ECE 1786

Creative Applications of Natural Language Processing

Lecture 6: Project Ideation & Approval & Proposal
## Project Timeline and Deliverables

<table>
<thead>
<tr>
<th>Date</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>04-Oct-22</td>
<td>Project Discussion in Class</td>
</tr>
<tr>
<td>11-Oct-22</td>
<td>Team Forming Deadline</td>
</tr>
<tr>
<td>27-Oct-22</td>
<td>Approval-in-Principle of Project Topic complete</td>
</tr>
<tr>
<td>31-Oct-22</td>
<td>Project Proposal Document Due</td>
</tr>
<tr>
<td>31-Oct-22</td>
<td>Project Proposal Slides Due</td>
</tr>
<tr>
<td>01-Nov-22</td>
<td>In-Class Proposal Presentations + <strong>Extra Class in Evening</strong></td>
</tr>
<tr>
<td>21-Nov</td>
<td>Progress Report Due</td>
</tr>
<tr>
<td>05-Dec</td>
<td>Final Presentation Slides Due</td>
</tr>
<tr>
<td>06-Dec-22</td>
<td>Final Presentations - <strong>Extra Class in Evening</strong></td>
</tr>
<tr>
<td>13-Dec-22</td>
<td>Final Report Due</td>
</tr>
</tbody>
</table>
There are extra course hours for Proposal Presentations
- There are **two** sessions on November 1\textsuperscript{st}
  - 10-12 noon and 6:00pm-8:30pm in GB 221

You’ll present during one of these sessions
You’ll be doing a **peer review** in the other session
- I’d like everyone to see all the presentations; since they are all different, there will be something to learn from each

So, be sure you’re available 6:00pm-8:30pm November 1
- In GB 221
- I will provide dinner @6pm; let me know of any food restrictions (e.g. vegetarian – will need to know how many)
Project Ideation
Where Do Ideas Come From?

1. From yourselves! Something you are interested in.
2. Look at projects on the internet
   - e.g. Stanford CS 224n Natural Language Processing with Deep Learning
     - https://web.stanford.edu/class/cs224n/project.html
3. A few curated topics from call for projects
Ideas from Yourself
Your Own Ideas

- What do you care about?
  - Peace, happiness, 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
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

- Georgia Tech:
  https://www.cc.gatech.edu/classes/AY2020/cs7650_spring/slides/5_project_info.pdf
Curated Ideas

Received in Call for Projects
Just a Few
Create Sci-Fi Story Synopsis

- Shane Saunderson - https://www.linkedin.com/in/shanesaunderson/
- For this project idea, I want to create a web-based tool that generates SciFi shorts.
- It would take in a variety of inputs (protagonist name, author style, key topics, technologies involved, etc.) and use that to generate a book title and either a short excerpt or brief synopsis of the story (200-300 words).
- This would require sourcing some open data sets of off-licence SciFi books, building a front end to gather different fields from the user, and of course generating and presenting the output.
- Reminds me of ideation for a play? See: https://www.youtube.com/watch?v=FTuQYgVHyg4&t=64s
Medical Publications Database Apps

- Chris Meaney from UofT Medicine
  - https://www.linkedin.com/in/christopher-meaney-23891314/
  - a Scopus based bibliometric dataset of papers

- His ideas for using the database:
  - Neural topic modelling (e.g. BERTopic, Top2Vec, etc.)
  - Neural information retrieval (e.g. can you extract relevant research articles given a query vector)
  - Abstractive/extractive summarization of research abstracts (e.g. could you generate a short representative sentence/paragraph synthesizing the core concepts of a research abstract)
Tagging research articles (e.g. given an abstract as input, could you generate a variable-length list of keywords associated with the article)

Recommender systems (e.g. given information in abstract +/- co-authorship information, for any given author)

could you recommend potential undeveloped co-authorship/research-collaboration opportunities)
What I think about: Therapeutic Chabots

- Particularly behaviour change
  - More specifically smoking cessation
  - 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

5. One-down converter
   - Convert statement to make it non-hierarchical, non-expert?
     - Input: “You’re afraid of judgement from your family”
     - Output: “Is it possible that you’re worried about being judged by your family?”
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 email

Due by October 27, but sooner is much better.

1. **What & Why:** 2-3 sentences that describe what the project is and how it is motivated. *(Not 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

Should be the completion of this sentence:

“The goal of our project is to …”
Project 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
4. Architecture of the model
   – Rough guesses of type and structure of model
   – Describe other parts of software that are involved if any

5. Baseline Model
   – Describe a simple baseline model that you’ll compare against
   – Good practice to always start simple
   – Could be a hand-coded heuristic
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
    - These words (the count & penalty) not included in count

- Due Monday October 31 at 9pm.
- Upload under Assignment – Project Proposal Document
  - Just one per group; Quercus will know your group
  - (Be sure to respond to my email request for a name)
Proposal Presentations

November 1, 2022
Proposal Presentation

Similar structure but not same as Document:

1. Introduction – what and why
2. Datasets, Data Collection and Processing
3. Anticipated Neural Net Architecture and Baseline Comparison
4. Risks
   - What might go wrong & mitigation
5. What You’ll Have completed by November 21
   - At progress report time
   - And the steps to get there
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 2-3 times
  - Make sure you make sense to yourself and team
  - Both team members must speak, roughly equally
Slides Due

- Slides due **Monday October 31 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
Proposal Presentation

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

- Two possible times to present, on Tuesday Nov 1:
  - During Regular Class: 10am-12 noon
  - Extra Evening Class: 6:30pm-8:30pm also GB 221
    - Dinner available at 6pm.
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 the opposite time period i.e. morning presentations will do peer review in evening