A Supplemental Material

This part first provides detailed derivation of Equation (8) and (11) from Equation (7) and (10), since our uniform bridge distribution and language-model bridge distribution have closed-form solutions given a fixed uniform distribution and a language model as constraints. Then, we give explanation of Equation (13), the objective function of coaching bridge, where the constraint is the inverse KL compared with previous two bridges and then give detailed derivation of the gradient update Equation (14).

Derivation of Equation (8)

$$\begin{aligned}
&L_B(\eta)\\
&=\mathop{\mathbb{E}}_{Y\sim p_\eta} -\frac{S(Y,Y^*)}{\tau} + KL(p_\eta(Y|Y^*)||U(Y))\\
&=\int_Y -p_\eta(Y|Y^*)\log\exp(\frac{S(Y,Y^*)}{\tau})\\
&\quad+\int_Y p_\eta(Y|Y^*)\log\frac{p_\eta(Y|Y^*)}{U(Y)}\\
&=\int_Y p_\eta(Y|Y^*)\log\frac{p_\eta(Y|Y^*)}{\exp(\frac{S(Y,Y^*)}{\tau})\cdot U(Y)}\\
&=\int_Y p_\eta(Y|Y^*)\log\frac{p_\eta(Y|Y^*)}{\exp(\frac{S(Y,Y^*)}{\tau})\cdot\frac{1}{|\mathcal{Y}|}}\\
&=\int_Y p_\eta(Y|Y^*)\log\frac{p_\eta(Y|Y^*)}{\exp\frac{S(Y,Y^*)}{\tau}}\\
&\quad+\log|\mathcal{Y}|\int_Y p_\eta(Y|Y^*)\\
&=\int_Y p_\eta(Y|Y^*)\log\frac{p_\eta(Y|Y^*)}{\exp\frac{S(Y,Y^*)}{\tau}}+Const\\
&=\int_Y p_\eta(Y|Y^*)\log\frac{p_\eta(Y|Y^*)}{\frac{\exp\frac{S(Y,Y^*)}{\tau}}{Z}}+Const'\\
&\quad=KL(p_\eta(Y|Y^*)||\frac{\exp\frac{S(Y,Y^*)}{\tau}}{Z})+Const'
\end{aligned}\tag{18}$$

Here, the Y^* related constant Z is needed to transform a unnormalized similarity score to a probability:

$$Z(Y^*)=\int_Y \exp\frac{S(Y,Y^*)}{\tau}\tag{19}$$

Derivation of Equation (11)

$$\begin{aligned}
&L_B(\eta)\\
&=\mathop{\mathbb{E}}_{Y\sim p_\eta} -\frac{S(Y,Y^*)}{\tau} + KL(p_\eta(Y|Y^*)||p_{LM}(Y))\\
&=\int_Y -p_\eta(Y|Y^*)\log\exp(\frac{S(Y,Y^*)}{\tau})\\
&\quad+\int_Y p_\eta(Y|Y^*)\log\frac{p_\eta(Y|Y^*)}{p_{LM}(Y)}\\
&=\int_Y p_\eta(Y|Y^*)\log\frac{p_\eta(Y|Y^*)}{\exp(\frac{S(Y,Y^*)}{\tau})\cdot p_{LM}(Y)}\\
&=\int_Y p_\eta(Y|Y^*)\log\frac{p_\eta(Y|Y^*)}{\frac{\exp\frac{S(Y,Y^*)}{\tau}\cdot P_{LM}(Y)}{Z}}+Const\\
&=KL(p_\eta(Y|Y^*)||\frac{\exp\frac{S(Y,Y^*)}{\tau}\cdot P_{LM}(Y)}{Z})+Const'
\end{aligned}\tag{20}$$

Here, the Y^* related constant Z is needed to transform a unnormalized *weighted* similarity score to a probability:

$$Z(Y^*)=\int_Y \exp\frac{S(Y,Y^*)}{\tau}\cdot P_{LM}(Y)\tag{21}$$

Explanation of Equation (13) This equation is the objective function of our coaching bridge, which uses an inverse KL term[5] as part of its objective. The use of inverse KL is out of the consideration of computational stability. The reasons are two-fold: 1). the inverse KL will do not change the effect of the constraint; 2). the inverse KL requires sampling from the generator and uses those samples as the target to train the bridge, which has the same gradient update ad MLE, so we do not need to consider baseline tricks in Reinforcement Learning implementation.

Gradient derivation of Equation (13)

$$\begin{aligned}
&\nabla L_B(\eta)\\
&=\nabla_\eta \mathop{\mathbb{E}}_{Y\sim p_\eta(Y|Y^*)} -\frac{S(Y,Y^*)}{\tau} + \nabla_\eta KL(p_\theta(Y|X)||p_\eta(Y|Y^*))\\
&=\mathop{\mathbb{E}}_{Y\sim p_\eta(Y|Y^*)} -\frac{S(Y,Y^*)}{\tau}\nabla_\eta\log p_\eta(Y|Y^*)\\
&\quad+\nabla_\eta \mathop{\mathbb{E}}_{Y\sim p_\theta(Y|X)}\log p_\eta(Y|Y^*)\\
&=\mathop{\mathbb{E}}_{Y\sim p_\eta(Y|Y^*)} -\frac{S(Y,Y^*)}{\tau}\nabla\log p_\eta(Y|Y^*)\\
&\quad+\mathop{\mathbb{E}}_{Y\sim p_\theta(Y|X)}\nabla\log p_\eta(Y|Y^*)
\end{aligned}\tag{22}$$

[5] That is the use of $KL(p_\theta||p_\eta)$ instead of $KL(p_\eta||p_\theta)$.