LING48 Computational Linguistics
Dartmouth College. Spring 2021

Final project: Parts 3 and 4

PART THREE HAS TWO DIFFERENT DEADLINES: ONE TO SUBMIT A VIDEO, AND ONE TO COMMENT ON YOUR CLASSMATES' PROJECTS. THESE ARE RIGID (AND GRADED) DEADLINES BECAUSE THEY INVOLVE OTHER PEOPLE IN THE CLASS. PLEASE READ THOSE CAREFULLY.

PART FOUR HAS A HARD DEADLINE. THIS DEADLINE IS IMPOSED BY THE COLLEGE (THE LAST DAY OF THE QUARTER), SO YOU CANNOT TURN IN ANYTHING AFTER THAT. IF YOU CANNOT MAKE THAT DEADLINE, YOU NEED TO CONTACT YOUR STUDENT DEAN, WHO WILL THEN CONTACT ME.

Part 3: Class presentation and peer feedback

In the third part of your project you will present a summary of your work to your classmates. You will also provide feedback for at least two other projects. You need to do the following:

(1) Submit a video summary of your final project before 3:00 am EDT of Tuesday, June 1st, 2020 (around midnight of Monday). Please record a short video (5 minutes or less).

Your video should include a short summary of the following:

(i) Your project as a whole
(ii) Your data
(iii) Your methodology (what kind of algorithms you are using to solve the problem)
(iv) Your preliminary results (what results have you gotten so far)
(v) Preliminary discussion of your results. (Is the data is behaving the way you expect it?)
(vi) Whether there are any ethical considerations in processing your data or using your work.

If you are working in a group, I encourage you to include everyone in the recording. It is okay to have it just as one person, but if everyone is in the recording, it will help us meet everyone in the class.

You will upload a link to your video into a discussion on Canvas. You can record the video and then (i) upload it as an unlisted video on YouTube, (ii) upload it as an MP4 to GoogleDocs/Dropbox and then share us the link, or (iii) use some other way that will make it easy for us to watch it. You can use any recording software you want (in case it helps, the one I'm using is https://obsproject.com/).

When you post your video onto Canvas, please include a one sentence description of your project so that others may decide to watch it.

(2) Please give at least fifty words of feedback on **at least two videos** of your classmates before 11:59 pm EDT of Wednesday, June 2nd, 2020 (the last minute of the last day of class). Please provide constructive feedback; tell them what you like about their project and whether you understood their goals clearly. If you do have suggestions for improvements, please propose them in a respectful manner.

Grading:

6%	Video (will be graded according to the qualitative scale on page 4 of this document)
1% x 2	Feedback for two other videos
1%	Length of video is correct (0.5 if longer than 6 mins; 0 if longer than 7 mins).
0.5%	Submitting video on schedule (no later than 3 am EDT of Tuesday, June 1st)
0.5%	Submitting feedback on schedule (no later than 11:59 pm EDT of Wednesday, June 2nd)

Part 4: Writeup and final code

You need to submit at least FOUR things. If you are working in a group, you need to submit FIVE things:

(1) Your code. For most of you, this will be a Python program that implements the final project you proposed. (If we have discussed some other format, then deliver that format). You can reuse any code from the templates and exercises. If you are going to use a technique that we haven't studied in class, please contact me so we can make sure that you have all of the resources and the code you need to implement it. Be sure that your code has metadata (your name, your e-mail, the date when it was made, and information about what the code is supposed to do).

(2) The data you used for training and testing. If you cannot send me the data for copyright/privacy reasons, please send me a small subsample that is representative of the way you trained the program.

(3) Screenshots of the output of your program in a PDF/LibreOffice/Word file.

(4) If you're in a group, please submit a document that explains how the work was split between the different people in the group. If you are working by yourself, you do not need to submit this.

(5) Your writeup. You need to submit a paper of **MAXIMUM 4 PAGES**, summarizing your work. You can use an additional fifth page for references if necessary. You need to use the ACL (Association of Computational Linguistics) template to write your project (LaTeX and Word files here). Make sure you include:

   a. A title that clearly explains what the project is about, as well as your name and e-mail address.

   b. An abstract of maximum 200 words. Make sure that this is an effective summary of your project. You can use the *Annotated Nature Abstract* template to guide your work.

   c. An introduction section, where you explain your goal and you present a summary of previous work. Your references to related work should go on this section. In general, your introduction should answer the question: "Why is this interesting or relevant"?

   d. A methodology section. Here is where you present your data. How much data do you have? How did you obtain it? Did you have to transform it in some way? You also need to explain how your machine learning algorithm works. What algorithm are you using? (e.g. SVMs, word2vecs, RNNs, BERTs). Be sure to include references to papers where these techniques were proposed.

e. A results section, where you report the performance of your program (or the results of your research). If you calculated things like accuracy, BLEU, %-positive, it should be reported here.

f. A discussion section. Here, you will discuss the results. Does your program/data behave the way you expected it? If your program is doing something new, how can it be applied to your field? Are there any ethical concerns related to how the program will be used?

g. A conclusions section. This has a summary of your results and of their significance. It also proposes "future work", which are ideas for how the program can be extended or improved.

h. A reference section. You can use any citation system, as long as you are consistent. (You also need in-line citations throughout your text above).

You need to submit these four/five files (code, data, screenshots, ACL writeup and group work contributions) to your Canvas site. You need to do this before 11:59 pm EDT of Tuesday, June 8th, 2021. This is a **hard deadline** because this is the last minute of the quarter as set by the Registrar. If you will not be able to turn the project in by this time, please contact your student dean to request an extension. Your student dean will then contact me.

Grading:

- 20% Write-up
  - 5% Intro and methodology (qualitative scale)
  - 5% Results (qualitative scale)
  - 5% Discussions and Conclusions (qualitative scale)
  - 5% Format
    - 2% Document within page limits
    - 1% References (both in-line and in a final reference section)
    - 1% Used the correct ACL format
    - 1% Well-written abstract
- 5% Code
  - 1% Does it run?
  - 2% Does it show similar results to those in the write-up (general "sanity" check)
  - 1% Metadata (your name, e-mail, etc.) and comments (concise and clear descriptions of functions; short descriptions of important pieces of code).
  - 1% Code cleanliness and quality (variable and function names are consistent and descriptive, code is orderly)

Qualitative scale:

"Completely clear" 100% of the score: Content is presented in a concise, clear and straightforward way. Relevant figures/examples/tables/citations are provided.

"Mostly clear" 75% of their score: Content is mostly concise, clear and straightforward. The explanation could be easier to understand or use more examples. Figures/examples/tables/citations are present, but they are not easy to understand.

"Somewhat clear" 50% of their score: Content is answered in a cursory way, or without delving into the concepts presented. Few relevant examples are present. Figures/examples/tables/citations are present, but they are not directly relevant to the contents.

"Unclear" 25% of their score: The explanations show misunderstandings of the concepts presented. The sections are not clearly connected to the rest of the paper. Few or no figures/examples/tables/citations.

"No answer" 0% of their score. The student wrote no answer.