Finding and Generating a Missing Part for Story Completion

Yusuke Mori¹  Hiroaki Yamane²,¹  Yusuke Mukuta¹,²  Tatsuya Harada¹,²

¹The University of Tokyo, ²RIKEN
Background – Story Understanding/Generation

• Today, thanks to the Internet, anybody can freely publish their original stories.

• However, writing a story is not easy.

• To write/generate a good story, a human/an AI model has to know what a good story is.
  • Contrary, writing a story makes us understand more about the secret of creating a story.
• Story writers have various problems…
  • can’t complete one’s story
  • works don't sell well
  • balance with other work (if they are part-time writers)
  • etc.

• Recent progress in natural language processing makes it feasible to support human creative endeavors.

• To assist writers in creating stories, it is essential to train computers to understand and generate stories.
• Creation method is different for media, genre, and even for each creators. Above is just an example.
• Sometimes it goes back to the previous step and starts over.
• Creation method is different for media, genre, and even for each creators. Above is just an example.

• Sometimes it goes back to the previous step and starts over.
Objective

- **Finding a missing part in the flow of a story** to complete an incomplete story
  - For Story Understanding, Story Generation, and Creative Support\(^*1\)

- Note that the incompleteness of a story is not only “missing”, but also there are other kind of incompleteness.

- **Writing a story and supporting it is a complex and difficult task.**
  We tackled it first by focusing on one important step.

\(^*1\): We initially devised this task for creative support, but an anonymous reviewer kindly advised that our contribution should be told in the context of narrative understanding rather than creative support.
Challenge

• Tasks of completing incomplete stories
  • Story Ending Generation [Zhao et al., NLPCC 2018]
  • Story Completion [Wang and Wan, IJCAI 2019]

• These studies require a user to have prior knowledge of the missing parts. The case where the missing parts is not known was remain untouched.

• Our Missing Position Prediction aims to fill this gap.

“From Plots to Endings: A Reinforced Pointer Generator for Story Ending Generation” [Zhao et al., NLPCC 2018]
“Transformer-Based Conditioned VAE for Story Completion” [Wang and Wan, IJCAI 2019]
Main Contributions

• Propose “Missing Position Prediction (MPP)” task
  • Predicts the position of a missing part of an incomplete story; this has significance in the context of support for the creation of stories.

• Propose a novel method for Missing Position Prediction
  • An analysis shows that highly accurate predictions can be obtained when the missing part is the beginning or end.

• Story Completion with Missing Position Prediction
  • It is possible to restore a story comparable with the original human-written story in 26% of cases.
Related Task: Story Completion

• Wang and Wan [IJCAI 2019] proposed “Story Completion (SC)” task in the field of generating and understanding stories.

• Given any four sentences of a five-sentence story, the objective of SC is to generate the sentence that is not given.

Jennifer has a big exam tomorrow. She got so stressed, she pulled an all-nighter. She went into class the next day, weary as can be. _______________. Jennifer felt bittersweet about it.

Jennifer has a big exam tomorrow. She got so stressed, she pulled an all-nighter. She went into class the next day, weary as can be. Her teacher stated that the test is postponed for next week. Jennifer felt bittersweet about it.
Jennifer has a big exam tomorrow. She got so stressed, she pulled an all-nighter. She went into class the next day, weary as can be. Jennifer felt bittersweet about it.

• **Existing “Story Completion”**
  • The position of a missing part is given.

• **Our New Task**
  • Predicting the position of a missing part from a remaining context.
Ronda was at the fair craving something buttery. When she had her bag, she began taking bites. To her dismay this popcorn was sweet, which she hated. She ended up giving the popcorn to her daughter instead.

lowercased

ronda was at the fair craving something buttery. she was excited to eat some popcorn. when she had her bag, she began taking bites. to her dismay this popcorn was sweet, which she hated. she ended up giving the popcorn to her daughter instead.
Experiments

• **Experiment 1:**
  Missing Position Prediction
  • First, we worked only on the proposed task.
  • Investigated the part of the proposed method that excludes LM, to examine the performance of Context Encoder.

• **Experiment 2:**
  Missing Position Prediction + Story Completion
  • Based on the results of Experiment 1, we tackled the combined task of Missing Position Prediction and Story Completion.
  • Conducted Human Evaluation with Amazon Mechanical Turk (MTurk).
Experiments

• **Experiment 1: Missing Position Prediction**
  • First, worked only on the proposed task.
  • Investigated the part of the proposed method that excludes LM, to show the desired Context Encoder.

• **Experiment 2: Missing Position Prediction + Story Completion**
  • Based on the results of Experiment 1, we tackled both Missing Position Prediction and Story Completion.
  • Conducted Human Evaluation with AMT.
Experiment 1: Dataset

- ROCStories [Mostafazadeh et al., 2016]

<table>
<thead>
<tr>
<th>set</th>
<th>#stories</th>
<th>missing position</th>
</tr>
</thead>
<tbody>
<tr>
<td>train</td>
<td>78,528</td>
<td>Given randomly during training</td>
</tr>
<tr>
<td>dev</td>
<td>9,816</td>
<td>Given when creating dataset</td>
</tr>
<tr>
<td>test</td>
<td>9,817</td>
<td>Given when creating dataset</td>
</tr>
<tr>
<td>total</td>
<td>98,161</td>
<td></td>
</tr>
</tbody>
</table>

- It is widely used in story-generation tasks:
  - Story Ending Generation [Zhao et al., 2018; Li et al., 2018; Guan et al., 2019]
  - Story Completion [Wang and Wan, 2019]

- We randomly split the dataset in the ratio of 8:1:1 to obtain the train, dev, and test sets. We randomly removed one sentence from a story.

“A corpus and cloze evaluation for deeper understanding of commonsense stories” [Mostafazadeh et al., NAACL 2016]
Jennifer has a big exam tomorrow. She got so stressed, she pulled an all-nighter. She went into class the next day, weary as can be. Jennifer felt bittersweet about it.

**Model**
- **Sentence Encoder**
  - Sentence-BERT
    [Reimers & Gurevych, EMNLP 2019]
- **Context Encoder**
  - GRU Context
  - Max-pool Context (comparison method)

“Sentence-BERT: Sentence embeddings using Siamese BERT-networks”
[Reimers & Gurevych, EMNLP 2019]
Experiment 1: Results

• Overall Accuracy

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max-pool Context</td>
<td>35.0±0.334</td>
</tr>
<tr>
<td>GRU Context</td>
<td><strong>52.2±0.220</strong></td>
</tr>
</tbody>
</table>

• Prediction accuracy, shown as mean ± std. It is a five-class classification task, so the chance rate is 20%.
Experiment 1: Results

- Accuracy for each position
  - The performance is lower when $k = 2, 3, 4$ than when $k = 1, 5$.

Figure 3: Heat maps showing the results of the (a) GRU context and (b) Max-pool context. The ground truth (GT) label is shown on the x-axis and the predicted label is on the y-axis. The squares on the diagonal line denote correct cases. The ratios of the predicted label to the GT label are shown numerically.
Experiments

• **Experiment 1:**
  Missing Position Prediction
  • First, worked only on the proposed task.
  • Investigated the part of the proposed method that excludes LM, to show the desired Context Encoder.

• **Experiment 2:**
  Missing Position Prediction + Story Completion
  • Based on the results of Experiment 1, we tackled both Missing Position Prediction and Story Completion.
  • Conducted Human Evaluation with MTurk.
Experiment 2: Settings

• Dataset
  • ROCStories (same as Experiment 1)

• Model
  • Sentence Encoder
    • Sentence-BERT
  • Context Encoder
    • GRU Context
  • Language Model
    • BERT [Devlin et al., NAACL 2019]

“BERT: Pre-training of deep bidirectional transformers for language understanding” [Devlin et al., NAACL 2019]
Experiment 2: Human Evaluation

- Qualification Test
  - Choose workers with the ability in evaluating stories.
  - Used randomly selected ten questions from Story Cloze Test.

- Pair-wise Evaluation Task
  - The qualified workers were given two similar short stories and asked to choose which story gave the impression of being a complete story.
  - Five MTurk workers evaluated each story pair. The most popular answer of five workers was considered as an agreement.
Experiment 2: Human Evaluation

- We used 200 stories. The most frequently chosen answers by five workers were considered as their agreement.

- **Our proposed method can generate a story that is as good as or better than a GT story with 26% probability.**
  - Regarding a simpler task, SEG, we had conducted a similar pairwise evaluation [Mori et al, 2019]. The generated endings were as good as or better than GT with a 10.5% ratio then. Thus, we believe that the 26% for this more difficult task is noteworthy.

<table>
<thead>
<tr>
<th>Proposed</th>
<th>GT</th>
<th>both</th>
<th>neither</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>148</td>
<td>44</td>
<td>0</td>
</tr>
</tbody>
</table>

“Toward a better story end: Collecting human evaluation with reasons”
[Mori et al., INLG 2019]
### Experiment 2: Generation Example

<table>
<thead>
<tr>
<th>Context</th>
<th>GT</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>since the questions were complicated. i was extremely nervous. despite believing that i’ve failed, i turned the exam in. the teacher handed the exams back to us the next day. i ended up receiving a b.</td>
<td>i took my class final in math today. since the questions were complicated. i was extremely nervous. despite believing that i’ve failed, i turned the exam in. the teacher handed the exams back to us the next day. i ended up receiving a b.</td>
<td>my teacher gave us a test. since the questions were complicated. i was extremely nervous. despite believing that i’ve failed, i turned the exam in. the teacher handed the exams back to us the next day. i ended up receiving a b.</td>
</tr>
</tbody>
</table>

#### Answers with Reasons (A: GT, B: Ours)

<table>
<thead>
<tr>
<th>Reason</th>
<th>Context</th>
<th>GT</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>both both are complete sentences</td>
<td>tom was at a local park. there was an egg hunt for the kids. tom decided to pick some eggs up. he enjoyed the treats in them.</td>
<td>tom was at a local park. it was easter. there was an egg hunt for the kids. tom decided to pick some eggs up. he enjoyed the treats in them.</td>
<td>tom was at a local park. there was an egg hunt for the kids. tom decided to pick some eggs up. tom was able to get many eggs. he enjoyed the treats in them.</td>
</tr>
<tr>
<td>neither both doesn’t make sense</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ours A is jumbled and does not make sense. B is logically arranged as a story.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ours In &quot;A,&quot; it wouldn’t make sense that a final exam was handed back in class the next day.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ours B was more appropriate since it is having a continuous flow than A</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Experiment 2: Generation Example

<table>
<thead>
<tr>
<th>Context</th>
<th>timothy loved to dance. timothy didn’t have much confidence in himself. it took everything he had to dance with all of his self doubt. everyone loved his dancing and he won the contest.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT</td>
<td>timothy loved to dance. there was a dance contest that was coming up soon. timothy didn’t have much confidence in himself. it took everything he had to dance with all of his self doubt. everyone loved his dancing and he won the contest.</td>
</tr>
<tr>
<td>Ours</td>
<td>timothy loved to dance. he decided to take dance lessons. timothy didn’t have much confidence in himself. it took everything he had to dance with all of his self doubt. everyone loved his dancing and he won the contest.</td>
</tr>
</tbody>
</table>

### Answers with Reasons (A: Ours, B: GT)

| GT | Only B makes sense and a complete story. |
| GT | A is more correct and arranged            |
| GT | Story A doesn’t mention the contest which Timothy ends up winning, therefore misses an important piece of the story. |
| GT | Story B mentions that there was a dance contest at the start and that he won it at the end. Story A only mentions a contest abruptly at the end making it seem out of place. |
| GT | B is more good                                   |
Discussion

• An analysis showed that highly accurate predictions can be obtained when the missing part is the beginning or the end.
  • It suggests that if a story has a specific beginning and end, they play important roles.
Conclusion

• Proposed “Missing Position Prediction” and a method for it.
• Examined the prediction accuracy for each missing position and found that a prediction is easier, if the beginning or the end of a story is missing.
• Tackled the combined task of missing position prediction and story completion. In human evaluation, our proposed method showed promising results.
• We believe that our proposed task would serve as a new test for models’ abilities for understanding the flow of a story. It is also an important step towards human creativity assistance.
Thank you!

- Our code is publicly available at https://github.com/mil-tokyo/missing-position-prediction

- Contact
  - Yusuke Mori (corresponding author)
    - mori [at] mi.t.u-tokyo.ac.jp