A Few Notes on Custom Final Projects

Last updated on January 27, 2022

So you heard you can do a *custom* final project for 224n. Sounds pretty special! But what does doing a custom final project really mean? Here are some points to keep in mind if you're considering a custom final project. Let's start with the non-negotiables.

If you do a custom final project, you **must** submit a clear and complete project proposal document (especially important for custom projects!). Your custom project **must** be human-language related. For example, applying Transformers to a genetics dataset is not allowed unless the purpose is to, e.g., to analyze the effect of different human language pre-training datasets/tasks on downstream performance on the genetics task and your final report discusses this effect in detail. Your custom project **must** be scoped to lead to some sort of non-trivial scientific or engineering deliverable by the end of the quarter. Your project **must** be tractable to complete on a budget of a few hundred dollars on Azure; we have only a little flexibility here. You can check per-hour Azure GPU prices here. Assume that any additional compute needed will have to be your own.

A successful final project **should** have clearly-defined metrics to measure the outcome of whatever experiment(s) you run; exact-match accuracy and F1 score on either in-distribution or out-of-distribution test data are used to evaluate experiments in the default final project. Human evaluation scores (as opposed to automated/objective metrics) can work too, but you should allocate time specifically for developing a rubric to ensure evaluation consistency and for performing the evaluation itself. This constraint may not apply to some theoretical or exploratory projects, but even for these, if possible, it's still useful to think about how you would measure the effects of your theoretical/conceptual insight if you were to apply them to a practical model (would it increase accuracy? reduce training time? reduce number of training steps?). A successful final project also **should** have a short, specific guiding question or hypothesis. An example of good guiding questions 'Does fine-tuning some BERT layers work better than others, and is the best layer to fine-tune the same for all tasks?' or 'Does pre-training in one language still help if we fine-tune on a different language?' Guiding questions like 'How does BERT work?' may be too vague to be useful for your final project. Questions of the form 'Will X method work better than state-of-the-art (SOTA) for Y problem?' can work, but often the answer is no. That is okay, as long as you can analyze or discuss some potential

reasons for why.

A successful custom final project **may** develop a novel method/algorithm, but studying existing methods or applying them to new problems can also lead to interesting results without inventing something new. Re-implementing an interesting model that someone else developed recently can lead to an interesting project- just remember that the best projects always push things further in some way, by considering improvements to the model, providing ablations or experimental studies that the original paper did not provide, or by running it on different datasets that illuminate novel questions. Separately, a successful final project **may** become a longer research project afterwards, but this is certainly not a requirement. A successful custom final project **may** improve on some existing SOTA, but well-executed projects that perform thoughtful experiments and produce negative results, projects that just aim to build a fun system, or projects that answer an interesting research question without advancing SOTA are equally valuable.

In terms of evaluating final projects, we will generally consider the 'value add' of the project to the NLP community. For example, if the final result of the project is to re-implement an interesting, non-trivial model from scratch because the source isn't available, or only in Tensorflow v1, etc., then that can already be a good achievement and a real value add. But downloading a package from GitHub and running it on the same data the original authors did definitely isn't a value add. For examples of successful final projects, check out last year's prizes for best custom project reports here.

If you're looking for more inspiration for your custom project, consider looking at the custom project proposals that have been written by members of the Stanford research community here. If you are interested in a project in the proposal list, send an email to the designated mentor contact. Note that mentors may not have the capacity to mentor the same project for multiple teams, so they may be assigned first-come, first-served. In other words, if you're interested in a project, reach out to the listed mentor sooner rather than later!

Of course, this guide is certainly not comprehensive, so if you have any additional questions about the custom or default final project, feel free to ask on Ed or in office hours. Good luck!