Consistency-Enhanced Story Generation

Team 39

Bhavani Chalasani 2022101014

Nishita Kannan 2022101112

Archit Narwadkar 2022111032

Problem Statement

- Story generation by pretrained architectures tend to struggle with maintaining consistency and coherency.
- The referred paper uses a two-stage story generation model that first generates an outline of the story from a given prompt and then by using this outline to generate a story.
- We aimed to implement this model and analyze its performance using the recommended evaluation metrics.

Datasets Used

- Book 8 - This is a discourse marker dataset that was used to fine tune a pre-trained BERT model used for the discourse coherency task.

- Writing Prompts - This is the dataset that was used to train and test both the transformers, as well as the entire pipeline. It consists of both prompts and target stories.

Approach

As per the paper, we have implemented four different parts to make the entire model:

- First transformer - this transformer generates outlines, given a prompt as an input.
- Second transformer - this transformer generates stories, given an outline as input.
- Discourse coherency model - this consists of building a model (encoder with two layer MLP) that can predict the connector between two sentences, in order to tell us how coherent the generated stories are.
- Coreference consistency - this consists of generating coreference chains and coreference clusters in order to understand how consistent the generated stories are.

Implementation - First Transformer

- The first transformer was trained using the input prompts from the writingprompts dataset, and was compared to an outline generated using the TextRank and RAKE algorithms.

- We have implemented the transformer both from scratch, and from a config of the GPT2LMHead model.

- We encoded the inputs, passed them through the transformer, compared the generated outputs to the expected targets, and updated the model.

Implementation - Second Transformer

- The second transformer was trained using the input prompts from the writingprompts dataset and the outline generated using the TextRank and RAKE algorithms, and was compared to the stories in the writingprompts dataset.

- We have implemented the transformer both from scratch, and from a config of the GPT2LMHead model.

- We encoded the inputs, passed them through the transformer, compared the generated outputs to the expected targets, and updated the model using a hybrid loss.

- The hybrid loss used the loss generated by the discourse coherency and coreference consistency elements to ensure that the generated story is both coherent and consistent.

Implementation - Discourse Coherency Model

- We trained a discourse coherency model (encoder with two layer MLP) along with the second transformer.
- This discourse coherency model takes two consecutive sentences from the generated story, predicts a discourse marker between the two sentences, and compares it with the expected marker.
- It then generates a loss based on the predicted marker, which is used while updating the second transformer.
- The expected marker is generated by the GoldenBERT model, which is fine tuned using the Book 8 dataset.

Implementation - Coreference Consistency

- The generated story is passed through the coreference_loss function, which returns a score related to the consistency of the story.

- This loss is then used to update the second transformer.

- The function generates coreference clusters and uses the mentions from each cluster to generate an average score, which represents how well words in the generated story relate to each other and maintain consistency.

Evaluation Metrics

We have analyzed the performance of the model based on the following evaluation metrics:

- Perplexity
- Distinct-1 an Distinct-2 scores
- Averaged coreference chains
- Percentage of unknown labels

Analysis

First Transformer

Method	Perplexity	Distinct-1 Score (%)	Distinct-2 Score (%)
From Keyword	293.083	0.1083	1.5276
From Abstract	209.794	0.1025	1.7136

Analysis

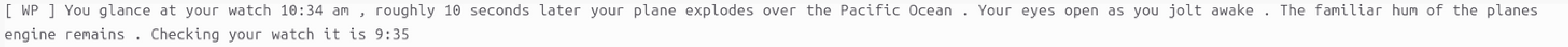
Second Transformer

Method	Perplexity	Distinct-1 Score (%)	Distinct-2 Score (%)	Unknown (%)	Coref Chains
From Keyword	720.436	0.333	0.335	50	0.23
From Abstract	651.756	0.312	0.238	50	0.37

Outputs

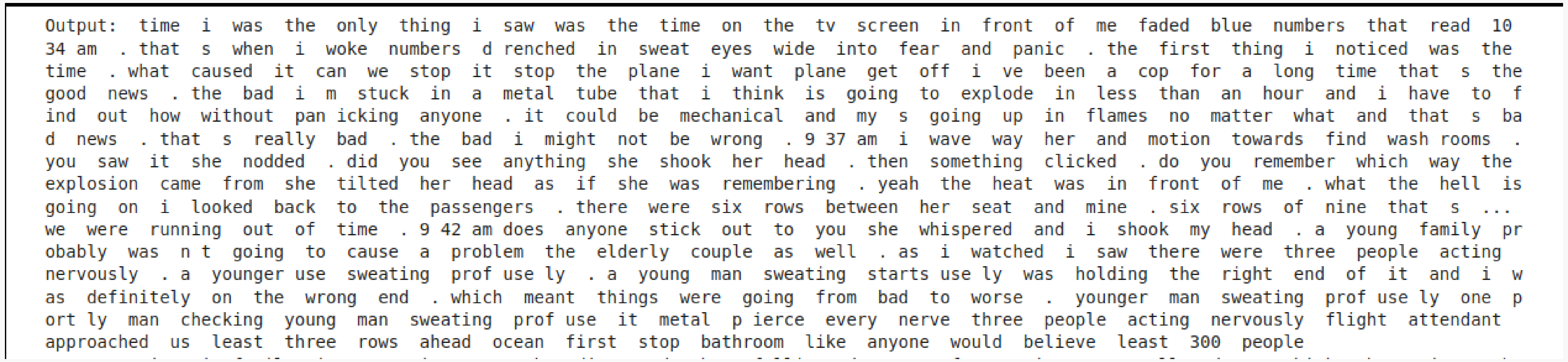
First Transformer

Input Prompt



```
[ WP ] You glance at your watch 10:34 am , roughly 10 seconds later your plane explodes over the Pacific Ocean . Your eyes open as you jolt awake . The familiar hum of the planes
engine remains . Checking your watch it is 9:35
```

Generated Outline

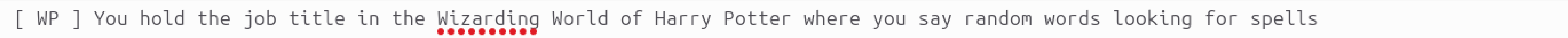


```
Output: time i was the only thing i saw was the time on the tv screen in front of me faded blue numbers that read 10
34 am . that s when i woke numbers d renched in sweat eyes wide into fear and panic . the first thing i noticed was the
time . what caused it can we stop it stop the plane i want plane get off i ve been a cop for a long time that s the
good news . the bad i m stuck in a metal tube that i think is going to explode in less than an hour and i have to f
ind out how without pan icking anyone . it could be mechanical and my s going up in flames no matter what and that s ba
d news . that s really bad . the bad i might not be wrong . 9 37 am i wave way her and motion towards find wash rooms .
you saw it she nodded . did you see anything she shook her head . then something clicked . do you remember which way the
explosion came from she tilted her head as if she was remembering . yeah the heat was in front of me . what the hell is
going on i looked back to the passengers . there were six rows between her seat and mine . six rows of nine that s ...
we were running out of time . 9 42 am does anyone stick out to you she whispered and i shook my head . a young family pr
obably was n t going to cause a problem the elderly couple as well . as i watched i saw there were three people acting
nervously . a younger use sweating prof use ly . a young man sweating starts use ly was holding the right end of it and i w
as definitely on the wrong end . which meant things were going from bad to worse . younger man sweating prof use ly one p
ort ly man checking young man sweating prof use it metal p ierce every nerve three people acting nervously flight attendant
approached us least three rows ahead ocean first stop bathroom like anyone would believe least 300 people
```

Outputs

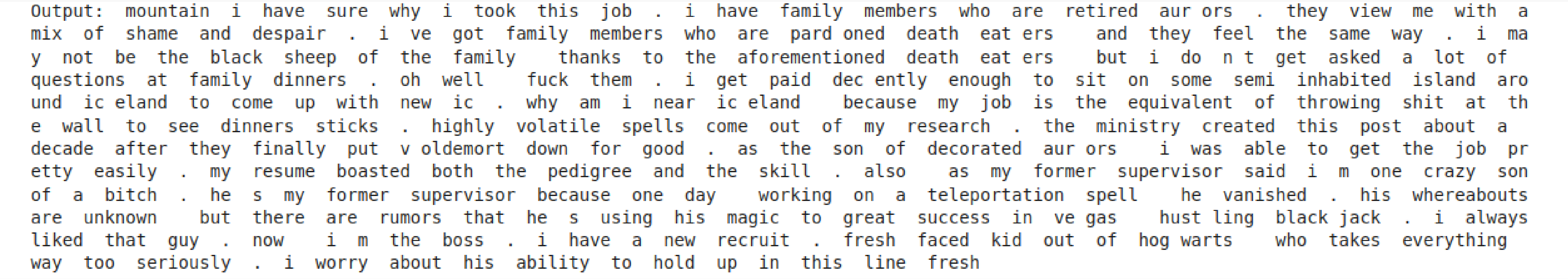
Second Transformer

Input Prompt



```
[ WP ] You hold the job title in the Wizarding World of Harry Potter where you say random words looking for spells
```

Generated Story



```
Output:  mountain  i  have  sure  why  i  took  this  job  .  i  have  family  members  who  are  retired  aur ors  .  they  view  me  with  a
mix  of  shame  and  despair  .  i  ve  got  family  members  who  are  pard oned  death  eat ers    and  they  feel  the  same  way  .  i  ma
y  not  be  the  black  sheep  of  the  family    thanks  to  the  aforementioned  death  eat ers    but  i  do  n t  get  asked  a  lot  of
questions  at  family  dinners  .  oh  well    fuck  them  .  i  get  paid  dec ently  enough  to  sit  on  some  semi  inhabited  island  aro
und  ic eland  to  come  up  with  new  ic  .  why  am  i  near  ic eland    because  my  job  is  the  equivalent  of  throwing  shit  at  th
e  wall  to  see  dinners  sticks  .  highly  volatile  spells  come  out  of  my  research  .  the  ministry  created  this  post  about  a
decade  after  they  finally  put  v oldemort  down  for  good  .  as  the  son  of  decorated  aur ors    i  was  able  to  get  the  job  pr
etty  easily  .  my  resume  boasted  both  the  pedigree  and  the  skill  .  also    as  my  former  supervisor  said  i  m  one  crazy  son
of  a  bitch  .  he  s  my  former  supervisor  because  one  day    working  on  a  teleportation  spell    he  vanished  .  his  whereabouts
are  unknown    but  there  are  rumors  that  he  s  using  his  magic  to  great  success  in  ve gas    hust ling  black jack  .  i  always
liked  that  guy  .  now    i  m  the  boss  .  i  have  a  new  recruit  .  fresh  faced  kid  out  of  hog warts    who  takes  everything
way  too  seriously  .  i  worry  about  his  ability  to  hold  up  in  this  line  fresh
```

Thank you