Fusion Label Enhancement for Multi-Label Learning

Appendix

A.1 Proof for Theorem 1

Lemma 1. *(Log sum inequality) Let $a_1, a_2, ..., a_n$ and $b_1, b_2, ..., b_n$ be non-negative numbers, Denote $a = \sum_{i=1}^{n} a_i$ and $b = \sum_{i=1}^{n} b_i$. There is*

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq a \log \frac{a}{b} \tag{A.1.1}$$

with equality if and only if a_i/b_i are equal for all i.

Proof. Notice that after setting $f(x) = xlogx$, we have

$$\begin{aligned}
&\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \\
&= \sum_{i=1}^{n} b_i f\left(\frac{a_i}{b_i}\right) = b \sum_{i=1}^{n} \frac{b_i}{b} f\left(\frac{a_i}{b_i}\right) \\
&\geq bf\left(\sum_{i=1}^{n} \frac{b_i}{b} \frac{a_i}{b_i}\right) = bf\left(\frac{1}{b}\sum_{i=1}^{n} a_i\right) \\
&= bf\left(\frac{a}{b}\right) = a \log \frac{a}{b}
\end{aligned} \tag{A.1.2}$$

where the inequality follows from Jensen's inequality [Jensen, 1906] since $b_i/b \geq 0$, $\sum_{i=1}^{n} (b_i/b) = 1$ and f is convex. □

Theorem 1. *$\mathcal{L}_{LD}$ gives an upper bound for cross-entropy loss.*

Proof. Cross-entropy loss is a wide-used loss function for classification, which can be defined as

$$\mathcal{L}_{CE} = \log\left(1 + \sum_{n\in\Omega_{neg}, p\in\Omega_{pos}} e^{s_n - s_p}\right). \tag{A.1.3}$$

where Ω_{pos} and Ω_{neg} are the sets of relevant and irrelevant labels respectively [Zhang *et al.*, 2021; Su, 2020]. Using Ω_1=$\Omega_{neg} \cup \{s_0\}$ and Ω_2=$\Omega_{pos} \cup \{s_0\}$ to replace the variables in Eq.(A.1.3), and the objective of Eq.(A.1.3) is consistent with:

$$\mathcal{L}_{CE} = \log \sum_{n\in\Omega_1, p\in\Omega_2} e^{s_n - s_p}. \tag{A.1.4}$$

According to Carlson inequality [Carlson, 1934], we have

$$\log \sum_{n\in\Omega_1, p\in\Omega_2} e^{s_n - s_p} \leq \log \sum_{n\in\Omega_1} e^{s_n - s_0} \sum_{p\in\Omega_2} e^{s_0 - s_p}, \tag{A.1.5}$$

in which we introduce an additional category 0, hoping that the scores of the target category are all greater than s_0 and the scores of the non-target categories are all less than s_0. Minimizing $\log \sum_{n\in\Omega_1} e^{s_n - s_0} \sum_{p\in\Omega_2} e^{s_0 - s_p}$ is equivalent with minimizing

$$\mathcal{L}_B = \log \sum_{j\in\Omega} \left(e^{s_j - s_0}\right)^{\Pi(s_j \le s_0)}, \tag{A.1.6}$$

where $\Omega = \Omega_1 \cup \Omega_2$ and

$$\Pi\left(s_j \le s_0\right) = \begin{cases} 1, & s_j \le s_0 \\ -1, & \text{otherwise} \end{cases}, \tag{A.1.7}$$

Setting $a_i = d^{(t)}$, $\sum_{i=1}^{n} a_i = \sum_t d^{(t)}$ and $b_i = 1/\sum_{j\in\Omega}\left(e^{s_j - s_0}\right)^{\Pi(s_j \le s_0)}$, according to Lemma 1, we have:

$$\begin{aligned} &\sum_t d^{(t)} \log d^{(t)} \sum_{j\in\Omega} \left(e^{s_j - s_0}\right)^{\Pi(s_j \le s_0)} \\ &\ge \log \sum_{j\in\Omega} \left(e^{s_j - s_0}\right)^{\Pi(s_j \le s_0)} \\ &\ge \log \sum_{n\in\Omega_1, p\in\Omega_2} e^{s_n - s_p} \end{aligned} \tag{A.1.8}$$

since $\sum_t d^{(t)} = 1$.

Replacing $\sum_j \left(e^{s_j - s_0}\right)^{\Pi(s_j \le s_0)}$ by $\sum_j \left(e^{s_j - s_t}\right)$, we can find that $\mathcal{L}_{LD}$ gives an upper bound for $\mathcal{L}_{CE}$.

$\square$

References

[Carlson, 1934] F. Carlson. Une inégalité. *Ark. Math. Astron. Fys.*, 25B, no.1:1–5, 1934.

[Jensen, 1906] Johan Ludwig William Valdemar Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta mathematica*, 30(1):175–193, 1906.

[Su, 2020] Jianlin Su. Extend "softmax+cross entropy" to the multi-label classification problem. https://kexue.fm/archives/7359, 2020.

[Zhang *et al.*, 2021] Ningyu Zhang, Xiang Chen, Xin Xie, Shumin Deng, Chuanqi Tan, Mosha Chen, Fei Huang, Luo Si, and Huajun Chen. Document-level relation extraction as semantic segmentation. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pages 3999–4006, 2021.