ECE 1786

Creative Applications of Natural Language Processing

Lecture 6 Part 2: Project Ideation & Approval & Proposal

Project Timeline and Deliverables

Date	Item
03-Oct	Project Discussion in Class
10-Oct	Team Forming Deadline – 38 teams received; only 73 A2's
26-Oct	**Approval-in-Principle/Uniqueness of Project Topic form**
30-Oct	**Project Proposal Document Due**
30-Oct	**Project Proposal Slides Due**
31-Oct/Nov 1	In-Class Proposal Present + Extra Class in Evening + Next Eve
20-Nov	Progress Report Due
04-Dec	Final Presentation Slides Due
05-Dec/Dec 6	Final Presentations + Extra Class in Evening + Next Eve
12-Dec	Final Report Due

Note on Extra Course Presentation Hours

- **There are extra course hours for Proposal Presentations**
  - There are **two extra** sessions: October 31st & November 1
  - That is three total sessions:
  - Oct 31: 10-12 noon and 6pm-9pm; Nov 1 6-9pm; in GB 221
- **You'll present during one of these sessions**
- **You'll be doing a peer review in another session**
  - I'd like everyone to see many of the presentations;
- **So, be sure you're available 6pm-9pm October 31 and November 1; let me know if not**
  - Bring food in the evening!

Project Ideas: from Yourself

Your Own Ideas

- **What do you care about?**
  - Peace, <u>happiness</u>, success, health?

- **What makes you happy?**
  - Maybe humour

- **Humour often comes from language**
  - Jokes
    - Joke generation; perhaps in one or more specific genres
    - Joke classification
  - Generation of funny stories; how to stimulate latest models to produce humour?

Ideas from Internet

Internet

- **Stanford CS 224n Natural Language Processing with Deep Learning**

  - https://web.stanford.edu/class/cs224n/project.html

  - Multiple years

  - More research-oriented, but plenty of applications too

- **Look at last year's projects for inspiration:**

  - https://www.eecg.utoronto.ca/~jayar/ece1786.2022/

Therapeutic Chatbots

Therapeutic Chabots

- **Make GPT-4 become a counselor for anxiety, or depression,**
  - or something more specific – test anxiety

- **First step is easy: ask it to become one**
  - Need to find out what that means – need some source of expertise on the subject
  - See if it works
  - Figure out how to measure if it works well or not – the hard part, and where you'll need to source/create some data/labels

Therapeutic Chabots

■ **Particularly behaviour change**

- More specifically using 'motivational interviewing'

- Gives rise to many sub-problems:

1. Value detection – detect a human value in a statement

   - e.g., peace, success, intellectual growth, money, fame ….

2. Curiosity generation –ask a curious question based on statements from a person

More Sub-Problems in Behaviour Change

3. Change Talk classifier

    – is person indicating preference towards change

4. Sustain Talk classifier

    – is person indicating preference away from change

Next Steps: After You've got an Idea

Do the Following

1. Look for prior work – in the literature
   - Will need to quote 2-3 relevant papers in proposal
   - General search; look at ACL conferences/journals.
   - This will help you see what has been done, what is hard, what is possible

2. Look for a Dataset
   - https://datasetsearch.research.google.com
   - https://www.kaggle.com

3. Contemplate if/how you can add to the data collection and/or labelling task
   - Recall this is a (tricky) requirement of the project

Next Step: Approval-in-Principle (AIP)

Request for Approval-in-Principle form

- Due by October 26, but sooner is better.

- https://forms.office.com/r/C9XHivcMUS

1. **What & Why:** 2-3 sentences that describe what the project is and how it is motivated. (<u>Not</u> **How)**

2. **Data Source:** Your initial thoughts on where you will find relevant data, and what role you plan to take in the collection/labelling

3. **Name:** Give your Project a Name
   - name should convey the essence of project; used for tracking
   - Creates your group identity!   Logos also welcomed!

How to Describe Your Topic?

- **Key is to say what & why**
  - engineers tend to think about *how* too soon, be warned
  - You will need to think about how to make the *what feasible*, but not in first description for someone else to understand

- **What** should be the completion of this sentence:

  > "**The goal of our project is to …**"

- Another way to force yourself to focus on **what:**
  - The inputs are:
  - The outputs are:

Project Proposal Document

Proposal Document

Document **must** have the Following sections:

1. Introduction
   - What and why (i.e. motivation)

2. Background
   - Describe 2-3 related papers you've found

3. Source of Data and Processing
   - Where will you get the data for part of project?
   - Requirement: some collection/labeling the data
     - But can't take up a big chunk of the project either

Proposal Document, cont'd

4. Architecture of the model/Structure of System

   Case 1, Old style:

   – Rough guesses of type and structure of model

   – Describe other parts of software that are involved if any

   Case 2: GPT-4 style

   – Describe the interactions of the LLM you're going to make

5. Comparison

   Case 1, Old style:

   – Describe a simple baseline model that you'll compare against

   – Simple model or hand-coded heuristic

   Case 2: GPT-4 style

   – You'll need to work on a metric that tells you if you're succeeding

   – Often evaluated by hand; perhaps also by GPT-4 classifier

Proposal Document, cont'd

6. Plan
   – Discuss how you're going to work together
     • Especially important if you don't know each other well
   – List of sub-tasks
   – Your guess as to how much time each task will take
   – Use to create estimate of end-to-end time
7. Risks
   – Predict what might go wrong & how you'd recover

■ Document also graded on structure, grammar and mechanics

Proposal Document, cont'd

- **Hard Limit of 1200 words total**
  - Doesn't count pictures or references
  - 1% penalty for every word in excess of 1200
  - Put word count and compute penalty on front cover of proposal
    - 5% penalty if this is missing
    - These words (the count & penalty) not included in count

- **Due Monday October 30 at 9pm.**
- **Upload under Assignment – Project Proposal Document**
  - Just one per group
  - Quercus will know your group, it will be the name you selected in your Approval-in-Principle

Proposal Presentations

November 1, 2022

Proposal Presentation

Similar structure **but not same** as Document:

– You choose the best order:

1. Introduction and Illustration
2. Data Collection and Processing
3. Architecture/Structure and Comparison
4. Risks
5. What You'll Have completed by November 20

– For the progress report
– Giving you a target to shoot at, that **is not** the end

Proposal Presentation

- **4 minutes maximum to present**
  - **Timer** will be set & presentation ended at 4 mins.
  - 8 Slides **maximum** (including title slide)
  - Font size **minimum** 20

- This is difficult: must choose essential messages

- Urge you to practice the talk at least 2-3 times
  - Make sure you make sense to yourself and team
  - All team members must speak, roughly equally

Slides Due

- Slides due **Monday October 30 at 9pm**
  - Uploaded to Quercus 'Assignment' Proposal Presentation
  - Must be either **powerpoint (pptx or ppt)** or **PDF**
  - No google doc web links, must convert to pdf/ppt;
    - check that it is correct

Proposal Presentation

- I will put up the schedule of which team is presenting in which time slot

- Three possible times to present
  - Oct 31: During Regular Class: 10am-12 noon
  - Oct 31: Extra Evening Class: 6:00pm-9:30pm also GB 221.
  - Nov 1: Extra Evening Class: 6:00pm-9:30pm room TBD

Peer Review of Proposals/Presentations

- You will be asked to review another group's document and presentation

- You'll be scheduled to do that in a different time period, one of the three

- If you have a hard conflict with one of the periods, you must email me what it is and why.

Questions?
Time for discussion with myself & TAs