Dynamic Multimodal Information Bottleneck for Multimodality Classification: Supplementary Materials

1. Preliminary Definitions

Given continuous random variables X, Y, Z, supported on $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ with probability distributions p_X, p_Y, p_Z:

(i) The definition of mutual information of X and Y and its relation to information entropy:

$$
\begin{aligned}
I(X;Y) &\equiv \mathbb{E}\left[\log \frac{p_{X,Y}(X,Y)}{p_X(X)p_Y(Y)}\right] \\
&= \int_{\mathcal{X},\mathcal{Y}} p_{X,Y}(x,y) \log \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)} dxdy \\
&= \mathbb{E}[\log p_{X,Y}(X,Y)] - \mathbb{E}[p_X(X)] - \mathbb{E}[p_Y(Y)] \\
&= -H(X,Y) + H(X) + H(Y) \\
&= H(Y) - H(Y|X) \\
&= H(X) - H(X|Y)
\end{aligned} \tag{1}
$$

where

$$
\begin{aligned}
H(X) &\equiv \mathbb{E}[-\log p_X(X)] = -\int_{\mathcal{X}} p_X(x) \log p_X(x) dx \\
H(X,Y) &\equiv \mathbb{E}[-\log p_{X,Y}(X,Y)] \\
&= -\int_{\mathcal{X},\mathcal{Y}} p_{X,Y}(x,y) \log p_{X,Y}(x,y) dxdy \\
H(Y|X) &\equiv \mathbb{E}[-\log p_{Y|X}(Y|X)] \\
&= -\int_{\mathcal{X},\mathcal{Y}} p_{X,Y}(x,y) \log p_{Y|X}(y|x) dxdy
\end{aligned}
$$

(ii) The conditional mutual information of X and Y given Z is defined as:

$$
\begin{aligned}
I(X;Y|Z) &\equiv \mathbb{E}\left[\log \frac{p_{X,Y|Z}(x,y|z)}{p_{X|Z}(x|z)p_{Y|Z}(y|z)}\right] \\
&= \int_{\mathcal{X},\mathcal{Y},\mathcal{Z}} p_{X,Y|Z}(x,y|z) p_Z(z) \\
&\quad \log \frac{p_{X,Y|Z}(x,y|z)}{p_{X|Z}(x|z)p_{Y|Z}(y|z)} dxdydz
\end{aligned} \tag{2}
$$

2. Proof of $I(f;f^*) = I(f;f^*|y) + I(y;f^*)$

Proof. We have $I(f^*;y) = H(f^*) - H(f^*|y)$ by Eq.(1). Furthermore, since f^* is obtained (deterministically) from f, we have $p_{f,f^*}(f,f^*) = p_f(f)$. Therefore, $I(f;f^*) = \mathbb{E}\left[-\log p_{f^*}(f^*)\right] = \mathbb{E}\left[\log \frac{p_{f,f^*}(f,f^*)}{p_{f^*}(f^*)p_f(f)}\right] = H(f^*)$ and similarly, $I(f;f^*|y) = H(f^*|y)$. Combining these, we have the desired result. ☐

3. Proof of Proposition

$$
KL\left[p(y|f) \| p(y|f^*)\right] = 0 \implies I(y;f) - I(y;f^*) = 0
$$

Proof.

$$
\begin{aligned}
& I(y;f) - I(y;f^*) = \\
& -\int p(f^*)\, p(y|f^*) \log p(y|f^*) df^* dy \\
& \quad + \int p(f) p(y|f) \log p(y|f) df dy \\
& = -\int p(f^*)\, p(y|f^*) \log\left[\frac{p(y|f^*)}{p(y|f)} p(y|f)\right] df^* dy \\
& \quad + \int p(f) p(y|f) \log\left[\frac{p(y|f)}{p(y|f^*)} p(y|f^*)\right] df dy \\
& = -\int p(f^*)\, KL\left[p(y|f^*) \| p(y|f)\right] df^* \\
& \quad - \int p(f^*)\, p(y|f^*) \log p(y|f) df^* dy \\
& \quad + \int p(f) KL[p(y|f) \| p(y|f^*)] df \\
& \quad + \int p(f)\, p(y|f) \log p(y|f^*)\, df dy \\
& = \mathbb{E}_f\left[KL[p(y|f) \| p(y|f^*)]\right] - \mathbb{E}_{f^*}\left[KL[p(y|f^*) \| p(y|f)]\right] \\
& \quad + \int p(y) \log \frac{p(y|f^*)}{p(y|f)} dy \\
& \leq \mathbb{E}_f\left[KL[p(y|f) \| p(y|f^*)]\right] + \int p(y) \log \frac{p(y|f^*)}{p(y|f)} dy.
\end{aligned}
$$

Using Jensen's inequality and the fact that $-\log$ is strictly convex, we can show that the KL-divergence is always non-negative and the equality only holds when the distributions are equal almost-everywhere, which is proven as

Table 1. Clinical variables in ITAC and percentage of missing data

#	1	2	3	4	5	6	7	8	9
Variable	Age	Oxygen saturation	Platelets	Measured saturation oxygen	Respiratory rate	PO2	D-Dimer	Cough	Dyspnea
Missing (%)	0	7.60	4.42	35.69	42.76	31.45	44.52	N/A	N/A
#	10	11	12	13	14	15	16	17	18
Variable	Diabetes	Neurological disease	Other CV disease	Admitted to ICU	Glucose	Urea	eGFR	GOT	PCR
Missing (%)	N/A	N/A	N/A	N/A	15.72	7.24	7.77	24.56	16.43

below:

$$\begin{aligned} KL[P\|Q] &= \mathbb{E}\left[-\log\frac{Q}{P}\right] \\ &\geq -\log\mathbb{E}\left[\frac{Q}{P}\right] \ \text{(by Jensen's inequality)} \\ &= -\log\int_{\mathcal{X}}\frac{Q(x)}{P(x)}P(x)dx = 0 \end{aligned} \tag{3}$$

where P and Q are two arbitrary distributions supported on $\mathcal{X}$. We have $KL[P\|Q] \geq 0$.

Hence, when $KL\left[p(y|f)\|p(y|f^{*})\right] = 0$, we have $p\left(y|f^{*}\right) = p\left(y|f\right)$ almost everywhere (follows from Eq. (3)), which implies $\int p(y)\log\frac{p(y|f^{*})}{p(y|f)}dy = 0$ and hence $I(y;f) - I(y;f^{*}) \leq 0$. We also have $I(y;f) - I(y;f^{*}) \leq 0$, therefore $KL\left[p(y|f)\|p(y|f^{*})\right] = 0 \implies I(y;f) - I\left(y;f^{*}\right) = 0$. □

4. Summary of clinical variables in ITAC

The overview of the missing data in the clinical variables in ITAC is given in Table 1. We simply fill the missing value by the mean value calculated from the overall datasets.