



De-summarization

Saad Obaid and Patrícia Schmidtová

May 27, 2021

Project objective:

Use de-summarization to generate summaries of theatre plays, acts or entire plays.





Phases of the project

Training on readily available data

Use what is available

- Pre-trained models
- Reversed summarization datasets

Fine-tuning on domain data

Make it sound like a story

- Take the models from the first phase
- Theatre data
- Movie and TV show data

Evaluation

Is the output good?

- Automatic metrics
 - deterministic
 - non-deterministic
- Human evaluation



Models

- Pegasus
- T5
- Bart
- GPT-2



Data

- De-summarization in general
 - Xsum
- Shorter summaries
 - Wikipedia summaries
 - Imdb summaries
- Shorter and longer summaries
 - Fanwikis of TV shows
 - Sparknotes
- Scripts
 - Fan-transcribed scripts
 - Theatre plays with expired author's rights



Fine-tuning

So far, we

- managed to produce plots if given a prompt
- tried to use the outlines to produce the plays directly

We aim to try:

- producing longer summaries
- using the longer summary to train another script model



GPT2 examples

Murder in the Night street

In a small town in the south of England, a group of young men, led by Mr. and Mrs. Witherspoon, are preparing to rob a bank. They are interrupted by the arrival of the police, who arrive to arrest them. The police are reluctant to let them go, since they believe that they are part of a gang, and that they may be involved in a murder. However, they are able to convince the police to allow them to go free.

The haunted house mystery

The film is set in a haunted house in Los Angeles, California. The film begins with a young woman, Laura, waking up in the middle of the night to find her dead body in the living room of her house. She is covered in blood, and her screams are heard by a disembodied voice. The disembodied voice tells Laura that she is possessed by the spirit of her deceased mother, and that she must find a way to stop the spirit. Laura and the disembodied voice go to the haunted house, where Laura finds her mother's body in the living room.



Pegasus examples

Murder in the Night street

During the reign of Queen Elizabeth I, England is concerned by the impending arrival of the Spanish Armada. In 1588, relations between Spain and England are at breaking point. Elizabeth's chief advisers are the Lord Treasurer, Lord Burleigh, and her longtime admirer, Robert Dudley, Earl of Leicester. Burleigh's 18-year-old granddaughter Cynthia is one of Elizabeth's ladies-in-waiting, and the ageing queen is plagued by jealousy of the girl's beauty and vivacity.

Jan is raising money for Tonda's Butterfly Disease, or how will it end?

As the UK prepares to vote on whether or not to remain in the European Union, BBC News takes a look at some of the key stories that will have an impact on the future of the country. In the UK, there are more than 60 different butterfly species, but only a handful can be found in the UK. One of those is Tonda, who has been dying of the disease for more than 30 years. Her body has been destroyed by the disease, and she is now being cared for by her daughter, Lisa, who is also suffering from the same disease.



Bart examples

Three and two is five

The contest is set to take place in New York City. Three and two are expected to win a trip to Hollywood to take part in the even bigger "Miss Drag Queen of America Pageant." Before they depart, the organizers are asked to write love letters to Hollywood stars for Rajan a teenage boy who also has a long-standing crush on Mel Gibson.

While on the road they are pulled over by the racist, homophobic, and sexist Sheriff Dollard.

An empty bag and a dead Eve in the Tunnel

The Tunnel Troll has been searching for a new identity since it was pulled over by the racist, homophobic, and sexist Sheriff Dollard. He discovers the truth behind the infamous "drag princess" incident in the tunnel.

While searching for the new identity of the drag queen, Dollard is ridiculed for her actions by her colleagues, who believe she was a drag queen.



Bad Example

The film follows two young friends, Kiran, a Hindu, and Delilah, a Catholic, from their first meeting as young children to young adulthood, when they realize their love for each other. At first, **Kirans** is asked to write love letters to **Delilahs** for Rajan a teenage boy who also has a long-standing crush on **Delilaha**. Kiran does so as it allows her to express her love to Delillah without having to be ostracized by her family, friends and culture. Eventually **Delilahu** discovers the truth behind the letters and poetry, and admits her mutual love to Kiran. This begins a delicate love affair.



Prompt = “Lord of the Rings”

The future of civilization rests in the fate of the One Ring, which has been lost for centuries. Powerful forces are unrelenting in their search for it. But fate has placed it in the hands of a young Hobbit named Frodo Baggins, who inherits the Ring and steps into legend. A daunting task lies ahead for Frodo when he becomes the Ringbearer - to destroy the One Ring in the fires of Mount Doom where it was forged.

The film follows two young friends, Kiran, a Hindu, and Delilah, a Catholic, from their first meeting as young children to young adulthood, when they realize their love for each other. At first, **Kirans** is asked to write love letters to **Delilahs** for Rajan a teenage boy who also has a long-standing crush on **Delilaha**. Kiran does so as it allows her to express her love to Delillah without having to be ostracized by her family, friends and culture. Eventually **Delilahu** discovers the truth behind the letters and poetry, and admits her mutual love to Kiran. This begins a delicate love affair.



ROUGE Evaluation

	Rouge-1			Rouge-2			Rouge-L		
	P	R	F	P	R	F	P	R	F
GPT-2	26.65	30.83	22.12	4.92	4.96	3.80	21.005	19.50	16.47
PEGASUS	26.70	6.04	8.35	3.10	0.50	0.76	19.86	5.32	7.39
BART	32.14	13.30	15.44	3.31	1.00	1.29	21.98	10.63	12.34



Things we tried and did not work

- T5 finetuning
 - Adafactor optimizer, recommended parameters
 - We tried both T5-base and T5-small
- GPT2 sentence of a summary -> scene
 - Trained on scene-wise summaries to scene scripts
 - Will hopefully improve with more robust preprocessing and better data selection
 - Also possible to use two sentences
- Bart-base
 - Outputs comparable to distilBart, but much longer training times

Questions?

