Assignment 2: Generation Evaluation Metrics

Due Date: April 29

In this assignment we will analyze the state-of-the-art evaluation metrics for NLG, especially those that check consistency.

1. [15pts] Check the code from TACL2021 paper: “SummaC: Re-Visiting NLI-based Models for Inconsistency Detection in Summarization”

<https://github.com/tingofurro/summac>

Run the code on the data given in this repository, and point out at least three categories that this metric fails to capture, and give examples for each category. Then propose concrete solutions on addressing the remaining challenges. Page limit: 1-4 pages.

2. [Up to 5pts] Implement one of your proposed solutions. Please note that if you are already doing generation metric related term projects, you should propose different ideas from your term projects.