

App ID #: 898

Mentor: Bunescu, Razvan

Email: rbunescu@uncc.edu

Title: Associate Professor

Department: Computer Science

Co-mentor: No

Community engaged research: No

Title: AI Models for Nonlinear Rewriting of Narratives

Description: Stories are often told in a non-linear fashion, in order to increase suspense, emotional involvement, engagement, and overall enjoyment. This is done across the various mediums used to tell the story, ranging from text (novels) to audio (radio, podcasts), to audio-visual (movies and TV series). However, the deadline-driven environment where creators need to generate new content at a very high pace leads to a non-linear storytelling that feels rushed, that detracts from understanding and enjoying the story as a whole, and that does not optimize the true potential of the non-linear technique. Additionally, readers and audiences enjoy different levels of non-linearity. In this context, we expect that a tool that can optimize non-linear story telling would have significant impact. However, for lack of access to the original linear version of a story, a pre-requisite task is that of taking as input a story and creating a linear version of it.

In this project, we propose to (1) develop LLM-based narrative linearization methods that take as input the text of a story and produce as output a linearized version of it. (2) models for nonlinear rewriting of narratives starting from their linear versions.

Looking for: Only 160 hours over an academic semester (~10h/wk)

of positions available: 1

Anticipated Student Learning Outcomes: A successful completion of this project would consist at a minimum in a dataset of manually linearized stories and a prototype system that is evaluated on this dataset. In terms of technical approaches, we plan to use LLMs provided through an MS Research Azure grant, in few-shot and possibly fine-tuning scenarios.

Upon successful completion of research, the student will have working knowledge of how to use and even fine tune LLMs through APIs, how to formulate a natural language processing problem, how to create a dataset with examples for fine-tuning and evaluation, how to run empirical comparisons between different models on the same task. These skills are highly sought after in both research and industry settings.

Training Description: No training required.

Training hours: 0

Mentoring plan: We will have weekly in-person meetings (faculty-student), with meeting reports submitted on <https://cci-git.charlotte.edu/>.

Outline of the tasks:

Literature search and documentation of existing approaches and state of the art.

Manual annotation of dataset of short stories or fragments from stories or TV show transcripts.

Annotation guidelines document and estimation of inter-annotator agreement.

Development of system prototype and comparison with baselines.

Writing of project report and potential paper for peer-review at an NLP conference.

Applicant Requirements: Student is expected to be comfortable with coding and have working knowledge of data structures and algorithms. Student needs to be self-motivated and have a passion for story telling / literature / narratives, which includes very good English / communication skills.

Applicant Preferences: Completion of (or registration into) a natural language processing course (e.g. ITCS 4101) or machine learning course (e.g. ITCS 3156) would be a significant plus.

Specific Time considerations/conflicts: No conflicts.