Appendix of Named Entity Recognition using Positive-Unlabeled Learning

A Proof of Theorem 1

Let denote L_M the Lipschitz constant that $L_M > \frac{\partial \ell(w,y)}{\partial w}, \forall w \in \mathbb{R}$, denote $C_0 = \max_y \ell(0, y)$, and denote $\mathcal{H}$ a Reproducing Kernel Hilbert Space (RKHS). For each given $R > 0$, we consider as hypothesis space $\mathcal{H}_R$, the ball of radius R in $\mathcal{H}$. Let $N(\epsilon)$ be the covering number of $\mathcal{H}_R$ following Theorem C in (Cucker and Smale, 2002). Then we have the following theorem.

Theorem 1. If ℓ is bounded by $[0, M]$, then for any $\epsilon > 0$,

$$\begin{aligned} &\mathrm{P}\{\mathrm{S} \in \mathcal{D} | \sup_{\mathrm{f} \in \mathcal{H}_\mathrm{R}} |\mathrm{R}_\ell - \hat{\mathrm{R}}_\ell| \le \epsilon\} \\ &\ge 1 - 2N(\frac{\epsilon}{4(1+2\pi_p)L_M})e^{-\frac{\min(n_p,n_u)\epsilon^2}{8(1+2\pi_p)^2B^2}}, \end{aligned} \tag{1}$$

where $B = L_M M + C_0$.

Proof 1 *Let denote $\hat{R}^s_\ell(f)$ the empirical estimation of $R_\ell(f)$ with k randomly labeled examples. Since ℓ is bounded, C_0, M, and B are finite. According to the Lemma in (Rosasco et al., 2004) we have:*

$$\begin{aligned} &\mathrm{P}\{\mathrm{S} \in \mathcal{D} | \sup_{\mathrm{f} \in \mathcal{H}_\mathrm{R}} |\mathrm{R}_\ell(\mathrm{f}) - \hat{\mathrm{R}}^\mathrm{s}_\ell(\mathrm{f})| \le \epsilon\} \\ &\ge 1 - 2N(\frac{\epsilon}{4L_M})e^{-\frac{k\epsilon^2}{8B^2}}. \end{aligned} \tag{2}$$

Then, the empirical estimation error of $R_\ell(f) - \hat{R}_\ell(f)$ in PU learning can be written as:

$$\begin{aligned} &R_\ell(f) - \hat{R}_\ell(f) \\ &= \left(\mathbb{E}_\mathbf{X}\ell(f(\boldsymbol{x}), 0) - \frac{1}{n_u}\sum_{i=1}^{n_u}\ell((f(x^u_i), 0)\right) \\ &+ \pi_p\left(\mathbb{E}_{\mathbf{X}|\mathrm{Y}=1}\ell(f(\boldsymbol{x}), 1) - \frac{1}{n_p}\sum_{i=1}^{n_p}\ell(f(x^p_i), 1)\right) \\ &- \pi_p\left(\mathbb{E}_{\mathbf{X}|\mathrm{Y}=1}\ell(f(\boldsymbol{x}), 0) - \frac{1}{n_p}\sum_{i=1}^{n_p}\ell(f(x^p_i), 0)\right) \end{aligned} \tag{3}$$

Thus,

$$\begin{aligned} &|R_\ell(f) - \hat{R}_\ell(f)| \\ &\le \left|\mathbb{E}_\mathbf{X}\ell(f(\boldsymbol{x}), 0) - \frac{1}{n_u}\sum_{i=1}^{n_u}\ell((f(x^u_i), 0)\right| \\ &+ \pi_p\left|\mathbb{E}_{\mathbf{X}|\mathrm{Y}=1}\ell(f(\boldsymbol{x}), 1) - \frac{1}{n_p}\sum_{i=1}^{n_p}\ell(f(x^p_i), 1)\right| \\ &+ \pi_p\left|\mathbb{E}_{\mathbf{X}|\mathrm{Y}=1}\ell(f(\boldsymbol{x}), 0) - \frac{1}{n_p}\sum_{i=1}^{n_p}\ell(f(x^p_i), 0)\right| \end{aligned} \tag{4}$$

Let $\mathrm{I}_\ell(\mathbf{X}, 0)$ denote

$$\mathbb{E}_\mathbf{X}\ell(f(\boldsymbol{x}), 0) - \frac{1}{n_u}\sum_{i=1}^{n_u}\ell((f(x^u_i), 0).$$

According to Eq. 2, we have:

$$\begin{aligned} &\mathrm{P}\{\mathrm{S} \in \mathcal{D} | \sup_{\mathrm{f} \in \mathcal{H}_\mathrm{R}} |\mathrm{I}_\ell(\mathbf{X}, 0)| \le \epsilon\} \\ &\ge 1 - 2N(\frac{\epsilon}{4L_M})e^{-\frac{n_u\epsilon^2}{8B^2}} \end{aligned} \tag{5}$$

Similarly, let $\mathrm{I}_\ell(\mathbf{X}|\mathrm{Y} = 1, 1)$ denote

$$\mathbb{E}_{\mathbf{X}|\mathrm{Y}=1}\ell(f(\boldsymbol{x}), 1) - \frac{1}{n_p}\sum_{i=1}^{n_p}\ell(f(x^p_i), 1),$$

and $\mathrm{I}_\ell(\mathbf{X}|\mathrm{Y} = 1, 0)$ denote

$$\mathbb{E}_{\mathbf{X}|\mathrm{Y}=1}\ell(f(\boldsymbol{x}), 0) - \frac{1}{n_p}\sum_{i=1}^{n_p}\ell(f(x^p_i), 0),$$

we have:

$$\begin{aligned} &\mathrm{P}\{\mathrm{S} \in \mathcal{D} | \sup_{\mathrm{f} \in \mathcal{H}_\mathrm{R}} |\mathrm{I}_\ell(\mathbf{X}|\mathrm{Y} = 1, 1)| \le \epsilon\} \\ &\ge 1 - 2N(\frac{\epsilon}{4L_M})e^{-\frac{n_p\epsilon^2}{8B^2}}, \end{aligned} \tag{6}$$

and

$$\begin{aligned}&\mathrm{P}\{\mathrm{S} \in \mathcal{D} | \sup_{\mathrm{f} \in \mathcal{H}_{\mathrm{R}}} |\mathrm{I}_\ell(\mathbf{X}|\mathrm{Y}=1,0)| \leq \epsilon\} \\ &\geq 1 - 2N(\frac{\epsilon}{4L_M})e^{-\frac{n_p\epsilon^2}{8B^2}},\end{aligned} \tag{7}$$

Therefore,

$$\begin{aligned}&\mathrm{P}\{\mathrm{S} \in \mathcal{D} | \sup_{\mathrm{f} \in \mathcal{H}_{\mathrm{R}}} |\mathrm{R}_\ell(\mathrm{f}) - \hat{\mathrm{R}}_\ell(\mathrm{f})| \leq (1+2\pi_{\mathrm{p}})\epsilon\} \\ &\geq \min\Bigg(1 - 2N(\frac{\epsilon}{4L_M})e^{-\frac{n_p\epsilon^2}{8B^2}}, \\ &\qquad 1 - 2N(\frac{\epsilon}{4L_M})e^{-\frac{n_u\epsilon^2}{8B^2}} \\ &= 1 - 2N(\frac{\epsilon}{4L_M})e^{-\frac{\min(n_p,n_u)\epsilon^2}{8B^2}}\end{aligned} \tag{8}$$

The theorem follows replacing ϵ with $\frac{1}{1+2\pi_p}\epsilon$.

References

Felipe Cucker and Steve Smale. 2002. On the mathematical foundations of learning. *Bulletin of the American mathematical society*, 39(1):1–49.

Lorenzo Rosasco, Ernesto De Vito, Andrea Caponnetto, Michele Piana, and Alessandro Verri. 2004. Are loss functions all the same? *Neural Computation*, 16(5):1063–1076.