**Conclusions**

- We propose C2F-StoryTree as a cost-effective, controllable, and transparent common summary writing method. Compared to the typical story generation framework, our method is more controllable and transparent and can lead to high-quality stories that align well with the user’s interest.
- We propose a cost-effective framework for realizing the entailment hierarchy, a text data structure for realizing the framework, using human-in-the-loop and in the loop and existing public datasets in the creative writing domain.
- We propose a new common summary writing task, which avoids the resulting summary/primitive from being too specific or general. In the task, we also measure the similarity of the propensity of the summary using the similarity of the two input stories, which allows us to test the performance of different models given different prompt specifications.
- We propose a novel method to generate the entailment hierarchy by summarizing two stories using joint decoding and reranker. We demonstrate that our method is much better than GPT-3.5 few-shot and GPT-3.5 refers to text-davinci-003 and GPT-3.5 refers to davinci.
- Our model allows us to build a high-quality 2-layer entailment hierarchy on top of 5-city summaries and stories.

**Figures**

1. The evaluation of C2F-StoryTree in the work. (a) We construct the entailment hierarchy (EH) by leveraging existing resources and machines in the-loop techniques. After fine-tuning the models using the corresponding CEL, we generate summaries for ROC and WritingPrompts (WP). (b) We train seq2seq models to generate lower-layer text and utilize the coarse-to-fine (C2F) generation.

2. General Summary/Hypothesis as the Input Prompt: “A girl achieved something impressive.” Short Story from CG (Conditional Generation): “A girl was very tall with a long tender tone. But she wanted to do something with it. Her parents bought her a tender tone.” The girl sat on the stool and stared at it. She had finally completed something that would be impressive to her mother.”

3. Table 1: An example generated by our C2F-StoryTree framework and the conditional generation baseline. We highlight the selected text for generating more specific text. RS refers to the reranker scores.

4. Figure 4: (a) Comparison of options for constructing the entailment hierarchy. (b) The task and our model of generating a common summary. We input story A, story B, and their concatenation into a summarization model and during the decoding time, we merge their generation probabilities word by word in order to get the common summary of both story A and story B.

5. Figure 5: sBERT similarity between story A and B is a city measurement of their common summary. The model is a classifier that takes two stories as input and outputs a common summary. We input story A, story B, and their concatenation into a summarization model and during the decoding time, we merge their generation probabilities word by word in order to get the common summary of both story A and story B.

6. Figure 6: The linear trendlines of the short story scores from different LMs given the prompt with different specificity.