1. 关于判别器和生成器的损失函数问题

论文中最小化判别器D的损失函数：

$$ \mathop{\min}{\theta_D} \ L\theta(\theta_D) = -log P_{\theta_D}(y=1|z_t) -log P_{\theta_D}(y=0|z_s) $$ (3)

等价于

$$ \mathop{\max}{\theta_D} \ P{\theta_D}(y=1|z_t) + P_{\theta_D}(y=0|z_s) $$

而论文中提到，z_t是source content representation, 即fake的表示，而z_s是gold的表示，要现在要最大化$P_{\theta_D}(y=1|z_t)$，即要判别器尽可能的将z_t识别为gold，貌似与论文前述相反了。直观来看(3)式中的$y=1$和$y=0$的位置应交换，这样比较合理。生成器的损失函数亦是如此。

2. 如果损失函数确实有错，是否影响训练

跟同学讨论了一下，得出一个结论，1中的损失函数无论是否出错，即无论将z_t标记为gold还是fake，只要最终生成器和判别器的优化目标相反，则不影响训练，因为其本身还是个对抗训练的过程。是否如此呢？