

Tackling hallucinations in neural chart-to-text generation

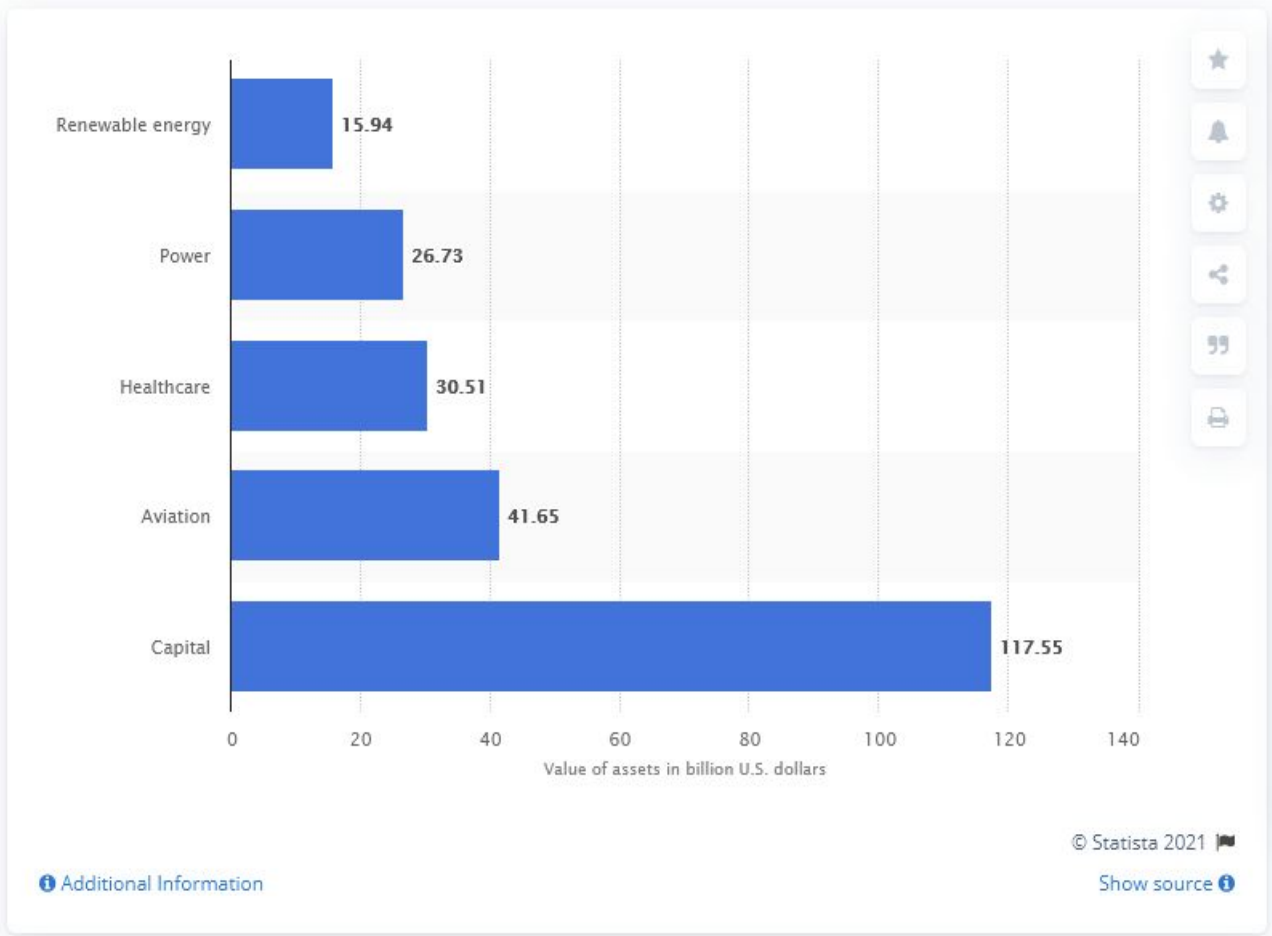
Saad Obaid ul Islam | saadob12@gmail.com

Overview

- In this poster, I present ongoing work that I am doing for my thesis.
- In the first section we look at the task at hand and the problem of hallucination. In the second and third section we look at previous work and datasets related to the task. In third, fourth, fifth section we discuss our approach to solve the problem, report and discuss the results.

Task and Problem

Chart-to-text is the task of generating textual descriptions of informations visualizations like bar chart, line chart, etc. This task can be modelled as an image captioning task and data-to-text task. We model this as a data-to-text task where the underline table of the chart is given.



Segment	Value of assets in billion U.S. dollars
Renewable energy	15.94
Power	26.73
Healthcare	30.51
Aviation	41.65
Capital	117.55

Title: General Electric 's total assets in FY 2019 , by segment (in billion U.S. dollars)

This statistic represents General Electric 's total assets in the fiscal year of 2019 , with a breakdown by segment . In its healthcare segment , the company had assets to the value of around 30.5 billion U.S. dollars .

Chart-to-text comes under the umbrella of natural language generation (NLG). There is a common problem found in NLG models that they often hallucinate. Hallucinations in NLG means generating unfaithful or meaningless text. Causes of hallucinations are source-reference divergence, training-modelling choices, and decoding strategies.

Related work

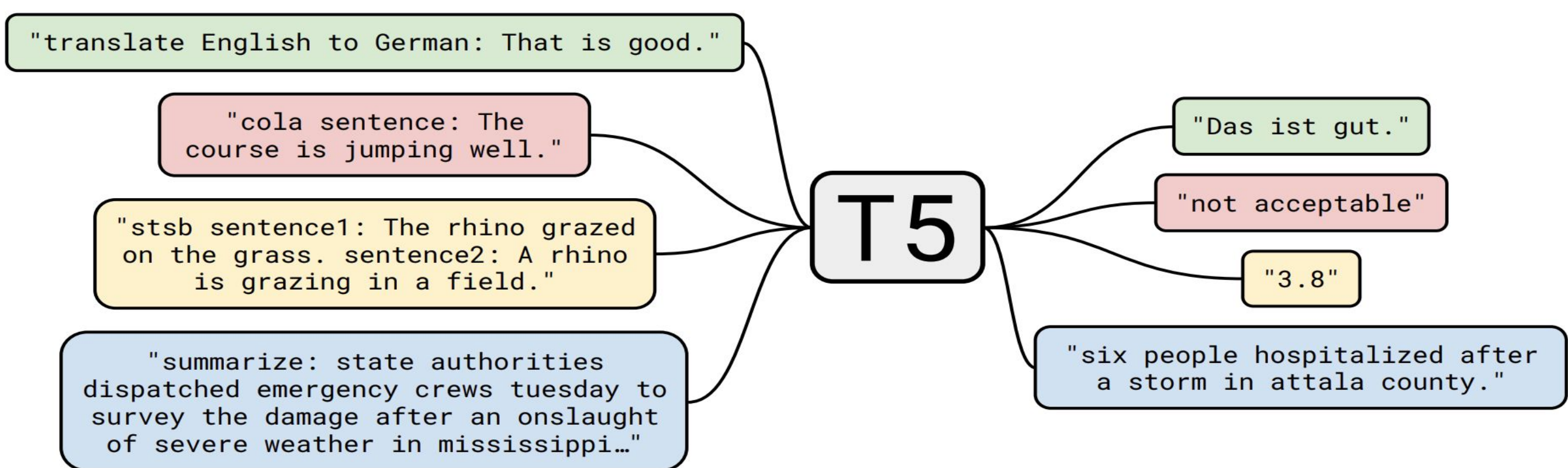
- Obeid et al., (2020) created chart-data-summarization dataset from statista.com. They adapted a data-to-text transformer model proposed by Gong et al., (2019) to chart summarization task by modifying the input embedding and introducing a data variable substitution method to tackle hallucinations. The model is trained on linearized table with x-y labels and modified delexicalized summaries which are later lexicalized after inference using data variable substitution.
- Kanthara el., (2022) created a chart-data-summarization dataset from statista.com. They experimented on Obeid et al, (2020) model, field infusion model, T5, and BART. The models are trained on linearized table with the title and summaries.

Datasets

- Chart2Text dataset by Obeid et al., 2020.
 - 8147 chart-data-summary examples.
 - Training: 5702, Validation: 1222, Test: 1222
- Chart-to-text dataset by Kanthara et al., 2022.
 - 34811 chart-data-summary examples.
 - Training: 24367, Validations: 5222, Test: 5222

Approach

- Improve linearized input representation of the table.
 - <title>, <x-y label(s)> , <x-y values>
- Append title and x-y labels to the input representation.



- Leverage pre-trained transformers like T5 to generate summaries.
- Evaluate model on BLEU, BLEURT, Perplexity and NUBIA.

Results

Dataset	Model	BLEU Score ↑	BLEURT ↑	Perplexity ↓
Chart2Text dataset by Obeid et al., 2020	Obeid et al., transformer	18.54	-	-
	Our T5	23.8	-	-
Chart-to-text dataset by Kanthara et al., 2022	Kanthara et al., T5	37.01	0.15	10.0
	Our T5*	39.82	0.17	8.27

*We fine-tune T5 for 15 epochs as compared to Kanthara el al., who fine-tune T5 for 200 epochs.

Dataset	Semantic Similarity ↑	Logical Agreement ↑	Contradiction ↓	Neutrality	NUBIA ↑
Chart2Text dataset by Obeid et al., 2020	3.5/5	33.27	22.31	44.41	46.9
Chart-to-text dataset by Kanthara et al., 2022	3.68/5	39.34	21.36	39.28	55.67

Chart	References	Hypothesis
	This statistic depicts the operating profit of the H & M Group worldwide from 2009 to 2019 . In 2019 , the global operating profit of the H & M Group was about 1.8 billion U.S. dollars.H & M is a leading global fashion company with strong values and a clear business concept . H & M constantly strives to have the best customer offering in each individual market – which includes giving customers the best price .	This statistic shows the operating profit of the H & M Group worldwide from 2009 to 2019. In 2019, the global operating profit of the H & M Group amounted to approximately 1.81 billion U.S. dollars. H & M is a leading global fashion company with strong values and a clear business concept.
	Lee Min-ho was ranked the most popular South Korean celebrity on Facebook , with around 17.9 million fans as of September 2020 . Min-ho is a South Korean actor and singer , initially gaining attention in 2009 with his role as Gu Jun-pyo in ‘ Boys Over Flowers ’ . It comes as no surprise that the actor has so many fans , with the South Korean film industry booming not only locally , but on a global scale as well .	This statistic presents a ranking of the most popular celebrities on Facebook from South Korea as of September 2020. According to the source, Lee Minho was the most popular celebrity on the South Korean social network with around 17.87 million Facebook fans. BTS and BIGBANG followed behind with approximately 12.13 and 10.63 million fans respectively.

Discussion

- Adding more chart related information in the input representation improves n-gram precision, BLEURT, and perplexity.
- Our BLEU score is almost 3 points higher than BLEU score obtained by Kanthara et al., 2022 even though we trained for much fewer epochs. Improving input representation drastically reduces training time as well.
- To check faithfulness, we compute NUBIA which computes entailment and semantic similarity. Low contradiction and high agreement score indicates that most reference text is entailed in the hypothesis.
- We also tested various sampling decoding methods like top-k and nucleus. Sampling decoding resulted with a lower BLEU and BLEURT score compared to search decoding.
- During our error analysis, we observed that lots of reference summaries in the training data contained text that can be categorized as extrinsic hallucination meaning that such text cannot be verified from the chart. It only provides outside information. We hypothesize that these extrinsic hallucinations only act as a noise and leads the model towards generating unfaithful text.

References

- Obeid, Jason, and Enamul Hoque. "Chart-to-text: Generating natural language descriptions for charts by adapting the transformer model." arXiv preprint arXiv:2010.09142 (2020).
- Kanthara, Shankar, et al. "Chart-to-Text: A Large-Scale Benchmark for Chart Summarization." arXiv preprint arXiv:2203.06486 (2022).
- Raffel, Colin, et al. "Exploring the limits of transfer learning with a unified text-to-text transformer." J. Mach. Learn. Res. 21.140 (2020): 1-67.
- Papineni, Kishore, et al. "Bleu: a method for automatic evaluation of machine translation." Proceedings of the 40th annual meeting of the Association for Computational Linguistics. 2002.
- Sellam, Thibault, Dipanjan Das, and Ankur P. Parikh. "BLEURT: Learning robust metrics for text generation." arXiv preprint arXiv:2004.04696 (2020).
- Kane, Hassan, et al. "NUBIA: NeUral based interchangeability assessor for text generation." arXiv preprint arXiv:2004.14667 (2020).



UNIVERZITA KARLOVA



UNIVERSITÄT
DES
SAARLANDES



With the support of the
Erasmus+ Programme
of the European Union

