

# FairySum

Omar Bayoumi

Sapienza University of Rome

Elective in Artificial Intelligence - Narrative Understanding And Storytelling (NUANS)

Minihomework 2

bayoumi.1747042@studenti.uniroma1.it

## 1 HUMAN-HUMAN comparison

In the story **bn:00173084n***The Old Dame and her hen* I noticed that the dialogue *'Your hen trips inside the hill'* was often included, often foreshadowing the death of one of the sisters. However, I found this dialogue not very useful since it is always the same and a potential model would go to learn dialogue like this instead of a description like *'She thought this strange, so she too went to see what it was, and fell through the trap-door too, deep, deep down, into a vault.'* to describe this event.

The main problem I had is with the **introduction to the characters** which were often placed in somewhat intuitive sentences and that is why they were not included but that is not why they should be considered useless parts because if that name appears later in the text, you have no reference to the character.

One thing I noticed is that in general there are **more ways to get the same summary result** such as, again in the same story, the sentence *'66: But you must mind and not look into the sack.'* and *'67: Well! he said he would carry the sack; and he said, too, that he would not look into it; but when he had gone a little way, he thought the sack got awfully heavy; and when he had gone a bit farther he said to himself- "Come what will, I must see what's inside this sack, for however sharp her eyes may be, she can't see me all this way off."* can express the same concept, in this case, the fact of not looking inside the sack. In these cases, my choice has always been to take the shorter sentence.

This is an example of consecutive sentences, however, it also happened with **non-consecutive sentences** such as in the story **bn:03214242n***Polaris*: *'27: It was my friend Alos who spoke, and his speech was one that pleased my soul, for it was the speech of a true man and*

*patriot.'* and *'31: Alos, my friend, was commander of all the forces of the plateau, and in him lay the last hope of our country.'* In this case, two different concepts are expressed however the character is introduced so that it is possible to exclude one of the two if I want to introduce the character of *Alos*.

Other types of phrases that I omitted but others put in were some **descriptive** phrases that however in my opinion should not be added in a summary. Since I am not the best at summaries, I realized only after reading the differences with others that some sentences, instead, were inconsistent and either needed to be removed or had nothing to do with the discourse.

I also received a summary from a student that I suspect was not made by him, but by a generator, or else his intent was to rewrite the story of **bn:03214242n***Polaris*. Let me explain... in this received summary many of the events related to the dream he had in the story were completely omitted except for two sentences, one in the middle of the story and one at the end. The two randomly placed sentences make me suspect an automatic tool, however, perhaps his idea was to reinterpret the story in such a way that the dream was not an important part of the story but the fact of sleeping looking at the stars, and waking up reflecting on this dream that was completely removed from the story and thus forgotten. This way of thinking actually sums up the story to the bone, and from one point of view could be seen as a very good summary or a very bad summary since you cut out the whole middle.

## 2 HUMAN-SYSTEM comparison

### 2.1 Model used

The model used is a **Sentence Transformer** with a checkpoint at *'all-MiniLM-L6-v2'* with which an embedding is created from which with **cosine similarity** and **LexRank** it is possible to extract the n

most central sentences in the discourse, understood as of importance [Reimers and Gurevych (2019)]. The number of sentences to be generated is equal to the number of sentences reconstructed from an abstractive summary generated by **Bert Summarizer** with a checkpoint at 'paraphrase-MiniLM-L6-v2' [Miller (2019)].

## 2.2 Scores

The scores used are **Rouge1, Rouge2, RougeL, RougeLsum** to compute the similarity between the n-grams of the sentences. A summary was generated with the same number of hand-generated summaries by me and the other students, and then these scores were computed considering the hand-generated summary as a prediction and the model-generated one as a reference.

The results were very low, although that of the other students were better than mine [Table 1]. As mentioned in the section *HUMAN-HUMAN comparison*, one of my problems was to include sentences that could be completely eliminated. In fact, these are not present within the model-generated summaries and as a result, many points in the scores are lost there.

## 2.3 Manual analyses

I also performed a manual analysis of the summaries generated by the system **automatically** (without choosing the number of sentences, *Model used*).

I noticed that the automatic model **does not make optimal summaries with non-long stories**. In fact, very often, some sentences are cut off which makes the text unintelligible, and without them, it is not clear who is being talked about. For example, these two consecutive sentences are generated: '15: *Then the Man o' the Hill got so angry that he took her up and wrung her head off, and threw both head and trunk down into the cellar.*' and '22: *There she went from room to room, and in the innermost one the Man o' the Hill came to her and asked if she would be his sweetheart?*'. The first sentence mentions a she who dies and the second sentence mentions a she who goes from room to room but who is the second she?

In contrast on **long stories**, such as **bn:02439324n\_Micromégas**, it performs much better in my opinion, perhaps because the model is trained on newspaper articles and Reddit comments and thus perhaps is not optimized

enough for generating summaries based on non-long, narrative stories.

## 3 Conclusion

In conclusion, my summaries were **slightly shorter** than those of the other students despite having lower Rouge scores [Table 1]. However, I do not feel that I made worse summaries because most likely the scores were lowered because of the "extra", without context, sentences I could have removed, and the sub-optimal performance of the model itself.

| My summaries             |         |         |         |            |
|--------------------------|---------|---------|---------|------------|
| Story id                 | Rouge 1 | Rouge 2 | Rouge L | Rouge Lsum |
| bn_00173084n             | 0.29    | 0.14    | 0.23    | 0.23       |
| bn_02439324n             | 0.16    | 0.05    | 0.13    | 0.13       |
| bn_03214242n             | 0.26    | 0.12    | 0.23    | 0.23       |
| Other students summaries |         |         |         |            |
| Story id                 | Rouge 1 | Rouge 2 | Rouge L | Rouge Lsum |
| bn_00173084n             | 0.35    | 0.21    | 0.30    | 0.30       |
|                          | 0.30    | 0.16    | 0.25    | 0.25       |
| bn_03214242n             | 0.39    | 0.27    | 0.37    | 0.36       |
|                          | 0.33    | 0.20    | 0.30    | 0.30       |

Table 1: All the Rouge scores.

## References

- Derek Miller. 2019. [Leveraging bert for extractive text summarization on lectures](#).
- Nils Reimers and Iryna Gurevych. 2019. [Sentence-BERT: Sentence embeddings using Siamese BERT-networks](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.