Supplementary Materials of Frustratingly Easy Model Ensemble for Abstractive Summarization

Hayato Kobayashi
Yahoo Japan Corporation / RIKEN AIP
hakobaya@yahoo-corp.jp

A Proof of Theorem 1

Proof. First, we will prove the following equation.

$$\tilde{p}(y) = \max_{s \in S} \tilde{p}(s), \tag{1}$$

where S and y are the output candidates and the selected output, respectively, in Algorithm 1 with $K(s, s') = \cos(s, s')$, and $\tilde{p}$ is the first order Taylor series approximation of the kernel density estimator p based on the von Mises-Fisher kernel.

From the definition of the von Mises-Fisher kernel, we have

$$\begin{aligned} p(s) &= \frac{1}{|S|} \sum_{s' \in S} K_{\text{vmf}}(s, s') && (2) \\ &= \frac{1}{|S|} \sum_{s' \in S} C_q(\kappa) \exp(\kappa \cos(s, s')) && (3) \\ &\propto \sum_{s' \in S} \exp(\kappa \cos(s, s')), && (4) \end{aligned}$$

where $C_q(\kappa)$ and κ are the normalization constant and concentration parameter of the von Mises-Fisher kernel. Using the first order Taylor series approximation at 0 of $\exp(x)$, i.e., $\exp(x) \approx 1+x$, we have $\tilde{p}(s) \propto \sum_{s' \in S}(1 + \kappa \cos(s, s'))$. Therefore, the definition of y yields

$$\begin{aligned} y &= \frac{1}{|S|} \operatorname*{argmax}_{s \in S} \sum_{s' \in S} \cos(s, s') && (5) \\ &= \operatorname*{argmax}_{s \in S} \sum_{s' \in S} 1 + \kappa \cos(s, s') && (6) \\ &= \operatorname*{argmax}_{s \in S} \tilde{p}(s). && (7) \end{aligned}$$

This proves Eq. (1).

Next, we consider the following equation.

$$p(y^*) - p(y) \le C_q(\kappa) \kappa^2 \exp(\kappa)(\sigma^2 + \mu^2), \tag{8}$$

where y^* is the ideal output that maximizes the von Mises-Fisher kernel, i.e., $y^* = \operatorname{argmax}_{s \in S} p(s)$, and μ and σ^2 are the maximum values of the mean and variance of the cosine similarities $\cos(s, s')$ with respect to an output candidate s, defined as

$$\mu = \max_{s \in S} \mathbb{E}_{s'}[\cos(s, s')] \tag{9}$$

$$\sigma^2 = \max_{s \in S} \mathbb{V}_{s'}[\cos(s, s')]. \tag{10}$$

The Lagrange error bound $R_n(x)$ of the n-th Taylor series approximation of $f(x)$ is defined as

$$R_n(x) = \frac{\max_{x'} f^{(n+1)}(x')}{(n+1)!} x^{n+1}. \tag{11}$$

In our case, the error bound $\tilde{R}(x)$ is calculated for the first order approximation of $\exp(x)$, where $x = \kappa \cos(s, s')$, and $-\kappa \le x \le \kappa$, and thus, we obtain the upper bound as

$$\begin{aligned} \tilde{R}(x) &= \frac{\max_{x'} \exp(x')}{2!} x^2 && (12) \\ &\le \frac{\exp(\kappa)}{2} x^2. && (13) \end{aligned}$$

Here, we define the approximation error between $p(s)$ and $\tilde{p}(s)$ with respect to an output s as $R'(s)$. This error can be bounded as follows.

$$\begin{aligned} R'(s) &= |p(s) - \tilde{p}(s)| && (14) \\ &\le \frac{1}{|S|} \sum_{s' \in S} C_q(\kappa) \tilde{R}(\kappa \cos(s, s')) && (15) \\ &\le \frac{1}{|S|} \sum_{s' \in S} C_q(\kappa) \frac{\exp(\kappa)}{2} (\kappa \cos(s, s'))^2 && (16) \\ &= C_q(\kappa) \frac{\kappa^2 \exp(\kappa)}{2} \frac{1}{|S|} \sum_{s' \in S} \cos^2(s, s') && (17) \\ &= C_q(\kappa) \frac{\kappa^2 \exp(\kappa)}{2} (\sigma_s^2 + \mu_s^2), && (18) \end{aligned}$$

where $\mu_s = \frac{1}{|S|} \sum_{s' \in S} \cos(s, s')$, and $\sigma_s^2 = \frac{1}{|S|} \sum_{s' \in S} \cos^2(s, s') - \mu_s^2$.

From the approximation error of $\tilde{p}(y^*)$, we obtain the following.

$$p(y^*) - R'(y^*) \le \tilde{p}(y^*) \tag{19}$$

Similarly, from the approximation error of $\tilde{p}(y)$, we obtain the following.

$$\tilde{p}(y) \leq p(y) + R'(y) \tag{20}$$

Using the optimality of y with respect to $\tilde{p}$, i.e., $\tilde{p}(y^*) \leq \tilde{p}(y)$, we can connect the above two inequalities as

$$p(y^*) - p(y) \leq R'(y^*) + R'(y) \tag{21}$$

$$\leq 2 \max_{s \in S} R'(s) \tag{22}$$

$$= C_q(\kappa)\kappa^2 \exp(\kappa) \max_{s \in S}(\sigma_s^2 + \mu_s^2). \tag{23}$$

This concludes the theorem. □