R}^*_{\text{bdy}}(f) - \mathcal{R}_{\text{bdy}}(f) \\
&= \mathcal{R}^*_{\text{bdy}}(f) \\
&- \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon), \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) = \pi_y(\boldsymbol{q}, \boldsymbol{d}_i)],
\end{aligned} \tag{19}$$

where $\pi_f(\boldsymbol{q}, \boldsymbol{d}_i)$ and $\pi_y(\boldsymbol{q}, \boldsymbol{d}_i)$ are the predicted ranking position and the ground truth position, respectively.

Based on Eq. 18 in the above proof of Theorem 1, $\mathcal{R}^*_{\text{bdy}}(f)$ is denoted as,

$$\begin{aligned}
\mathcal{R}^*_{\text{bdy}}(f) &= \mathbb{E}_{\boldsymbol{d}_i \sim \mathcal{D}} \max_{\boldsymbol{d}_i' \in \mathbb{B}(\boldsymbol{d}_i, \epsilon)} \mathbb{I}\left\{\pi_f(\boldsymbol{q}, \boldsymbol{d}_i) \neq \pi_f(\boldsymbol{q}, \boldsymbol{d}_i')\right\} \\
&= \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon)].
\end{aligned} \tag{20}$$

To simplify Eq. 19, we introduce the model predicted rankings for document $\boldsymbol{d}_i$ to compose a joint probability. According to the definition in Eq. 4, the neighborhood of the ranking decision boundary for a ranking model f is denoted as,

$$\begin{aligned}
&\mathbb{B}(\text{DB}(f), \epsilon) := \{\boldsymbol{d}_i \sim \mathcal{D} : \exists\, \boldsymbol{d}_i' \in \mathbb{B}(\boldsymbol{d_i}, \epsilon) \text{ s.t.} \\
&[\pi_f(\boldsymbol{q}, \boldsymbol{d}_i) - \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)] \cdot [\pi_f(\boldsymbol{q}, \boldsymbol{d}_i') - \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)] \leq 0\},
\end{aligned} \tag{21}$$

where $\pi_n(\boldsymbol{q}, \boldsymbol{d}_i) = \pi_y(\boldsymbol{q}, \boldsymbol{d}_i) - 1$ is the neighborhood ranking of $(\boldsymbol{q}, \boldsymbol{d}_i)$, with the attacker's objective being to achieve a higher ranking than it.

According to the definition of the neighborhood of the ranking model decision boundary in Eq. 21, we can further expand Eq. 20 as follows,

$$\begin{aligned}
\mathcal{R}^*_{\text{bdy}}(f) &= \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon)] \\
&= \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon), \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) < \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)] \\
&+ \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon), \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) \geq \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)].
\end{aligned} \tag{22}$$

Considering $\pi_n(\boldsymbol{q}, \boldsymbol{d}_i) = \pi_y(\boldsymbol{q}, \boldsymbol{d}_i) - 1$, we have,

$$\begin{aligned}
\mathcal{R}^*_{\text{bdy}}(f) &= \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon), \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) < \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)] \\
&+ \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon), \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) = \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)] \\
&+ \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon), \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) = \pi_y(\boldsymbol{q}, \boldsymbol{d}_i)] \\
&+ \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon), \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) > \pi_y(\boldsymbol{q}, \boldsymbol{d}_i)].
\end{aligned} \tag{23}$$

According to Eq. 19 and Eq. 23, we can derive the gap $\mathcal{G}$ between the boundary ranking error and the upper bound of boundary ranking error. Each term in the two equations includes a joint probability, which consists of: (i) the probability that document $\boldsymbol{d}_i$ within the neighborhood of the ranking decision boundary $\mathbb{B}(\text{DB}(f), \epsilon)$, and (ii) the probability that the predicted ranking π_f of ranking models has a specific positional relationship with the neighborhood ranking π_n or ground truth ranking π_y.

By considering the impact of the second part alone, we can reduce the joint probability into a marginal probability. Further, this allows us to bound the gap $\mathcal{G}$ as follows:

$$\begin{aligned}
&\mathcal{G}\left(\mathcal{R}^*_{\text{bdy}}(f), \mathcal{R}_{\text{bdy}}(f)\right) \\
&= \mathcal{R}^*_{\text{bdy}}(f) - \mathcal{R}_{\text{bdy}}(f) \\
&= \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon), \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) < \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)] \\
&+ \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon), \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) = \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)] \\
&+ \mathbb{P}[\boldsymbol{d}_i \in \mathbb{B}(\text{DB}(f), \epsilon), \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) > \pi_y(\boldsymbol{q}, \boldsymbol{d}_i)] \\
&\leq \mathbb{P}[\boldsymbol{d}_i \in \mathcal{D}, \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) < \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)] \\
&+ \mathbb{P}[\boldsymbol{d}_i \in \mathcal{D}, \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) = \pi_n(\boldsymbol{q}, \boldsymbol{d}_i)] \\
&+ \mathbb{P}[\boldsymbol{d}_i \in \mathcal{D}, \pi_f(\boldsymbol{q}, \boldsymbol{d}_i) > \pi_y(\boldsymbol{q}, \boldsymbol{d}_i)] \\
&= \mathbb{E}_{\boldsymbol{d}_i \sim \mathcal{D}}\, \mathbb{I}\left\{\pi_f(\boldsymbol{q}, \boldsymbol{d}_i) \;\neq\; \pi_y(\boldsymbol{q}, \boldsymbol{d}_i)\right\}.
\end{aligned} \tag{24}$$

As a result, we deduce that the upper limit of the gap is the sum of three independent probabilities: for document $\boldsymbol{d}_i$, (i) the ranking position predicted by the ranking model is higher than the neighborhood ranking; (ii) the ranking position predicted by the ranking model is equal to the neighborhood ranking; (iii) the ranking position predicted by the ranking model is lower than the ground-truth ranking.

Therefore, the upper limit of the gap $\mathcal{G}$ is equivalent to the expectation that the ranking position predicted by the ranking model is not equal to the ground-truth ranking position. For this gap, its upper limit (note as $\mathcal{G}^*$) can be guaranteed due to the expectation being closely related to the natural ranking loss of a ranking model given by Eq. 9.

In summary, we have demonstrated that the gap $\mathcal{G}$ between the boundary ranking error $\mathcal{R}_{\mathrm{bdy}}(f)$ and its upper bound $\mathcal{R}^*_{\mathrm{bdy}}(f)$ is bounded, reflecting the tightness of the upper bound we provided. Moreover, $\mathcal{G}^*$, the upper limit of gap $\mathcal{G}$, is consistent with the natural ranking error, suggesting that this gap can potentially be optimized by the natural ranking loss. In the following, we will present a detailed analysis to prove that the gap $\mathcal{G}$ is indeed optimizable under the natural ranking loss.

The natural optimizability of the gap. In this work, the natural ranking error is optimized through the natural ranking loss embedded within the proposed regularized surrogate loss as defined in Eq. 8. Our goal is to show that the upper limit of the ga $\mathcal{G}$p, denoted as $\mathcal{G}^*$, can be optimized through the natural ranking loss. This is achieved by demonstrating that the natural ranking error can be effectively optimized by the natural ranking loss.

Indeed, Calauzenes, Usunier, and Gallinari (2012) have shown the effectiveness of natural surrogate losses, like the hinge loss, in optimizing classification errors. This insight was further extended by Bartlett, Jordan, and McAuliffe (2006), who showcased that the pairwise loss function, acting as a natural ranking loss, optimizes the natural ranking error towards its infimum in conjunction with itself. In this paper, the pairwise loss we use for the natural ranking loss is also included. Hence, in the scenario of this paper, we have

$$\begin{aligned} \mathcal{L}_{\mathrm{nat}}\left(\mathbb{D}, f\right) \rightarrow \inf_f \mathcal{L}_{\mathrm{nat}}(\mathbb{D}, f) \\ \Rightarrow \mathcal{R}_{\mathrm{nat}}\left(\mathbb{D}, f\right) \rightarrow \inf_f \mathcal{R}_{\mathrm{nat}}(\mathbb{D}, f), \end{aligned} \tag{25}$$

where the infima are taken over all data from the ranking dataset with the distribution $\mathbb{D}$.

According to Eq. 25, the natural ranking error can be effectively optimized by our natural ranking loss. Since $\mathcal{G}^*$, the upper limit of gap $\mathcal{G}$, is consistent with the natural ranking error. It can also be naturally optimizable by the natural ranking loss.

The tightness of the upper bound of the boundary ranking error. Assume that the natural ranking error of the ranking model f on the set of document candidates $\mathcal{D}$ is represented as,

$$\mathbb{E}_{\boldsymbol{d}_i \sim \mathcal{D}}\, \mathbb{I}\left\{\pi_f(\boldsymbol{q}, \boldsymbol{d}_i) \;\neq\; \pi_y(\boldsymbol{q}, \boldsymbol{d}_i)\right\} = \eta, \tag{26}$$

where $\eta \geq 0$ is the expectation of the natural ranking error associated with the ranking model f. In this way, the upper limit of gap $\mathcal{G}$ can be determined as,

$$\begin{aligned} &\mathcal{G}\left(\mathcal{R}^*_{\mathrm{bdy}}(f), \mathcal{R}_{\mathrm{bdy}}(f)\right) \\ &\quad \leq \mathcal{G}^*\left(\mathcal{R}^*_{\mathrm{bdy}}(f), \mathcal{R}_{\mathrm{bdy}}(f)\right) \\ &\quad = \eta. \end{aligned} \tag{27}$$

Finally, according to Eq. 26 and Eq. 25, the upper limit of the gap $\mathcal{G}$ between the boundary ranking error $\mathcal{R}_{\mathrm{bdy}}(f)$ and its upper bound $\mathcal{R}^*_{\mathrm{bdy}}(f)$ is a deterministic value. And it is confident that the deterministic value will be continuously reduced during training with the natural ranking loss.

Hence, given the presence of the deterministic value, optimizing the upper bound of the boundary ranking error $\mathcal{R}^*_{\mathrm{bdy}}(f)$ leads to effective optimization of the boundary ranking error $\mathcal{R}_{\mathrm{bdy}}(f)$ as well. Consequently, the upper bound of the boundary ranking error in Theorem 1 proves to be sufficiently tight.

References

Bartlett, P. L.; Jordan, M. I.; and McAuliffe, J. D. 2006. Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 101(473): 138–156.

Calauzenes, C.; Usunier, N.; and Gallinari, P. 2012. On the (non-) existence of convex, calibrated surrogate losses for ranking. *Advances in Neural Information Processing Systems*, 25.