Supplementary Materials of Consistency-based Semi-supervised Active Learning: Towards Minimizing Labeling Cost

Mingfei Gao[1*], Zizhao Zhang[2], Guo Yu[3], Sercan Ö. Arık[2], Larry S. Davis[1], and Tomas Pfister[2]

[1]University of Maryland [2]Google Cloud AI [3]University of Washington

1 Proof of Proposition

Proof. Denote $\mathcal{X}$ as the feature space and $\{1, \ldots, J\}$ as the label space. Note that by Baye's formula and the law of total probability, we have

$$
\begin{aligned}
&R_H[p(\hat{Y}|X)] = \mathrm{E}_X \left\{ H\left[p(Y|X), p(\hat{Y}|X) \right] \right\} \\
=& -\sum_{x\in\mathcal{X}} \sum_{y=1}^{J} p(Y=y|X=x) \log p(\hat{Y}=y|X=x) p(X=x) \\
=& -\sum_{y=1}^{J} \sum_{x\in\mathcal{X}} p(X=x, Y=y) \log \left[\frac{p(\hat{Y}=y) p(X=x|\hat{Y}=y)}{p(X=x)} \right] \\
=& -\sum_{y=1}^{J} \sum_{x\in\mathcal{X}} p(X=x, Y=y) \log p(\hat{Y}=y) \\
& -\sum_{y=1}^{J} \sum_{x\in\mathcal{X}} p(X=x, Y=y) \log \left[\frac{p(X=x|\hat{Y}=y)}{p(X=x)} \right] \\
=& -\sum_{y=1}^{J} p(Y=y) \log p(\hat{Y}=y) - \sum_{x\in\mathcal{X}} \sum_{y=1}^{J} p(X=x, Y=y) \log \left[p(X=x|\hat{Y}=y) \right] \\
& + \sum_{x\in\mathcal{X}} \sum_{y=1}^{J} p(X=x, Y=y) \log \left[p(X=x) \right] \\
=& H\left[p(Y), p(\hat{Y}) \right] + \sum_{x\in\mathcal{X}} p(X=x) \log \left[p(X=x) \right] \\
& - \sum_{x\in\mathcal{X}} \sum_{y=1}^{J} p(X=x, Y=y) \log \left[p(X=x|\hat{Y}=y) \right]
\end{aligned}
$$

* Work done while the author was an intern at Google; now at Salesforce Research. Email: `mgao@cs.umd.edu`

$$=H\left[p(Y), p(\hat{Y})\right] - H\left[p(X)\right] - \sum_{x \in \mathcal{X}} \sum_{y=1}^{J} p(X = x, Y = y) \log \left[p(X = x | \hat{Y} = y)\right]. \tag{1}$$

We first give a lower bound. Note that $p(X = x|\hat{Y} = y) \leq 1$ for any $(x, y) \in \mathcal{X} \times [J]$, so Eq. 1 implies that

$$\mathrm{E}_X \left\{H\left[p(Y|X), p(\hat{Y}|X)\right]\right\} \geq H\left[p(Y), p(\hat{Y})\right] - H\left[p(X)\right].$$

To prove the upper bound, denote $\min_{(x,y)\in\mathcal{X}\times[J]} p(X = x|\hat{Y} = y) = \hat{Z} \in (0, 1)$ where $(x, y) \in \mathcal{X} \times [J]$. Then from Eq. 1

$$\begin{aligned}\mathrm{E}_X \left\{H\left[p(Y|X), p(\hat{Y}|X)\right]\right\} \leq & H\left[p(Y), p(\hat{Y})\right] \\ & - H\left[p(X)\right] - \log \hat{Z} \sum_{x\in\mathcal{X}} \sum_{y=1}^{J} p(X = x, Y = y) \\ = & H\left[p(Y), p(\hat{Y})\right] - H\left[p(X)\right] - \log \hat{Z}.\end{aligned}$$

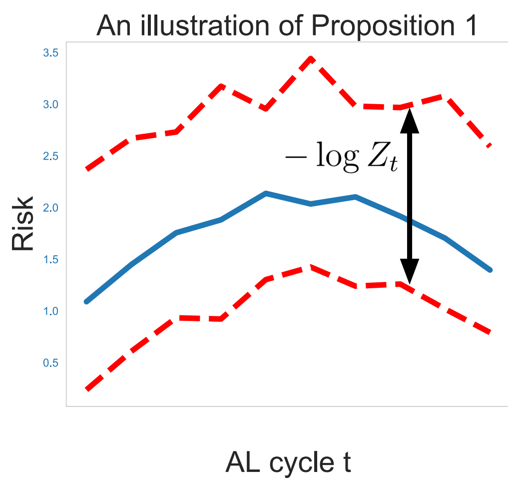

Fig. A1. An illustration of Proposition: the blue curve represents the (expected) cross-entropy, and the two red curves are the lower and upper bounds. The value $-\log \hat{Z}_t$ characterizes the range of the bounds.

2 Consistency-based Selection with Other SSL methods

To investigate the effectiveness of our method, we combine our selection method with two more SSL methods, *i.e.*, Pi-Model [1] and VAT [2]. We consider Pi-

Table A1. Comparison between our method and k-center trained with different SSL methods on CIFAR-10. The reported results are averaged over 3 trials

Methods	Selection	# of labeled samples in total				
		1000	1500	2000	2500	3000
Pi-Model[1]	k-center	67.82±1.34	71.72±0.39	74.56±0.36	75.98±0.53	77.7±0.32
	Ours		**72.06±0.30**	**74.96±0.20**	**77.05±0.50**	**78.64±0.48**
VAT[2]	k-center	80.52±0.32	82.71±0.46	84.51±0.25	86.03±0.08	86.61±0.19
	Ours		**85.22±0.20**	**87.05±0.25**	**88.32±0.19**	**89.1±0.13**

Model and VAT, since they use consistency-related regularization which matches our assumption. The experiments are conducted against our strongest baseline, k-center, on CIFAR-10. All models start from 1000 labels to avoid cold start problems. We follow the experimental setting of CIFAR-10 in our main paper. In each AL cycle, the model is initialized with the model trained in the previous cycle. We use the implementation of these SSL methods provided by MixMatch [1]. As shown in Table A1, our consistency-based selection works consistently better than k-center with these SSL methods.

[1] https://github.com/google-research/mixmatch

References

1. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. ICLR (2017)
2. Miyato, T., Maeda, S.i., Koyama, M., Ishii, S.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. TPAMI (2018)