Comp541 Final Project Log

<Your Name> <Date>  
<Link/citation to the paper you are replicating>

**Table of Contents:**

[2017.03.05 Project setup](#_7bak70lidjlp)

[2017.03.05 General writing advice](#_knv9prj40ivq)

# 

# 

# 2017.03.05 Project setup

For your final project, you will replicate the results of the paper you picked using Knet, run further experiments to understand and improve upon the results, prepare a report and a presentation. The main requirement to pass this course is to understand and replicate the experiments in the paper you chose, however I will consider significant extra credit for those who try and improve upon these results using creative experiments. I expect some of your work to turn into actual publications.

Here are the items you are expected to set-up by March 12:

1. **Research Log:** You will use this Google Doc as your research log for the remainder of the semester. Write your name and paper title (with link) below the main title. As you work on your project and read new papers, understand some math, implement an algorithm, get new results, etc. please create new entries in this log so we can follow your progress. For each entry, use a Heading 1 title in the “YYYY.MM.DD Title” format, and they will appear in the Table of Contents on the first page when you click to refresh it. I prefer newest entries at the top, but up to you -- it is your research log. **Your first entry (besides this one) should contain the work you have done for the first two reading passes.** Feel free to copy and paste your paper summary, glossary of terms, an annotated bibliography of related papers you have read etc. Your subsequent entries should describe your work in human readable form (e.g. nice graphs with explanatory captions are ok, big ugly tables aren’t). For each experiment, in addition to a good description, make sure to include a data table (where you provide relevant statistics on your train/dev/test data), a parameter table (where you give all the hyperparameters you used during training necessary to replicate your experiment), and a results table/figure. We will track your progress using your research log, so update it often.
2. **Data Sheet:** For large tables and other experimental raw data that is not easily “human readable”, create an associated Google Sheet. All the experiments reported in your Research Log can have their raw data in this data sheet. For example, if you are going to provide a learning curve, put the data you used to produce the graph in the Data Sheet, and a nice graph (which may also be prepared using Google Sheets) in the Research Log entry. Provide links into your Data Sheet in your Research Log wherever necessary. **As a start, summarize the datasets used in the experimental section of your paper in your data sheet.**  Indicate details like number of instances, number of features, number of sentences/words (if text), dimensions (if image), etc. Also provide the location of the dataset you have downloaded.  
     
   **Link to your data sheet:  
   Location of your data:**
3. **GitHub Repo:** Create a GitHub repo for all the software for your experiments. If you are not familiar with GitHub, Google “GitHub tutorial” and browse a couple of tutorials this week. **As a start, describe your project in the README.md file of your GitHub repo.** Make sure you know how to clone your repo, add new code, commit and push new code to GitHub.   
     
   **Link to your GitHub repo:**
4. **Presentation Slides:** You will use Google Slides to prepare your presentation. **For now set up a blank presentation and add your title slide.** You can update the presentation with important results during the project.

**Link to your presentation slides:**

1. **Technical Report:** You will use <https://www.overleaf.com> to write your final report using LaTeX. If you have not used LaTeX before, Google “LaTeX tutorial” and browse some this week, trying what you have learnt on overleaf. This will eventually become an 8-10 page report in a conference proceedings format (similar in format to the paper you are replicating) complete with an abstract, introduction, related work, model, experiments, analysis, contributions, and bibliography. It should introduce, motivate, and give a detailed background on the problem you are working on, and it should have enough detail on your model and experiments to make replication possible. See the next section for more detailed writing advice. **For now just write the introduction section summarizing your last two reading assignments, and add a table of evaluation results from previously published work on your task.**  
     
   **Link to your final report:**

# 2017.03.05 General writing advice

Here is some general writing advice for your final report (or anything else technical you write, for that matter):

1. **Empathy:** This is the single most important principle of technical writing. *Try reading what you write from the perspective of somebody who has not spent the last few months working on your problem.* Better yet, find such a person and see if they understand everything you are talking about. Don’t just take their word for it, ask them to tell you what they understand in their own words. See where they struggle and debug your paper: Do they get lost in too much detail and miss the main point? Do they get disoriented because you jump around too much? Are there terms they do not understand? Fix the paper using the following techniques until a dedicated freshman can understand all the important points.
2. **Winston’s Onion Rule:** The document should state the most important points first, and expand on them gradually. It is a mistake to keep any important points until the end of the paper. Only details and supporting material should be left to the end. *If I stop reading the document at any point, everything I haven't read so far should be less important than everything I have read up to this point*:
   1. The title should be descriptive of the main point.
   2. The first sentence should state the main point.
   3. The first paragraph should expand on the first sentence.
   4. The first section should expand on the first paragraph.
   5. The first chapter should expand on the first section.
   6. The whole paper/thesis should expand on the first chapter, etc.
3. **Yuret’s Fractal Rule:** Parts at every level of your document, down to each paragraph, should have its own introduction / conclusion to keep the reader oriented (i.e. stop them from asking “What is this person talking about now, and why?”):
   1. The first chapter of a paper/thesis should state the topic of the paper/thesis and the last chapter should summarize its point.
   2. The first section of a chapter should state the topic of the chapter and the last section should summarize its point.
   3. The first paragraph of a section should state the topic of the section and the last paragraph should summarize its point.
   4. The first sentence of a paragraph should state the topic of the paragraph and the last sentence should summarize its point.
4. **No undefined terms:** Any technical term your nine year old niece would not understand should be defined before first use. Any acronym should first be given in parentheses next to its long form before first use. All variables in equations, all axes in graphs should be explained at the first opportunity. Tables and Figures should have descriptive captions that can be understood stand-alone without referring to the text. Technical terms and mathematical notation should be used consistently, with no confusing variations (i.e. calling the same thing context vector somewhere and word context vector elsewhere will confuse the reader into thinking these are two separate things).
5. **Replicability**: Science is based on replicatable results. Your paper should provide enough detail (possibly in the appendices), and links to all its code and data, to replicate each of its results. In particular, for each set of experiments you should have:
   1. Data table: e.g. in a natural language processing experiment, things like number of words and sentences in train, dev, test; vocabulary size, tagset size, tag frequencies, out-of-vocabulary rate, average sentence length, i.e. any data statistic relevant to the task should go to a data table.
   2. Parameter table: things like the model structure, the training algorithm used, the hyperparameters used, number of training epochs, and any other details related to experimental replication should go to a table.
   3. Result table: the results (table or plot) should clearly indicate the evaluation metric, sensible lower bound baselines, upper bounds (e.g. inter-annotator agreement) if available, and current state of the art in published work to put your results in perspective.
6. **Presentation:** You will have a 15 minute presentation slot, divided into 10 minutes of presentation and 5 minutes of questions. This means 10-15 slides should be sufficient (about 1 slide per minute). Start with a good description, motivation, and background of your problem. Describe your model, your experiments and any new understanding / improvements you have achieved. *Do not ever read from your slides!*  That puts everybody to sleep. Slides are for visualization of things that are hard to describe in words, not speaker notes. If you have too much text on your slides, you should revise them and use more graphics. Test the presentation on your friends beforehand. Make sure you can finish in time. Most importantly, quiz them to see if they understood what you are talking about, go back to the drawing board if they don’t. You should have the first draft of your slides ready at least one week before the deadline.