9 Appendix

A Derivation of $\mathcal{B}ottle\mathcal{S}um^{Ex}$ Loss

This is a derivation of the loss equation (3) used in section 2, starting with the Information Bottleneck (IB) loss given in eq 2:

$$I(\tilde{S};S) - \beta I(\tilde{S};Y), \tag{4}$$

The goal here is to consider how to interpret this equation when we only have access to single values of source S and relevance variabel Y at a time. In the original IB formulation, a distribution $p(\tilde{S}|S)$ to go from sources to summaries is learned by optimizing this expression across the distribution of source-target pairs (s, y). In the case of $\mathcal{B}ottle\mathcal{S}um^{Ex}$, the goal is to consider this expression on a case-by-case basis, not requiring training over a large distribution of pairs.

First, we consider an alternate form of the equation above:

$$\begin{aligned} &I(\tilde{S};S) - \beta I(\tilde{S};Y) \\ &= \mathop{\mathbb{E}}_{S,\tilde{S}}[pmi(\tilde{S};S)] - \beta \mathop{\mathbb{E}}_{Y,\tilde{S}}[pmi(\tilde{S};Y)] \end{aligned} \tag{5}$$

where $pmi(x, y) = \frac{p(x,y)}{p(x)p(y)}$ denotes pointwise mutual information.

$$= \mathop{\mathbb{E}}_{S,\tilde{S}}\Big[pmi(\tilde{S};S)\Big] - \beta \mathop{\mathbb{E}}_{Y,\tilde{S}}\Big[pmi(\tilde{S};Y)\Big] \tag{6}$$

As stated above, we want to consider this for only single values of s and y at a time, so for these values we can investigate the applicable terms of these expectations:

$$= \sum_{\tilde{s}} \Big[p(s,\tilde{s})pmi(\tilde{S};S) - \beta p(y,\tilde{s})pmi(\tilde{S};Y)\Big] \tag{7}$$

This is the expression we are then aiming to optimize, as it covers all terms in the original IB objective that we have access to on a case-by-case basis.

As in the original IB problem, we can think of learning a distribution $p(\tilde{s}|s)$. However, we are now only taking an expectation over $\tilde{S}$ and so we simply collapse all probability onto the setting of $\tilde{s}$ that optimizes this expression. Simply:

$$p(\tilde{s}|s) = 1 \text{ for chosen summary, 0 otherwise} \tag{8}$$

This results in finding $\tilde{s}$ that optimizes:

$$\begin{aligned} &p(s,\tilde{s})pmi(\tilde{S};S) - \beta p(y,\tilde{s})pmi(\tilde{S};Y) \\ &= p(s,\tilde{s})\log\frac{p(s,\tilde{s})}{p(s)p(\tilde{s})} - \beta p(y,\tilde{s})\log\frac{p(y,\tilde{s})}{p(y)p(\tilde{s})} \end{aligned} \tag{9}$$

Any terms that rely only on s and y will be constant and so can be collected into coefficients. As well, remembver that we set $p(\tilde{s}|s) = 1$. Doing rearranging:

$$\begin{aligned} &= p(\tilde{s}|s)p(s)\log\frac{p(\tilde{s}|s)}{p(\tilde{s})} - \beta p(y|\tilde{s})p(\tilde{s})\log\frac{p(y|\tilde{s})}{p(y)} \\ &= c_1\log\frac{1}{p(\tilde{s})} - \beta p(y|\tilde{s})p(\tilde{s})\log p(y|\tilde{s}) - c_2 \end{aligned} \tag{10}$$

This is equivalent to optimizing:

$$\log\frac{1}{p(\tilde{s})} - \beta_1 p(y|\tilde{s})p(\tilde{s})\log p(y|\tilde{s}) \tag{11}$$

for some positively signed β_1.