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

Mingfei Gao\textsuperscript{1}, Zizhao Zhang\textsuperscript{2}, Guo Yu\textsuperscript{3}, Sercan ¨O. Ark\textsuperscript{2}, Larry S. Davis\textsuperscript{1}, and Tomas Pfister\textsuperscript{2}

\textsuperscript{1}University of Maryland \textsuperscript{2}Google Cloud AI \textsuperscript{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

$$R_H[p(\hat{Y}|X)] = E_X \left\{ H \left[ p(Y|X), p(\hat{Y}|X) \right] \right\}$$

$$= - \sum_{x \in \mathcal{X}} \sum_{y=1}^{J} p(Y = y|X = x) \log p(\hat{Y} = y|X = x) p(X = x)$$

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

$$= - \sum_{y=1}^{J} \sum_{x \in \mathcal{X}} p(X = x, Y = y) \log p(\hat{Y} = y)$$

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

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

$$+ \sum_{x \in \mathcal{X}} \sum_{y=1}^{J} p(X = x, Y = y) \log \left[ p(X = x) \right]$$

$$= H \left[ p(Y), p(\hat{Y}) \right] + \sum_{x \in \mathcal{X}} p(X = x) \log \left[ p(X = x) \right]$$

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

\textsuperscript{*} Work done while the author was an intern at Google; now at Salesforce Research.

Email: mgao@cs.umd.edu
\[ H[p(Y), p(\hat{Y})] - H[p(X)] - \sum_{x \in X} \sum_{y=1}^J p(X = x, Y = y) \log[p(X = x|\hat{Y} = y)]. \]

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

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

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

\[ \mathbb{E}_X \left\{ H[p(Y|X), p(\hat{Y}|X)] \right\} \leq H[p(Y), p(\hat{Y})] - H[p(X)] - \log \hat{Z} \sum_{x \in \mathcal{X}} \sum_{y=1}^J p(X = x, Y = y) \]

\[ = H[p(Y), p(\hat{Y})] - H[p(X)] - \log \hat{Z}. \]

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 \( -\log \hat{Z} \) 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.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Selection</th>
<th># of labeled samples in total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1000</td>
</tr>
<tr>
<td>Pi-Model[1]</td>
<td>k-center</td>
<td>67.82±1.34</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td>72.06±0.30</td>
</tr>
<tr>
<td>VAT[2]</td>
<td>k-center</td>
<td>80.52±0.32</td>
</tr>
<tr>
<td></td>
<td>Ours</td>
<td>85.22±0.20</td>
</tr>
</tbody>
</table>

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 MixMatch1. 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