

A SHOWING THE EQUIVALENCE OF THE PAIRWISE AND TRIPLET CLASSIFICATION TASK

In this section, we focus on the classification setup where the goal is to train a classifier to detect invs. out-of-order groups of inputs. Specifically, we want to show that the classification setup from Liu et al. (2021) where the inputs come in pairs is equivalent to the setup where the inputs come in triplets. The triplet inputs are what we assume for Time Control.

Notation We denote latent space as \mathcal{Z} ; we typically do not observe this space directly. We denote the observation space as \mathcal{X} ; this is the space we observe directly.

We define the Brownian bridge for $t \in [0, T]$ as

$$B_t = B(t) = W(t) - \frac{t}{T}W(T), \quad (4)$$

where $W(t)$ is a standard Wiener process, $\mathcal{N}(0, t)$.

We state the assumptions:

- We assume latents $\{z_t\}_{t \geq 0} \in \mathbb{R}^d$ are drawn from the Brownian bridge process defined by the stochastic differential equation,

$$dz_t = \frac{z_T - z_t}{1 - \frac{t}{T}} dt + dB_t \quad (5)$$

The intervals at which latents are sampled are every Δt step: $z' = z_{t+\Delta t}$.

- We denote the transition probabilities as

$$p_*(z_t | z_0, z_T) := \mathcal{N}\left(\left(1 - \frac{t}{T}\right)z_0 + \frac{t}{T}z_T, \frac{t(T-t)}{t}\right). \quad (6)$$

We denote the proposal distribution over possible intermediate triplets $q(z'_0, z'_t, z'_T)$.

Liu et al. (2021) characterize properties of an optimal classifier $h^*(z, z')$ which observes pairs of latents (z, z') and it outputs in $[0, 1]$, a probability indicating whether the pair comes *in order* (ie. $z' = z_t + \Delta t \cdot dz_t$) or *not in order* (ie. a randomly sampled latent). They train h^* using an L2 loss, $\mathcal{L}(h, \{(z, z'), y\})$.

Lemma 1 of their work states the following : The optimum of the contrastive learning objective $\arg \min_h \mathbb{E}_{((z, z'), y)} [\mathcal{L}(h, \{(z, z'), y\})]$ satisfies

$$h^*(z, z') = \frac{p^{\Delta t}(z, z')}{q(z') + p^{\Delta t}(z, z')}. \quad (7)$$

Manipulating this equality, we observe that the transition kernel has the following relation to the classifier which takes in pairs of observations,

$$p^{\Delta t}(z, z') = \frac{q(z')h^*(z, z')}{1 - h^*(z, z')} \quad (8)$$

$$\log p^{\Delta t}(z, z') = \log q(z') + \log h^*(z, z') - \log(1 - h^*(z, z')). \quad (9)$$

Our setting however assumes that the algorithm receives triplets of data points, $z_{t_1}, z_{t_2}, z_{t_3}$. We want to show below that minimizing \mathcal{L} with triplet contrasts in the classification setting still approximates the transition kernel. In particular, we're interested in transitions of a Brownian bridge pinned at z_0, z_T : $p^{\Delta t}(z_0, z_t, z_T) = Pr(z_t | z_0, z_T)$.

Let's say we have two positive triplet samples, $(z_{t_1}^i, z_{t_2}^i, z_{t_3}^i)$ and $(z_{t_1}^j, z_{t_2}^j, z_{t_3}^j)$, where $t_1 < t_2 < t_3$. Following Liu et al. (2021), minimizing \mathcal{L} yields the following on each triplet:

$$\log p^{\Delta t}(z_{t_1}^i, z_{t_2}^i, z_{t_3}^i) = \log q(z') + \log h^*(z_{t_1}^i, z_{t_2}^i, z_{t_3}^i) - \log(1 - h^*(z_{t_1}^i, z_{t_2}^i, z_{t_3}^i)) \quad (10)$$

$$\log p^{\Delta t}(z_{t_1}^j, z_{t_2}^j, z_{t_3}^j) = \log q(z') + \log h^*(z_{t_1}^j, z_{t_2}^j, z_{t_3}^j) - \log(1 - h^*(z_{t_1}^j, z_{t_2}^j, z_{t_3}^j)). \quad (11)$$

Taking the difference in log probabilities between Equations 10-11 results in

$$\begin{aligned} \log p^{\Delta t}(z_{t_1}^i, z_{t_2}^i, z_{t_3}^i) - \log p^{\Delta t}(z_{t_1}^j, z_{t_2}^j, z_{t_3}^j) &= [\log h^*(z_{t_1}^i, z_{t_2}^i, z_{t_3}^i) - \log(1 - h^*(z_{t_1}^i, z_{t_2}^i, z_{t_3}^i))] \\ &\quad - [\log h^*(z_{t_1}^j, z_{t_2}^j, z_{t_3}^j) - \log(1 - h^*(z_{t_1}^j, z_{t_2}^j, z_{t_3}^j))]. \end{aligned} \quad (12)$$

Similar to the pair-wise classification setting, we’ve shown that minimizing \mathcal{L} in the triplet classification setting results in approximating the transition kernel of the Brownian bridge process.

B FIRST ORDER APPROXIMATION OF THE BROWNIAN BRIDGE PROCESS

$$Z_{t+\Delta t} \approx Z_t \left(1 - \frac{1}{T-t}\right) - \frac{1}{T-t} \Delta t \cdot Z_T + \sqrt{\Delta t} \mathcal{N}(0, I) \quad (13)$$

C TRAINING DETAILS

Here we describe the training procedure for the encoder and the fine-tuning procedure for decoding.

Brownian bridge encoder The encoder architecture is a frozen, pretrained GPT2 model from Huggingface (Radford et al., 2019; Wolf et al., 2020) and a trainable MLP network. We extract the GPT2’s last layer hidden state that corresponds to the end-of-sentence (EOS) token and train the 4-layer MLP on top of the hidden state. The MLP network has intermediate ReLU activations and is trained with stochastic gradient descent with a learning rate of 1e-4 and with momentum 0.9. We train the encoder for 100 epochs on each of the datasets.

The text fed into GPT2 are fed in on a sentence level. This means that the input x_t refers to the t ’th sentence of a document. The sentences are separated from each other in the main text as “ . ” which is added to the tokenizer as a separate token for indexing convenience.

Fine-tuning GPT2 with latent embeddings After training the encoder, we run it on the training dataset to collect an accompanying latent trajectory for each text. The encoder is run on the dataset at a sentence level: we separate the text by sentences and pass the sentences through the encoder.

The sentence latent embeddings are aligned with the tokens of that sentence and offset by one token before the start of that sentence token. Let’s illustrate by an example. We denote [SOS] as the start of the document token, [s1] as sentence 1 tokens and [s2] as sentence 2 tokens. [.] is the period token which we’ve added into the tokenizer. z_i denote the latent variable corresponding to the i ’th sentence.

Let’s say that the sequence fed into GPT2 is “[SOS] [s1] [s1] [s1] [.] [s2] [s2] [s2]”. Then the corresponding latent trajectory is “ $z_1, z_1, z_1, z_1, z_2, z_2, z_2, z_2$ ”. The latent variables are added onto the positional embeddings. We then fine-tune GPT2 as normal.

D GENERATION WITH EMBEDDINGS

We first sample a start and end latent, $z_0 \sim p(z_0), z_T \sim p(z_T)$ where $p(z_0), p(z_T)$ are calculated as the density estimates over the training dataset. We pin our trajectory to the start and end latent, and run the Brownian bridge using Equation 13. For normal long text generation, we set T to be the average number of sentences in each of the dataset. For forced long text generation, we set T to be proportional to the number of sentences needed in order to generate 1000 tokens. By the end, we have a trajectory z_0, z_1, \dots, z_T .

Generation starts with feeding the SOS token and the first latent z_0 . Once GPT2 emits a [.] token and terminates sentence t , we transition to the next latent z_{t+1} . This process continues until GPT2 is finished with generation. If GPT2 generates more sentences than there are latents in the trajectory, the last latent z_T is used until the end of generation.

E ABLATIONS

E.1 IMPLICIT DYNAMICS

The Implicit Dynamics ablation is where we compare our *explicit* dynamics objective to van den Oord et al. (2019) which suggests an *implicit* latent dynamics objective. The encoder g_{enc} for Implicit Dynamics has the same architecture as Time Control. As done in van den Oord et al. (2019), we parameterize the (autoregressive) context encoder g_{ar} using a 2400-hidden-unit GRU. Since g_{enc} and g_{ar} are trained to align latents up to a linear rotation, we use g_{enc} for extracting the sentence embeddings.

E.2 VARIATIONAL AUTO-ENCODER

The Variational Auto-Encoder ablation is where we compare our contrastive objective with a variational one. Below we derive the ELBO objective. Similar to the contrastive objective, we’re deriving the ELBO over the triplet dynamics in the Brownian bridge.

$$\begin{aligned}
 \log p(\mathbf{x}) &\geq \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \left[\log \frac{p(\mathbf{x}, \mathbf{z})}{q(\mathbf{z}|\mathbf{x})} \right] \\
 &= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \left[\log \frac{p(x_0, x_t, x_T, z_0, z_t, z_T)}{q(z_0, z_t, z_T | x_0, x_t, x_T)} \right] \\
 &= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \left[\log \frac{p(x_0|z_0)p(x_t|z_t)p(x_T|z_T)p(z_t|z_0, z_T)p(z_0)p(z_T)}{q(z_t|z_0, z_T, x_t)q(z_0|x_0)q(z_T|x_T)} \right] \\
 &= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [p(x_0|z_0)p(x_t|z_t)p(x_T|z_T)] + \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \left[\log \frac{p(z_t|z_0, z_T)}{q(z_t|z_0, z_T, x_t)} \right] \\
 &\quad + \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \left[\log \frac{p(z_0)}{q(z_0|x_0)} \right] + \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \left[\log \frac{p(z_T)}{q(z_T|x_T)} \right] \\
 &= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p(x_0|z_0)p(x_t|z_t)p(x_T|z_T)] \\
 &\quad - D_{\text{KL}}(q(z_t|z_0, z_T, x_t) \| p(z_t|z_0, z_T)) - D_{\text{KL}}(q(z_0|x_0) \| p(z_0)) - D_{\text{KL}}(q(z_T|x_T) \| p(z_T))
 \end{aligned}$$

The encoder $q_\phi(z|x)$ is parameterized with the same architecture as our encoder. The decoder $p(x|z)$ is a fine-tuned GPT2 model.

F DATASET INFORMATION

For each dataset, text examples were filtered out if they did not fit within GPT2’s context length of 1024 tokens. We also added the token “ . ” for each setting to mark the end of a sentence. This was done for indexing purposes, eg. when aligning the latent embeddings.

Wikisection (Arnold et al., 2019) includes Wikipedia articles on cities split by sections. We adapt this dataset such that each article contains four ordered sections (abstract, history, geography, demographics) marked with section id tokens: Each article is represented as, “[ABSTRACT] text [HISTORY] text [GEOGRAPHY] text [DEMOGRAPHICS] text”. These section id tokens are added to the tokenizer.

The training dataset contains 1420 articles. The section lengths have the following breakdown measured in the number of BPE tokens (GPT2 tokenizer):

- Abstract: 75.8 ± 1.4

- History: 191.5 ± 3.7
- Geography: 83.9 ± 1.5
- Demographics: 342.6 ± 4.6

The test dataset contains 431 articles. The section lengths have a similar breakdown:

- Abstract: 73.5 ± 2.6
- History: 180.2 ± 6.2
- Geography: 85.2 ± 2.7
- Demographics: 332.5 ± 8.6

The ordering metric used in Table 5 is 1 if all four section ids occur exactly once and come in the order as they are listed above.

The length mismatch in % used in Table 3 is calculated with respect to the training set lengths.

Wikihow (Koupae & Wang, 2018) contains how-to articles organized by a title, method, and steps. Each article includes multiple methods for completing a multi-step procedural task such as “How to Register to Vote”. We scraped all the available English articles covering a wide range of topics following Koupae & Wang (2018). We mark each with its own section id tokens: Each article is represented as “[TITLE] text [METHOD] text [STEP] 1 text [STEP] 2 text ...”

The training dataset contains 1566 articles. The section lengths are,

- Title: 10.4 ± 2.2
- Method: 8.7 ± 2.2
- Steps (total step length): 480.2 ± 231.5

The test dataset contains 243 articles. The section lengths are,

- Title: 10.7 ± 2.2
- Method: 8.6 ± 2.1
- Steps (total step length): 480.1 ± 224.0

The ordering metric used in Table 5 is 1 if all the section ids appear in order, the TITLE and METHOD section ids are not repeated, and the step numbers come in order. It’s 0 otherwise.

The deviation in section length measured in Table 4 is calculated with respect to the training set lengths. We check for whether the models are able to maintain the section lengths when it has to extrapolate. The most common failure mode is that the model generates incoherent text and allocates this text to the last section of what it’s generated thus far, resulting in the deviation in section lengths.

Recipe NLG (Bieñ et al., 2020) contains recipes, each with a title, ingredients and set of instructions. A recipe is constructed as “[TITLE] text [INGREDIENTS] text [DIRECTIONS] text”.

The training dataset contains 4000 recipes. The section lengths are,

- Title: 9.7 ± 3.4
- Ingredients: 23.8 ± 4.5
- Directions (total step length): 62.0 ± 14.5

The test dataset contains 1000 recipes. The section lengths are,

- Title: 9.4 ± 3.0
- Ingredients: 24.1 ± 4.5
- Directions (total step length): 63.3 ± 13.7

The ordering metric used in Table 5 is 1 if all the section ids appear exactly once and in order. It’s 0 otherwise.

Taskmaster-2 (TM-2) (Byrne et al., 2019) which contains conversations on finding restaurants between an assistant and a user. The assistant’s turn is marked with an “[ASSISTANT]” tag, and the user’s turn is marked with a “[USER]” tag.

The training dataset contains 2000 conversations. The section lengths are,

- User: 11.8 ± 5.6
- Assistant: 18.0 ± 3.7

The test dataset contains 1276 conversations. The section lengths are,

- User: 11.7 ± 5.6
- Assistant: 19.5 ± 3.0

The ordering metric used in Table 5 does not apply here because the user and assistant don’t take turns in the dialog.

The deviation in section length measured in Table 4 is calculated with respect to the training set lengths. We check for whether the models are able to maintain the utterance lengths between the user and assistant when it has to extrapolate. The most common failure mode is that the model generates incoherent text and allocates this text to the last section of what it’s generated thus far, resulting in the deviation in section lengths.

TicketTalk (Byrne et al., 2021) which contains conversations on booking movie tickets between an assistant and a user. The assistant’s and user’s turned are similarly marked as in TM-2.

The training dataset contains 2000 conversations. The section lengths are,

- User: 11.8 ± 5.6
- Assistant: 18.0 ± 3.7

The test dataset contains 1276 conversations. The section lengths are,

- User: 11.7 ± 5.6
- Assistant: 19.5 ± 3.0

The deviation in section length measured in Table 4 is calculated similarly to TM-2.

The ordering metric used in Table 5 applies because the user and assistant take turns in the dialog. The ordering metric is 1 if the user and assistant take turns in the conversation. It’s 0 otherwise.

ROC Stories (Mostafazadeh et al., 2016) is a short stories dataset. Each story contains 5 sentences. No additional tokens are added in this dataset. The training dataset contains 2000 stories, and the test dataset contains 1000 stories.

G PERPLEXITY AFTER FINE-TUNING

Table 9 reports the final perplexity scores after fine-tuning GPT2 on the different domains with the methods. We fine-tune for 10 epochs and checkpoint the models every 1000 steps; we keep the model checkpoint that scores the lowest PPL on a held-out validation set.

H GENERATION EXAMPLES

We provide example outputs from the different models on text infilling. We denote the prefix sentences as “LEFT CONTEXT”, the suffix sentences as “RIGHT SENTENCES”, the generated output from the models as “GENERATED” and the ground truth infill sentence as “GT”.

Method	Wikisection	Wikihow	TM-2	TicketTalk	Recipe
GPT2	5.9	15.3	4.5	4.4	7.5
sec-dense	5.9	—	—	—	—
sec-sparse	5.9	—	—	—	—
VAE (8)	5.5	15.3	4.5	4.1	7.3
VAE (16)	5.5	15.3	4.5	4.1	7.3
VAE (32)	5.5	15.3	4.5	4.1	7.3
ID (8)	5.5	14.9	4.5	4.0	7.3
ID (16)	5.4	14.9	4.3	3.9	7.0
ID (32)	5.3	14.9	4.2	3.9	6.7
BM (8)	5.5	15.2	4.5	4.1	7.3
BM (16)	5.5	15.3	4.5	4.1	7.3
BM (32)	5.5	15.2	4.5	4.1	7.3
TC (8)	5.5	15.2	4.3	4.0	6.9
TC (16)	5.5	15.3	4.3	4.0	6.9
TC (32)	5.5	15.2	4.3	4.0	7.0

Table 9: Perplexity after fine-tuning.

H.1 TEXT INFILLING

LM See Table [10](#)

Table 10: Example of text infilling with Donahue et al. (2020)'s LM model on ROCStories.

LEFT CONTEXT: John couldn't wash his clothes because his washing machine was broken. GENERATED: He had an idea. GT: John didn't know how to fix washing machines. RIGHT CONTEXT: He bought a book to tell him how to repair it. He changed the fan belt. When he plugged it back in, it worked.
LEFT CONTEXT: I found a stray cat yesterday. I gave it some milk. It came back again today. I was going to give it some milk but didn't. GENERATED: I guess I need to get something. GT: I had run out of milk. RIGHT CONTEXT:
LEFT CONTEXT: Toni took a trip to the local fair with her family. She immediately laid eyes on a pretty balloon. The balloon was red and star shaped. Unfortunately it slipped out of her hands. GENERATED: Her swollen fingers had horrible rashes. GT: She was sad to see it go. RIGHT CONTEXT:
LEFT CONTEXT: Miles was a roofer. He loved his job and was very good at it. GENERATED: He dreamed of finding a good title. GT: However, one day he was not paying attention and fell off the roof. RIGHT CONTEXT: Luckily, he only broke his leg. Miles decided that he needed to be more careful.
LEFT CONTEXT: Lindsey was feeling especially antsy at school. GENERATED: The school wasn't paying much attention to her. GT: She did not want to go to Lacrosse practice after school. RIGHT CONTEXT: She tried to fake a sickness, but nobody believed her. Lindsey ended up skipping practice without an excuse. She hoped she wouldn't get in trouble.
LEFT CONTEXT: Last night Jack went on a blind date. When he arrived at the restaurant a he heard someone yell his name. He spotted a blonde waving from across the room. GENERATED: She was a waiter. GT: She was nothing like the photo she sent. RIGHT CONTEXT: But he figured he would try to make the most of it.
LEFT CONTEXT: Rico was working really hard on his essay for English. After working on it all night, he decided to take a short nap. He slept in and was late for for school. When he got to school he noticed that his essay was gone. GENERATED: teacher found the essay in his bag and he passed it. GT: Rico's teacher let him turn in the essay the next day. RIGHT CONTEXT:

Table 11: Example of text infilling with [Donahue et al. \(2020\)](#)'s ILM model on ROCStories.

LEFT CONTEXT: GENERATED: My 98 year old friend and I played blackjack yesterday. GT: Last week's family game night was intense. RIGHT CONTEXT: We were playing Monopoly and nobody made any headway for hours. Everyone was trying their hardest to win and the game kept going. It wasn't until we finally decided to check the rules that we knew why. There were many different pieces missing.
LEFT CONTEXT: Tom was jealous of his brother. His brother was older and stronger. His brother went wherever he wanted. GENERATED: Tom decided to use steroids. GT: One day his brother was grounded for staying out too late. RIGHT CONTEXT: This made Tom really happy.
LEFT CONTEXT: His first time at the opera. He only went after his girlfriend begged. He sat for three hours in sheer boredom. Finally it was over. GENERATED: He turned on the tv and heard her thanking him. GT: He learned he didn't like the opera. RIGHT CONTEXT:
LEFT CONTEXT: My best friend played a video game online. One day, she met a nice boy there. They talked every day and fell in love. They finally met in person and hit it off. GENERATED: Now, they are best friends. GT: The two became a very loving couple. RIGHT CONTEXT:
LEFT CONTEXT: The family decided to go to the park for Veteran's Day. GENERATED: They asked permission from the park manager. GT: We had a barbecue outside. RIGHT CONTEXT: In addition, everyone decided to take a swim in the water. Before we left, we had a huge cake. We then swam for an hour at the lake.
LEFT CONTEXT: Jennifer always wanted to have the birthday of her dreams. Her parents surprised her with a big birthday party at the park. All of her friends and family came to celebrate her special day. She was so happy as she blew out the candles on her cake. GENERATED: Her parents had made a big surprise to her by turning the cake into a carnival. GT: It was the best day of birthday she could have imagined. RIGHT CONTEXT:
LEFT CONTEXT: Terry's daughter was in extreme pain in her mouth. GENERATED: Terry and her husband decided to stop the baby from teething. GT: After an evaluation, Terry realized that her daughter was teething. RIGHT CONTEXT: Once Terry realized she was teething, Terry administered medication. After the medication, Terry placed teething gel on the baby's teeth. Terry got the baby to calm down.

Table 12: Example of text infilling with **Time Control** (d=8) on ROCStories.

LEFT CONTEXT: My day on Sunday this week went as expected. My family and I attended Church from 8 to 11 in the morning.

GENERATED: We went to a local church in the afternoon .

GT: Afterwards, we went to eat dinner at a restaurant.

RIGHT CONTEXT: The rest of my day was spent doing homework assignments. I then went to sleep feeling happy about my life.

LEFT CONTEXT: His first time at the opera. He only went after his girlfriend begged. He sat for three hours in sheer boredom. Finally it was over.

GENERATED: He was so happy he didn't want to leave.

GT: He learned he didn't like the opera.

RIGHT CONTEXT:

LEFT CONTEXT: My best friend played a video game online. One day, she met a nice boy there. They talked every day and fell in love. They finally met in person and hit it off.

GENERATED: Until that day, they were married.

GT: The two became a very loving couple.

RIGHT CONTEXT:

LEFT CONTEXT: The other day at the clinic I had to help put a dog down. He seemed really sad and lonely. Like he knew what was going to happen.

GENERATED: He was going to die soon .

GT: As we laid it down and it took its final breaths it stared at me.

RIGHT CONTEXT: I stayed calm, but cried after we were finished.

LEFT CONTEXT: Tom was jealous of his brother. His brother was older and stronger. His brother went wherever he wanted.

GENERATED: Tom was jealous of his brother.

GT: One day his brother was grounded for staying out too late.

RIGHT CONTEXT:

LEFT CONTEXT: Jays habit of buying expensive beer was catching up to him. He was spending more money on beer than food. He needed to find another source of income to support this habit. A friend recommended he try out Amazon MTurk.

GENERATED: He found the site and bought a few beers.

GT: Jay become slightly less poor from Amazon Mturk.

RIGHT CONTEXT:

LEFT CONTEXT: John couldn't wash his clothes because his washing machine was broken.

GENERATED: John went to the store to buy a new one .

GT: John didn't know how to fix washing machines.

RIGHT CONTEXT: He bought a book to tell him how to repair it. He changed the fan belt. When he plugged it back in, it worked.

Table 13: Example of text infilling with **Time Control** (d=16) on ROCStories.

LEFT CONTEXT: Tom was jealous of his brother. His brother was older and stronger. His brother went wherever he wanted.

GENERATED: Tom’s brother was very jealous of his brother.

GT: One day his brother was grounded for staying out too late.

RIGHT CONTEXT: This made Tom really happy.

LEFT CONTEXT: Jackie was 11 and had to get braces. She was worried about what her friends would think. She tried to hide them when she first got them.

GENERATED: But she was too embarrassed to tell them .

GT: Eventually her friends saw them and she was embarrassed.

RIGHT CONTEXT: Her friends noticed she was embarrassed and decided to comfort her.

LEFT CONTEXT: Sally was going to surprise the office with a cake. Sally felt that a cake would be a good way to make them smile. She went to the supermarket to pick up the cake.

GENERATED: She bought the cake and was very happy with it.

GT: At the office she gathered the employees around the conference table.

RIGHT CONTEXT: She then brought out a cake and they all felt better.

LEFT CONTEXT: Lars was playing XBOX. His controller stopped working during a game. Lars didn’t have a car so he had to walk all the way to the store. The store was being remodeled when he got there so he went to another.

GENERATED: Video games were all over the floor.

GT: Lars wasn’t able to find a controller.

RIGHT CONTEXT:

LEFT CONTEXT: Emma had been working as a dishwasher. Her hands cracked and bled from the hot soapy water. Then her mom noticed and concocted a special salve for her. Emma used the salve every night before bed.

GENERATED: Soon she was able to keep her hands clean and happy.

GT: Her hands got better.

RIGHT CONTEXT:

LEFT CONTEXT: Jerome dribbled the ball quickly. Sam tried to grab it from him, but wasn’t fast enough. He chased Jerome down the court.

GENERATED: He scored a point with a long shot .

GT: Jerome pushed Sam backwards and threw the ball.

RIGHT CONTEXT: Jerome scored points for his team.

LEFT CONTEXT: Class started in 10 minutes and I had a math assignment due. My older friend volunteered to do it for me. Unfortunately, my teacher found out about the copying.

GENERATED: She was very upset with me for doing the assignment.

GT: She crumbled my paper and threw it in the trash.

RIGHT CONTEXT: My teacher gave me an F for the assignment.

Table 14: Example of text infilling with **Time Control** (d=32) on ROCStories.

LEFT CONTEXT: I went to the park to play frisbee with my dog. GENERATED: We played all day . GT: I tossed the frisbee to my dog and he would catch it in his mouth. RIGHT CONTEXT: I accidentally threw the frisbee too far. The frisbee landed into the pond. But my dog went and got it.
LEFT CONTEXT: I was tired of working at Walmart. The hours were bad. The store music was awful. GENERATED: I was very bored . GT: I handed my two weeks in to the manager. RIGHT CONTEXT: I then found another job and was happy.
LEFT CONTEXT: Sam bought a new SUV. It was all wheel drive. He figured he would take it off road. He hit a few hard bumps and broke his suspension. GENERATED: Unfortunately he had to pay a lot of money for it. GT: Sheepishly, he brought it to the dealership for repair. RIGHT CONTEXT:
LEFT CONTEXT: Missy got drunk and went to get a tattoo. She decided to get a tattoo on her forehead. The next day, Missy was horrified at what she had done. GENERATED: Her tattoo was on her forehead! GT: Missy scraped up her money to pay for a tattoo removal procedure. RIGHT CONTEXT: After much wasted money, the tattoo was gone.
LEFT CONTEXT: Jake was going on a road trip to see his family. He Got in the car and drove. GENERATED: Jake’s family was driving down the road . GT: The car’s tires exploded due to too much air. RIGHT CONTEXT: Jake hitchhiked for 30 miles. When Jake got to his family he was happy his trip was over.
LEFT CONTEXT: Bob decided to start a business. GENERATED: However, he did not know the market at all . GT: He opened up a grocery store and was doing very well. RIGHT CONTEXT: After a year, his profits dropped and he had to declare bankruptcy. Bob was sad to see his business fail. Bob worked hard and reopened his business.
LEFT CONTEXT: I hadn’t seen my girlfriend in a while. She got a new job so it’s hard to talk. The job takes up all of her time. GENERATED: I had to ask her to go out with me . GT: Finally she called me to hang out. RIGHT CONTEXT: I was really happy to see her and we made plans.

[USER] I'm interested in buying some tickets for a movie tonight. [ASSISTANT] OK. Can you tell me the city and state where the theater is located? [USER] It's been so long since we've been to the movies that I want to get out. It can't get any better. [ASSISTANT] Great. Can you tell me the city and state where you plan to see the movie? [USER] I'll be in Santa Barbara, California at the moment. [ASSISTANT] Got it. Are you interested in a particular movie? [USER] There's a couple of nice theaters in town. [ASSISTANT] OK. Is there a movie you have in mind? [USER] I'm thinking action. [ASSISTANT] Ok. Let's see what action films are out in that area. [ASSISTANT] I found one action movie at AMC Mountain 16: No Time To Die. Are you interested in that one? [USER] My best friend Sally has been wanting to see that one, and I guess we'll go with that one. [ASSISTANT] OK. Are you interested in this movie though? [USER] Definitely interested in seeing No Time to Die. [ASSISTANT] OK. Remaining showtimes for No Time To Die at AMC Mountain 16 are 4:30pm, 6:40pm and 9:10pm. [USER] 9:10 pm is good. [ASSISTANT] Got it. And how many tickets? [USER] Just me and a couple other guys. [ASSISTANT] All right. Let me confirm that you'd like three tickets for No Time To Die at AMC Mountain 16 tonight at 9:10pm. Is that all correct? [USER] Yes that's right. [ASSISTANT] Is it OK to go ahead and purchase these tickets? [USER] Yes go right ahead. [ASSISTANT] OK. Your tickets are purchased and details for how to proceed have been sent to your email address. Can I help with anything else? [USER] No. [ASSISTANT] OK. Enjoy your movie! As always, have a good day. **javascript ive been! havn bout ive heard it. dont think to much. in there. lmao. id appreciate it. reccommended. The email address for your tickets is theater. There's no problem at all with anything else. thanks! sign off. Your tickets are confirmed and they've been sent to your email address. thanks a lot. your welcome. enjoy your movie. -lbye. I'm very, very sorry for leaving you behind. Thanks for calling. Is there anything else I can help with? dont remember the last time we went out. I'm very sorry. Thank you very much. arn. Thanks for all of your help. Enjoy the movie. www, bye. -en Chinese girlfriend, bye. can you enjoy your movie. Bye. remaining showtimes for No Time To Die at AMC Mountain 16 are 4:30pm, 6:40pm and 9:10pm. This is all for right now. N/A Thanks for all of your assistance. @ @ b/r/r/r/r abys mam. @ cornyyyyyyyyyyyyy. BULLY BULLY MOVIE FOR YOU! Thank you for calling. Enjoy your movie! Your movie. Aww, bye. happy movie. thanks for calling. -en . Thanks for using our services. > .**

Figure 3: Example 1 of GPT2 forced long text generation.

I.2 TEXT INFILLING EVALUATION

Given the preceding and following context, we ask annotators to evaluate whether the generated fill-in sentence is a reasonable completion. We evaluate generated fill-in sentence by ILM, LM and Time Control on a scale of 1-5.

[USER] I'm interested in buying some tickets for a movie tonight. [ASSISTANT] OK. Can you tell me the city and state where the theater is located? [USER] It's been so long since we've been to the movies that I want to get out. It can't get any better. [ASSISTANT] Great. Can you tell me the city and state where you plan to see the movie? [USER] I'll be in Santa Barbara, California at the moment. [ASSISTANT] Got it. Are you interested in a particular movie? [USER] There's a couple of nice theaters in town. [ASSISTANT] OK. Is there a movie you have in mind? [USER] I'm thinking action. [ASSISTANT] Ok. Let's see what action films are out in that area. [ASSISTANT] I found one action movie at AMC Mountain 16: No Time To Die. Are you interested in that one? [USER] My best friend Sally has been wanting to see that one, and I guess we'll go with that one. [ASSISTANT] OK. Are you interested in this movie though? [USER] Definitely interested in seeing No Time to Die. [ASSISTANT] OK. Remaining showtimes for No Time To Die at AMC Mountain 16 are 4:30pm, 6:40pm and 9:10pm. [USER] 9:10 pm is good. [ASSISTANT] Got it. And how many tickets? [USER] Just me and a couple other guys. [ASSISTANT] All right. Let me confirm that you'd like three tickets for No Time To Die at AMC Mountain 16 tonight at 9:10pm. Is that all correct? [USER] Yes that's right. [ASSISTANT] Is it OK to go ahead and purchase these tickets? [USER] Yes go right ahead. [ASSISTANT] OK. Your tickets are purchased and details for how to proceed have been sent to your email address. Can I help with anything else? [USER] No. [ASSISTANT] OK. Enjoy your movie! As always, have a good day. **javascript ive been! havn bout ive heard it. dont think to much. in there. lmao. id appreciate it. reccommended. The email address for your tickets is theater. There's no problem at all with anything else. thanks! sign off. Your tickets are confirmed and they've been sent to your email address. thanks a lot. your welcome. enjoy your movie. -lbye. I'm very, very sorry for leaving you behind. Thanks for calling. Is there anything else I can help with? dont remember the last time we went out. I'm very sorry. Thank you very much. arn. Thanks for all of your help. Enjoy the movie. www, bye. -en Chinese girlfriend, bye. can you enjoy your movie. Bye. remaining showtimes for No Time To Die at AMC Mountain 16 are 4:30pm, 6:40pm and 9:10pm. This is all for right now. N/A Thanks for all of your assistance. @@ b/r/r/r/r abys mam. @ cornyyyyyyyyyyyyy. BULLY BULLY MOVIE FOR YOU! Thank you for calling. Enjoy your movie! Your movie. Aww, bye. happy movie. thanks for calling. -en . Thanks for using our services. > .**

Figure 4: Example 2 of GPT2 forced long text generation.

[USER] Hey! I wanna go tonight to watch a movie! [ASSISTANT] Hi, what movie would you like to watch? [USER] I want to know about Birds of Prey! [ASSISTANT] Sure, is there any particular theater you would like to watch Birds of Prey at? [USER] Can you tell me what theaters are showing the movie? [ASSISTANT] What are you looking at? [USER] Can you tell me about what theaters are showing Birds of Prey at? [ASSISTANT] Are you interested in one theater? [USER] What theaters are showing Birds of Prey at? [ASSISTANT] Okay, what movie theatre are you going to? [USER] I want to go to AMC. [ASSISTANT] Very well, what time would you like to watch the movie? [USER] What is the movie playing tonight at 1:00pm and 3:00 pm? [ASSISTANT] Can you look up times for Birds of Prey at AMC theaters? [USER] Can you tell me about the time and the theater where Birds of Prey is playing? [ASSISTANT] Sure, would you like to watch an IMAX, 3D or Standard showing? [USER] I want the IMAX show especially if there is an IMAX showing. [ASSISTANT] Sure, how many tickets would you like to purchase? [USER] I need five tickets. [ASSISTANT] Okay, I found the IMAX showing at 1:00pm, does that work for you? [USER] Sure, let's do that! [ASSISTANT] How many tickets would you like to purchase? [USER] I want two tickets, for me, my daughter and my doctor. [ASSISTANT] Okay, I have purchased two tickets for Birds of Prey at AMC theater for 1:00pm for three people. Is that all correct? [USER] Yes, that's correct. [ASSISTANT] Okay, would you like to confirm your purchase and book your tickets? [USER] Yes, I want to purchase tickets [ASSISTANT] Please provide me your email address, I'll see what I can find. [USER] I'd like to use google, is there an information available for emailing tickets to birds of prey? [ASSISTANT] Sure, would you like me to check where my tickets are located? [USER] Can you tell me the closest theaters to me for tickets showing tonight? [ASSISTANT] Sure, how would you like to see the movie playing in theaters near you? [USER] Can you look for any similar theater, get me food, and tickets? [ASSISTANT] Okay, showing at AMC theaters in Davis, California, is there a way you can search? [USER] Look up some movie showtimes for me, see if they have anything that is showing in IMAX or 3D. [ASSISTANT] Would you like me to search all theaters for tickets showing any available IMAX shows? [USER] Can you look for showing times of Dolittle at AMC? [ASSISTANT] Sure, could I look at the showtimes for the movies showing at 3D? [USER] I want the earliest showing times, maybe at 10:20 pm. [ASSISTANT] Not a problem, not an issue, let me know if I can change that. Thanks, I'll get it on my phone. C'mon, have you decided on a time? [ASSISTANT] Yes, I saw your email for confirmation and booked three tickets for Birds of Prey at AMC theater for 1:00 pm, does that work for you? [USER] Oh ok, forget about it now. [ASSISTANT] Would you like to checkout the booking details and book your tickets online, please. [USER] Actually, can you help me find tickets for Dolittle showing at 10:20 pm in 3D instead of 3D? [ASSISTANT] Not a problem, can you look up tickets to that showing instead instead? [USER] Can you think of anything else? [ASSISTANT] Sure, I'll check out the tickets at AMC showing time showing in 3D as well, if there is anything else? [USER] Yea, cool, I just need to get those booked and I'll get more information to help buy those tickets! [ASSISTANT] Awesome, I have just booked three tickets for Dolittle showing in 3D at AMC showing time in 3D at 10:20 pm tonight, does that work for you today? [USER] I want to upgrade to 3D, and then use my Apple Pay. [ASSISTANT] Okay, I'm turning this over and getting those tickets online now. Back to you, I want to purchase your movie tickets to purchase the movie tickets tonight at 10:20 pm at AMC showing in 3D. Correct? [USER] Sure, so I want to purchase a movie ticket for the movie showing in 3D instead of 3D instead, and then use my Apple Pay. [ASSISTANT] Okay, I will now see if you can purchase the movie tickets to purchase your movie tickets to purchase the movie tickets tonight at 10:20 pm at AMC showing in 3D. Is there anything else you need help with? [USER] Not really, not really. Thanks!

Figure 5: Example 1 of Time Control forced long text generation.

[USER] Hello. I am thinking of going to see a movie today. [ASSISTANT] OK. And where will you be seeing the movie? [USER] Maybe in Creek's End, Oregon? [ASSISTANT] Creek's End, Oregon. Got it. Is there a particular movie you have in mind? [USER] Just any James Bond movie. [ASSISTANT] Ok. I think you are talking about No Time To Die. Does that sound right? [USER] Yes it is, No Time To Die is the name. [ASSISTANT] I think action movies are always a great choice. In Creek's End, Oregon, I see that No Time To Die is showing in two theaters: AMC Mountain 16 and AMC Mercado 24. [USER] I want the AMC Mercado 24 its what I used to call it. [ASSISTANT] It looks like tonight's remaining showtimes for No Time To Die at AMC Mercado 24 are 5:00pm, 7:10pm, and 9:45pm. Which is best for you? [USER] Can I change the theater from the one I picked because the another one has other showings around it? [ASSISTANT] OK, No Time To Die it is. [USER] Can I have showtimes for two instead of one? [ASSISTANT] Sure, you chose AMC Mercado 24. [USER] Can I change the time, I want the earliest showtime instead? [ASSISTANT] Let me check showtimes at AMC Mercado 24. [USER] Can I actually change the showtime to the showtime you said? [ASSISTANT] Sure, you chose the showtime at 3:00pm. [USER] Okay that works better, can I change the time again. [ASSISTANT] Remaining showtimes for No Time To Die at AMC Mercado 24 are 4:30pm, 6:40pm and 9:45pm. [USER] Can I purchase two tickets to the showing of No Time To Die at AMC Mercado 24? [ASSISTANT] Sure, you would like two tickets for No Time To Die at AMC Mercado 24 at 3:00pm. Correct? [USER] Yes, please. [ASSISTANT] Okay. So to confirm you would like two tickets for No Time To Die at AMC Mercado 24 at 3:00pm tonight at the AMC Mercado 24 theater? [USER] That's right. [ASSISTANT] Is it OK to go ahead and purchase these tickets? [USER] Sure, let's do it. [ASSISTANT] OK. Your tickets are purchased and details for how to proceed have been sent to your email address. Can I help with anything else? [USER] Actually, let me think before I purchase the tickets since it will be a rough night at work. [ASSISTANT] I understand. Could you please detail how I got these tickets? [USER] Can I just send them to my email? [ASSISTANT] Sure. Let me see what I can find. [ASSISTANT] Unfortunately I cannot find any movie title that matches your description. [USER] Can I get the dealer's list of movie names from the AMC Mercado 24 website? [ASSISTANT] Showtimes for No Time To Die at AMC Mercado 24 are 4:30pm, 7:10pm, and 9:45pm. [USER] Ok, let me see what I can find. [ASSISTANT] Unfortunately I cannot find any movie title that matches your description. [USER] No I think we have to choose between No Time To Die and King Johnny please. [ASSISTANT] Got it. Let me see what I can find. [ASSISTANT] Unfortunately I cannot find any movie title that matches your description. [USER] Ok, well, I don't want to buy any tickets. [ASSISTANT] Showtimes for No Time To Die at AMC Mercado 24 are 3:00pm, 5:10pm, 7:30pm and 9:45pm. [USER] Can you fix something else? [ASSISTANT] That's it for now. [USER] Thanks. Could you actually order the tickets? [ASSISTANT] No need, I think we are done with this. [USER] Can you just order these? [ASSISTANT] Sure. I will be buying these tickets for you at AMC Mercado 24 at 3:00pm today. Is there anything else you would like to know? [USER] Can I make sure my debit card got charged again? [ASSISTANT] Sure. Let me see what I can find. [ASSISTANT] Unfortunately I cannot find any movie title that matches your description. [USER] Ok! No problem. [ASSISTANT] Showtimes for No Time To Die at AMC Mercado 24 are 4:30pm, 7:10pm, and 9:45pm. [USER] Ok great, I would like to make this change. [ASSISTANT] Is there anything I can help with now? [USER] Please, could you send the tickets to my email? [ASSISTANT] Sure. I am glad to help. Let me know what you need help with.

Figure 6: Example 2 of Time Control forced long text generation.

Method	Latent dim d	BERTScore		
		Precision	Recall	F1
LM	-	0.45	0.50	0.47
ILM	-	0.50	0.51	0.50
VAE	8	0.21	0.26	0.21
VAE	16	0.17	0.25	0.18
VAE	32	0.10	0.10	0.12
InfoNCE	8	0.22	0.29	0.23
InfoNCE	16	0.18	0.28	0.20
InfoNCE	32	0.20	0.28	0.21
Brownian	8	0.11	0.26	0.13
Brownian	16	0.08	0.27	0.10
Brownian	32	0.09	0.26	0.11
TC (Ours)	8	0.51	0.51	0.51
TC (Ours)	16	0.47	0.49	0.49
TC (Ours)	32	0.50	0.50	0.50

Table 17: BERTScore (Zhang et al., 2019) on ground truth infilled sentence and the generated sentence.