NLP 2023: Seminar-style Requirements

Rowan University

January 17, 2023

Instructions: A large part of graduate studies in computer or data science requires reading, digesting, re-implementing, questioning and adapting current research to new problems. The way through which this osmosis of information is best achieved (in my opinion) is through having to present that very material back to your fellow students. We will tackle this beginning on or after Week 5 of the tentative plan in the syllabus. Each week, students will sign up on the schedule to read, review and present an academic paper in the NLP field. After reading, students will be asked to put together a 30 minute presentation (PPT/PDF) detailing the paper and its impact. The following are points you should touch on in your presentation.

1. **What problem is being solved?** This can mostly be gleaned through the abstract and conclusion, but please give us the frame of reference through which this paper was written.

2. **What datasets are used for evaluation?** There are numerous open-source datasets used in NLP research and these will be a common thread in our readings. Please discuss which is/are used in the study and their descriptive statistics.

3. **What is novel about this research?** Does this paper represent a paradigm shift in NLP research, or have the authors found a clever way to improve on existing research?

4. **What existing research do the authors compare themselves to?** Much of machine learning research is treated like a competition, comparing new models to existing baselines. What models do the authors compare themselves to, and why? Do the authors compute standard statistical tests to validate that their results are not due to random chance?

5. **Do the authors provide open-source code to re-create their experiments?** Is the code for the experiments available and easy to run? Which frameworks do the authors use? Is their data included, or readily available elsewhere?

The purpose of this exercise is three-fold; first, it is important to learn how to read, structure and write academic papers, **even if you intend to work in industry!** Second, it teaches the traps to look out for, including lack of statistical support, code or data. Finally, presenting the material means that you fully understand that material, and public speaking is a skill data and computer scientists must practice!

If this is your first experience with reading technical academic papers, this link should provide you with some hints on getting started.